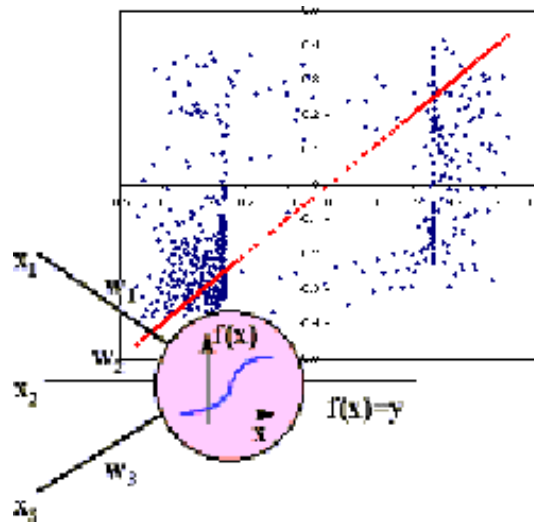


Міністерство освіти і науки України
Національна академія наук вищої освіти України
Інститут проблем штучного інтелекту (Україна)
Донбаська державна машинобудівна академія (Україна)
Apeiron University in Banja Luka, (Bosnia and Herzegovina)
DAAAM International Vienna

Mechanical Engineering Faculty in Slavonski Brod, JJ Strossmayer University of Osijek (Croatia)
University of Montenegro Faculty of Mechanical Engineering
University of Zielona Góra (Poland)
"American Jurnal Neural Network and Aplication" (USA)
Тернопільський національний технічний університет імені Іва-на Пулюя
Вінницький національний аграрний університет (Україна)
Вінницький національний технічний університет (Україна)
Проблемна лабораторія мобільних інтелектуальних технологічних машин (Україна)



ЗБІРНИК НАУКОВИХ ПРАЦЬ

XXII Міжнародної наукової конференції

«НЕЙРОМЕРЕЖНІ ТЕХНОЛОГІЇ ТА ЇХ ЗАСТОСУВАННЯ НМТЗ-2023»

м.м. Краматорськ-Вінниця-Тернопіль - 2023

Міністерство освіти і науки України
Національна академія наук вищої освіти України
Інститут проблем штучного інтелекту (Україна)
Донбаська державна машинобудівна академія (Україна)
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)
"American Journal Neural Network and Application" (USA)
Тернопільський національний технічний університет імені Івана Пулюя
Вінницький національний аграрний університет (Україна)
Вінницький національний технічний університет (Україна)
Проблемна лабораторія мобільних інтелектуальних технологічних машин (Україна)

НЕЙРОМЕРЕЖНІ ТЕХНОЛОГІЇ ТА ЇХ ЗАСТОСУВАННЯ НМТЗ-2023

ЗБІРНИК НАУКОВИХ ПРАЦЬ

XXII Міжнародної наукової конференції

за заг. ред. д-ра техн. наук., проф. С. В. Ковалевського і
Hon.D.Sc., prof. Dasic Predrag

м.м. Краматорськ-Вінниця-Тернопіль - 2023

УДК 004.032.26+621(061.3)
Н46

Рецензенти:

Рамазанов С.К., докт.техн.наук, докт.экон.наук, 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 30.11.2023)

Н46 Нейромережні технології та їх застосування НМТЗ-2023: збірник наукових праць ХХІІ
Міжнародної наукової конференції «Нейромережні технології та їх застосування НМТЗ-
2023» / [за заг. ред. д-ра техн. наук., проф. С. В. Ковалевського і Hon.D.Sc., prof. Dasic Predrag].
- Краматорськ: ДДМА, 2023. – 142 с.

ISBN 978-617-7889-58-7

У збірнику праць представлені перспективні теоретичні та практичні розробки в області нейромережних технологій, виконані в 2023 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 першого/другого/третього (бакалаврського/магістерського/освітньо-наукового) рівнів вищої освіти, наукових працівників широкого профілю та фахівців.

ISBN 978-617-7889-58-7

©ДДМА, 2023

Програмний комітет конференції

Dašić Predrag – Hon.D.Sc., Prof., Academy of Professional Studies Šumadija - Department in Trstenik (Serbia);

Jenek Mariusz – Dr. inz (Polska, Uniwersitet Zielonogorski);

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

Sandra Poirier – Doctor of Education, CFCS, LD/N Professor (Middle Tennessee State University, USA);

Dorđević Milan, president – Dr.Sc., Prof., Academy of Professional Studies Šumadija - Kragujevac (Serbia);

Guida Domenico – Dr.Sc., Prof., University of Salerno, Department of Industrial Engineering (DIIn), Fisciano (Italy);

Zdravko Krivokapić – Dr.Sc., Prof., Faculty of Mechanical Engineering, Podgorica, Montenegro;

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);

Zhelezarov S. Piya, rector – Dr.Sc., Prof., Technical University of Gabrovo, Gabrovo (Bulgaria);

Yevhenii Shkvar - Dr.Sc., Prof., College of Engineering, Zhejiang Normal University, Department of Mechanical Design and Automaton (China);

Ковалевська О.С. – к.т.н., доц., ДДМА, м.Краматорськ-Тернопіль (Україна);

Ковалевський С.В. – д.т.н., проф., ДДМА, м.Краматорськ-Тернопіль (Україна);

Ковальов В.Д. – д.т.н., проф., ДДМА, м.Краматорськ-Тернопіль (Україна);

Новіков Ф.В. – д.т.н., проф., ХНЕУ, м.Харків (Україна);

Петров О.В.- к.т.н., доцент (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-Тернопіль (Україна);

Хаєцька О.П. – к.е.н., доц., ВНАУ, м.Вінниця (Україна);

Шевченко А.І. – д.т.н., проф., ППШ НАНУ, м.Київ, (Україна);

Шевчук О.Ф. – к.ф-м.н., доц., ВНАУ, м.Вінниця (Україна).

1. Ковалевський С.В. (<i>Донбаська державна машинобудівна академія, м. Краматорськ-Тернопіль, Україна</i>) XXII МІЖНАРОДНА НАУКОВА КОНФЕРЕНЦІЯ «НЕЙПРОМЕРЕЖНІ ТЕХНОЛОГІЇ ТА ЇХ ЗАСТОСУВАННЯ - НМТІЗ-2023».	8
2. Christopher Bulte, Max Kleinebrahm, Hasan Umitcan Yilmaz (<i>Institute for Industrial Production, Chair of Energy Economics, Karlsruhe Institute of Technology, Germany</i>) NEURAL NETWORK-BASED DATA IMPUTATION FOR IMPROVED ENERGY TIME SERIES ANALYSIS	9
3. Derrick Hines, Philipp Bekemeyer (<i>DLR German Aerospace Center, Lilienthalpl. Brunswick, Germany</i>) NEURAL NETWORK-DRIVEN EXPLORATION IN AIRCRAFT DESIGN	12
4. Enrico Collini, Paolo Nesi, Gianni Pantaleo (<i>Distributed Systems and Internet Technologies Lab, Department of Information Engineering, University of Florence, Florence, Italy</i>) DEEP LEARNING APPROACHES IN TOURISM MANAGEMENT	15
5. Farhan A. Alenizi, Amir Masoud Rahmani (<i>Future Technology Research Center, National Yunlin University of Science and Technology, Taiwan</i>) AI IN INDUSTRY 4.0: COMPREHENSIVE DESIGN AND QUALITY MANAGEMENT	18
6. H. Ruppert, A. Krug (<i>Faculty of ElectriRal Engirteering, Heilbronn Univrsity of Applied Sciences Heilbronn, Germany</i>) DESIGNING A NEURAL NETWORK BASED ON ADAPTABILITY AND NOISE SUPPRESSION	21
7. Hamid Reza Attar, Nan Li (<i>Dyson School of Design Engineering, Imperial College London, UK</i>) A DEEP LEARNING PLATFORM FOR SHEET STAMPING GEOMETRY OPTIMIZATION	23
8. K.L. Keung, C.K.M. Lee (<i>Department of Industrial and Systems Engineering, The Hong Kong Polytechnic University, China</i>) INTEGRATING CPRMFS FOR TOOLS STORAGE IN SMART MANUFACTURING	25
9. Koliesnyk O. O, Antonenko O.S. (<i>Odesa I. I. Mechnikov National University, Odesa, Ukraine</i>) COMPARISON OF TECHNOLOGIES FOR FULL CYCLE DEVELOPMENT OF MACHINE LEARNING MODELS	27
10. Kshitij Bhatta, Qing Chang (<i>Department of Mechanical and Aerospace Engineering, University of Virginia, USA</i>) FLEXIBLE MANUFACTURING SYSTEMS: HGNN-MARL STRATEGY	35
11. Maciej Klimas, Dariusz Grabowski (<i>Department of Electrical Engineering and Computer Science, Silesian University of Technology, Gliwice, Poland</i>) LSTM FOR ELECTRIC ARC FURNACE MODELING	38

12. **Marco Forgiione, Aneri Muni, Dario Piga** (*Dalle Molle Institute for Artificial Intelligence, IDSIA USI-SUPSI, Lugano-Viganello, Switzerland*) **ADAPTIVE TRANSFER LEARNING FOR DYNAMIC SYSTEM MODELING WITH RECURRENT NEURAL NETWORKS** 42
13. **Marija Kostic, Vuk Batanovic, Bosko Nikolic** (*University of Belgrade, Innovation center of the School of Electrical Engineering, Belgrade, Serbia*) **NEURAL NETWORK-DRIVEN MULTILINGUAL CODE COMMENT CLASSIFICATION** 45
14. **Marta Lipnicka, Andrzej Nowakowski** (*Faculty of Math and Computer Sciences, University of Lodz, Poland*) **NEW PARADIGM DEEP LEARNING WITH OPTIMAL CONTROL** 47
15. **Martin Greguric, Filip Vrbanic, Edouard Ivanjko** (*University of Zagreb, Faculty of Transport and Traffic Sciences, Zagreb, Croatia*) **INTELLIGENT ESTIMATION OF SPEED ON HIGHWAYS WITH THE HELP OF AI** 50
16. **Miho Klaic, Danko Brezak, Toimslav Staroveski, Dora Bagaric** (*University of Zagreb, Faculty of Mechanical Engineering and Naval Architecture, Zagreb, Croatia*) **INTELLIGENT TOOL WEAR MONITORING IN STONE DRILLING** 53
17. **Nikolas Sacchi, Antonella Ferrara** (*Dipartimento di Ingegneria Industriale e dell'Informazione, University of Pavia, Italy*) **NEURAL NETWORKS TRANSFORMING ENERGY FORECASTING IN METALLURGICAL PRODUCTION** 56
18. **Sergiy Kovalevskyy** (*Donbass State Engineering Academy, Ukraine*), **Julia Volodchenko** (*Legal Counsel in QuartSoft Ukraine*) **NEURAL NETWORK TECHNOLOGIES FOR UKRAINE'S RECOVERY AND DEVELOPMENT** 58
19. **Tao Li, Feng Xie, Xin Guo, Qingchun Feng** (*School of Agricultural Engineering, Jiangsu University, China*) **AUTONOMOUS MULTI-ARM ROBOTIC SYSTEMS** 60
20. **V. Giannella, R. Sepea** (*Department of Industrial Engineering, University of Salerno, Italy*) **REAL-TIME PROBABILISTIC STRUCTURAL LIFE PREDICTION USING NEURAL NETWORKS AND MONTE CARLO** 63
21. **Бабаш А.В.** (*Україна, м. Краматорськ, ДДМА*) **ВИКОРИСТАННЯ ЗАСОБІВ ШТУЧНОГО ІНТЕЛЕКТУ У МОБІЛЬНИХ ДОДАТКАХ** 67
22. **Бодянський Є.В., Костюк С.О.** (*Харківський національний університет радіоелектроніки, Харків, Україна*) **ПАРАМЕТРИЧНА ІДЕНТИФІКАЦІЯ АДАПТИВНИХ АКТИВАЦІЙНИХ ФУНКЦІЙ В ПОПЕРЕДНЬО НАВЧЕНИХ МОДЕЛЯХ ШТУЧНИХ НЕЙРОННИХ МЕРЕЖ** 72
23. **Воробель А. О.** (*Національний університет «Львівська політехніка», Україна*) **ЗАСТОСУВАННЯ НЕЙРОМЕРЕЖЕВИХ ТЕХНОЛОГІЙ ДЛЯ ВИЯВЛЕННЯ ТА КЛАСИФІКАЦІЇ ПЕРЕЛОМІВ НА РЕНТГЕНІВСЬКИХ ЗНІМКАХ** 79

24. Ізонін І.В., Бляхар Р. С., Ткаченко Р. О. (Національний університет «Львівська політехніка», Львів, Україна) **НЕЙРОМЕРЕЖЕВИЙ МЕТОД ПОСЛІДОВНОЇ ГЛОБАЛЬНО-ЛОКАЛЬНОЇ АПРОКСИМАЦІЇ КОРОТКИХ НАБОРІВ МЕДИЧНИХ ДАНИХ** 87
25. Кириченко М. Г. (Національний університет «Львівська політехніка», Україна) **ІНТЕЛЕКТУАЛЬНІ МЕТОДИ ОПТИМІЗАЦІЇ РОБОТИ ІНТЕРНЕТ-МАГАЗИНІВ** 90
26. Ковалевський С.В., Ковалевська О.С. (Донбаська державна машинобудівна академія, м.Краматорськ-Тернопіль, Україна) **ОСВІТНЬО-КОНСУЛЬТАЦІЙНИЙ ЦЕНТР ШТУЧНОГО ІНТЕЛЕКТУ ДЛЯ ВСІХ** 98
27. Майстренко О.С., Ключин Д.А. (Київський національний університет імені Тараса Шевченка м. Київ, Україна) **ДІАГНОСТИКА РАКУ МОЛОЧНОЇ ЗАЛОЗИ ЗА ДОПОМОГОЮ ЗГОРТКОВОЇ НЕЙРОМЕРЕЖІ І ФРАКТАЛЬНОГО АНАЛІЗУ ЯДЕР БУКАЛЬНОГО ЕПІТЕЛІУ** 100
28. Мельников О.Ю., Капелешук А.О. (Донбаська державна машинобудівна академія, м. Краматорськ, Україна) **ЗАСТОСУВАННЯ ЗГОРТКОВОЇ НЕЙРОННОЇ МЕРЕЖІ ДЛЯ КЛАСИФІКАЦІЇ МОНЕТ АНТИЧНОГО СВІТУ** 106
29. Музика М.В. (Національний університет «Львівська політехніка», м. Львів, Україна) **АНАЛІЗ СТАНУ ЯКОСТІ ПОВІТРЯ ТА ВИЗНАЧЕННЯ ЧИННИКІВ, ЩО НА НЬОГО ВПЛИВАЮТЬ** 111
30. Нетилько С.А. (Національний університет «Львівська політехніка», Україна) **ЗАСТОСУВАННЯ МЕТОДІВ МАШИННОГО НАВЧАННЯ ДЛЯ ДІАГНОСТИКИ ПНЕВМОНІЇ** 118
31. Нечволода Л.В., Крикуненко К.М., Макаров С.І. (Донбаська державна машинобудівна академія, м. Краматорськ-Тернопіль, Україна) **ВИКОРИСТАННЯ НЕЙРОННОЇ МЕРЕЖІ ДЛЯ ПРОГНОЗУВАННЯ ТРЕНДІВ ОНЛАЙН ІГОР** 124
32. Олійник С.Ю., Онищук С.Г., Тулупов В.І. (Донбаська державна машинобудівна академія, м. Краматорськ, Україна) **ІНТЕГРАЦІЯ ШТУЧНОГО ІНТЕЛЕКТУ ТА НЕЙРОННИХ МЕРЕЖ У КУРСІ «ОСНОВИ ТЕХНІЧНОЇ ТВОРЧОСТІ»** 127
33. Павлюк О.М., Медиковський М.О., Міщук М.В. (Національний університет «Львівська політехніка», Львів, Україна) **МЕТОД ДАТА-МАЙНІНГУ У НЕЙРОМЕРЕЖЕВОМУ ПРОГНОЗУВАННІ ЗАЛИШКОВОГО ЗАРЯДУ АКУМУЛЯТОРНОЇ БАТАРЕЇ КЕРОВАНОЇ РОБОТОТЕХНІЧНОЇ ПЛАТФОРМИ** 129
34. Слюсар В.І. (Центральний науково-дослідний інститут озброєння та військової техніки ЗС України), Слюсарь І.І. (Полтавський державний аграрний університет, Україна) **КЛАСИФІКАЦІЯ ЗИМОВОГО ТА ЛІТНЬОГО СЕЗОНІВ З БОРТА БПЛА** 134

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

XXII МІЖНАРОДНА НАУКОВА КОНФЕРЕНЦІЯ «НЕЙРОМЕРЕЖНІ ТЕХНОЛОГІЇ ТА ЇХ ЗАСТОСУВАННЯ - НМТЗ-2023».

Чергова двадцять друга Міжнародна науково-технічна конференція «Нейромережні технології та їх застосування НМТЗ-2023» проходить у складних умовах агресії росії, яка розв'язала проти нашої країни нічим не спровоковану, жорстоку, варварську, загарбницьку війну. Ворог поставив за мету знищення України, її інтелекту, культури, ідентичності. Народ України мужньо протистоїть імперським планам загарбника і робить усе, щоб вистояти у цьому протистоянні. І ми відчуваємо величезну підтримку та допомогу всього демократичного світу, який вирізняється високими моральними та інтелектуальними якостями.

Перед вітчизняною наукою стоять завдання, вирішення яких є життєво важливим з урахуванням викликів, аналогів яким у світі немає. Необхідно вирішувати завдання безпосередньо пов'язані з використанням вкрай обмежених ресурсів у найкоротші терміни. Причому ці рішення мають бути оптимальними у кожний момент часу та у довгостроковій перспективі. А технічні, економічні, організаційні заходи, що вживаються, покликані безперервно формувати соціально-економічне середовище, в якому гармонійно і безпечно розвивається як суспільство в цілому, так і кожна людина.

Сьогодні об'єднанню інструментарію дослідження та побудови глибоко комп'ютеризованого середовища з метою багатокритеріальної оптимізації сприяють розробки в галузі перспективних засобів моделювання різних об'єктів та процесів із високим ступенем ідентифікації таких моделей реального світу. До таких засобів у першу чергу можна віднести нейромережні технології як процеси ідентифікації нейромережних моделей, що наближають, до процесів системної взаємодії живої та неживої природи на основі інтелектуальних обчислень.

У зв'язку з викладеним природним розвитком системного моделювання є розробки у сфері розвитку штучного інтелекту як найвищого рівня моделювання довкілля та оптимального управління нею. При цьому досвід застосування нейромережних технологій для вирішення різного кола практичних завдань сприяє розумінню та служить подальшому розвитку наукомістких засобів та методів побудови середовища безпечної життєдіяльності кожної людини окремо та всього суспільства загалом.

Чергова двадцять друга науково-технічна конференція «Нейромережні технології та їх застосування НМТЗ-2022» також сприятиме цим перерахованим вище цілям.

Christopher Bulte, Max Kleinebrahm, Hasan Umitcan Yilmaz (*Institute for Industrial Production, Chair of Energy Economics, Karlsruhe Institute of Technology, Germany*)

NEURAL NETWORK-BASED DATA IMPUTATION FOR IMPROVED ENERGY TIME SERIES ANALYSIS

Abstract: *This study introduces a novel approach, named MVTSI (Multivariate Time Series Imputation), leveraging neural network-based models for enhanced imputation of missing data in energy time series. The proposed method incorporates Long Short-Term Memory (LSTM) networks to learn temporal dependencies and capture cross-dimensional correlations in multivariate data. The first stage focuses on learning the distribution of missing values, allowing for the generation of synthetic data. Subsequently, an attention-based neural network estimates the actual missing values, enabling simultaneous consideration of temporal and cross-dimensional dynamics. Applied to European energy data from the ENTSO-E transparency platform, MVTSI surpasses benchmark methods, demonstrating its effectiveness in improving energy time series analysis. The study underscores the pivotal role of neural networks in advancing imputation techniques, contributing to more accurate and adaptable energy system models.*

Анотація: *Це дослідження представляє новий підхід під назвою MVTSI (багатомірна імпутація часових рядів), який використовує моделі на основі нейронних мереж для вдосконаленої імпутації відсутніх даних у часових рядах енергії. Запропонований метод включає в себе мережі довгострокової короткочасної пам'яті (LSTM) для вивчення часових залежностей і захоплення міжвимірних кореляцій у багатомірних даних. Перший етап зосереджений на вивченні розподілу відсутніх значень, що дозволяє генерувати синтетичні дані. Згодом нейронна мережа на основі уваги оцінює фактичні відсутні значення, що дозволяє одночасно враховувати часову та міжвимірну динаміку. Застосовуючи європейські енергетичні дані з платформи прозорості ENTSO-E, MVTSI перевершує еталонні методи, демонструючи свою ефективність у покращенні аналізу часових рядів енергії. Дослідження підкреслює ключову роль нейронних мереж у вдосконаленні методів імпутації, сприяючи створенню більш точних і адаптованих моделей енергетичних систем.*

Multivariate time series, prevalent in energy applications, frequently encounter missing values, impacting downstream analysis. Classical imputation methods focus on either temporal dynamics or cross-dimensional correlations. This paper presents MVTSI, utilizing a two-step approach with an attention-based neural network, demonstrating flexibility and adaptability to the specific characteristics of the energy data. *Motivation:* Missing values pose challenges in energy data analysis, affecting models crucial for proposing future energy systems, reducing CO₂ emissions, and estimating investment expenditures. This paper addresses the importance of imputing missing values in energy data, emphasizing the necessity for accurate imputation methods in ensuring reliable energy system models. A comprehensive review of related work is presented, covering the significance of missing values in energy data and various methods for time series analysis. *Importance of Missing Values in Energy Data:* The impact of missing data on energy system models is discussed, highlighting the variability in results and planning strategies. The insufficiency of simple imputation methods is underscored, especially for long-term planning models and short-term forecasting models in renewable energy contexts. This section provides an overview of the ENTSO-E transparency platform, a key data source for the proposed imputation model. It details the characteristics of the data, including electricity demand, prices, and energy production from diverse technologies.

Extensive experiments and evaluations are conducted on the performance of the proposed MVTSI model. The results are compared to several benchmarks and assessed on different subsets of the ENTSO-E data, demonstrating the model's superiority in imputing missing values. While the proposed MVTSI model shows promising results, it is essential to acknowledge its limitations. One limitation lies in the assumption of learning the distribution prior to imputation. The model's performance may be sensitive to the quality and representativeness of the training data. Future research should explore methods to enhance the robustness of the model to variations in data quality and distribution.

Additionally, the current model is designed for application to energy data from the ENTSO-E platform. Extending its applicability to other domains and datasets warrants investigation. The model's adaptability to different energy system structures and dynamics, as well as its scalability to handle larger datasets, should be explored in future studies. A comprehensive comparison with existing imputation methods is crucial for validating the efficacy of MVTSI. This involves benchmarking against classical imputation techniques and other deep learning-based approaches. Furthermore, evaluating the model's performance on diverse datasets beyond energy data will contribute to understanding its generalizability and versatility.

The interpretability of the MVTSI model's decisions is a vital aspect, especially in domains like energy where decision-makers rely on model outcomes for critical planning. Future work should focus on developing methodologies to interpret the attention mechanism and provide insights into why certain imputation decisions are made. Explainability is paramount for building trust in the model's recommendations. As the MVTSI model progresses towards practical deployment, considerations regarding computational efficiency, real-time processing, and integration with existing energy system analysis frameworks become imperative. Future research should delve into optimizing the model for deployment in real-world scenarios, ensuring it aligns with the operational requirements of energy stakeholders.

The use of advanced imputation models raises ethical considerations, particularly regarding the potential biases introduced during the learning process. Researchers should proactively address ethical concerns related to the imputation of missing values, ensuring fairness and transparency in the decision-making processes influenced by the model. In conclusion, the MVTSI model presents a significant advancement in addressing missing values in multivariate time series data, particularly in the context of energy systems. The interdisciplinary nature of this research, combining deep learning and energy domain expertise, positions it at the forefront of data imputation methodologies. As the model evolves, addressing the outlined considerations will contribute to its robustness, applicability, and ethical soundness. The importance of the MVTSI model extends beyond its technical contributions, impacting the field of energy analytics profoundly. Imputing missing values in multivariate time series data is pivotal for ensuring the reliability and accuracy of energy system models. Several key aspects highlight the significance of this research. Accurate energy system models are paramount for informed decision-making in energy planning. The MVTSI model, by effectively handling missing values, contributes to the generation of more reliable forecasts, reducing uncertainties in long-term energy planning. This, in turn, aids policymakers and energy stakeholders in devising robust strategies for sustainable energy development.

Efficient allocation of resources is a critical aspect of energy system management. By providing accurate imputations, MVTSI aids in optimizing resource allocation for electricity generation, distribution, and demand management. This optimization is essential for enhancing energy system efficiency, reducing waste, and minimizing environmental impacts. In the context of increasing reliance on renewable energy sources, accurate imputation becomes imperative. The intermittent nature of renewable generation, such as wind and solar, requires precise forecasting. MVTSI's ability to handle missing data in energy production time series contributes to reliable predictions, facilitating the seamless integration of renewable energy into the grid. The economic implications of data imputation in energy analytics are substantial. Inaccurate or incomplete data can lead to suboptimal investment decisions, impacting the financial viability of energy projects. MVTSI's role in improving the quality of data ensures more reliable economic assessments, fostering a conducive environment for sustainable energy investments.

Inaccuracies in energy demand and supply predictions can lead to grid instabilities with potential cascading effects. MVTSI, by providing more accurate imputations, contributes to mitigating grid instabilities. This is particularly crucial in the context of modern energy grids, where stability is essential for ensuring continuous and reliable electricity supply. The societal and environmental impacts of robust energy analytics are profound. Reliable energy system models, facilitated by accurate imputation, support the transition towards cleaner and more sustainable energy

practices. This aligns with global efforts to reduce carbon emissions and address climate change, emphasizing the broader societal and environmental relevance of MVTSI. In conclusion, the MVTSI model transcends its technical intricacies to play a pivotal role in shaping the future of energy analytics. By addressing the challenge of missing values in multivariate time series data, MVTSI contributes to enhanced decision-making, resource optimization, renewable energy integration, economic viability, and the overall stability of energy grids. The societal and environmental impacts underscore the transformative potential of this research, emphasizing its enduring significance in the evolving landscape of energy analytics.

The integration of neural networks, particularly Long Short-Term Memory (LSTM) networks, in the MVTSI model is a crucial aspect that elevates the imputation process to a new level of sophistication. Neural networks play a pivotal role in addressing the inherent challenges of missing values in multivariate time series data within the energy domain. LSTM networks are adept at capturing long-term dependencies in sequential data, making them well-suited for modeling temporal dynamics within multivariate time series. In the context of MVTSI, the LSTM-based model contributes by learning the intricate temporal relationships among variables in the presence of missing values. This capability enables the model to discern patterns and trends critical for accurate imputation.

The first stage of MVTSI, centered around the LSTM-based model, focuses on learning the distribution of missing values. Neural networks excel at capturing complex distributions, allowing the model to gain insights into the statistical characteristics of the missing data. This distribution learning facilitates the generation of synthetic missing values, enhancing the model's adaptability to the nuances of the multivariate time series. The attention-based neural network in the second stage of MVTSI introduces a critical element for handling cross-dimensional correlations. The attention mechanism enables the model to selectively emphasize relevant features and relationships across different dimensions of the multivariate data. This attention mechanism is instrumental in simultaneously capturing temporal dependencies within variables and cross-dimensional correlations, providing a holistic approach to imputation.

Neural networks, being highly flexible and adaptable, empower the MVTSI model to generalize well to diverse energy datasets. The ability to learn from data without relying on predefined assumptions allows the model to dynamically adjust to the specific characteristics of the energy data it encounters. This adaptability is crucial in real-world scenarios where energy data exhibits varying patterns and complexities. In contrast to classical imputation methods that often struggle with capturing intricate relationships and patterns in multivariate time series, neural networks offer a data-driven approach. The capacity to automatically extract features and dependencies from data makes neural networks particularly effective in scenarios where the underlying dynamics are complex and non-linear.

The incorporation of neural networks, as witnessed in the MVTSI model, represents a paradigm shift in imputation techniques. By leveraging the capacity of neural networks to learn from data, MVTSI not only overcomes the limitations of classical methods but also sets a new standard for accuracy and adaptability in multivariate time series imputation. As neural network-based imputation gains prominence, future research directions may explore advancements in model architectures, regularization techniques, and the fusion of different types of neural networks for improved performance. The interdisciplinary collaboration between data scientists and domain experts will continue to drive innovation in harnessing the power of neural networks for handling missing values in diverse domains, including energy analytics.

Derrick Hines, Philipp Bekemeyer (DLR German Aerospace Center, Lilienthalpl. Brunswick, Germany)

NEURAL NETWORK-DRIVEN EXPLORATION IN AIRCRAFT DESIGN

Abstract: *This study delves into the realm of computational fluid dynamics (CFD) simulations, emphasizing the crucial role of graph neural networks (GNNs) in advancing aerodynamic predictions for modern aircraft design. The exploration navigates through challenges posed by high-fidelity simulations, advocating for efficient surrogate methods. Methodologies, including proper orthogonal decomposition (POD) and multilayer perceptron, are dissected alongside a pioneering GNN approach. Test cases featuring the NLR7301 airfoil and the NASA Common Research Model underscore the GNN's potential in predicting surface pressure distributions across diverse operational parameters. The findings contribute insights into balancing accuracy and computational efficiency, paving the way for transformative impacts on aircraft design workflows.*

Анотація: *це дослідження заглиблюється в сферу моделювання обчислювальної гідродинаміки (CFD), підкреслюючи вирішальну роль графових нейронних мереж (GNN) у вдосконаленні аеродинамічних прогнозів для сучасного дизайну літаків. Дослідження проходить через виклики, пов'язані з моделюванням високої точності, захищаючи ефективні сурогатні методи. Методології, включно з правильним ортогональним розкладанням (POD) і багатошаровим перцептроном, розбираються разом із піонерським підходом GNN. Тестові випадки, що включають аеродинамічний профіль NLR7301 і Загальну дослідницьку модель NASA, підкреслюють потенціал GNN у передбаченні розподілу поверхневого тиску за різними робочими параметрами. Отримані результати допомагають зрозуміти баланс між точністю та обчислювальною ефективністю, прокладаючи шлях до трансформаційного впливу на робочі процеси проектування літаків.*

Computational fluid dynamics (CFD) simulations play a crucial role in modern aircraft design, offering highly accurate flow field predictions for in-depth insights into aircraft behavior. These simulations, typically solving the Reynolds-averaged Navier-Stokes (RANS) equations, are fundamental for assessing aerodynamic performance, structural loads, and handling qualities at various design stages. However, their high computational cost limits their feasibility in early design phases and time-sensitive environments, such as aerodynamic load estimation.

To address this challenge, surrogate methods are essential for efficiently predicting flow fields. Data-driven models, particularly proper orthogonal decomposition (POD) combined with interpolation methods, have gained attention. While POD is effective for linear phenomena, its limitations in capturing local nonlinearities, especially in transonic flow conditions, motivate exploration of alternative methods.

Recent research has turned to deep learning (DL) models for their ability to extract hierarchical data features. These DL models have shown success in predicting aerodynamic coefficients, surface pressure distributions, and unsteady forces. However, existing studies often focus on 2D airfoils with structured grids, whereas industrial applications involve more complex geometries with unstructured grids.

This paper introduces a graph neural network (GNN) approach for predicting distributed quantities on large-scale unstructured grids. The method is applied to predict surface pressure distributions for the NASA Common Research Model (NASA CRM) transport aircraft, featuring transonic flows with shocks and boundary layer separation.

The organization of the paper includes a task description and available geometric information, followed by a detailed presentation of data-driven methods, with a focus on GNN development. Two test cases, the NLR7301 airfoil and the NASA CRM configuration, are discussed, and the results are presented for evaluating prediction performance and generalization capabilities of the models.

In conclusion, this study demonstrates the application of a GNN approach for predicting surface pressure distributions on complex 3D aircraft configurations with large-scale unstructured grids, showcasing its potential for industrial use in aerodynamic simulations. The presented

comparison with other methods contributes to understanding their prediction performance and computational efficiency.

Continuing with the exploration of aerodynamic simulations, it's crucial to emphasize the ongoing challenges and advancements in the field. Despite the promising capabilities of data-driven models and deep learning techniques, several considerations merit attention.

One notable challenge lies in the extension of deep learning methods to handle more complex 3D aircraft configurations with unstructured grids. While existing studies have demonstrated success in predicting aerodynamic quantities for 2D airfoils, the transition to intricate geometries common in industrial applications demands further investigation. The effectiveness and scalability of deep learning approaches on unstructured grids with millions of points require careful scrutiny.

Additionally, the comparison of different methods, such as proper orthogonal decomposition (POD) coupled with interpolation, multilayer perceptron, and graph neural network (GNN) simulators, provides valuable insights into their prediction performance and computational efficiency. Understanding the trade-offs between accuracy and computational cost is essential for selecting the most suitable method based on the specific requirements of the aerodynamic simulation task.

Furthermore, the discussion on test cases, including the NLR7301 airfoil and the NASA CRM transport aircraft, underscores the importance of evaluating models in diverse scenarios. The ability to predict surface pressure distributions across various operational parameters is a critical aspect of industrial relevance. Assessing the generalization capabilities of models based on testing samples not seen during training and validation adds robustness to the findings.

As the aerospace industry continues to evolve, with an increasing focus on efficiency, sustainability, and innovation, the role of computational methods in aerodynamic design becomes even more pronounced. The integration of advanced modeling techniques, such as graph neural networks, into the design and optimization processes holds the potential to streamline and enhance the accuracy of aerodynamic simulations for next-generation aircraft.

In conclusion, this study contributes to the ongoing dialogue in aerodynamic simulations by presenting a novel approach using graph neural networks for predicting surface pressure distributions on complex 3D aircraft configurations. The findings underscore the importance of leveraging data-driven models and deep learning techniques while acknowledging the nuances and challenges associated with their application in industrial settings. Future research directions may involve refining these models, expanding to even more complex scenarios, and exploring hybrid approaches that combine the strengths of different methodologies.

Continuing the exploration of aerodynamic simulations and the role of computational methods in modern aircraft design, it's essential to delve into the broader implications and potential future directions in the field.

The adoption of advanced techniques, such as graph neural networks (GNNs), represents a paradigm shift in how aerodynamic simulations are approached. GNNs offer a promising avenue for handling complex 3D configurations with unstructured grids, addressing a long-standing challenge in the application of deep learning to real-world aerospace scenarios. As research in this area progresses, it becomes imperative to fine-tune these models to accommodate the intricacies of diverse geometries and flow conditions encountered in practical aircraft design.

Moreover, the discussion on different methodologies, including proper orthogonal decomposition (POD) coupled with interpolation and multilayer perceptron, adds nuance to the understanding of model selection. While deep learning methods exhibit remarkable capabilities, it's crucial to recognize scenarios where computationally efficient approaches like PODI may still find utility. Balancing accuracy with computational cost remains a pivotal consideration, especially in time-sensitive environments and early design stages.

The presented test cases, featuring the NLR7301 airfoil and the NASA CRM transport aircraft, serve as benchmarks for evaluating the models' performance. The emphasis on predicting surface pressure distributions across diverse operational parameters underscores the practical applicability of

these models in industrial settings. Future studies could expand these test cases to include even more challenging scenarios, reflecting the evolving demands of the aerospace industry.

Looking ahead, the integration of advanced computational methods into the design workflow holds the potential to revolutionize aircraft development. The continuous evolution of these techniques, driven by advancements in artificial intelligence and machine learning, aligns with the industry's goals of enhancing efficiency, reducing environmental impact, and pushing the boundaries of innovation.

In conclusion, the intersection of aerodynamics and computational methods is a dynamic and evolving space. The study's findings contribute valuable insights, but they also invite further exploration and refinement of models to meet the growing complexity of aerospace challenges. As researchers and industry practitioners collaborate, the synergy between cutting-edge computational methods and traditional aerodynamic principles will pave the way for transformative advancements in aircraft design and performance.

In summary, this extensive discussion has revolved around the pivotal role of computational methods, particularly focusing on graph neural networks (GNNs), in advancing aerodynamic simulations for modern aircraft design. The exploration encompassed the challenges of high-fidelity simulations, the need for efficient surrogate methods, and the evolving landscape of data-driven models.

The study presented a comprehensive comparison of methodologies, ranging from proper orthogonal decomposition (POD) coupled with interpolation to multilayer perceptron and, notably, the innovative GNN approach. Each method was scrutinized for its predictive performance, computational efficiency, and applicability to real-world scenarios.

Test cases featuring the NLR7301 airfoil and the NASA Common Research Model (CRM) transport aircraft provided practical contexts for evaluating the models' capabilities in predicting surface pressure distributions across diverse operational parameters. The results not only showcased the potential of GNNs in handling complex 3D configurations with unstructured grids but also underscored the ongoing need to balance accuracy with computational cost.

Looking forward, the integration of advanced computational methods into the aircraft design workflow holds promise for transformative impacts. The continuous evolution of these techniques, driven by artificial intelligence and machine learning advancements, aligns with industry goals of enhancing efficiency, reducing environmental impact, and fostering innovation.

In conclusion, this exploration at the intersection of aerodynamics and computational methods signifies a dynamic and evolving space. The study's findings contribute valuable insights and set the stage for future research aimed at refining models to meet the growing complexity of aerospace challenges. Collaborative efforts between researchers and industry practitioners are poised to drive transformative advancements in aircraft design and performance.

Enrico Collini, Paolo Nesi, Gianni Pantaleo (*Distributed Systems and Internet Technologies Lab, Department of Information Engineering, University of Florence, Florence, Italy*)

DEEP LEARNING APPROACHES IN TOURISM MANAGEMENT

Abstract: *This research applies neural network technologies to analyze and forecast tourism aspects, focusing on social media, especially Twitter, for sentiment analysis. Utilizing deep learning algorithms, including RNNs, TCNs, and TFT, the study assesses sentiment and predicts reputation. The research also explores forecasting visitor numbers, considering methods like RF, XGBoost, TCN, and TFT. Emphasis is placed on analyzing causal relationships using reinforcement learning and causal learning algorithms. Addressing diverse data sources, the study underscores deep learning's efficacy in multimodal analysis of text, images, and other information in tourism management. The conclusion highlights the contribution of various deep learning methods to innovative tourism management technologies, offering a promising avenue for destination analysis and optimization.*

Анотація: *Це дослідження застосовує технології нейронних мереж для аналізу та прогнозування аспектів туризму, зосереджуючись на соціальних медіа, особливо Twitter, для аналізу настроїв. Використовуючи алгоритми глибокого навчання, включаючи RNN, TCN і TFT, дослідження оцінює настрої та прогнозує репутацію. Дослідження також вивчає прогнозування кількості відвідувачів, враховуючи такі методи, як RF, XGBoost, TCN і TFT. Акцент робиться на аналізі причинно-наслідкових зв'язків за допомогою навчання з підкріпленням і алгоритмів причинно-наслідкового навчання. Звертаючись до різноманітних джерел даних, дослідження підкреслює ефективність глибокого навчання в мультимодальному аналізі тексту, зображень та іншої інформації в управлінні туризмом. Висновок підкреслює внесок різних методів глибокого навчання в інноваційні технології управління туризмом, пропонуючи багатообіцяючий шлях для аналізу та оптимізації напрямків.*

Tourism plays a crucial role in the development of historical and cultural cities. In the context of Smart Cities, leveraging data from various sources in the tourism domain is essential for monitoring and predicting different indicators related to tourist locations and attractions. This paper introduces a framework that utilizes social media and big data for forecasting online reputation and the presence of tourists at attractions. Machine learning, deep learning, causality assessment, and explainable Artificial Intelligence techniques are employed to identify relevant variables for each prediction. The proposed framework includes a multilingual sentiment analysis tool for social media data based on transformers, comparing data sources such as Trip Advisor and Twitter. Causality analysis evaluates the temporal impact of social media posts and other factors on the number of tourist presences. Developed within the Herit-Data project, funded by the European Commission, and based on the Snap4City infrastructure, the paper presents solutions for major European touristic locations, focusing on Florence, Italy, and Pont du Gard, France.

Introduction: Tourism, as an experiential phenomenon, relies heavily on communication. Social media serves as an ideal instrument for implementing word-of-mouth communication, gaining significance in the tourism field for its role in propagating information about attractions and cultural heritage sites. The success of touristic destinations is intricately linked to their reputation, with platforms like Trip Advisor collecting user comments and scores as explicit reputation assessments. This paper addresses the growing relevance of social media content in the tourism sector and the need for effective analysis to understand its impact on destination choices.

State of the Art and Related Works: The emphasis on providing predictive tools for decision-makers in the tourism industry is growing, with a focus on online reputation and the number of presences at touristic attractions. Predictive models, driven by data-driven approaches like machine learning and data analytics, aim to empower decision-makers with the ability to anticipate changes in reputation and forecast visitor presences.

Reputation Assessment: Numerous research papers have explored Tourism Reputation Assessment, particularly on online platforms. Social media content on tourism platforms, such as Trip Advisor, influences the perception of attractions. This paper highlights the importance of

understanding the relationships between different platforms, such as Twitter and Trip Advisor, for reputation assessment.

Reputation Forecasting: Online Reputation Management systems play a crucial role in assessing and predicting the online reputation of touristic attractions. The complexity of reputation prediction, akin to predicting stock values, requires innovative solutions. This paper discusses the use of Trip Advisor review data and explores the changing nature of reputation over time as a time-series data.

Tourist Presence Forecasting: The prediction of the number of people at touristic attractions is a valuable application within the tourism domain. Various data sources, including Wi-Fi sniffing, laser counting, thermal cameras, and Twitter data, contribute to predicting tourist presences. The paper delves into predictive models and methodologies, such as seasonal autoregressive integrated moving average (SARIMA), for estimating and forecasting the number of tourists.

In summary, the proposed framework integrates social media and big data to forecast online reputation and tourist attraction presences, offering insights into the interplay between reputation, social media, and visitor numbers. Developed within the Herit-Data project, the solutions presented for specific European locations contribute to the advancement of predictive tools in the tourism sector.

In the conducted research, much attention is focused on the analysis of feedback in social networks, such as Twitter, in order to evaluate the orientation of tweets, determine the quality of services, form recommendations for visiting places, and predict the online reputation of tourist destinations. Neural network technologies such as sentiment analysis algorithms are used to achieve these goals.

These methods are based on deep learning architectures such as recurrent neural networks (RNNs), temporal convolutional networks (TCNs) or temporal transformers (TFTs). They are designed to process large volumes of textual information, analyze sentiments and use them to predict reputation.

In addition to the analysis of social networks, the forecasting of the number of visitors is also investigated. Different methods are used here, such as deep learning models for working with time series. This can include ensemble methods such as random forest (RF) and gradient boosting (XGBoost), or specialized deep learning models for time series forecasting such as TCN or TFT.

A study of causal relationships between reputation, presence and posts in social networks is also proposed. Reinforcement learning methods or causal learning algorithms can be used for such analysis.

Given the diversity of data sources, there is a need for multimodal analysis that encompasses text, images, and other data. Deep learning can be an effective tool for processing various types of information.

This study reflects the practical application of deep learning methods in tourism management, where a large amount of processed data is taken into account and forecasting models are developed. In conclusion, it can be noted that the work contributes to the development of innovative technologies in the field of tourism management and can have a significant impact on the development of the industry.

1. Using deep learning for personalized recommendations: One potential perspective is the application of deep learning to develop personalized recommendations for tourists. Models that take into account personal preferences and visit history can significantly improve service quality and guarantee tourist satisfaction.

2. Taking spatio-temporal features into account: Extending the analysis to spatio-temporal characteristics can contribute to more effective forecasting of the popularity of different places in different periods. It is possible to consider network architectures capable of simulating the dynamics of visits depending on the season, festive events and other factors.

3. Improving the analysis of feedback and emotions: Developing more accurate models for analyzing feedback and emotions can contribute to the improvement of recommendations and the

formation of a positive image of tourist destinations. Considering architectures that more accurately recognize sentiment and context in text is an urgent task.

4. Application of reinforcement learning in the management of tourist facilities: Reinforcement learning models can be used to optimize the management and organization of tourist facilities. Algorithms that learn from feedback and visit dynamics can significantly improve service and increase tourist satisfaction.

5. Impact of technology on the development of tourism: It is necessary to conduct research on the impact of the use of advanced technologies, such as artificial intelligence and neural networks, on the development of tourism in general. How can these technologies increase the competitiveness of tourist destinations and ensure the sustainability of the industry?

These areas of research in the field of using artificial intelligence in tourism have great potential for improving service and increasing the satisfaction of tourists. The development of models based on deep learning and reinforcement learning can open new opportunities for innovative management of tourism resources.

Farhan A. Alenizi, Amir Masoud Rahmani (*Future Technology Research Center, National Yunlin University of Science and Technology, Taiwan*)

AI IN INDUSTRY 4.0: COMPREHENSIVE DESIGN AND QUALITY MANAGEMENT

Abstract: *This research explores the integration of Artificial Intelligence (AI) in Industry 4.0, addressing cost reduction, resource management, and production quality. Industry 4.0 principles guide the study, emphasizing AI's role in achieving industrial goals. Machine learning aids in data analysis, contributing to effective issue resolution. The survey classifies AI approaches, tools, and challenges, providing insights into their impact on societal and economic technologies. Despite limitations, the research offers a complete taxonomy, contributing to further advancements in AI in Industry 4.0.*

Анотація: *це дослідження досліджує інтеграцію штучного інтелекту (ШІ) в Індустрію 4.0, спрямовану на зниження витрат, управління ресурсами та якість виробництва. Дослідження ґрунтується на принципах Industry 4.0, які підкреслюють роль штучного інтелекту в досягненні промислових цілей. Машинне навчання допомагає в аналізі даних, сприяючи ефективному вирішенню проблем. Опитування класифікує підходи, інструменти та виклики штучного інтелекту, надаючи розуміння їх впливу на соціальні та економічні технології. Незважаючи на обмеження, дослідження пропонує повну таксономію, сприяючи подальшому розвитку ШІ в Індустрії 4.0.*

In recent years, industries and technology have advanced, focusing on cost reduction, resource management, and enhancing production efficiency and product quality. The principles of Industry 4.0 address these objectives. The advent of Artificial Intelligence (AI) aligns with Industry 4.0, offering significant improvements in production costs and addressing product-related challenges.

AI encompasses various applications, including security, automatic configuration, planning, control and monitoring, prediction and diagnosis, and decision-making based on information and algorithms. These applications aim to achieve industrial goals, such as replicating product processes and monitoring factory performance. Machine learning analyzes and predicts data, enhancing production efficiency, meeting high customer standards, and improving product quality. The data obtained serves as a reference for researchers and manufacturers in their transition to Industry 4.0.

The integration of AI brings profound changes to industrial operations, including improved maintenance and control, process monitoring, optimized production processes, and reduced complexity. AI approaches offer advantages such as managing different branches using smart systems, continuous system analysis, simulation for optimization, and quick pattern recognition for improved system management and reduced costs.

This paper conducts an in-depth survey of technologies and tools for AI approaches in Industry 4.0, emphasizing AI's essential role in the fourth industrial revolution. While literature explores the impact of this revolution on society, the specific contributions of AI to poverty reduction, infrastructure development, and sustainable goals remain under-researched. The 21st century has witnessed substantial AI investment, transforming it into a crucial factor in societal and economic technologies.

Drawing on AI advantages, this paper contributes a comprehensive view, providing a taxonomy of AI approaches in Industry 4.0. It classifies AI and Industry 4.0 integration approaches, tools, assessment methodologies, and criteria, offering an overview of practical techniques, challenges, and emerging issues. The research also addresses the holistic understanding of how AI contributes to the 4th industrial revolution's main goals.

The paper's contributions include delivering a concise tutorial and an integration architectural view for AI and Industry 4.0, designing a taxonomy for AI in Industry 4.0, and providing a comprehensive overview of practical techniques, challenges, and emerging issues related to AI and Industry 4.0. The research methodology involves systematic mapping study methods, presenting existing AI solutions in Industry 4.0, culminating in a taxonomy.

Previous studies explored AI solutions in Industry 4.0, focusing on specific domains or challenges. Research concentrated on data management using machine learning and deep learning, emphasizing AI's significance in handling various product life cycle levels. Another contribution delved into quality management across product, service, and production cycle stages, considering AI and related technologies from the perspective of customer satisfaction and product performance improvement.

Research Methodology outlines the steps in employing Systematic Mapping Study methods to survey AI approaches in Industry 4.0. It presents a taxonomy of AI approaches in Industry 4.0, examining the emergence of Industry 4.0 and its various aspects. The research categorizes selected articles based on AI's role in product quality management and its implementation in the industry's complete production cycle.

The study acknowledges limitations, including the exclusion of certain studies, impacting result interpretation. Despite these limitations, the research provides a complete taxonomy of AI approaches in Industry 4.0, contributing to further developments.

This paper systematically explores AI methods and solutions in the context of Industry 4.0, classifying AI approaches into two categories: quality management in production and product life cycle management. Factors such as cost, performance, time, reliability, security, data, and service accuracy are crucial in quality management, while the product lifecycle involves considerations such as data transmission, load balancing, and data reliability. The research concludes by summarizing the comprehensive design for developing artificial intelligence in the industry, emphasizing its potential impact on the goals of the 4th industrial revolution.

The research opens avenues for future exploration in the realm of AI and Industry 4.0. Notably, incorporating an in-service performance measure, such as stiffness or crashworthiness, alongside manufacturability considerations could enhance the platform's capabilities for more comprehensive design optimization. Developing a surrogate model that accurately predicts these performance measures would be a crucial step in this direction.

Additionally, the optimization of other variables influencing the stamping process, such as material characteristics, blank shape, or processing parameters, presents an opportunity for further refinement. Integrating these variables into the optimization platform could provide valuable insights for stamping process setup and contribute to more informed decision-making.

Considering measures of uncertainty in the surrogate model's predictions could enhance the robustness of the platform's results. This would enable more confident decision-making and contribute to the overall reliability of the optimization process.

Developing a user-friendly graphical user interface (GUI) for the platform could simplify its use and facilitate broader adoption in industry applications. Removing the need for end-users to grasp advanced machine learning concepts could streamline the interface and make the platform more accessible to a wider audience.

In conclusion, these suggested future directions aim to enhance the utility of the proposed platform, making it more versatile, robust, and user-friendly, ultimately contributing to its broader adoption in both industry and academia.

Moreover, the integration of advanced neural network architectures holds promise for further advancing the capabilities of the optimization platform. Exploring state-of-the-art architectures, such as transformer-based models or novel recurrent neural network structures, could lead to improvements in capturing complex relationships within sheet stamping geometries.

The use of attention mechanisms within neural networks could enhance the model's ability to focus on critical features and regions during the optimization process. This could lead to more targeted and efficient geometry updates, addressing specific challenges or constraints in a more nuanced manner.

Incorporating transfer learning techniques into the neural network models could leverage pre-trained models on related tasks, potentially accelerating the training process and improving the overall

performance of the optimization platform. This approach could be particularly beneficial when dealing with limited data scenarios.

Furthermore, exploring the integration of reinforcement learning principles could open up new avenues for adaptive and autonomous optimization. Reinforcement learning algorithms could enable the platform to learn from interactions with the optimization environment, adapting its strategies over time for improved efficacy.

The combination of neural networks with meta-learning approaches could enable the platform to adapt quickly to new stamping geometries or manufacturing constraints, reducing the need for extensive retraining when applied to diverse scenarios.

In conclusion, the incorporation of advanced neural network architectures, attention mechanisms, transfer learning, reinforcement learning, and meta-learning techniques could significantly enhance the optimization platform's capabilities, making it more adaptive, efficient, and versatile in addressing a wide range of sheet stamping challenges.

This study examines artificial intelligence methods and solutions in the context of Industry 4.0, classifying them into two areas: production quality management and product life cycle management. Important factors in quality management include cost, performance, time, reliability, security, data, and service accuracy, while aspects such as data transfer, load balancing, and data reliability are considered in the product lifecycle. The study highlights a comprehensive design for the development of artificial intelligence in industry, emphasizing its potential impact on the goals of the Fourth Industrial Revolution.

H. Ruppert, A. Krug (*Faculty of Electrical Engineering, Heilbronn University of Applied Sciences Heilbronn, Germany*)

DESIGNING A NEURAL NETWORK BASED ON ADAPTABILITY AND NOISE SUPPRESSION

Abstract: *This research introduces an innovative approach to neural network design based on inflection point analysis, automating the selection of the minimum required neurons. The methodology demonstrates adaptability to diverse system complexities, mitigating overfitting without additional regularization. Emphasizing noise reduction in data, the designed neural network acts as a filter, advantageous for real-world applications with noisy signals. The study explores practical implications, highlighting potential in system identification and offering insights for future integration into ensemble learning or deep neural network frameworks. Overall, this methodology showcases versatility and effectiveness in addressing challenges in neural network design and holds promise for practical implementation in various domains.*

Анотація: *У цьому дослідженні представлено інноваційний підхід до розробки нейронних мереж, заснований на аналізі точок перегину, що автоматизує вибір мінімально необхідних нейронів. Методологія демонструє адаптивність до різноманітних складних систем, пом'якшуючи надмірне оснащення без додаткової регуляризації. Підкреслюючи зменшення шуму в даних, розроблена нейронна мережа діє як фільтр, що є перевагою для реальних додатків із зашумленими сигналами. Дослідження досліджує практичні наслідки, висвітлюючи потенціал ідентифікації системи та пропонуючи ідеї для майбутньої інтеграції в ансамблеве навчання або структуру глибокої нейронної мережі. Загалом ця методологія демонструє універсальність та ефективність у вирішенні проблем у розробці нейронних мереж і є перспективною для практичного впровадження в різних сферах.*

The presented study introduces a novel approach for designing neural networks (NN) to approximate data sets. The methodology leverages inflection point analysis to determine the number of neurons in a feedforward neural network with one hidden layer, utilizing a hyperbolic tangent activation function. The distinctive advantage of this method lies in its ability to automatically select the smallest number of neurons without the need for regularization.

The simulation experiments conducted demonstrate the effectiveness of this approach in defining the minimal number of neurons required to approximate both linear and nonlinear functions with minimal error. This methodology proves applicable to a broad spectrum of linear and nonlinear systems, assuming the function possesses derivatives in the desired approximation domain.

It is noteworthy that, when employing the hyperbolic tangent as the activation function, a feedforward neural network with a single neuron in the hidden layer can adequately approximate the time behavior of nonoscillating system responses. Such systems typically represent solutions to linear and nonlinear first-order differential equations. However, for the approximation of linear or nonlinear differential equations of higher order, particularly those exhibiting oscillatory behavior, two or more neurons become necessary.

It is crucial to recognize that the choice of activation functions other than the hyperbolic tangent may yield diverse outcomes. The study concludes by outlining potential avenues for future research. Specifically, further investigation is needed to determine whether this method is applicable to a specific class of functions, especially those governed by differential equations with exponential functions as solutions. Additionally, the feasibility of automating the creation of neural networks based on this method for deployment as regression models in soft sensors requires validation. Subsequent works will explore the applicability of the proposed method to deep neural networks with multiple hidden layers.

Continuing the exploration of this methodology, it is imperative to highlight the significance of inflection point analysis in determining the optimal number of neurons. The approach demonstrated its robustness across diverse scenarios, showcasing its potential to handle both linear

and nonlinear systems. This adaptability stems from the automatic selection of the minimum required neurons, obviating the need for explicit regularization.

The simulations conducted in this study provided valuable insights into the performance of the neural network in approximating various functions. Notably, the utilization of hyperbolic tangent activation function in a single-neuron hidden layer proved effective for capturing the time behavior of nonoscillating system responses. This aligns with the solutions typically found in linear and nonlinear first-order differential equations.

However, the study acknowledges the inherent complexity of systems displaying oscillatory behavior or higher-order dynamics. In such cases, the methodology suggests the incorporation of additional neurons to achieve accurate approximations. This nuanced approach recognizes the diversity in system responses and tailors the neural network architecture accordingly.

A key takeaway from the study is the consideration of alternative activation functions. The choice of hyperbolic tangent was instrumental in the specific context explored, but the study acknowledges that different activation functions may yield different results. This insight underscores the importance of selecting an appropriate activation function based on the characteristics of the system under consideration.

Looking ahead, future research avenues include a more nuanced examination of differential equations with exponential functions as solutions. The method's applicability to a broader range of nonlinear differential equations warrants investigation. Additionally, the prospect of automating the creation of neural networks based on this methodology opens doors to practical applications, such as employing these networks as regression models in soft sensors.

Expanding on the methodology's implications, it's essential to delve into its potential applications in practical domains. The automatic selection of the minimum number of neurons, guided by inflection point analysis, not only streamlines the neural network design process but also addresses a common challenge in machine learning—overfitting. By inherently adapting to the complexity of the system under study, this method provides a tailored and efficient solution without the need for additional regularization techniques.

The study's emphasis on handling noisy data introduces a valuable dimension to the methodology's utility. The simulation results illustrate that the neural network, designed with the minimum required neurons, acts as a filter, effectively suppressing unwanted noise during system response approximation. This characteristic is particularly advantageous in real-world scenarios where data are often corrupted by various sources of interference.

Moreover, the study's acknowledgment of potential applications in system identification aligns with current trends in industry and technology. The ability of the proposed neural network to discern the true function in a noisy signal, given knowledge of the system response characteristics, positions it as a valuable tool for practical implementations. In environments where measured values are inherently noisy, such as in industrial processes, this method offers a robust means of approximating system behavior.

Looking forward, the research community could explore the integration of this methodology into broader frameworks, such as ensemble learning or hybrid models combining neural networks with other machine learning approaches. Additionally, investigating the scalability of this method to deep neural networks with multiple hidden layers would be a pertinent avenue for future exploration.

In summary, the presented methodology not only contributes to the field of neural network design but also holds promise for real-world applications. Its adaptability to noisy data, potential for system identification, and simplicity in implementation make it a compelling approach in the broader landscape of machine learning and computational modeling. As research progresses, the methodology's versatility and applicability across diverse domains are likely to become even more evident.

Hamid Reza Attar, Nan Li (Dyson School of Design Engineering, Imperial College London, UK)

A DEEP LEARNING PLATFORM FOR SHEET STAMPING GEOMETRY OPTIMIZATION

Abstract: *The research introduces a novel deep-learning-based optimization platform for sheet stamping geometries, addressing challenges in vehicle light weighting. By combining CNN-based surrogate models and implicit neural representations, the platform allows early-stage design exploration with a focus on manufacturing constraints. Demonstrated through case studies on Corners and Bulkheads geometries, the platform exhibits significant potential in reducing localized thinning, thereby enhancing manufacturing performance. While facing limitations related to initial geometry constraints, the platform opens avenues for future research, presenting a promising tool for improving efficiency, reducing development time, and cutting costs in the sheet metal stamping industry.*

Анотація: *Дослідження представляє нову платформу оптимізації на основі глибокого навчання для геометрії листового штампування, яка вирішує проблеми з легкістю транспортного засобу. Поєднуючи сурогатні моделі на основі CNN і неявні нейронні представлення, платформа дозволяє досліджувати дизайн на ранній стадії з акцентом на виробничі обмеження. Продемонстровано на тематичних дослідженнях геометрії кутів і перегородок, платформа демонструє значний потенціал у зменшенні локального стоншення, тим самим підвищуючи продуктивність виробництва. Незважаючи на обмеження, пов'язані з початковими геометричними обмеженнями, платформа відкриває шляхи для майбутніх досліджень, представляючи перспективний інструмент для підвищення ефективності, скорочення часу розробки та скорочення витрат у галузі штампування листового металу.*

The study addresses challenges in vehicle light weighting by introducing a deep-learning-based platform for optimizing sheet stamping geometries. Current stamping processes are promising but unfamiliar to industrial designers, leading to increased development time. The proposed platform combines CNN-based surrogate models and implicit neural representations to optimize stamping geometries, allowing for expressive topology changes. The platform features a novel backward pass enabling gradient-based optimization, considering manufacturing constraints. It extends CNN-based surrogate models and implicit neural representations to complex sheet stamping geometries, including those from multiple parameterization schemes. The platform defines objective and constraint functions based on predicted post-stamped physical fields. The study validates the platform through case studies, demonstrating its effectiveness in achieving expressive geometry changes and reducing severe localized thinning. Despite limitations related to initial geometry constraints and machine learning adoption, the platform presents significant potential for improving manufacturing performance, reducing development time, and lowering production costs in the sheet metal stamping industry."

The research presents a deep-learning-based optimization platform for sheet stamping geometries, aiming to address challenges in vehicle light weighting. The platform integrates CNN-based surrogate models and implicit neural representations, allowing for expressive topology changes and optimizing geometries based on manufacturing constraints. The study extends the application of these models to complex sheet stamping geometries and proposes a novel backward pass for gradient-based optimization. Validation through case studies demonstrates the platform's capability to achieve significant reductions in localized thinning and make expressive geometry changes, providing potential benefits for manufacturers in terms of improved manufacturing performance and reduced development time.

Despite notable achievements, the platform has limitations related to the requirement for initial geometries within the model training distribution, sensitivity to initial geometry, and the learning curve associated with machine learning adoption in the stamping industry. The study opens avenues for future research, suggesting the incorporation of in-service performance measures, optimization of additional variables affecting the stamping process, inclusion of uncertainty measures

in surrogate model predictions, and the development of a user-friendly interface. These directions could enhance the platform's utility and foster its adoption in industry applications.

In conclusion, the proposed optimization platform has the potential to contribute significantly to the sheet metal stamping industry. By enabling non-trivial geometric changes to suit new stamping processes, the platform offers opportunities for improved manufacturing performance, reduced development time, and cost savings. Furthermore, the platform's ability to optimize complex geometries with small-scale features positions it as a valuable tool for addressing intricate design challenges. Overall, the research contributes to innovation in the sheet metal stamping industry and holds promise for further growth in both industry and academia.

The deep-learning-based optimization platform for sheet stamping geometries, as presented in this research, signifies a crucial step toward addressing challenges in the domain of vehicle light weighting. By integrating CNN-based surrogate models and implicit neural representations, the platform enables the exploration of expressive topology changes in the early stages of component design, optimizing geometries with a focus on manufacturing constraints.

The study's notable contributions include the extension of CNN-based surrogate models to intricate sheet stamping geometries and the introduction of a unique backward pass for gradient-based optimization. Through case studies involving Corners and Bulkheads geometries, the platform showcased its ability to significantly reduce localized thinning, demonstrating its potential impact on improving manufacturing performance. The platform's capability to handle complex geometries with small-scale features suggests its suitability for addressing design challenges in various components.

Despite its achievements, the platform faces limitations related to the need for initial geometries within the model training distribution and potential sensitivity to the initial geometry. The learning curve associated with machine learning adoption in the stamping industry adds another layer of consideration. The research wisely outlines future research avenues, such as incorporating in-service performance measures, optimizing additional variables affecting the stamping process, and enhancing uncertainty measures in surrogate model predictions. The proposed development of a user-friendly interface further emphasizes the practical applicability of the platform.

In summary, this research not only introduces a powerful optimization tool for sheet stamping geometries but also lays the groundwork for further advancements. The potential benefits for manufacturers in terms of improved efficiency, reduced development time, and cost savings underscore the platform's significance in driving innovation in the sheet metal stamping industry. As the platform evolves, it has the potential to become a key player in shaping the future of vehicle light weighting and design optimization.

K.L. Keung, C.K.M. Lee (*Department of Industrial and Systems Engineering, The Hong Kong Polytechnic University, China*)

INTEGRATING CPRMFS FOR TOOLS STORAGE IN SMART MANUFACTURING

Abstract: *This study delves into the realm of smart manufacturing, exploring the integration of a Cyber-Physical Robotic Mobile Fulfillment System (CPRMFS) for tools storage in conjunction with multi-robots. The proposed system addresses the need for reducing human errors and enhancing operational efficiency within smart manufacturing environments. The research leverages contemporary technologies such as Cyber-Physical Production Systems (CPPS), the Internet of Things (IoT), and 3D printing, showcasing the adaptability of the CPRMFS across diverse manufacturing scenarios. The core focus lies in the simulation of decentralized multi-robot path planning, providing valuable insights into the system's performance under various instances and classification algorithms. The study concludes by underscoring the managerial implications, practical applicability, and potential avenues for future research, positioning the CPRMFS as a promising solution in the dynamic landscape of Industry 4.0.*

Анотація: *Це дослідження заглиблюється в сферу інтелектуального виробництва, досліджуючи інтеграцію Cyber-Physical Robotic Mobile Fulfillment System (CPRMFS) для зберігання інструментів у поєднанні з декількома роботами. Запропонована система спрямована на зменшення людських помилок і підвищення ефективності роботи в інтелектуальних виробничих середовищах. Дослідження використовує сучасні технології, такі як Cyber-Physical Production Systems (CPPS), Інтернет речей (IoT) і 3D-друк, демонструючи адаптивність CPRMFS до різноманітних сценаріїв виробництва. Основна увага зосереджена на моделюванні децентралізованого планування шляху з декількома роботами, надаючи цінну інформацію про продуктивність системи в різних випадках і алгоритмах класифікації. Дослідження завершується підкресленням управлінських наслідків, практичної застосовності та потенційних шляхів для майбутніх досліджень, позиціонуючи CPRMFS як багатобічне рішення в динамічному ландшафті Індустрії 4.0.*

Smart manufacturing, defined as the optimal utilization of labor, material, and energy for customized, high-quality products, is marked by characteristics such as context awareness, modularity, and interoperability. Cloud manufacturing, introduced in 2010, involves sharing manufacturing capabilities on a cloud platform for sustainable and robust routes. Cyber-physical production systems (CPPS) encompass layers like manufacturing resources, virtual services, global services, and applications. Digital twins on shop-floors enable real-time performance monitoring.

The integration of smart manufacturing with cloud, digital twins, and the Internet of Things (IoT) is a burgeoning trend. Industrial IoT (IIoT) integration under the cyber layer facilitates a plug-and-play gateway solution for data acquisition, communication, and storage. Machine learning, specifically deep transfer learning, proves effective for accurate process recognition. Robotics plays a significant role in reducing human involvement and enhancing operational efficiency. Robotic-assisted warehouse systems, like the Robotic Mobile Fulfillment System (RMFS), contribute to efficient material transfer.

The extension of robotic-assisted methods into manufacturing involves the Cyber-Physical Robotic Mobile Fulfillment System (CPRMFS). This system integrates mobile robots for material storage and retrieval, enhancing overall manufacturing efficiency. Additive manufacturing, incorporating 3D printing and robotic-based moving systems, further improves manufacturing effectiveness. The proposed CPRMFS, combining CPPS and IoT, offers a decentralized multi-robot path planning solution for smart manufacturing.

The article reviews the state-of-the-art in CPPS, presents the CPRMFS architecture, and introduces a simulation for decentralized multi-robot path planning. The study aims to reduce human errors and involvement in smart manufacturing, providing insights for practical applications.

The proposed CPRMFS architecture for smart manufacturing combines the capabilities of a Cyber-Physical Production System (CPPS) with the Internet of Things (IoT) to enhance tool storage efficiency. The system leverages digitalization, servitization, and mobile robots to reduce human

labor involvement and errors. The integration of a decentralized multi-robot path planning solution further optimizes manufacturing operations.

The simulation architecture demonstrates the connectivity of an IoT-enabled smart manufacturing system within the CPPS environment. This setup allows seamless storage and transfer of operational data to the cloud-based CPPS. The digitalization of 3D printing modules, mobile robots, and the Robotic Mobile Fulfillment System (RMFS) modules underlines the potential for improved efficiency and reduced human intervention in manufacturing processes.

In conclusion, the study introduces a novel approach, CPRMFS, to address contemporary challenges in smart manufacturing. By incorporating cutting-edge technologies such as IoT, CPPS, and decentralized multi-robot path planning, the proposed system offers a comprehensive solution for tool storage and retrieval. The findings have implications for industry practitioners seeking to enhance their manufacturing processes through the adoption of advanced robotic-assisted systems.

Furthermore, the study highlights the significance of the proposed CPRMFS in the context of Industry 4.0 and the ongoing evolution of manufacturing paradigms. The incorporation of 3D printing as a case study exemplifies the adaptability of the system to diverse manufacturing scenarios.

The simulation results, focusing on decentralized multi-robot path planning, provide valuable insights into the system's performance under various instances and classification algorithms. These results contribute to a deeper understanding of the practical applicability and potential challenges associated with the proposed CPRMFS architecture.

The research concludes by emphasizing the managerial implications and practical applicability of the proposed system. The comprehensive approach to tool storage, assisted by multi-robots and advanced technologies, positions CPRMFS as a promising solution for reducing errors and improving overall efficiency in smart manufacturing environments. The study also suggests avenues for future research to further refine and expand the proposed framework, keeping pace with the dynamic landscape of smart manufacturing.

COMPARISON OF TECHNOLOGIES FOR FULL CYCLE DEVELOPMENT OF MACHINE LEARNING MODELS

Abstract: Several approaches can be employed for the implementation of the full cycle development of machine learning models. Each of these approaches is supported by numerous technologies facilitating its execution. The primary challenge associated with approaches such as AutoML and similar ones lies in the fact that machine learning model developers often find themselves engaged in various configurations and manual actions, including data extraction and application deployment. Embracing the MLOps approach, the entire full cycle flow is segmented into stages, each of which can be automated. The primary drawbacks of the AutoML approach compared to the MLOps approach have been examined. The automation of the operational aspects enables data scientists to focus solely on model development, leading to both an acceleration of the development process and an enhancement of model quality through automated monitoring and overall system continuity. The paper introduces a set of technologies that facilitate the full-cycle development of machine learning models. Additionally, the convenience of utilizing these technologies is discussed, taking into account the specific tasks at hand. Numerous technologies exist to implement such a process, with AWS SageMaker standing out as one of the most convenient technologies, offering end-to-end capabilities within a single cloud service - from development to deployment.

Анотація: Для реалізації повного циклу розробки моделей машинного навчання можна використовувати декілька підходів. Кожен з цих підходів підтримується численними технологіями, які полегшують його виконання. Основний виклик, пов'язаний із підходами, такими як AutoML і подібні, полягає в тому, що розробники моделей машинного навчання часто знаходяться в роботі з різноманітними конфігураціями та ручними діями, включаючи вилучення даних та розгортання додатків. Застосування підходу MLOps дозволяє розбити весь цикл роботи на етапи, кожен з яких може бути автоматизований. В роботі були розглянуті основні недоліки підходу AutoML порівняно із підходом MLOps. Автоматизація операційних аспектів дозволяє вченим з даних зосередитися виключно на розробці моделі, що призводить як до прискорення процесу розробки, так і до покращення якості моделі завдяки автоматизованому моніторингу та загальній неперервності системи, над якою працює. У роботі представлено набір технологій, які полегшують повноциклову розробку моделей машинного навчання. Додатково розглянуто зручність їх використання в залежності від конкретних завдань. Існує багато технологій для реалізації такого процесу, причому AWS SageMaker виділяється серед найзручніших технологій, що пропонує повний спектр можливостей в рамках одного хмарного сервісу - від розробки до розгортання.

Introduction.

The traditional approach to developing a neural network typically involves a developer working on a local computer, utilizing programming languages (for example, Python) and additional packages (for example, TensorFlow) to design the neural network. This process is usually preceded by data collection, with the data then being split into training and testing sets. Subsequently, the neural network undergoes training and testing, yielding certain results. Traditionally, the outcomes of a specialist's work were the actual results — a form of classification or other data-related work, enabling analytical conclusions. In today's context, computational capabilities enable a more comprehensive approach to implementing neural networks, making them part of a larger data pipeline, possibly consisting of multiple such neural networks. With the advancement of web technologies, neural networks often take the form of web services integrated into a distributed system that utilizes another web service. Examples of this can be observed widely, with the most common being recommendation systems based on neural networks, utilized by almost all well-known e-commerce platforms or other internet services involved in content ranking.

When it comes to industrial applications of neural networks, it becomes evident that merely running them on a local computer is not a rational use of resources. As demand for this software product increases, engineers face challenges related to automation, deployment, and overall support of machine learning models, particularly neural networks. This mirrors the evolution seen in other software products — desktop, mobile, and web applications — written in various programming

languages. There are several paradigms that enable the establishment of a complete cycle for the development of machine learning models, and one of them is AutoML [1].

AutoML is a set of technologies and methods aimed at automating the process of developing machine learning models. Thanks to this technology, model development becomes more accessible, and the development process is significantly accelerated. AutoML allows for the automatic selection of models, hyperparameter tuning, and model evaluation. Such automation indeed speeds up the development process. However, this approach has several drawbacks - the automation of development itself leads to a limited choice of algorithms, complexities in integration with other systems, and possible errors in data processing. These drawbacks are evident for any approach that makes development more accessible and consequently less flexible. Moreover, the main problem with AutoML is in other thing. Although automation and efficiency are among its main selling points, this process still requires a surprising level of human involvement [2].

Another approach to machine learning model development is known as MLOps [3]. This approach encompasses the entire lifecycle of models, including development, testing, deployment, monitoring, and model updates. Its primary goal is to ensure efficient deployment, scalability, and management of machine learning models in production with minimal risks and maximum efficiency. Unlike AutoML, MLOps does not involve automated algorithm selection. Instead, it focuses on automating other processes such as deployment, model testing, and more.

The main advantage of this approach is that the individual creating machine learning models is not involved in configuration and other tasks unrelated to machine learning directly. To achieve this, the entire development cycle is divided into several parts, each of which is targeted for automation.

All these software products require certain processes and technologies that, broadly speaking, ensure continuous and deterministic delivery, update, and integration of different software products. This approach is known as DevOps [4]. DevOps, a portmanteau of "development" and "operations," is a software development method that extends the agile philosophy to rapidly produce software products and services and to improve operations performance and quality assurance.

If this philosophy and approach are applicable to software products such as web applications and mobile apps, it is logical to assume that the same holds true for the process of creating neural networks in general and machine learning in particular. This leads to the following tasks:

1. Define the fundamental flow of creating a neural network.
2. Identify the key logical points of this flow.
3. Identify various technological approaches that can be used.
4. Determine which of these approaches is best suited for a specific task.

The **purpose** of the work is to explore alternative to classical approaches like AutoML to the development of machine learning models, specifically focusing on MLOps. The article aims to provide insights into the advantages, drawbacks, and implications of these approaches in the context of neural network development.

In this case, we will consider a specific task: receiving a certain number of images, from which a 3D model is generated using a neural network. For this purpose, the 3D Recurrent Reconstruction Neural Network (3D-R2N2) is employed. This model uses recurrent layers for the step-by-step generation of 3D models from multiple 2D images [5].

Neural Network Development Flow

Let's outline the entire process flow of creating a neural network. The entire flow can be divided into three main stages — Design, Development, Deployment (see Figure 1).

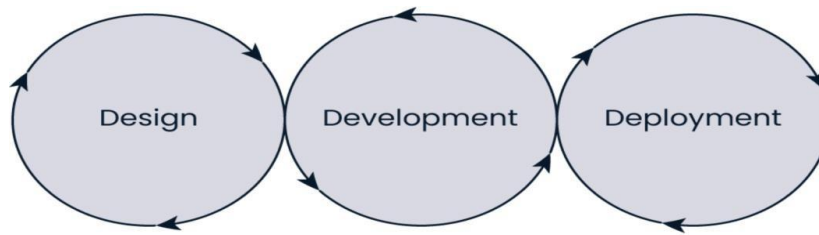


Figure 1. Three phases - Design, Development, Deployment

The first stage is pre-processing. Here, we articulate the problem and specify how it will be addressed through machine learning. Additionally, we establish key metrics to monitor the machine learning lifecycle. This stage also involves the collection of data necessary for the machine learning model.

Ultimately, the preparatory stage includes the following points:

1. context of the problem;
2. business/technical requirements;
3. key metrics;
4. data processing.

Context of the problem — the team should articulate the task and how the machine learning model will solve this problem.

Business/technical requirements, especially in the design phase, it is critical to consider the end user of the machine learning model, how well this model will meet business needs. Depending on the problem we are trying to solve, there could also be compliance and regulatory requirements for the usage of machine learning. For instance, in finance, when the law requires an explanation from the system. It is also important to consider technical requirements, such as the size of input data, defining conditions for the completion of the training process (e.g., achieving a certain accuracy on the validation dataset). Specifications for computational resources on the servers where training will take place are also important, including GPU or CPU performance, memory size, etc. The architecture of the neural network is also formulated at this stage.

Talking about key metrics, we need to see if the machine learning lifecycle progresses as expected, it is often wise to track the performance of the model. The data scientist looks at the accuracy of a model, how many times the algorithm is correct.

At the Data Processing stage, it is crucial to define the concept of data quality in a broad sense. Data Quality refers to both the characteristics associated with data, and the processes used to measure or improve the quality of data [6]. Data quality can be defined along four main dimensions, namely accuracy, completeness, consistency, and timeliness [7]. Accuracy describes the extent to which data is accurate or correct for the task at hand. Completeness is about to what extent the data fully describes the problem at hand. Timeliness is about in what time frame the data will be readily available.

In the development phase, we focus on developing the machine learning model. We do this by experimenting with a combination of data, algorithms, and hyperparameters in line with the implementation design. During the experiment, we train and evaluate one or more models in order to find the most suitable one. The goal of the development phase is to end up with the most suitable machine learning model that is ready for deployment.

In the end, the development stage includes processes such as:

- feature engineering;
- experiment tracking;
- model training & evaluation.

Feature engineering is the process of selecting, manipulating, and transforming raw data into features [8]. Feature is a variable. A feature store, as the name says, is a tool for storing commonly used features or variables relevant to the machine learning model. The feature store is the central

place where features can be managed. Using a feature store, a data scientist can find the appropriate features for their project, define new features, and use the features for training the model.

Part of machine learning model development is performing machine learning experiments. In a machine learning experiment, we train and evaluate multiple machine learning models to find the best one. As in any experiment, we test different configurations to see what works best.

During machine learning experiments, we can configure different machine learning models, for example, linear regression or deep neural networks. We can alter the model hyperparameters, like the number of layers in a neural network. We could use different versions of the data and different scripts to run the experiment. We can also use different environment configuration files per experiment, like what version of Python or R are used and what libraries. When altering each of these factors during experiments, the amount of different configurations can become huge. Each experiment also has a different outcome. This is why it is a good idea to track the configurations and the results of each experiment. Once the machine learning model is developed, we need to move the machine learning model into the production environment. In the production environment, the machine learning model will make predictions based on actual, incoming data.

In the deployment phase, we integrate the machine learning model we developed earlier into the business process. This might involve building a microservice from the machine learning model. A microservice is a small application that includes the machine learning model such that we can easily integrate it into the business process. We also aim to set up monitoring of the machine learning model. We can set up alerts when we encounter data drift or when our model does not output a prediction anymore. Data drift occurs when our data changes, which impacts the machine learning model. We'll look at these concepts in greater detail later in the course, where we'll walk through the different components of each phase.

Indeed, the deployment stage of machine learning model encompasses processes such as:

- 1. deploying the model in production;
- 2. monitoring the performance.

The implementation of each step

A diverse set of technologies is employed to implement each step in the machine learning model creation and deployment flow. For data processing, there are two primary paradigms: ETL (Extract, Transform, Load) and ELT (Extract, Load, Transform). In practice, this involves several Python scripts responsible for extracting and transforming data, along with a distributed file system where these data are stored (see Figure 2).

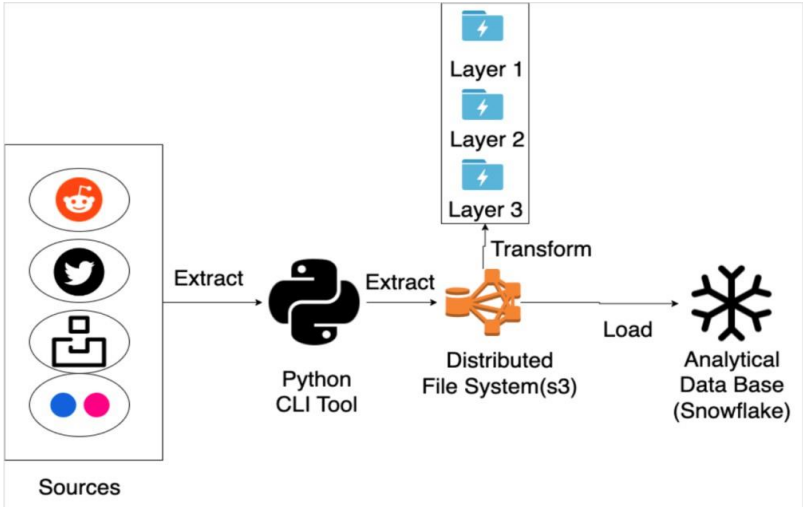


Figure 2. Architecture of ETL data processing

To implement a Feature Store, various technologies can be utilized, broadly categorized into two groups—some integrated into cloud infrastructure, while others exist as standalone software products. In the latter category, frameworks like Feast or Apache Hudi find prominence. In the former, Amazon SageMaker Feature Store, integrated with Amazon SageMaker [9], stands out. The advantage of this approach lies in its visualization capabilities (see Figure 3) and user-friendly nature, streamlining configuration processes.

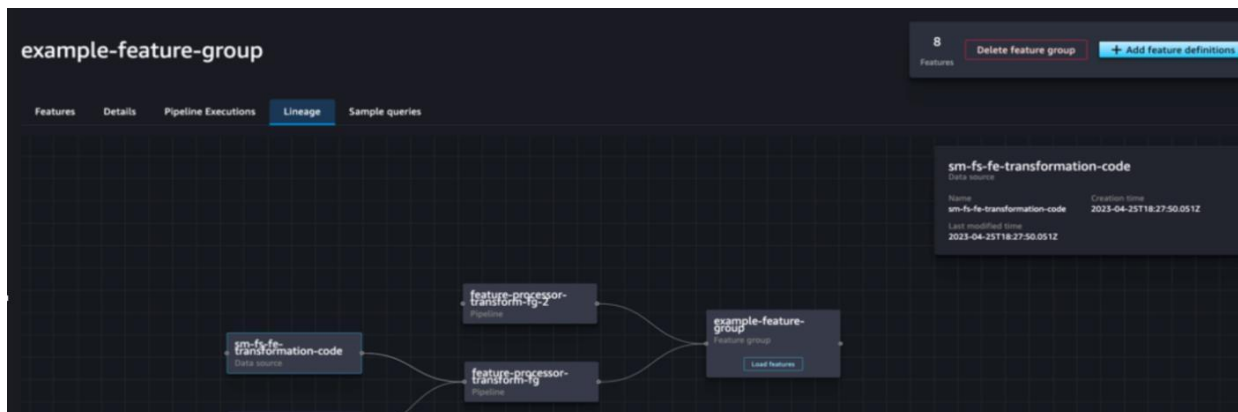


Figure 3. Example of visualization SageMaker feature store

For experiment tracking, the situation remains the same—either a standalone solution or a cloud-based one. The most popular tool for experiment tracking that is not part of the cloud is MIFlow (see Figure 4).

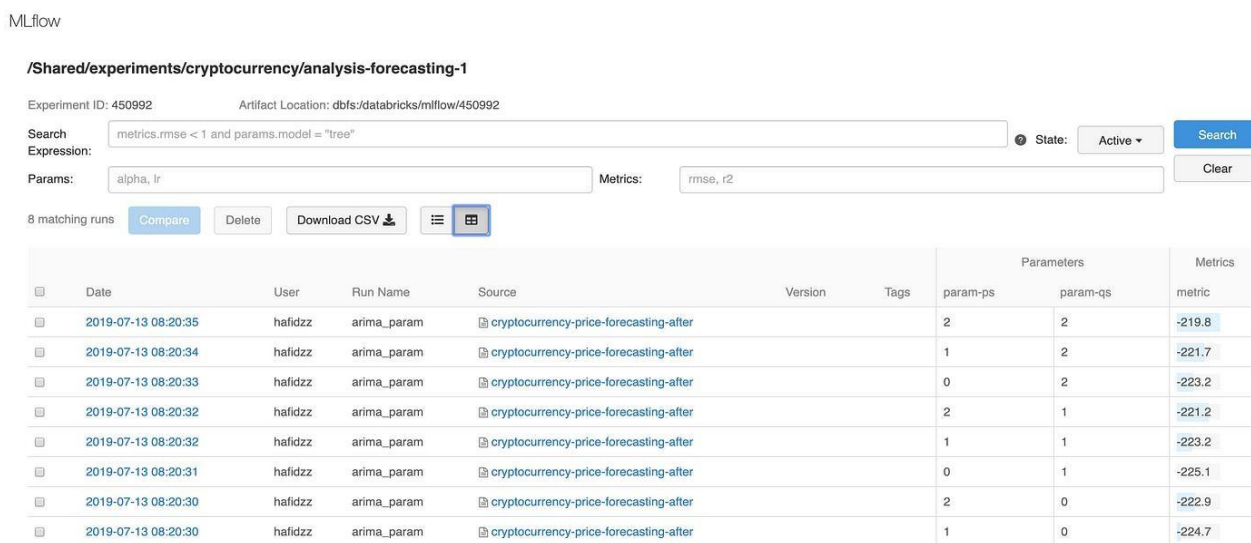


Figure 4. Example of interface MIFlow for experiment tracking

One popular cloud solution for this is SageMaker Experiments.

For model training and evaluation, two main Python frameworks are commonly used: TensorFlow and PyTorch.

There are various approaches to deploy machine learning models. Frameworks like TensorFlow or PyTorch often save machine learning models in files with extensions such as ".h5" or ".pb". A popular deployment solution involves wrapping the model in a Python framework that provides access to the model (via the file with the required extension). Web frameworks such as Flask or Django are typically used for this purpose. Subsequently, this application is deployed on a server,

functioning as a regular web application. Thus, the deployed machine learning model takes the form of a web service, accessible through some interface, often a REST API, allowing data to be passed to the model and obtaining results in return.

Other deployment methods exist, particularly with the use of cloud technologies. Consider AWS SageMaker as an example. If a file of type "h5" is available, it can be registered in the SageMaker Model Registry and deployed to a SageMaker Endpoint. The advantage of this approach is that there is no need to create a Flask application wrapper independently (see Figure 5).

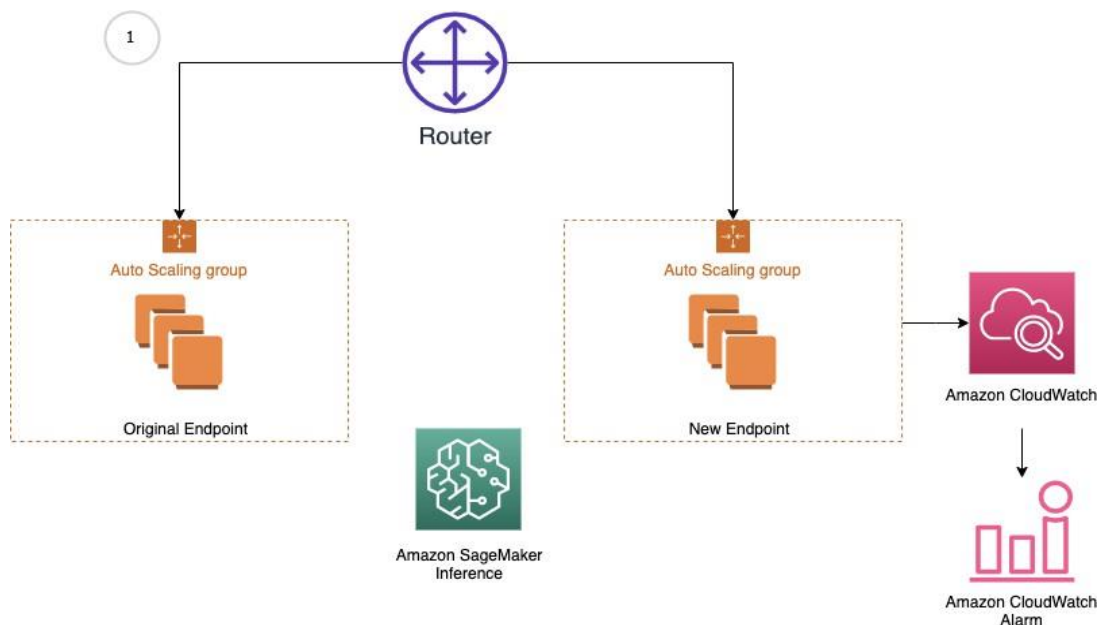


Figure 5. Architecture of SageMaker Endpoint

The outcome of this process will be an endpoint, providing access to the machine learning model. An additional benefit is that, under this approach, model monitoring becomes immediately available. In contrast, when employing a self-authored web wrapper with Flask, monitoring needs to be addressed separately. In SageMaker, this aspect is handled by another AWS service known as Amazon CloudWatch.

Best way

Hence, it is evident that for industrial development and the utilization of machine learning models, a myriad of diverse technologies can be employed. One may choose to develop and train a model on a local machine using MIFlow, tracking experiments there and utilizing Feast as a Feature Store. Subsequently, an interface for this model can be developed using Flask. Alternatively, machine learning model development can occur on a server, followed by deployment using SageMaker Endpoint. Moreover, deployment activities can be carried out manually or through Continuous Delivery (CD), which involves automated delivery and deployment. Various technologies, such as Jenkins, GitHub Actions, etc., can facilitate this process. The foundation of this continuous delivery will entail just a few command-line instructions (see Listing 1).

```
aws sagemaker create-model --model-name YourModelName

aws sagemaker create-model-version --model-name YourModelName --model-package-arn arn:aws:sagemaker:region:account-id:model-package/model-name/1.0.0

aws sagemaker create-endpoint --endpoint-name YourEndpointName --config-name YourEndpointConfig
```


Listing 1: Registration and deploying model in SageMaker

However, how can one determine the most appropriate approach in a given situation? It can be definitively asserted that in industrial development, adhering to the DevOps philosophy, minimizing manual interventions is imperative — the majority of the workflow should be automated. Such a DevOps approach applied to machine learning model development is referred to as MIOps. Based on this, MIOps maturity levels can be formulated (see Figure 6).

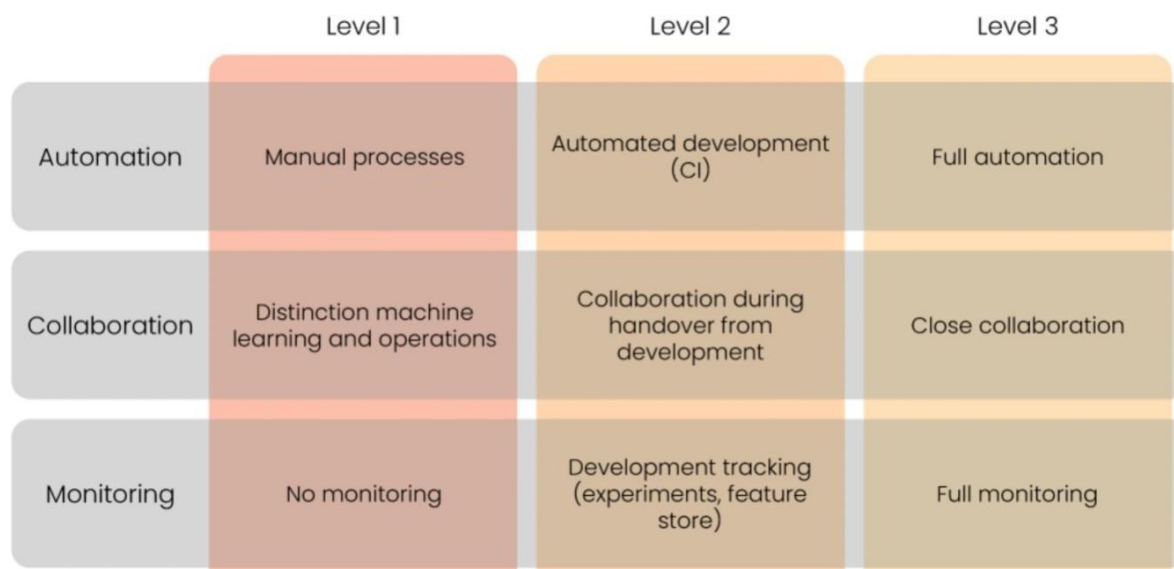


Figure 6. Level of MIOps maturity

Here, it is evident that irrespective of technologies, the benchmark is complete automation and monitoring. One of the most effective means to achieve this outcome is through a continuous, evolving, and automated infrastructure, leveraging cloud technologies. Among the good solutions within the AWS environment is AWS SageMaker, comprising several components. It enables the deployment of infrastructure, even for the development of machine learning models, based on another AWS service - EC2. For the "user" of such infrastructure, i.e., the data scientist, the interface is familiar - a Jupyter Notebook with a web interface and the required packages installed, such as Tensorflow, numpy, pandas, etc. Following development, the model can be automatically saved and logged, subsequently deployed with its endpoint. Additionally, the advantages of this approach include simplified integration with other AWS services, such as the distributed file system S3, where both input and output data can be stored. This approach streamlines the work of the machine learning model developer, eliminating the need to concern themselves with configurations and settings; they can focus solely on model development.

Beyond SageMaker, there are alternative methods to achieve similar outcomes. For instance, MIFlow facilitates the complete cycle of machine learning model development, but it comes with its limitations. Firstly, MIFlow lacks a built-in feature store, and secondly, the issue of file storage arises, requiring the separate provisioning of a distributed file system, either within Hadoop or utilizing other technologies, which necessitates additional time for configuration.

CONCLUSIONS

There are numerous ways to approach the full cycle development of machine learning models. Frameworks such as TensorFlow or PyTorch do not encompass the entire scope of machine learning model development; there are many steps before and after that require implementation, configuration, and deployment. The primary method to determine the suitability of a particular approach for a task

is the time required by a machine learning development specialist for manual tuning, configuration, and deployment. The more automated the entire process, the more efficient and faster the application of machine learning models on an industrial scale will be. MLOps way of creating machine learning model enables the automation of configuration management, thus expediting the development process. Cloud services such as AWS SageMaker good in addressing this challenge.

REFERENCES

1. Neutatz, F., Lindauer, M. & Abedjan, Z. *AutoML in heavily constrained applications*. *The VLDB Journal* (2023). <https://doi.org/10.1007/s00778-023-00820-1>
2. Shubhra Kanti Karmaker, Md. Mahadi Hassan, Micah J. Smith, Lei Xu, Chengxiang Zhai, and Kalyan Veeramachaneni. 2021. *AutoML to Date and Beyond: Challenges and Opportunities*. *ACM Comput. Surv.* 54, 8, Article 175 (November 2022), 36 pages. <https://doi.org/10.1145/3470918>
3. Leonhard Faubel and Klaus Schmid (2023): *An Analysis of MLOps Practices Hildesheimer Informatik-Berichte number / issue 1/2023, SSE 1/23/E. Software Systems Engineering, Institut für Informatik, Universität Hildesheim. Universitätsplatz 1, 31134 Hildesheim.*
4. Bayser, M.D., Azevedo, L.G., & Cerqueira, R.F. (2015). *ResearchOps: The case for DevOps in scientific applications*. 2015 IFIP/IEEE International Symposium on Integrated Network Management (IM), 1398-1404. Available online: https://www.researchgate.net/publication/283757158_ResearchOps_The_case_for_DevOps_in_scientific_applications
5. Choy, C.B., Xu, D., Gwak, J., Chen, K., Savarese, S.: *3d-r2n2: a unified approach for single and multi-view 3d object reconstruction*. In: *European conference on computer vision*, pp. 628–644. Springer (2016) https://doi.org/10.1007/978-3-319-46484-8_38 Available online: <http://3d-r2n2.stanford.edu/main.pdf>
6. DAMA International. 2011. *The DAMA Dictionary of Data Management, 2nd Edition: Over 2,000 Terms Defined for IT and Business Professionals (2nd. ed.)*. Technics Publications, LLC, Denville, NJ, USA.
7. Blake, Roger H. and Paul Mangiameli. “*The Effects and Interactions of Data Quality and Problem Complexity on Data Mining*.” *MIT International Conference on Information Quality* (2008). Available online: https://www.researchgate.net/publication/220918870_The_Effects_and_Interactions_of_Data_Quality_and_Problem_Complexity_on_Data_Mining
8. Tara Rawat and Vineeta Khemchandani, “*Feature Engineering (FE) Tools and Techniques for Better Classification Performance*”, *International Journal of Innovations in Engineering and Technology (IJJET)*, <http://dx.doi.org/10.21172/ijjet.82.024,2017>. Available online: https://www.researchgate.net/publication/333015077_Feature_Engineering_FE_Tools_and_Techniques_for_Better_Classification_Performance
9. Santiago Arboleda Quiroz “*Creating AWS SageMaker Pipelines for MLOps*” Available online: <https://medium.com/@sarboledaq/creating-sagemaker-pipeline-for-mlops-part-1-831c25b2110e>

Kshitij Bhatta, Qing Chang (*Department of Mechanical and Aerospace Engineering, University of Virginia, USA*)

FLEXIBLE MANUFACTURING SYSTEMS: HGNN-MARL STRATEGY

Abstract: *The work considers the integration of intelligent control technologies in production, quality control and maintenance planning in the conditions of growing industrial automation. It highlights the lack of unified research in these areas and proposes a holistic control strategy to improve production efficiency and quality control. Anticipating the transition to flexible manufacturing systems (FMS), the paper advocates a multidisciplinary, robot-driven FMS. The complexity lies in managing jobs with different skills, tools and dealing with problems such as tool wear, aging and failures, which requires a combined management strategy. This work extends an integrated control strategy to FMS using heterogeneous graph structure, heterogeneous graph neural network (HGNN) and multi-agent reinforcement learning (MARL) for coordinated control. Contributions include unified modeling, information aggregation using HGNNs, and problem formulation in a Decentralized Partially Observable Markov Decision Process (Dec-POMDP).*

Анотація: *У роботі розглядається інтеграція інтелектуальних технологій управління у виробництво, контроль якості та планування технічного обслуговування в умовах зростаючої промислової автоматизації. Він підкреслює відсутність єдиних досліджень у цих областях і пропонує стратегію цілісного контролю для підвищення ефективності виробництва та контролю якості. Очікуючи перехід до гнучких виробничих систем (FMS), документ виступає за мультидисциплінарну FMS, керовану роботами. Складність полягає в управлінні роботами з різними навичками, інструментами та вирішенні таких проблем, як знос інструментів, старіння та поломки, що вимагає комбінованої стратегії управління. Ця робота розширює інтегровану стратегію керування на FMS, використовуючи гетерогенну графову структуру, гетерогенну графову нейронну мережу (HGNN) і багатоагентне навчання з підкріпленням (MARL) для координованого керування. Внески включають уніфіковане моделювання, агрегацію інформації за допомогою HGNN і формулювання проблеми в децентралізованому частково спостережуваному марковському процесі прийняття рішень (Dec-POMDP).*

As industries undergo increasing automation, various aspects of the plant floor are integrating smart control technologies. Production system control, quality control, and real-time maintenance scheduling stand out as pivotal domains heavily influenced by automation. The literature on these subjects reflects a significant surge in the exploration of smart manufacturing, leveraging cutting-edge technologies such as image recognition, advanced robotics, and machine learning (ML). Despite this, there has been a limited number of studies comprehensively addressing all three areas within a unified smart control strategy. Given the inherent interconnectedness of each facet in the manufacturing system, the overall system's performance hinges on a consolidated control approach. Therefore, this paper adopts a holistic perspective on the manufacturing system and devises a control strategy to effectively manage the system's operation and maintenance. The goal is to achieve heightened production efficiency and quality control in the output.

Anticipating the imminent future of industry as Flexible Manufacturing Systems (FMS), the focus is on the adoption of a multi-skilled robot-operated FMS as the manufacturing system of interest. This system offers a notable advantage by not only embodying an FMS capable of adapting to sudden changes but also being transformable into a conventional serial manufacturing system. Consequently, it represents a diverse array of production systems. The system comprises multi-skilled robots, mobile entities equipped with quick-change end effectors enabling them to execute all tasks within the designated system. These robots operate in specific areas of the manufacturing system known as workstations, each representing distinct tasks that must be accomplished before the final part is retrieved from the end-of-line workstation.

Given the multi-skilled nature of robots, it is assumed they come equipped with various tools suitable for different workstations. Over time, these tools experience wear, and the robots themselves undergo degradation. Additionally, random events like human intervention or workstation/conveyor issues can introduce unforeseen disruptions in the system, each with the potential to significantly impact production performance. Aging-related robot downtime or random disruptions can result in

production losses, while tool wear may lead to the production of defective parts. Consequently, the combined control of robot assignment and maintenance scheduling becomes paramount.

The complexity of controlling such a system lies in the existence of multiple solutions, each with its own set of advantages and disadvantages. Aging-related robot downtime can be mitigated through preventive maintenance (PM), albeit at the cost of downtime. PM, with multiple levels, might extend the robot's lifespan at the expense of more time invested. To address tool wear, assigning robots to different workstations proves effective, yet this may necessitate tool changes, resulting in additional downtime. Exploring further possibilities, random disruption events could potentially offer opportunities for maintenance without extra production loss. Performing maintenance on multiple robots simultaneously might be a viable strategy. Amidst these maintenance decisions, effective coordination of robots is crucial for enhanced productivity and superior product quality.

Both multi-robot assignment and maintenance scheduling in a manufacturing system pose non-trivial challenges. Integrating these two challenges into a unified control strategy is even more daunting, necessitating the utilization of innovative and state-of-the-art control methods. Building on prior work, where an integrated machine-process-system control for a serial manufacturing system was developed using a homogeneous Graph Neural Network (GNN) structure and a Multi-Agent Reinforcement Learning (MARL) framework, this paper extends the combined control strategy to a Flexible Manufacturing System (FMS). It incorporates the critical aspect of maintenance scheduling. As robots are dynamically assigned to workstations in real-time, the changing graph structure leads to feature vectors of varying lengths in homogeneous graph neural networks. To address this, a method is proposed to model the system as a heterogeneous graph and employ a Heterogeneous Graph Neural Network (HGNN) to capture features of connected elements. HGNN operates on different types of node elements in a heterogeneous graph, facilitating multi-level information exchange for systems with diverse elements like robots, gantries, and workstations. Combined with MARL, this model enables a more knowledge-guided and coordinated control of the system, aligning with the primary objective of this paper.

The main contributions of this work encompass the following: (1) Integration of robot assignment, product quality, and maintenance scheduling into one unified model utilizing the heterogeneous graph structure. (2) Utilization of Heterogeneous Graph Neural Network (HGNN) for efficient information aggregation from the graph into node embeddings, facilitating a knowledge-guided control strategy. (3) Formulation of an integrated robot assignment and maintenance scheduling problem within the Decentralized Partially Observable Markov Decision Process (Dec-POMDP) framework, employing node embeddings as observations. The problem is subsequently addressed through Multi-Agent Reinforcement Learning (MARL), and the performance of the learned policy is compared against other policies.

In the evolving landscape of smart manufacturing, spurred by the advent of Industry 4.0 and the recent emergence of Industry 5.0, there is a surge in new literature. Research on innovative smart manufacturing systems is on the rise, driven by the demand for increased efficiency, adaptability, and sustainability in manufacturing. This surge is largely attributed to technological advances in Machine Learning, Robotics, Cloud Computing, and blockchain technology, among others.

A manufacturing system, as a stochastic dynamic system subject to random disruption events, poses significant complexity due to its highly non-linear nature. Finding a closed-form representation for such a multi-stage manufacturing system is virtually impossible. However, akin to any system, the production dynamics depend on the system state, control input, and external disturbances.

Various metrics, such as system throughput, system yield, and Permanent Production Loss (PPL), exist to gauge system performance. The metric choice depends on the analysis purpose. In this case, focusing on real-time control to enhance overall system efficiency requires a metric that communicates real-time performance. PPL proves effective for this purpose, as demonstrated in previous works.

In an ideal clean case, the control strategy is straightforward—robots are assigned in the ideal clean configuration. However, the system under consideration is subject to multiple disruption events

due to robot age and random disruptions. Control actions, like maintenance scheduling and robot assignment, can also trigger disruption events, as evident from the concept of effective disruption events. Thus, effective downtime is intricately linked to control input, making it a challenging aspect of the control strategy.

In this paper, a model of a multi-skilled robot-operated Flexible Manufacturing System (FMS) is developed using a heterogeneous graph structure, considering robot assignment, robot maintenance, and product quality. This model forms the basis for formulating an integrated real-time control problem within the Dec-POMDP framework. Information aggregation using HGNN and training using MARL are employed to achieve coordinated decision-making among the robots.

Maciej Klimas, Dariusz Grabowski (*Department of Electrical Engineering and Computer Science, Silesian University of Technology, Gliwice, Poland*)

LSTM FOR ELECTRIC ARC FURNACE MODELING

Abstract: *This study explores innovative models for electric arc furnaces (EAFs) in steel production, proposing a novel approach by integrating Long Short-Term Memory (LSTM) networks, a subset of deep learning. Addressing challenges in traditional methods, the research focuses on the dynamic and unpredictable nature of electric arcs, presenting LSTM networks as a solution. Utilizing datasets from different EAF stages, with emphasis on the melting phase's impact on power quality, the study employs LSTM networks to generate realizations of the stochastic process governing changes in arc equation coefficients. Another LSTM network simulates high-frequency components of voltage signals. Simulation results demonstrate the models' efficacy in capturing arc behavior, showcasing their potential for power system simulation. The adoption of neural networks, particularly LSTM, simplifies the modeling process and offers universality compared to existing methodologies. This research contributes to Industry 4.0 by incorporating advanced neural network techniques, ensuring transparency and trust in decision-making processes. As industries move towards intelligent systems, the exploration of diverse neural network architectures promises insights for optimizing complex industrial systems. This study signifies a crucial step in harnessing neural networks to enhance the resilience, sustainability, and efficiency of industrial processes.*

Анотація: *У цьому дослідженні досліджуються інноваційні моделі для електродугових печей (ЕДП) у виробництві сталі, пропонується новий підхід шляхом інтеграції мереж довготривалої короткочасної пам'яті (LSTM), підмножини глибокого навчання. Вирішуючи проблеми традиційних методів, дослідження зосереджено на динамічній і непередбачуваній природі електричних дуг, представляючи мережі LSTM як рішення. Використовуючи набори даних з різних етапів EAF, з акцентом на вплив фази плавлення на якість електроенергії, дослідження використовує мережі LSTM для створення реалізації стохастичного процесу, що керує змінами в коефіцієнтах рівняння дуги. Інша мережа LSTM імітує високочастотні складові сигналів напруги. Результати моделювання демонструють ефективність моделей у відображенні поведінки дуги, демонструючи їхній потенціал для моделювання енергосистеми. Застосування нейронних мереж, зокрема LSTM, спрощує процес моделювання та забезпечує універсальність порівняно з існуючими методологіями. Це дослідження робить внесок у Індустрію 4.0, використовуючи передові технології нейронної мережі, забезпечуючи прозорість і довіру в процесах прийняття рішень. Оскільки галузі рухаються до інтелектуальних систем, дослідження різноманітних архітектур нейронних мереж обіцяє розуміння для оптимізації складних промислових систем. Це дослідження означає важливий крок у використанні нейронних мереж для підвищення стійкості, стійкості та ефективності промислових процесів.*

Steel, as one of the most extensively used materials in the industry, faces challenges related to limited natural resources and environmental concerns, necessitating a focus on recycling. Electric arc furnaces (EAFs) represent a popular method for recycling-based steel production, but they exhibit drawbacks, notably in the application of the electric arc phenomenon for melting the furnace load. Issues such as high power consumption and the stochastic nature of the arc can adversely impact the power system, leading to power quality (PQ) problems like harmonics, flickering, and voltage variations. Addressing these PQ issues requires accurate models of connected devices.

This research aims to develop precise EAF models applicable in power system simulation software to ensure optimal operation of PQ improvement systems or EAFs. Modeling the electric arc proves challenging due to its dynamic, nonlinear, and unpredictable nature. Various approaches in the literature include linear piecewise functions, exponential, hybrid hyperbolic-exponential functions, and dynamic models based on differential equations. Additionally, stochastic components are introduced, such as hidden Markov models, Ornstein–Uhlenbeck processes, and deterministic chaos.

While some leverage machine learning, such as artificial neural networks, to simplify arc modeling, this research employs a deterministic equation derived from the instantaneous power

balance, enhanced by stochastic components generated with long short-term memory (LSTM) networks. The proposed approach eliminates the need for complex parameter identification and showcases the potential of LSTM networks in providing a universal and accurate solution.

The research utilizes current and voltage waveforms recorded during the industrial EAF work cycle for training, validation, and testing of LSTM networks. The application of LSTM networks is motivated by their ability to effectively represent stochastic changes in coefficient time series, simplifying the overall EAF modeling process.

Simulation results showcase the effectiveness of the proposed models in generating realizations of stochastic processes and reproducing high-frequency components of voltage signals. The LSTM-1 network generates stochastic process realizations, while the LSTM-2 network captures high-frequency components, contributing to a comprehensive EAF model.

In conclusion, this paper introduces two EAF models based on LSTM networks, addressing the challenges of stochastic arc behavior. The proposed models offer a promising avenue for accurate and efficient EAF simulation, with potential applications in power system optimization and steel production processes.

Continuing on the theme of electric arc furnace (EAF) modeling, the proposed LSTM-based models demonstrate significant potential for addressing the challenges associated with the dynamic and stochastic nature of the electric arc. The use of deep learning methods, specifically LSTM networks, brings advantages in effectively capturing the stochastic changes in the time series of coefficients, providing a more accurate representation of the complex and inherently unpredictable behavior of the electric arc.

The application of LSTM networks is particularly advantageous in scenarios where traditional deterministic models face limitations. The ability of LSTM networks to learn and reflect patterns in time series data reduces the need for intricate parameter identification processes, offering a more streamlined and efficient modeling approach. This is crucial in the context of EAFs, where power quality issues can have substantial consequences on equipment wear, power losses, and overall system performance.

Moreover, the two-tiered approach with LSTM-1 and LSTM-2 networks addresses different aspects of the EAF model. LSTM-1 focuses on generating realizations of the stochastic process describing changes in arc equation coefficients. This captures the inherent randomness in the system, providing a more realistic representation. On the other hand, LSTM-2 is designed to generate high-frequency components of the voltage signal, contributing to a more comprehensive model that aligns with the actual measured waveforms.

The significance of this research extends beyond the immediate application in EAF modeling. The success of the proposed LSTM-based models suggests a potential universal applicability of deep learning solutions in addressing challenges related to complex and dynamic systems in various industrial processes. The capability of LSTM networks to adapt and learn from data patterns opens avenues for their use in optimizing not only EAFs but also other interconnected systems within the broader industrial landscape.

The reliance on real industrial data, specifically current and voltage waveforms from different stages of the EAF work cycle, enhances the practical relevance of the developed models. The focus on the melting stage, known for its adverse impact on power quality, underscores the models' applicability in addressing critical operational challenges in steel production.

In conclusion, the integration of LSTM networks into EAF modeling represents a novel and promising approach, showcasing the potential of deep learning methods in enhancing the accuracy and efficiency of simulations. As industries continue to embrace digital transformation and advanced technologies, the presented models contribute to the growing body of knowledge on leveraging artificial intelligence for optimizing complex industrial processes. Further research and applications in this direction are expected to yield valuable insights and innovations for the sustainable and efficient operation of industrial systems.

Expanding further on the trajectory of neural networks in the context of industrial applications, particularly focusing on electric arc furnace (EAF) modeling, it's essential to acknowledge the broader implications of leveraging advanced machine learning techniques. Neural networks, and more specifically deep learning architectures like Long Short-Term Memory (LSTM) networks, present a paradigm shift in how we approach and comprehend complex and dynamic systems.

The success of LSTM networks in capturing the stochastic behavior of the electric arc opens up avenues for exploring other neural network architectures and methodologies. One notable aspect is the potential integration of convolutional neural networks (CNNs) in conjunction with LSTMs. CNNs are adept at capturing spatial patterns in data, and when combined with the temporal learning capabilities of LSTMs, they could offer a more comprehensive understanding of the intricate dynamics within industrial processes.

Additionally, the application of neural networks extends beyond mere modeling; it can be instrumental in predictive maintenance strategies. By training neural networks on historical data, it becomes feasible to predict potential equipment failures or deviations from optimal operating conditions. This proactive approach to maintenance not only enhances the longevity of machinery but also contributes to the overall efficiency and reliability of industrial processes.

The incorporation of explainable artificial intelligence (XAI) techniques becomes crucial in the industrial context. Understanding the decisions and predictions made by neural networks is paramount, especially when applied to critical systems. Interpretability ensures that the outcomes of the models align with domain expertise, facilitating trust and informed decision-making.

As industries progress towards Industry 4.0 and smart manufacturing, the role of neural networks becomes even more pivotal. The interconnectedness of various components in modern industrial systems necessitates intelligent and adaptive models. Reinforcement learning, a branch of machine learning where agents learn to make sequences of decisions, holds promise in optimizing complex processes and adapting to dynamic operational conditions.

Moreover, collaborative research endeavors between academia and industry can further propel the integration of neural networks into real-world applications. This collaborative approach allows for the development of tailored solutions that address industry-specific challenges, fostering a symbiotic relationship between theoretical advancements and practical implementations.

In conclusion, the journey into the realm of neural networks in industrial applications, exemplified by the EAF modeling using LSTM networks, marks a transformative era. Embracing the full spectrum of neural network architectures, from LSTMs to CNNs and reinforcement learning, opens doors to unparalleled insights and optimizations in industrial processes. The continued exploration of these methodologies, coupled with a commitment to transparency and collaboration, promises a future where artificial intelligence becomes an indispensable tool for ensuring the resilience, sustainability, and efficiency of industrial systems.

In conclusion, the integration of neural networks, particularly the application of Long Short-Term Memory (LSTM) networks in electric arc furnace (EAF) modeling, signifies a transformative leap in understanding and optimizing industrial processes. The success of LSTM networks in capturing the stochastic nature of the electric arc's behavior opens avenues for further exploration and application of advanced machine learning techniques.

As industries transition towards Industry 4.0, the role of neural networks, including convolutional neural networks (CNNs) and reinforcement learning, becomes pivotal. These technologies not only contribute to accurate modeling but also offer prospects for predictive maintenance, adaptive process optimization, and enhanced operational efficiency.

The collaborative synergy between academia and industry is paramount in tailoring neural network solutions to address industry-specific challenges. Moreover, the incorporation of explainable artificial intelligence (XAI) ensures transparency and aligns model outcomes with domain expertise, fostering trust in critical decision-making processes.

This journey into the realm of neural networks exemplifies a paradigm shift towards intelligent, adaptive, and interconnected industrial systems. Embracing a diverse range of neural network architectures and methodologies promises unparalleled insights, contributing to the resilience, sustainability, and efficiency of industrial processes.

As we advance, continued exploration and research in the field of neural networks, coupled with a commitment to collaboration and transparency, herald a future where artificial intelligence becomes an indispensable tool in shaping the future of industrial systems.

Marco Forgiione, Aneri Muni, Dario Piga (*Dalle Molle Institute for Artificial Intelligence, IDSIA USI-SUPSI, Lugano-Viganello, Switzerland*)

ADAPTIVE TRANSFER LEARNING FOR DYNAMIC SYSTEM MODELING WITH RECURRENT NEURAL NETWORKS

Abstract: *This paper introduces a transfer learning approach to enhance the adaptability and efficiency of Recurrent Neural Network (RNN) models for dynamical systems. The methodology begins with the identification of a nominal RNN model using available measurements. As the system dynamics change, leading to a degradation in performance, a correction term is introduced. This correction term is learned through Jacobian Feature Regression (JFR), utilizing features derived from the model's Jacobian concerning its nominal parameters. Additionally, a non-parametric extension using Recurrent Neural Tangent Kernel (RNTK-GP) for RNNs is proposed. Implementation details for the computation of the correction term and initial state estimation are discussed. Numerical examples demonstrate the efficacy of the methodology in handling significant system variations.*

Анотація: *У цьому документі представлено підхід до навчання з перенесенням для підвищення адаптивності та ефективності моделей рекурентної нейронної мережі (RNN) для динамічних систем. Методологія починається з ідентифікації номінальної моделі RNN за допомогою доступних вимірювань. Оскільки динаміка системи змінюється, що призводить до погіршення продуктивності, вводиться термін корекції. Цей термін корекції вивчається за допомогою регресії ознак Якобіана (JFR), використовуючи ознаки, отримані з Якобіана моделі щодо її номінальних параметрів. Крім того, пропонується непараметричне розширення з використанням рекурентного нейронного дотичного ядра (RNTK-GP) для RNN. Обговорюються деталі реалізації для розрахунку терміну корекції та оцінки початкового стану. Числові приклади демонструють ефективність методології в обробці значних змін системи.*

In the realm of Deep Learning (DL), intricate model structures are constructed using simple linear/non-linear building blocks, and model learning leverages gradient-based optimization and Automatic Differentiation (AD). While DL-based models for System Identification (SI) exhibit excellent performance on benchmarks, real-world systems often undergo dynamic variations. This paper addresses the challenge of adapting DL-based dynamical models over time, presenting a methodology inspired by transfer learning.

Several research contributions in DL for SI have adapted existing architectures to SI needs or introduced specialized model structures. DL models for SI have proven effective in standard benchmarks, yet adapting them to changing system dynamics remains an open challenge. Traditional model adaptation techniques in the systems and control field involve joint estimation of states and model parameters using stochastic descriptions and filtering techniques.

The proposed approach computes a local linear correction of the model dynamics by utilizing Jacobian features obtained through recursive operations. It extends transfer learning literature to dynamical systems modeled as RNNs, introduces an approach for initializing the RNN state based on past data, and provides a non-parametric view using RNTK-GP in a Gaussian Process framework.

The paper is organized to formulate the problem, present nominal model training, address RNN initial state estimation, and detail the proposed transfer learning approach. Efficient implementation aspects are discussed, including computational and memory costs. Numerical examples from chemical and electrical domains showcase the methodology's effectiveness in handling dynamic variations. The approach contributes to the evolving intersection of DL and SI, demonstrating potential applications in adaptive modeling for changing system dynamics.

Continuing in the same vein, the proposed transfer learning approach addresses a critical aspect of real-world systems—dynamic variations over time. The conventional practice of retraining a model from scratch in the face of system perturbations can be computationally expensive and data-intensive. In contrast, the introduced methodology leverages the existing knowledge captured in the nominal model and efficiently adapts it to the changing system dynamics.

The methodology's efficiency is particularly highlighted in scenarios where computational resources are constrained or only limited data points are available for adaptation. This is crucial in practical applications where continuous monitoring and adaptation to dynamic changes are essential for maintaining model accuracy.

The emphasis on the non-parametric view, incorporating Recurrent Neural Tangent Kernel (RNTK-GP), introduces a valuable perspective. This extension aligns with the current trend of exploring the intersection between neural networks and Gaussian Processes. The use of RNTK-GP enhances the adaptability of the methodology, showcasing its versatility across different domains and system complexities.

Moreover, the consideration of efficient initialization for the RNN state further contributes to the methodology's practical applicability. Accurate initialization becomes pivotal, especially in applications like Model Predictive Control, where capturing accurate transient predictions is crucial for effective system control.

The numerical examples presented in the paper provide concrete evidence of the methodology's effectiveness in handling significant system variations. These examples, drawn from chemical and electrical domains, serve as practical validations of the proposed approach's utility in real-world scenarios.

Looking ahead, the methodology opens avenues for further exploration, especially in refining its application to diverse system types and expanding its capabilities. Future research could delve into the integration of this approach with other adaptive modeling techniques or explore its potential in online learning scenarios where systems continuously evolve.

In conclusion, the presented transfer learning approach represents a significant stride in addressing the challenge of adapting deep learning-based models to dynamic system changes. Its computational efficiency, adaptability to limited data scenarios, and demonstrated effectiveness in numerical examples position it as a promising methodology with practical implications for a wide range of applications in dynamic system modeling and control.

Building on the foundation laid by the transfer learning approach, it is essential to underscore its broader implications and potential contributions to the fields of deep learning and system identification. The adaptability demonstrated by the methodology aligns with the evolving landscape of artificial intelligence and underscores its relevance in addressing challenges beyond static modeling.

The methodology's efficient handling of dynamic variations introduces a valuable paradigm for continuous learning in complex systems. This aspect becomes particularly crucial in applications where systems encounter diverse external conditions, aging, or other evolving factors. The ability to adapt without exhaustive retraining positions the approach as a viable solution for real-time applications, where responsiveness to changing conditions is paramount.

One notable aspect is the extension to the non-parametric view through the incorporation of Recurrent Neural Tangent Kernel (RNTK-GP). This not only enhances the methodology's flexibility but also opens avenues for exploring synergies between neural networks and Gaussian Processes in a recurrent setting. The implications of this extension go beyond the presented examples, paving the way for more sophisticated applications in diverse scientific and engineering domains.

Efficient initialization of the Recurrent Neural Network (RNN) state adds a practical dimension to the methodology. The accurate initialization of the state variable is often a critical factor in achieving reliable predictions, particularly in scenarios where early-time behavior significantly influences subsequent system dynamics. This aspect enhances the methodology's utility in applications that demand precise transient responses.

The numerical examples drawn from chemical and electrical domains serve as compelling illustrations of the methodology's real-world effectiveness. However, future research could explore its application across a broader spectrum of domains, including but not limited to biological systems, environmental modeling, and industrial processes. The generalizability of the approach to diverse system types remains an exciting avenue for exploration.

As the methodology continues to demonstrate its merit, it prompts considerations for integration into broader frameworks of adaptive modeling. Exploring how it synergizes with other adaptive techniques or ensemble learning approaches could yield insights into its scalability and potential in handling complex, multifaceted systems.

In conclusion, the transfer learning approach presented in this paper transcends its immediate applications and resonates with the overarching theme of continuous learning and adaptation in the realm of artificial intelligence. Its efficiency, adaptability, and demonstrated effectiveness mark it as a valuable addition to the toolkit of practitioners and researchers working at the intersection of deep learning and dynamic system modeling.

In conclusion, the transfer learning approach introduced in this paper represents a significant advancement in the realm of dynamic system modeling using deep learning techniques. The methodology's key strengths lie in its adaptability to changing system dynamics, computational efficiency, and applicability to scenarios with limited data.

The efficient adaptation of Recurrent Neural Network (RNN) models through transfer learning showcases a pragmatic solution to the challenge of continuous learning in real-world systems. By leveraging the knowledge captured in a nominal model and incorporating an additive correction term, the methodology mitigates the need for exhaustive retraining when the system undergoes perturbations.

The extension to a non-parametric view, incorporating Recurrent Neural Tangent Kernel (RNTK-GP), enhances the methodology's flexibility and opens avenues for exploring the synergy between neural networks and Gaussian Processes in recurrent settings. This not only broadens its applicability but also positions it at the forefront of research exploring novel intersections in deep learning.

The consideration of efficient initialization for the RNN state further enhances the methodology's practical utility, particularly in applications where capturing accurate transient responses is crucial. The demonstrated effectiveness of the approach in numerical examples from chemical and electrical domains validates its real-world applicability.

Looking forward, the methodology holds promise for further exploration across diverse domains, including biological systems, environmental modeling, and industrial processes. Its generalizability to different system types and its potential integration with other adaptive modeling techniques present exciting avenues for future research.

In summary, the transfer learning approach emerges as a valuable tool for practitioners and researchers engaged in dynamic system modeling. Its ability to adapt to evolving conditions, coupled with computational efficiency, positions it as a practical and impactful contribution to the evolving landscape of deep learning applications in dynamic systems.

Marija Kostic, Vuk Batanovic, Bosko Nikolic (*University of Belgrade, Innovation center of the School of Electrical Engineering, Belgrade, Serbia*)

NEURAL NETWORK-DRIVEN MULTILINGUAL CODE COMMENT CLASSIFICATION

Abstract: *This research explores the integration of neural networks into code commenting systems, aiming to enhance the efficiency and effectiveness of software documentation. Leveraging machine learning techniques, particularly neural networks, for code comment generation presents a promising avenue for automating and improving the quality of comments. The study delves into the training of neural models on multilingual datasets, considering both English and Serbian, and evaluates their performance across various programming languages. Additionally, the research addresses the classification of code comments, introducing a new taxonomy that accommodates different programming paradigms and languages. The proposed neural models demonstrate notable potential in enhancing code comprehension, supporting collaborative development efforts, and streamlining the documentation process. The findings underscore the significance of ongoing research in advancing the application of neural networks in code commenting for the benefit of software developers and the overall software engineering community.*

Анотація: *Досліджує інтеграцію нейронних мереж у системи коментування коду з метою підвищення ефективності та результативності програмної документації. Використання методів машинного навчання, зокрема нейронних мереж, для створення коментарів до коду є багатообіцяючим шляхом для автоматизації та покращення якості коментарів. Дослідження заглиблюється в навчання нейронних моделей на багатомовних наборах даних, враховуючи як англійську, так і сербську, і оцінює їх продуктивність різними мовами програмування. Крім того, дослідження стосується класифікації коментарів до коду, вводячи нову таксономію, яка враховує різні парадигми та мови програмування. Запропоновані нейронні моделі демонструють помітний потенціал у покращенні розуміння коду, підтримці зусиль спільної розробки та оптимізації процесу документування. Отримані результати підкреслюють важливість поточних досліджень у просуванні застосування нейронних мереж у коментуванні коду на користь розробників програмного забезпечення та спільноти розробників програмного забезпечення в цілому.*

Comments in source code represent a crucial element in contemporary software development methodologies, and their role can be explored through the application of neural networks in the field of computer science. These comments aim to provide additional information about the implementation of source code, maintain context, and facilitate collaboration among developers. Neural networks can enhance the process of comment writing and analysis. Deep learning models can learn the structure and syntactic features of code, as well as the contextual relationships between its various elements. The use of neural networks in this context can contribute to the automatic generation of comments that not only reflect the syntactic aspects of the code but also provide understandable information about its functionality and purpose.

Research in natural language processing and neural networks demonstrates that models can effectively analyze textual information and generate text based on learned dependencies in data. This approach can be applied to the analysis of source code and the automatic creation of detailed comments. However, it is essential to consider that the effectiveness of neural networks in this context may depend on the quality and volume of available training data. Additionally, the use of neural networks for comment generation requires careful tuning of model parameters and validation of results to ensure their adequacy and correctness.

In summary, the integration of neural networks into the process of creating and analyzing comments can automate routine development tasks and make code more accessible and understandable for developers. The application of neural networks to code comment analysis may also involve the detection and classification of different types of comments. Neural networks can automatically recognize functional comments describing the role and purpose of specific code fragments, as well as comments containing important hints regarding potential errors or implementation features.

One advantage of using neural networks is their adaptability and improvement over time based on new data. This means that models can learn from new comments and adapt to changing programming styles or software development practices. Potentially, the use of neural networks can improve not only the process of writing comments but also their quality and usefulness for developers. For example, models can identify the need for detailing specific code fragments and provide recommendations for including specific information to enhance understanding.

This approach to using neural networks in code commenting opens opportunities to improve communication among developers, reduce barriers for newcomers to understand the codebase, and enhance the quality of documentation. However, ongoing research in this area is crucial to address technical challenges and refine the methods of employing neural networks in software development.

An additional aspect of applying neural networks in code commenting is the potential for automatically detecting and adapting to changes in code structure. Networks can respond to modifications and generate new comment recommendations, considering the project's dynamic development. This can ensure the continuous relevance and accuracy of documentation, even in large and rapidly evolving projects.

It is important to note that the application of neural networks in code commenting also requires careful control and training on relevant datasets. Networks should learn various programming styles, consider the context of use, and interact with other developers.

Furthermore, a potential perspective involves the development of interactive tools for code commenting based on neural networks. Such tools can interact with developers, offering recommendations or automatically generating comments based on provided code.

All these research directions in the field of code commenting using neural networks indicate potentially significant contributions to improving software quality and easing the work of developers. Continuing research in this direction may define new standards and approaches to effective and understandable code commenting in modern programming.

Marta Lipnicka, Andrzej Nowakowski (*Faculty of Math and Computer Sciences, University of Lodz, Poland*)

NEW PARADIGM DEEP LEARNING WITH OPTIMAL CONTROL

Abstract: *This research introduces a novel paradigm in deep learning by integrating optimal control principles into the learning process. Leveraging the rich theoretical framework of optimal control theory, we propose a family of functions that parameterizes the initial neural network, allowing for a controlled and principled learning process. Optimal control, traditionally employed in engineering systems, is adapted to guide the learning of neural networks by minimizing a specified cost function subject to system dynamics and constraints. The incorporation of optimal control principles provides a systematic approach to learning that extends beyond conventional deep learning methods. By optimizing not only for observed data but also considering underlying system dynamics, our approach enhances adaptability and generalization capabilities. This methodology is applicable across diverse domains, including robotics, finance, and healthcare, showcasing its versatility and potential impact. Our findings demonstrate that the integration of optimal control principles enriches the deep learning paradigm, resulting in models that are not only accurate on observed data but also exhibit improved generalization to unseen scenarios. This research paves the way for advancing the field of deep learning and opens avenues for its application in complex and dynamic real-world systems.*

Анотація: *це дослідження вводить нову парадигму глибокого навчання шляхом інтеграції принципів оптимального контролю в процес навчання. Використовуючи багату теоретичну базу теорії оптимального керування, ми пропонуємо сімейство функцій, які параметризують початкову нейронну мережу, дозволяючи керувати та принципово виконувати процес навчання. Оптимальний контроль, який традиційно використовується в інженерних системах, адаптований для керування навчанням нейронних мереж шляхом мінімізації заданої функції витрат відповідно до системної динаміки та обмежень. Включення принципів оптимального контролю забезпечує системний підхід до навчання, що виходить за рамки звичайних методів глибокого навчання. Завдяки оптимізації не лише для спостережених даних, але й урахування базову динаміку системи, наш підхід покращує адаптивність і можливість узагальнення. Ця методологія застосовна в різних областях, включаючи робототехніку, фінанси та охорону здоров'я, демонструючи її універсальність і потенційний вплив. Наші висновки демонструють, що інтеграція принципів оптимального керування збагачує парадигму глибокого навчання, в результаті чого моделі не тільки є точними щодо спостережених даних, але й демонструють покращене узагальнення для невидимих сценаріїв. Це дослідження прокладає шлях для розвитку глибокого навчання та відкриває можливості для його застосування в складних і динамічних системах реального світу.*

Having examined data from experiments, artificial neural networks are frequently employed to model and describe the given data. The data, denoted as $(\mathbf{x}_0, \mathbf{y})$, $\mathbf{x}_0 \in \mathbb{R}^n$, $\mathbf{y} \in \mathbb{R}^m$, $n=1, \dots, N$, where $\{\mathbf{x}_0\}$ represents inputs to the neural network, and $\{\mathbf{y}\}$ represents the output of the feed-forward neural network. The focus is on supervised learning from this data, acknowledging the limitations of this type of training. In the best case, the trained neural network performs well only on observed data, losing accuracy beyond them. In domains like economics, biology, engineering, or medicine, where models are sought based on experimental data, mathematical expressions are often used.

The learned neural network defines a function, yet its correctness is typically assessed only on observed data. The question arises: can we infer anything about the behavior of that function for all points in its domain? This is crucial in fields like medicine or economics, where the method must work for all possible inputs, not just test cases. The challenge lies in estimating the error of the output for all possible inputs, given that the learned network is unknown.

This paper introduces a method to approximate a model for given observable data, recognizing that a finite set of data does not uniquely determine the function representing the learned neural network. To address this, a probability distribution function $\rho(\mathbf{x}_0, \mathbf{y})$, $(\mathbf{x}_0, \mathbf{y}) \in \mathbb{R}^n \times \mathbb{R}^m$, describing the distribution of observed data, is considered. The paper develops a mathematical approach to overcome the lack of uniqueness and knowledge about ρ . The proposed method assumes $\rho \subseteq \mathbb{R}^n \times \mathbb{R}^m$ and

ρ is continuous, leading to an ordinary differential equation representing a general artificial neural network.

The paper formulates an optimal control problem to find the best neural network among those trained on observable data, considering a probability distribution function and additional information. The goal is to optimize the functional among a family of functions and probability distribution functions, seeking the best approximation of the learned neural network according to the knowledge imposed into the action functional.

This approach is distinct from traditional optimization techniques for deep learning. The paper aims to answer whether the trained neural network is the one sought, introducing a methodology to study the quality of learning. The presented dual dynamic programming approach derives sufficient ε -optimality conditions for the neural network that approximates the learned network based on additional knowledge.

The contributions of the paper extend to forecasting therapeutic classes of medications, addressing the limitations of existing neural network outcomes. The approach provides a methodology to study the neural network received with the help of other methods, allowing for corrections based on additional knowledge. The paper emphasizes the importance of obtaining a neural network that performs well for all inputs in various fields, such as medicine, economics, or space shuttles.

In summary, the paper introduces a novel method for approximating models of neural networks trained on observable data, addressing issues of uniqueness, knowledge gaps, and the need for accuracy across all inputs. The proposed optimal control problem and dual dynamic programming approach contribute to enhancing the understanding and application of neural networks in diverse domains.

In our methodology, we aimed to enhance the learning process by introducing a family of functions that parameterizes the initial neural network with control. This approach creates a family of functions that we are familiar with. Presumably, the function (neural network) we are seeking is a member of this family, or at the very least, it contains a function that approximates our desired function well enough. However, as before, we still don't know whether the function we have a sense of, which approximates (given the provided data!), is the one defined by the observed data. The essence lies in implementing this type of knowledge into a deep neural network. Then, after deep learning, we obtain a refined neural network—a function. However, we still won't know anything about this learned function compared to our approach where we know an approximate function.

Through our analysis, we obtain a neural network that performs well for all points in the domain, not just for specific points (patterns). This means that we find a solution that is a function defined over the entire domain and provides a good answer for all inputs, not just for specific points.

This is crucial for fields like medicine, for predicting therapeutic classes of drugs a patient is taking, or for economics (forecasting production development) or space exploration. We propose a method that works for all conceivable cases, and we can compute the accuracy of the method irrespective of the observed data, i.e., for all possible inputs.

To integrate control into the learning process, we consider the family of functions derived from optimal control theory. Optimal control provides a powerful framework for introducing control into the learning process. The family of functions parameterizes the initial neural network, allowing for a more nuanced and controlled learning process.

In optimal control, the goal is to find the control inputs that minimize or maximize a certain objective function, subject to system dynamics and constraints. Translating this to the context of neural networks, we aim to find the optimal parameters that minimize a certain cost function, subject to the constraints imposed by the observed data.

The key advantage of integrating optimal control into the learning process is that it allows for a more principled and systematic approach. Instead of relying solely on the observed data, we leverage the principles of optimal control to guide the learning process towards a solution that is not only accurate for the observed data but also generalizes well to unseen data.

The application of optimal control in the context of neural networks has shown promise in various fields, including robotics, finance, and healthcare. By incorporating control principles, we enhance the adaptability and robustness of the learned models.

In summary, our approach extends beyond traditional deep learning by incorporating optimal control principles, resulting in a more controlled and principled learning process. This not only improves performance on observed data but also enhances the generalization capabilities of the learned models. The integration of optimal control opens new avenues for advancing the field of deep learning and its applications in diverse domains.

Martin Greguric, Filip Vrbanic, Edouard Ivanjko (*University of Zagreb, Faculty of Transport and Traffic Sciences, Zagreb, Croatia*)

INTELLIGENT ESTIMATION OF SPEED ON HIGHWAYS WITH THE HELP OF AI

Abstract: *This research redefines motorway safety assessment by integrating Connected and Automated Vehicles (CAVs) with Intelligent Speed Adaptation (ISA) systems. Utilizing dynamic, high-resolution data from CAVs in micro-grids, the study employs an explainable AI framework, specifically the Local Interpretable Model-agnostic Explanations (LIME) algorithm, to enhance transparency in safety classifications and interpret spatial features. Emphasizing proactive interventions based on AI insights, the study explores the interplay between CAV penetration rates and safety outcomes, allowing tailored interventions aligned with specific traffic demand profiles. Contributions include a micro-grid-based spatial analysis using real-time traffic flow data and the introduction of AI frameworks for transparent model interpretation. The envisioned future entails adaptive control strategies, predictive analytics, and innovation for a safer, smarter, and sustainable motorway safety landscape.*

Анотація: *Це дослідження переосмислює оцінку безпеки на автомагістралях шляхом інтеграції підключених і автоматизованих транспортних засобів (CAV) із системами інтелектуальної адаптації швидкості (ISA). Використовуючи динамічні дані високої роздільної здатності з CAV у мікромережах, у дослідженні використовується зрозуміла структура штучного інтелекту, зокрема алгоритм LIME (Local Interpretable Model-agnostic Explanations), щоб підвищити прозорість класифікації безпеки та інтерпретувати просторові характеристики. Наголошуючи на проактивних втручаннях, заснованих на розуміннях штучного інтелекту, дослідження досліджує взаємозв'язок між рівнем проникнення CAV і результатами безпеки, дозволяючи індивідуальне втручання, узгоджене з конкретними профілями попиту на трафік. Внески включають просторовий аналіз на основі мікро-сітки з використанням даних про транспортні потоки в реальному часі та впровадження фреймворків AI для прозорої інтерпретації моделі. Передбачене майбутнє передбачає адаптивні стратегії управління, прогнозу аналітику та інновації для безпечнішого, розумнішого та стійкого ландшафту безпеки на автомагістралях.*

Motorways near urban regions often experience recurrent congestion, influenced by local commuters and transit traffic. Predictions indicate a potential increase in congestion due to urban population growth and rising vehicle sales. Critical congestion points, especially near on- and off-ramps, pose safety risks on motorways. Speed differentials at these bottlenecks, coupled with increased traffic demand, can compromise safety. This study aims to validate an approach identifying critical regions on motorways for spatial safety. Simulations, considering close on- and off-ramps and recurrent congestion, represent realistic scenarios. Various traffic demand profiles are analyzed to validate the proposed approach.

Safety-critical regions are expected to be managed by Variable Speed Limit (VSL) control. This method, enhancing throughput and safety by controlling mainstream flow speeds, adjusts vehicle entry rates to bottlenecks. Assuming all vehicles are Connected and Automated Vehicles (CAVs), VSL control is enforced through a mandatory Intelligent Speed Adaptation (ISA) system. Different CAV penetration rates with ISA generate diverse traffic scenarios. VSL control widens the range for analyzing spatial safety. Assessing safety across various traffic scenarios is crucial for planning speed limit zones and other safety measures. The study also explores how different CAV penetration rates impact the precision of localizing safety-critical regions on motorways.

The main contribution lies in spatially assessing motorway safety using data from CAVs, providing more accurate assessments than macroscopic traffic data from static road sensors. CAVs, assumed to transmit data to the traffic control infrastructure, enable high-resolution spatial analysis through image-like representations and heat maps. Deep Neural Network (DNN) models, employing Convolution Neural Network (CNN) architecture, classify traffic safety based on these heat maps. The model's spatial features focus on vehicle clusters with lower speeds, considered critical for safety classification.

To explain how spatial features influence safety classification, an explainable Artificial Intelligence (xAI) framework, specifically the Local Interpretable Model-agnostic Explanations (LIME) algorithm, is introduced. This framework highlights critical regions on heat maps, aiding in understanding safety classifications. This xAI-based approach improves the accuracy of VSL zone placement and enhances safety measures.

Key contributions include the use of micro-grids for high-resolution spatial safety analysis, simulation-based analysis under various CAV penetration rates and traffic demand profiles, and the introduction of an xAI framework for explaining safety classifications. The study's focus on macroscopic traffic parameters for safety assessment and its application to realistic motorway control methods distinguish it from previous approaches using static historic crash data or lower-resolution spatial analyses.

This study advances the understanding of motorway safety by delving into the intricate dynamics of spatial safety assessment using data-driven insights from Connected and Automated Vehicles (CAVs). The utilization of micro-grids for spatial analysis sets this research apart, offering a significantly higher resolution compared to previous studies. By exclusively focusing on the motorway network and disregarding surrounding urban infrastructure, this approach provides valuable insights for planning localized control measures to enhance traffic safety on specific motorway segments.

The simulation-based environment allows for the examination of various penetration rates of CAVs equipped with Intelligent Speed Adaptation (ISA) systems under diverse traffic demand profiles. This controlled environment enables a comprehensive analysis of spatial safety, considering specific motorway control methods and varying traffic loads. The study's commitment to using macroscopic traffic parameters, such as the Time-To-Collision (TTC) safety parameter computed in real-time from traffic flow data, for spatial safety assessment distinguishes it from approaches relying on static historic crash data.

The introduction of an explainable Artificial Intelligence (xAI) framework, particularly the Local Interpretable Model-agnostic Explanations (LIME) algorithm, serves as a groundbreaking element. This xAI framework sheds light on critical regions within the analyzed motorway by providing explanations for the safety classifications determined by the Convolution Neural Network (CNN) model. The CNN model, based on the spatial image-alike representations formatted as heat maps, focuses on the spatial distribution and configuration of vehicle clusters with lower speeds. This attention to spatial features aligns with previous studies correlating increased density with crash potential.

This study's contributions extend further by exploring the impact of different CAV penetration rates on the focused localization of safety-critical regions. The proposed xAI framework offers a nuanced understanding of how varying degrees of CAV integration influence the identification and understanding of safety-related challenges on motorways. This forward-looking aspect contributes to the ongoing discourse on the integration of CAVs into existing traffic frameworks.

In summary, this research not only refines the methodology for spatial safety assessment on motorways but also introduces innovative elements such as micro-grids, simulation-based analyses, and an xAI framework. These elements collectively enhance the precision and applicability of safety assessments, providing a valuable foundation for future advancements in the realm of motorway traffic safety.

Looking ahead, the trajectory of motorway safety research is poised for transformative developments, propelled by the integration of cutting-edge technologies and novel methodologies. The incorporation of Connected and Automated Vehicles (CAVs) equipped with Intelligent Speed Adaptation (ISA) systems marks a pivotal shift toward a safer and more efficient motorway ecosystem.

The ongoing evolution of motorway safety assessment will likely witness an intensified focus on real-time data streams and dynamic analyses. The fusion of CAV-generated data, specifically the spatial image-alike representations derived from micro-grids, with advanced Artificial Intelligence

(AI) models, is expected to usher in an era of unparalleled precision in safety evaluations. This synergy presents an opportunity to not only identify safety-critical regions but also dynamically adapt control measures in response to evolving traffic conditions.

As the deployment of CAVs becomes more pervasive, the role of explainable AI (xAI) frameworks, exemplified by the Local Interpretable Model-agnostic Explanations (LIME) algorithm, will become increasingly crucial. These frameworks not only provide insights into safety classifications but also offer a transparent understanding of how the AI models interpret and prioritize spatial features. This transparency fosters trust in the decision-making processes of AI systems, a vital aspect as these systems become integral to motorway safety management.

The future landscape of motorway safety research is likely to witness a paradigm shift towards proactive interventions. The utilization of AI-driven insights to predict and prevent safety-related issues holds promise in mitigating risks before they escalate. Predictive analytics, coupled with the adaptability of control measures, may offer a pre-emptive approach to enhance motorway safety, ensuring a resilient and responsive system.

Moreover, the interplay between CAV penetration rates and safety outcomes opens avenues for tailored interventions based on specific traffic demand profiles. This nuanced approach acknowledges the dynamic nature of motorway environments, allowing for customized safety measures that align with the unique challenges posed by varying traffic scenarios.

In conclusion, the future of motorway safety research envisions a synthesis of advanced technologies, real-time data analytics, and adaptive control strategies. The proactive integration of CAVs, coupled with transparent xAI frameworks, holds the promise of not only redefining safety assessment methodologies but also establishing a blueprint for a safer, smarter, and more sustainable motorway infrastructure. This trajectory reflects a commitment to staying at the forefront of innovation to meet the evolving demands of a rapidly changing transportation landscape.

Miho Klaic, Danko Brezak, Toimslav Staroveski, Dora Bagaric (*University of Zagreb, Faculty of Mechanical Engineering and Naval Architecture, Zagreb, Croatia*)

INTELLIGENT TOOL WEAR MONITORING IN STONE DRILLING

Abstract: This research delves into the realm of intelligent tool wear monitoring for stone drilling processes, with a specific focus on the application of Radial Basis Function Neural Network (RBFNN). Neural networks, inspired by the human brain, play a pivotal role in processing multifaceted sensor data obtained during stone machining. The study emphasizes the adaptability and robustness of RBFNN in classifying tool wear conditions, showcasing its efficacy in real-world scenarios. Through a comprehensive exploration of neural network training, feature selection, and dynamic adaptation, the research contributes to the broader field of smart manufacturing. The findings advance the understanding of artificial intelligence's role in optimizing production outcomes and offer insights applicable to diverse manufacturing contexts.

Анотація: Це дослідження заглиблюється в сферу інтелектуального моніторингу зносу інструменту для процесів буріння каменю з особливим акцентом на застосуванні нейронної мережі радіальної базової функції (RBFNN). Нейронні мережі, натхненні людським мозком, відіграють ключову роль в обробці багатогранних даних датчиків, отриманих під час обробки каменю. Дослідження підкреслює адаптивність і надійність RBFNN у класифікації умов зносу інструменту, демонструючи його ефективність у реальних сценаріях. Завдяки всебічному дослідженню навчання нейронної мережі, вибору функцій і динамічної адаптації дослідження робить внесок у ширшу сферу інтелектуального виробництва. Отримані результати сприяють розумінню ролі штучного інтелекту в оптимізації результатів виробництва та пропонують розуміння, застосовне до різноманітних виробничих контекстів.

The development of robust and precise online tool wear monitoring systems is one of the most important segments in the development of smart machine tools. The tool wear process directly influences machine tool productivity, product quality, and energy consumption. Therefore, it is crucial to accurately identify tool wear dynamics during the cutting process, not only for diagnostic purposes but also as a prerequisite for implementing process control models to adapt machining parameters, such as cutting speed and feed rate.

However, online tool wear identification is a complex task because it is not possible to directly inspect the cutting edges during the cutting process. As a result, this problem must be addressed by using indirect process signals measured by different types of sensors. It is well-known that tool wear is a highly nonlinear and partially stochastic process, making it challenging to identify. This challenge becomes particularly evident when machining highly non-homogeneous and anisotropic materials with variable material properties.

Variable material properties significantly influence the selection of cutting parameters, and improper values of these parameters can negatively impact tool wear dynamics, potentially leading to tool or workpiece damage due to higher cutting forces. The study focuses on data-fusion analysis of features extracted from three types of signals: cutting forces, servomotor currents, and acoustic emission. These signals were measured during the drilling of three different types of non-homogeneous stone samples with drill bits worn to four different levels.

The features' capacity to classify the four tool wear conditions was analyzed using a Radial Basis Function Neural Network. The experimental setup involved drilling three types of stone samples with varying major constituents, employing nine combinations of cutting parameters. Measurements were randomly repeated seven times for each combination, each with four different flank wear levels. The drilling process utilized a custom-made triaxial milling machine retrofitted for stone drilling investigation.

Three types of process signals were measured during drilling: servomotor currents, cutting forces, and acoustic emission. Features were extracted from these signals and normalized between 0 and 1 for use as input data in the training and testing phases of the Radial Basis Function Neural

Network. The study aimed to classify tool wear conditions based on these features, considering different types of stone samples and machining parameters.

The results showed that features extracted from force signals generally outperformed those from current signals in terms of classification accuracy. Additionally, features extracted from acoustic emission signals, particularly in the frequency bandwidth of 50 kHz to 90 kHz, demonstrated promising results. The study explored various feature combinations and their impact on the classification accuracy, providing valuable insights into the effectiveness of different signals and their features for tool wear monitoring in stone drilling processes.

The research also delved into the importance of selecting appropriate features to enhance the accuracy of tool wear condition classification. Various combinations of features from cutting forces, servomotor currents, and acoustic emission signals were examined to identify the most effective set for the Radial Basis Function Neural Network. This feature selection process aimed to optimize the model's performance in differentiating tool wear conditions under diverse machining parameters and stone types.

Feature selection is a critical aspect of machine learning applications, as it influences the model's ability to generalize and adapt to varying conditions. In this study, the focus was not only on achieving high classification accuracy but also on ensuring the robustness of the model across different scenarios encountered in stone drilling processes.

The investigation included an in-depth analysis of the impact of cutting parameters on the tool wear process and the subsequent classification performance. Understanding the interplay between cutting speed, feed rate, and tool wear is essential for developing adaptive machining strategies. The study considered nine combinations of cutting parameters, reflecting real-world machining scenarios where optimal parameter selection is crucial for maximizing efficiency and minimizing tool wear-related issues.

Moreover, the study provided insights into the challenges posed by machining non-homogeneous and anisotropic materials. Stone samples with varying major constituents were chosen to simulate real-world conditions where the composition of workpieces can vary significantly. This variability introduces complexity to the tool wear identification task, as different materials may exhibit distinct wear patterns and responses to machining forces.

In conclusion, the research presented a comprehensive analysis of data-fusion techniques for tool wear monitoring in stone drilling processes. By exploring the synergy of cutting forces, servomotor currents, and acoustic emission signals, the study contributed valuable knowledge to the field of smart machining. The findings underscored the importance of feature selection, considering the superiority of certain features and their combinations in accurately classifying tool wear conditions.

The application of a Radial Basis Function Neural Network showcased its potential as a tool wear monitoring system, with the capability to adapt to diverse machining parameters and material compositions. This research not only addressed the specific challenges of stone drilling processes but also laid the groundwork for broader applications in smart manufacturing, where online monitoring of tool wear is integral to achieving optimal production outcomes.

The research prominently highlighted the pivotal role of neural networks, particularly the Radial Basis Function Neural Network (RBFNN), in the context of tool wear monitoring for stone drilling processes. Neural networks are a class of machine learning models inspired by the structure and functioning of the human brain. In this study, the application of RBFNN demonstrated its effectiveness in classifying tool wear conditions based on data fusion from cutting forces, servomotor currents, and acoustic emission signals.

The RBFNN, known for its ability to model complex relationships in data, was employed to enhance the accuracy of tool wear classification. Its architecture, characterized by radial basis functions as activation functions, proved advantageous in capturing intricate patterns inherent in the multifaceted sensor data collected during stone drilling operations.

The study delved into the intricacies of neural network training and optimization, emphasizing the significance of selecting appropriate features to input into the model. Feature selection, a crucial step in the machine learning pipeline, involves choosing relevant variables that contribute significantly to the model's predictive performance. The research underscored the importance of this process in tailoring the RBFNN to the specific challenges posed by stone machining, showcasing its adaptability to diverse machining parameters and material compositions.

Furthermore, the investigation highlighted the role of neural networks in addressing the dynamic nature of tool wear processes. The ability of RBFNN to learn and adapt to changing conditions during stone drilling is particularly noteworthy. This adaptability contributes to the model's robustness and reliability in real-world scenarios where machining parameters and material properties can vary.

In conclusion, the research provided a nuanced exploration of the application of neural networks, specifically the RBFNN, in the domain of smart manufacturing. By leveraging the power of neural networks for data fusion and classification, the study advanced the understanding of intelligent tool wear monitoring systems. The findings not only contribute to the field of stone drilling processes but also have broader implications for the implementation of neural network-based monitoring in diverse manufacturing contexts, emphasizing the role of artificial intelligence in optimizing production outcomes.

Nikolas Sacchi, Antonella Ferrara (*Dipartimento di Ingegneria Industriale e dell'Informazione, University of Pavia, Italy*)

NEURAL NETWORKS TRANSFORMING ENERGY FORECASTING IN METALLURGICAL PRODUCTION

Abstract: *The integration of neural networks in forecasting electricity consumption for metallurgical production represents a transformative leap. Neural networks, with their adaptability and non-linear modeling capabilities, excel in capturing complex relationships within diverse datasets. Their application extends beyond prediction, offering insights into optimizing energy usage in metallurgical processes. Ongoing efforts in model development, architectural optimizations, and diversifying datasets enhance adaptability across varying production conditions. Exploring alternative deep learning methodologies further augments real-time analysis capabilities. The utilization of neural networks signifies a commitment to precision, adaptability, and efficiency, fostering a sustainable approach to energy management in metallurgical production.*

Анотація: *Інтеграція нейронних мереж у прогнозування споживання електроенергії для металургійного виробництва є трансформаційним стрибком. Нейронні мережі з їх адаптивністю та можливостями нелінійного моделювання чудово справляються зі фіксацією складних взаємозв'язків у різноманітних наборах даних. Їх застосування виходить за рамки прогнозування, пропонуючи розуміння оптимізації використання енергії в металургійних процесах. Постійні зусилля з розробки моделей, оптимізації архітектур і диверсифікації наборів даних підвищують адаптивність у різних виробничих середовищах. Вивчення альтернативних методологій глибокого навчання ще більше розширює можливості аналізу в реальному часі. Використання нейронних мереж означає прагнення до точності, адаптивності та ефективності, сприяючи стійкому підходу до управління енергією у виробництві сталі.*

The development of a deep learning model for predicting electricity consumption in metallurgical production can be a significant research direction in terms of optimizing energy efficiency and reducing costs in manufacturing processes. Analyzing a large amount of data covering various production conditions allows identifying key parameters influencing electricity consumption.

The obtained results suggest the possibility of creating an effective model that considers the relationship between different factors and electricity consumption. The use of deep learning methods, such as neural networks, enables adapting the model to changing production conditions and providing accurate predictions of electricity consumption.

Such an approach can be beneficial for metallurgical enterprises aiming to optimize production processes and efficiently use electricity. Further research may include refining the model, considering additional factors, and validating the results in practice.

This model development approach effectively utilizes deep learning methods to address current challenges in energy efficiency in the metallurgical sector. Incorporating various factors and utilizing neural networks enhance the model's accuracy and adaptability to changing production conditions. This can make a significant contribution to the practical optimization of electricity consumption in the metallurgical industry.

Additional development of this model may involve optimizing the neural network architecture, considering the non-stationarity of parameters, and expanding the dataset to improve the model's overall performance in different production conditions.

Furthermore, considering the possibility of using other deep learning methods, such as deep convolutional networks or recurrent neural networks, for real-time analysis of electricity consumption and accounting for dynamic changes in production processes, is worth exploring.

Applying mathematical formulas to formalize and optimize models can deepen the theoretical foundations of this research. Mathematical modeling can be an effective tool for analyzing the model's sensitivity to different parameters and determining optimal values for maximum prediction accuracy.

Expanding the scope and optimizing the model are crucial for the further development of this research. Involving other deep learning methods and using mathematical models can significantly enhance the overall contribution of this approach to the practice of metallurgical production.

The integration of robotics, additive manufacturing, and the use of renewable energy sources in metallurgy opens up broad possibilities for improvement and sustainable development in this industry. Ensuring occupational safety, optimizing production processes, and reducing environmental impact are key aspects in the future development of metallurgical production.

The utilization of neural networks in predicting electricity consumption within metallurgical production epitomizes a cutting-edge approach that transcends traditional modeling methodologies. Neural networks, as a subset of deep learning, exhibit unparalleled capabilities in discerning intricate patterns within vast and complex datasets. Their intrinsic ability to adapt and learn from data, particularly in the context of metallurgical processes characterized by dynamic conditions, positions them as potent tools for precise electricity consumption forecasting.

The primary advantage of neural networks lies in their capacity to capture non-linear relationships and dependencies among diverse factors influencing electricity consumption. Traditional models may falter in handling the nuanced interplay of variables inherent in metallurgical production, whereas neural networks excel in extracting nuanced patterns that elude conventional modeling techniques.

Moreover, the adaptive nature of neural networks allows them to evolve and recalibrate in response to changes in production conditions. This adaptability is crucial in the context of metallurgy, where factors influencing electricity consumption can vary widely over time. The neural network's ability to generalize from historical data and fine-tune predictions based on real-time inputs enhances its accuracy and reliability.

In the realm of metallurgical production, where optimizing energy consumption is paramount, the neural network's proficiency in capturing the complex relationships between input parameters and electricity consumption is invaluable. Its application extends beyond mere prediction, as neural networks can offer insights into the nuanced dynamics of energy usage, identifying optimal strategies for minimizing waste and maximizing efficiency.

As we delve deeper into the development and refinement of neural network models, considerations should encompass not only architectural optimizations but also addressing challenges such as non-stationarity of parameters. Continuous efforts to expand and diversify datasets further enhance the model's adaptability to a broad spectrum of production scenarios, ensuring its robust performance across varying conditions.

Furthermore, exploring alternative deep learning methodologies, including convolutional or recurrent neural networks, may provide avenues for enhancing real-time analysis capabilities, accommodating dynamic shifts in metallurgical production processes.

In conclusion, the integration of neural networks in forecasting electricity consumption for metallurgical production represents a transformative leap, empowering the industry with a tool capable of navigating the intricacies of its operational landscape. The ongoing development and optimization of neural network models underscore a commitment to precision, adaptability, and efficiency, fostering a sustainable approach to energy management in metallurgical production.

Sergiy Kovalevskyy (*Donbass State Engineering Academy, Ukraine*), **Julia Volodchenko** (*Legal Counsel in QuartSoft Ukraine*)

NEURAL NETWORK TECHNOLOGIES FOR UKRAINE'S RECOVERY AND DEVELOPMENT

Abstract: *This work underscores the imperative for Ukraine to navigate post-war recovery and development amid intricate challenges. Limited resources necessitate exploring alternative funding, including support from Western countries and private investors. Beyond financial considerations, active restoration of territories, especially in threat-prone regions like Donetsk Oblast, demands strategic planning and comprehensive data analysis. Neural network technologies emerge as pivotal tools for recovery, aiding in infrastructure optimization, innovative technology development, investment attraction, and enhancing quality of life. The article advocates for open-source neural network models, emphasizing their potential to revolutionize civic life and governance. Immediate action is urged for successful implementation, recognizing this task's national significance.*

Анотація: *Ця робота підкреслює необхідність для України йти післявоєнного відновлення та розвитку серед складних викликів. Обмежені ресурси зумовлюють необхідність пошуку альтернативного фінансування, включаючи підтримку західних країн та приватних інвесторів. Крім фінансових міркувань, активне відновлення територій, особливо в таких загрозливих регіонах, як Донецька область, вимагає стратегічного планування та комплексного аналізу даних. Технології нейронних мереж стають основними інструментами для відновлення, сприяючи оптимізації інфраструктури, розвитку інноваційних технологій, залученню інвестицій та покращенню якості життя. Стаття виступає за моделі нейронної мережі з відкритим вихідним кодом, наголошуючи на їхньому потенціалі революціонізувати громадянське життя та управління. Необхідно вжити негайних дій для успішної реалізації, визнаючи національну значущість цього завдання.*

Ukraine is confronted with the formidable challenge of rebuilding and advancing in the midst of exceedingly intricate difficulties. The nation's recovery demands substantial resources, which, regrettably, are finite. Predicting potential reparations from the aggressor country proves to be a complex task that necessitates time. However, alternative funding sources, such as support from Western nations and private investors, can be explored. Convincing these stakeholders that their investments will be utilized efficiently and transparently is crucial. Additionally, assurances need to be provided regarding the safeguarding of investments from potential risks.

Beyond the financial dimension, it is also crucial to actively engage in the restoration of territories and facilitate the return of people to normal life. This is particularly vital for societal rejuvenation. A key focus is the formulation of a strategy for the recovery of regions and territorial communities. A clear vision of the importance of this, especially for regions like Donetsk Oblast, which remain under threat, is essential. Comprehensive data analysis and infrastructure planning are avenues to achieve this.

In this process, neural network technologies can emerge as a vital tool to bolster Ukraine's recovery and development. They can aid in intricate analytical calculations and the construction of models for territorial development strategies. However, this necessitates access to diverse data sources, including private ones. Therefore, artificial intelligence can become a crucial asset in supporting Ukraine's recovery and fortifying its position.

Neural network technologies can play a pivotal role in Ukraine's recovery and development post-war, addressing a spectrum of critical tasks:

1. Optimization of infrastructure projects: Analyzing extensive data with neural network technologies to identify the most effective locations for constructing infrastructure objects like roads and bridges.
2. Development and implementation of innovative technologies: Assistance from neural networks in developing new technologies contributing to Ukraine's economic stability and growth.
3. Attracting investments: Utilizing neural networks to create an appealing environment for investments, fostering new business opportunities, and enhancing the quality of life.

4. Improving the quality of life for people: Leveraging neural network technologies to enhance education, healthcare, employment, and transportation infrastructure.
5. Forecasting workforce needs: Predicting labor force needs at various levels through the analysis of economic and social data with neural networks.
6. Competency analysis: Examining skills and experience of individuals to identify their competencies using neural networks.
7. Personalized training: Developing individualized training programs considering each student's skills and abilities with the help of neural networks.
8. Market trends forecasting: Anticipating future trends by analyzing labor market data with neural networks.
9. Optimization of educational programs: Evaluating the effectiveness of educational programs and materials using neural networks.
10. Individualization and personalization: Adapting educational material to the needs of each student through neural networks.
11. Automated assessment and reporting: Analyzing students' responses to assess their level of understanding with neural networks.
12. Improving teaching methods: Assessing the effectiveness of different teaching methods using neural networks.
13. Automated creation of educational materials: Generating educational content considering the needs of students through neural networks.
14. Support for inclusive education: Facilitating teachers in adapting education for students with different needs, providing more accessible learning with neural networks.

In summary, neural network technologies hold the potential to offer numerous perspectives for supporting Ukraine's recovery and development. However, their successful implementation requires substantial investments and qualified personnel. Neural technologies have the potential to revolutionize civic life by fostering increased civic engagement, people's involvement in the political process, and the creation of more effective and accountable governing bodies. Concrete examples include:

1. Creating platforms for civic engagement: Developing platforms with neural technologies to facilitate communication, idea exchange, and coordination of citizens' actions.
2. Improving access to information: Streamlining citizens' access to information on issues of concern through the application of neural technologies.
3. Engaging people in the political process: Activating citizens' participation in political life using neural technologies, including personalized messages encouraging participation in political initiatives.
4. Establishing more efficient and accountable governing bodies: Enhancing the efficiency and accountability of governing bodies with neural technologies, such as identifying effective government policy programs.

To succeed in implementing neural network technologies, it is imperative to channel scientific potential toward creating open-source neural network models for the mentioned tasks and provide ongoing support to users of these artificial intelligence tools. This is a nationally significant task, and its resolution demands immediate action.

Tao Li, Feng Xie, Xin Guo, Qingchun Feng (School of Agricultural Engineering, Jiangsu University, China)

AUTONOMOUS MULTI-ARM ROBOTIC SYSTEMS

Abstract: This study delves into the realm of precision agriculture, exploring the development of autonomous multi-arm robotic systems for orchard fruit harvesting. To address the challenges posed by fruit occlusions, a novel deep convolutional neural network algorithm is employed to enhance fruit recognition and localization. Additionally, a task planning method based on multi-agent reinforcement learning optimizes harvesting efficiency. Field trials with a four-arm robotic system demonstrate promising success rates and cycle times. The research highlights the transformative potential of robotic systems in orchard harvesting, paving the way for increased efficiency, reduced labor costs, and sustainable fruit production practices.

Анотація: це дослідження заглиблюється в сферу точного землеробства, досліджуючи розробку автономних багаторукових робототехнічних систем для збору плодів у садах. Щоб вирішити проблеми, пов'язані з оклюзіями плодів, використовується новий алгоритм глибокої згорткової нейронної мережі для покращення розпізнавання та локалізації плодів. Крім того, метод планування завдань, заснований на багатоагентному навчанні з посиленням, оптимізує ефективність збирання. Польові випробування роботизованої системи з чотирма руками демонструють багатообіцяючі показники успіху та тривалість циклу. Дослідження підкреслює трансформаційний потенціал роботизованих систем у збиранні врожаю в садах, прокладаючи шлях до підвищення ефективності, зниження витрат на робочу силу та практики сталого виробництва фруктів.

The labor-intensive harvesting process in fruit production incurs high costs, prompting the need for autonomous and unmanned harvest solutions to address industry bottlenecks. Robotic harvesters, leveraging advancements in computer and sensor technologies, have made significant progress in fruit detection, localization, path planning, and mechanical design. Various harvesting robots for fruits like apples, oranges, tomatoes, and pineapples have been developed in recent years.

The cultivation mode of dwarfing and dense planting in apple orchards has gained popularity due to its advantages in facilitating mechanized operations. Picking robots, especially those with multiple arms, have shown positive progress in enhancing operating efficiency. Researchers have integrated multiple operational units, enabling robots to perform large-scale collaborative tasks and improve efficiency. Some multi-arm robots have been successfully commercialized, proving their effectiveness in enhancing harvesting efficiency.

However, multi-arm robots face technical challenges, requiring meticulous planning and control methods due to overlapping working spaces and cooperation needs. Task planning becomes crucial for harvesting collaboration, resembling decision-making problems like the multiple traveling salesperson problem (mTSP) or vehicle routing problem (VRP). Optimizing task planning for inter-arm collaboration poses challenges due to constraints like inter-arm motion coupling, workspace, shortest neighboring distance, and multiple optimization objectives.

Another challenge lies in the perception of robots for fruit targets, especially in complex environments with varying lighting conditions and occlusions. While learning-based image processing algorithms have improved visual perception, many existing studies struggle with occluded fruits during harvesting operations. Robotic harvesters need to cope with varying degrees of occlusions, demanding advanced image recognition and 3D localization for pixel-level fruit segmentation and estimating positions of occluded fruits.

To address these challenges, this work presents:

1. A multi-task deep convolutional neural network-based algorithm using multiple stereo cameras for fruit recognition and localization, aiming to enhance recognition rates and localization accuracy of occluded fruits.

2. A task planning method based on a multi-agent reinforcement learning model to optimize harvesting time for multiple fruits with multiple arms, improving overall operational efficiency and reducing average cycle time.

3. A highly integrated four-arm apple harvesting robotic system demonstrated in field trials with a harvest success rate ranging from 71.28% to 80.45% and an average cycle time from 5.8% to 6.7%, showcasing its performance under different apple tree growth conditions.

The past decade has witnessed the rapid development of dwarf dense planting in orchards, making it the mainstream in fruit production. Dwarf dense planting orchards, with their visibility and reachability advantages, have become the primary focus of harvesting robots. The integration of advanced technologies and collaborative multi-arm systems holds promise for addressing the challenges and improving the efficiency of fruit harvesting in orchards.

Furthermore, the development of dwarf dense planting in orchards has become the mainstream trend in fruit production over the past decade. This cultivation pattern offers advantages such as well-defined visibility and accessibility, making it an ideal scenario for the deployment of harvesting robots. Researchers and engineers have directed their efforts toward enhancing the capabilities of these robots to meet the specific challenges posed by dwarf dense planting orchards.

A notable trend in the field involves the integration of advanced technologies into robotic systems, particularly those with multiple arms. The use of multi-arm robots has shown promise in improving the overall operational efficiency of fruit harvesting in orchards. These robots, equipped with sophisticated sensors, computers, and artificial intelligence, demonstrate enhanced information perception, intelligent decision-making, precision control, and execution of harvesting operations.

In response to the concentrated ripening and large-scale harvesting operations characteristic of apple production, researchers have focused on optimizing the comprehensive operating efficiency of robotic harvesters. The integration of multiple operational units in these robots allows them to perform large-scale collaborative tasks, paving the way for innovative solutions to improve efficiency. Multi-arm collaborative task planning methods have been explored, demonstrating a highly promising technological approach.

However, the adoption of multi-arm robots introduces additional technical challenges. Unlike traditional single-arm robots, multi-arm systems require meticulous planning and control methods due to overlapping working spaces and the necessity for cooperation. Factors such as the sequencing of robots picking multiple fruits can significantly impact operation duration, emphasizing the importance of task planning. The scheduling of interarm movements plays a crucial role in the efficiency of harvesting operations, making task planning a pivotal aspect of the robotic harvesting process.

Addressing the challenges associated with multi-arm harvesting robots in orchards, this work proposes innovative solutions:

1. A multi-task deep convolutional neural network-based algorithm employs multiple stereo cameras to enhance fruit recognition and localization accuracy, particularly for occluded fruits. This algorithm aims to improve the overall effectiveness of the robotic harvesting process by increasing recognition rates and accuracy.

2. A task planning method based on a multi-agent reinforcement learning model is introduced to optimize the time required for harvesting multiple fruits with multiple arms. The objective is to enhance the overall operational efficiency of the robot and reduce the average cycle time. This approach leverages advanced artificial intelligence techniques to make intelligent decisions in a dynamic orchard environment.

3. A highly integrated four-arm apple harvesting robotic system is developed and tested in field trials conducted in apple orchards in the Haidian and Changping Districts of Beijing, China. The orchards feature a Simple, Narrow, Accessible, and Productive (SNAP) fruiting-wall canopy architecture, creating a conducive environment for robotic harvesting. The field trials demonstrate the robot's successful harvest rates ranging from 71.28% to 80.45%, showcasing its performance under various apple tree growth conditions.

In conclusion, the rapid evolution of orchard cultivation practices, coupled with advancements in robotic technologies, presents exciting opportunities for improving the efficiency of fruit harvesting. The integration of multi-arm robotic systems, advanced algorithms, and artificial intelligence holds the key to addressing the specific challenges posed by modern orchard environments. This research contributes valuable insights and practical solutions to propel the field of robotic fruit harvesting towards increased effectiveness and widespread adoption in commercial orchards.

The success of multi-arm robotic systems in orchard environments is contingent upon their ability to adapt to dynamic and complex conditions. The challenges of fruit occlusion, varying lighting environments, and the intricate canopy structures demand innovative solutions to further enhance the capabilities of harvesting robots.

The proposed multi-task deep convolutional neural network-based algorithm represents a significant advancement in addressing the issue of occluded fruits. By leveraging multiple stereo cameras, the algorithm aims to achieve pixel-level fruit segmentation and accurately estimate the position of occluded fruits based on partial information. This not only improves recognition rates under challenging conditions but also contributes to the overall robustness of the robotic harvesting system.

Task planning remains a critical aspect of optimizing the efficiency of multi-arm robotic harvesters. The introduction of a task planning method based on a multi-agent reinforcement learning model signifies a shift towards intelligent decision-making in real-time harvesting scenarios. This approach considers factors such as inter-arm collaboration, minimizing idle time, and traversing distances, contributing to a more streamlined and efficient operation.

The field trials conducted in apple orchards employing the highly integrated four-arm robotic system provide valuable insights into the practical applicability of the proposed solutions. The success rates ranging from 71.28% to 80.45% and average cycle times from 5.8% to 6.7% demonstrate the robot's adaptability to different growth conditions of apple trees. These results underscore the potential of advanced robotic systems to revolutionize fruit harvesting practices, especially in environments characterized by dense planting and concentrated ripening.

Looking forward, the continued development and refinement of multi-arm robotic systems for orchard harvesting will likely involve ongoing research in several key areas. Further advancements in perception technologies, including image processing algorithms and 3D localization techniques, will play a crucial role in overcoming challenges related to occlusions and varying environmental conditions.

Additionally, the integration of advanced learning-based approaches, such as reinforcement learning, will contribute to the adaptive and autonomous decision-making capabilities of robotic harvesters. This can lead to more efficient task planning and resource allocation, ultimately enhancing the overall performance of the robotic system.

The success of this research in addressing the specific challenges of multi-arm robotic harvesting in orchards opens avenues for future exploration and innovation. As technology continues to evolve, the collaborative efforts of researchers, engineers, and orchard operators will likely drive the widespread adoption of robotic systems, marking a transformative shift in fruit harvesting practices worldwide. The potential benefits include increased efficiency, reduced labor costs, and improved overall productivity, contributing to the sustainability and competitiveness of the fruit industry.

V. Giannellaa, R. Sepea (Department of Industrial Engineering, University of Salerno, Italy)

REAL-TIME PROBABILISTIC STRUCTURAL LIFE PREDICTION USING NEURAL NETWORKS AND MONTE CARLO

Abstract: This study presents a novel methodology for real-time probabilistic residual life predictions of structures, combining Finite Element Method (FEM), Monte Carlo (MC), and Neural Networks (NNs). The "H1-L1" neural network proves optimal, demonstrating high accuracy and stability with minimal data. The methodology's application to a thin plate with a central hole showcases its effectiveness in real-world scenarios. This approach eliminates the need for prior information on input parameter distributions, making it a valuable tool for structural health assessments in uncertain conditions. The study bridges traditional and advanced methods, providing a promising avenue for future research in structural engineering.

Анотація: У цьому дослідженні представлено нову методологію ймовірнісного прогнозування залишкового ресурсу конструкцій у режимі реального часу, яка поєднує метод скінченних елементів (FEM), Монте-Карло (MC) і нейронні мережі (NN). Нейронна мережа «H1-L1» виявилася оптимальною, демонструючи високу точність і стабільність з мінімальними даними. Застосування методології до тонкої пластини з центральним отвором демонструє її ефективність у сценаріях реального світу. Цей підхід усуває потребу в попередній інформації про розподіли входних параметрів, що робить його цінним інструментом для оцінки стану конструкції в невизначених умовах. Дослідження поєднує традиційні та передові методи, забезпечуючи багатообіцяючий шлях для майбутніх досліджень у будівельній інженерії.

In this study, neural networks were employed to create models for predicting residual fatigue life, enabling real-time probabilistic life predictions of damaged structures under stochastically varying input parameters. The research involved a detailed comparison of five different neural network architectures, assessing accuracy, computational runtimes, and the minimum number of samples required for training. The objective was to identify the architecture with optimal generalization power. The neural networks were trained, validated, and tested using fatigue life predictions obtained through simulations employing Finite Element Method (FEM) and Monte Carlo methods.

A noteworthy outcome was the identification of the "H1-L1" neural network as the most effective model, exhibiting high accuracy (Mean Square Error of $4.8e-7$ on the test dataset) and stability, particularly when dealing with reduced amounts of data. The model's potential applicability for Structural Health Monitoring (SHM) in cost-effective GPU devices was emphasized due to its minimal parameter requirements.

The study addressed the critical issue of predicting the residual fatigue life of structures, considering uncertainties and variability in input data. The research demonstrated the applicability of neural networks, especially the identified optimal architecture, in providing accurate life predictions even when faced with uncertain and highly variable input conditions. A real-world case study illustrated the effectiveness of the proposed approach in delivering precise life predictions under such challenging circumstances.

The integration of advanced computational methods, including FEM and Monte Carlo simulations, with neural networks showcased a robust methodology for structural life prediction. The study's methodology, validated through a case study involving a thin plate with a central hole, presented a versatile approach that could be extended to study larger and more complex industrial components. Additionally, the ability of the constructed neural networks to provide real-time forecasts on cost-effective GPU devices positioned them as viable solutions for SHM applications in practical engineering scenarios.

The comprehensive discussion and results presented in the study contribute to the advancement of methodologies for probabilistic residual life predictions, emphasizing the synergy between traditional simulation techniques and state-of-the-art neural network approaches. The findings have implications for enhancing safety, reducing maintenance costs, and improving the

design of fatigue-resistant products across various engineering sectors. Further research and application of the proposed methodology in diverse industrial contexts could lead to significant advancements in structural health monitoring and predictive maintenance strategies.

In conclusion, this research represents a noteworthy contribution to the field of structural life prediction, showcasing the potential of neural networks in addressing complex challenges associated with uncertainties in input data and variability in structural conditions. The identified optimal neural network architecture, "H1-L1," stands out for its accuracy, stability, and suitability for practical applications, especially in the context of cost-effective GPU devices for real-time monitoring.

Opinion: The presented research underscores the significance of integrating neural networks into the realm of structural life prediction, showcasing their efficacy in handling uncertainties and variability. The identified neural network architecture, "H1-L1," demonstrates promising potential for practical implementation, particularly in cost-effective GPU devices for real-time Structural Health Monitoring. The meticulous comparison of different architectures and optimizers, along with the validation through a real-world case study, enhances the credibility of the study's findings. The methodology's versatility and applicability to larger and more complex industrial components open avenues for future research and practical implementation in diverse engineering sectors.

The proposed approach addresses a critical need in the engineering domain by tackling the challenging task of predicting the residual fatigue life of structures, a vital aspect for preventing safety accidents and minimizing economic losses. The study's focus on the crack nucleation process and fatigue crack growth (FCG) underlines its relevance to diverse manufacturing processes, from welding to additive manufacturing technologies.

A key strength of the research lies in recognizing the inherent variability in FCG processes, attributed to factors such as geometrical variations, material scattering, and statistical uncertainties. By leveraging advanced tools, including Finite Element Method (FEM) and Monte Carlo simulations, the study successfully navigates the complexities associated with predicting structural life under uncertain and variable conditions.

The integration of Neural Networks (NNs) into this framework represents a paradigm shift in structural life prediction methodologies. The study acknowledges the limitations of traditional computational methods, especially in scenarios requiring robust and probabilistic predictions. The rise of NNs, capable of handling large and complex datasets, offers a promising avenue for more precise assessments under highly variable and uncertain experimental conditions.

The significance of the "H1-L1" neural network architecture emerges as a key finding, demonstrating exceptional accuracy and stability, particularly when faced with limited data. The model's potential for Structural Health Monitoring (SHM) on small, cost-effective Graphics Processing Unit (GPU) devices positions it as an attractive solution for real-world engineering applications.

The investigation not only fills existing literature gaps but also sets a precedent for future research directions. The seamless combination of traditional computational methodologies with deep learning approaches provides a holistic framework for tackling structural life prediction challenges. The study's real-world case study involving a thin plate with a central hole serves as a practical demonstration of the proposed methodology's effectiveness.

The outlined approach, requiring no a priori information about distribution functions and ranges of variation in input parameters, aligns with real-life scenarios where such information is often unknown. This characteristic enhances the methodology's practicality, saving valuable time and costs typically associated with experimental evaluations on specific structures.

The research's innovative methodology, validated through rigorous comparisons, not only advances the field of structural life prediction but also holds implications for enhancing competitiveness in designing fatigue-safe products. The ability to provide real-time forecasts on cost-effective GPU devices further extends its applicability to Structural Health Monitoring, offering a valuable tool for assessing and updating predictions in real-world engineering scenarios.

In summary, this research represents a pioneering effort in integrating neural networks with traditional computational methods for probabilistic residual life predictions. The identified "H1-L1" neural network architecture stands out as a reliable model, showcasing its potential for addressing uncertainties and variability in structural conditions. The study's findings contribute significantly to the evolution of methodologies for structural health monitoring, predictive maintenance, and the design of fatigue-resistant products across diverse engineering sectors.

Moving forward, the research methodology unfolds with a comprehensive exploration of the fatigue crack propagation simulation using Finite Element Method (FEM). The case study, centered around a thin plate with a central hole, serves as a tangible illustration of the proposed approach. The subsequent incorporation of Monte Carlo (MC) simulations accelerates the generation of a substantial volume of life predictions, essential for training the Neural Networks (NNs).

The comparisons among MC results and different NN architectures provide valuable insights into accuracy and computational efficiency. The inclusion of probabilistic life assessment exemplifies the practical application of the proposed approach, showcasing its utility in real-world scenarios.

The concluding remarks encapsulate the essence of the research. The methodology, marrying advanced FEM crack-growth simulations, MC methods, and sophisticated multi-level NNs, emerges as a novel and effective means for probabilistic residual life predictions. The validation process, involving benchmarking of different NN architectures and optimizers, reinforces the robustness of the approach.

The research underscores the potential of Neural Networks in offering precise assessments under conditions of high variability and uncertainty. The "H1-L1" neural network, identified as the optimal model, not only replicates Monte Carlo reference data with exceptional accuracy but also demonstrates stability and efficiency, particularly with reduced datasets. Its shallow architecture, coupled with minimal parameter requirements, enhances its attractiveness for Structural Health Monitoring on cost-effective GPU devices.

The final phase of the study involves the practical application of the best-performing model, "H1-L1," for a probabilistic fatigue life assessment. This demonstration serves as a testament to the approach's viability in real-world scenarios, where uncertainty in data and measurements is the norm rather than the exception.

In essence, this research represents a groundbreaking endeavor in developing models for real-time probabilistic residual life predictions under uncertain input conditions. By seamlessly integrating traditional simulation methods with advanced Neural Networks, the study pioneers a holistic approach to structural life prediction. The identified neural network architecture, along with the overarching methodology, stands as a beacon for future research endeavors seeking to address the complexities of predicting structural integrity in dynamic and uncertain environments.

In conclusion, this research advances the frontier of structural life prediction by introducing a pioneering methodology that combines traditional simulation techniques with state-of-the-art Neural Networks (NNs). The focus on real-time probabilistic residual life predictions under uncertain input conditions is paramount in enhancing the robustness and applicability of structural health assessments.

The developed approach, showcased through a comprehensive exploration of fatigue crack propagation in a thin plate with a central hole, leverages Finite Element Method (FEM) simulations, Monte Carlo (MC) methods, and intricate multi-level NNs. The synergy of these elements culminates in a methodology capable of delivering accurate life predictions even in the face of uncertain and highly variable input data.

The meticulous benchmarking of five distinct NN architectures and various optimizers underscores the significance of model selection in achieving optimal accuracy and computational efficiency. The standout performer, the "H1-L1" neural network, not only replicates Monte Carlo reference data with precision but also exhibits stability and efficiency, particularly with reduced datasets. Its shallow architecture, coupled with minimal parameter requirements, positions it as an attractive solution for real-time Structural Health Monitoring on cost-effective GPU devices.

The practical application of the identified model in a probabilistic fatigue life assessment serves as a tangible demonstration of its effectiveness in real-world scenarios. The approach's ability to forego the need for a priori information about the distribution functions and ranges of variation in input parameters aligns with the challenges posed by the unpredictability of structural integrity in practical applications.

This research contributes to the evolving landscape of structural engineering by bridging the gap between traditional simulation methods and the capabilities offered by advanced NNs. The identified neural network architecture and the overarching methodology represent a promising avenue for future research endeavors. As the demand for accurate, real-time predictions in the face of uncertain conditions continues to grow, the integration of advanced computational methods with Neural Networks stands out as a key enabler for ensuring the safety and reliability of engineering structures.

Бабаш А.В. (Україна, м. Краматорськ, ДДМА)

ВИКОРИСТАННЯ ЗАСОБІВ ШТУЧНОГО ІНТЕЛЕКТУ У МОБІЛЬНИХ ДОДАТКАХ

***Анотація.** Наведено опис можливостей класу CameraX для захоплення та обробки зображень, отриманих з камери мобільного пристрою. Описано режим роботи Image Analysis для захоплення та одночасної обробки зображень за допомогою моделей машинного навчання. Описано процедуру інтеграції можливостей CameraX у проект за допомогою gradle. Наведено процедуру інтеграції моделі машинного навчання для розпізнавання об'єктів з зображень у проект мобільного додатку. Показано результати роботи моделі машинного навчання у проекті мобільного (Android) додатку.*

***Abstract.** A description of the capabilities of the CameraX class for capturing and processing images obtained from the camera of a mobile device is provided. Image Analysis for capturing and simultaneously processing images using machine learning models is described, and the procedure for integrating CameraX capabilities into a project using gradle is described. The procedure for integrating a machine learning model for recognizing objects from images into a mobile application project is presented. The results of the machine learning model work, in form of mobile app project, are given here*

Використання засобів штучного інтелекту дозволяє вирішити багато різних задач. Наприклад, можна розпізнавати об'єкти зі статичних фотографій або відео потоку.

Також за допомогою нейромереж можна перетворювати чорно-біле зображення у кольорове. Сучасні охоронні системи використовують штучний інтелект для розпізнавання обличчя.

У сучасних смартфонах засоби штучного інтелекту використовуються для покращення якості зображень, отриманих з фотокамерах (зменшення шуму та ін.).

Для розпізнавання об'єктів використовуються моделі машинного навчання. Однією з таких моделей є TensorFlow Lite Object Detection [1]. Така модель легко інтегрується в проект Android додатку.

Мета роботи – аналіз, дослідження та використання засобів штучного інтелекту у Android додатках для визначення об'єктів на зображеннях, захоплених камерою мобільного пристрою.

Задачі дослідження:

-вивчення можливостей TensorFlow Lite Object Detection для визначення об'єктів з зображень;

-вивчення можливостей CameraX для захоплення зображень з камери мобільного пристрою;

-підключення та використання готової моделі машинного навчання для розпізнавання об'єктів з зображень, захоплених камерою мобільного пристрою;

-розробка Android додатку для демонстрації роботи моделі машинного навчання для розпізнавання об'єктів.

Об'єкт дослідження – модель машинного навчання TensorFlow Lite Object Detection.

Предмет дослідження – аналіз, дослідження та використання моделі у проекті Android додатку для розпізнавання об'єктів з зображень, захоплених камерою мобільного пристрою.

ОПИС МОЖЛИВОСТЕЙ КЛАСУ CAMERAX ДЛЯ ЗАХОПЛЕННЯ ТА ОБРОБКИ ЗОБРАЖЕНЬ

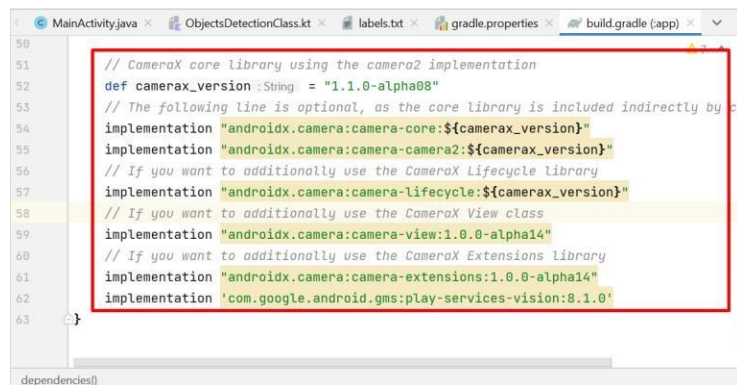
Клас CameraX [2] має багато можливостей для захоплення та обробки зображень з камери. Він значно спрощує процес розробки та є ефективним рішенням для отримання та обробки зображень з камери у реальному часі. CameraX підтримує декілька можливостей для

роботи з зображеннями. А саме Preview, Image Analysis, Image Capture та Video Capture. Для розпізнавання об'єктів було використано режим роботи Image Analysis.

Цей режим роботи дозволяє захоплювати зображення та одночасно їх ефективно обробляти за допомогою різних бібліотек, або з використанням засобів моделей машинного навчання.

При цьому не потрібно створювати якихось додаткових потоків для окремо для захоплення та обробки зображень з камери мобільного пристрою.

Для використання можливостей CameraX у проекті мобільного додатку з використанням середовища IntelliJ Idea [3] необхідно у gradle додати наступні залежності та зробити синхронізацію проекту (рисунок 1).



```
50
51 // CameraX core library using the camera2 implementation
52 def camerax_version :String = "1.1.0-alpha08"
53 // The following line is optional, as the core library is included indirectly by cc
54 implementation "androidx.camera:camera-core:${camerax_version}"
55 implementation "androidx.camera:camera-camera2:${camerax_version}"
56 // If you want to additionally use the CameraX Lifecycle library
57 implementation "androidx.camera:camera-lifecycle:${camerax_version}"
58 // If you want to additionally use the CameraX View class
59 implementation "androidx.camera:camera-view:1.0.0-alpha14"
60 // If you want to additionally use the CameraX Extensions library
61 implementation "androidx.camera:camera-extensions:1.0.0-alpha14"
62 implementation 'com.google.android.gms:play-services-vision:8.1.0'
63
dependencies{
```

Рисунок 1 – Необхідні залежності gradle для роботи з CameraX

Фрагмент програмного коду для захоплення зображення та його подальшої обробки за допомогою засобів CameraX наведено на рисунку 2.



```
10
11 imageAnalysis.setAnalyzer(ContextCompat.getMainExecutor( context: MainActivity.this),
12 new ImageAnalysis.Analyzer() {
13 @Override
14 public void analyze(@NonNull ImageProxy image) {
15 Image img = image.getImage();
16 bitmap = translator.
17 translateYUV(img, context: MainActivity.this).
18 copy(Bitmap.Config.ARGB_8888, isMutable: true);
19 preview.setRotation(image.getImageInfo().getRotationDegrees());
20
21 ObjUtils.getPredictions(getApplicationContext(), bitmap, preview);
22
23
24 image.close();
25
26
27
28
29 cameraProvider.bindToLifecycle( lifecycleOwner: MainActivity.this, cameraSelector, imageAnalysis, imageCapture);
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
```

Рисунок 2 –Фрагмент програмного коду для захоплення та обробки зображення з використанням CameraX

ІНТЕГРАЦІЯ МОДЕЛІ МАШИННОГО НАВЧАННЯ У ПРОЕКТ МОБІЛЬНОГО ДОДАТКУ ДЛЯ РОЗПІЗНАВАННЯ ОБ'ЄКТІВ

Для розпізнавання об'єктів на зображеннях, захоплених за допомогою камери мобільного пристрою була використана модель машинного навчання `ssd_mobilenet_v1_1_metadata_1.tflite`. Її інтеграція в проект доволі проста. Необхідно додати сам файл моделі в теку `ml` проекту мобільного додатку (рисунок 3).

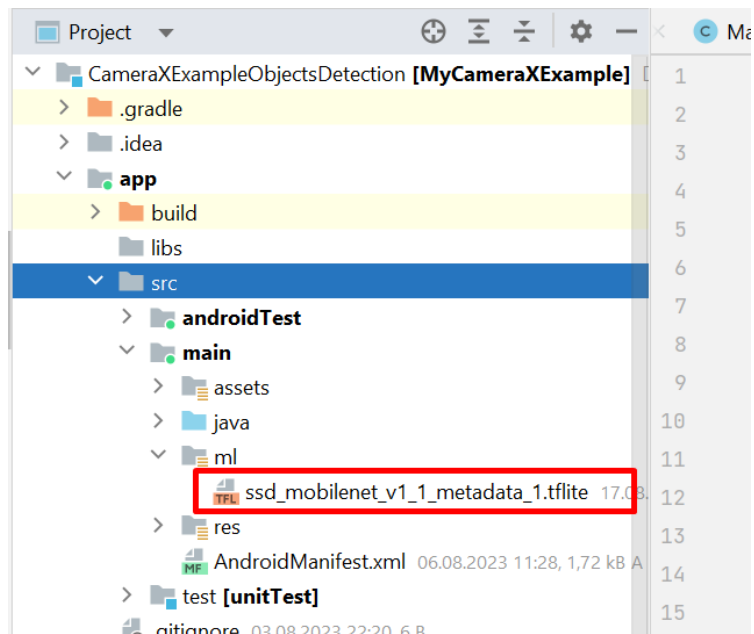


Рисунок 3 – Додавання моделі до проекту

Також необхідно додати опис об'єктів в assets (рисунок 4).

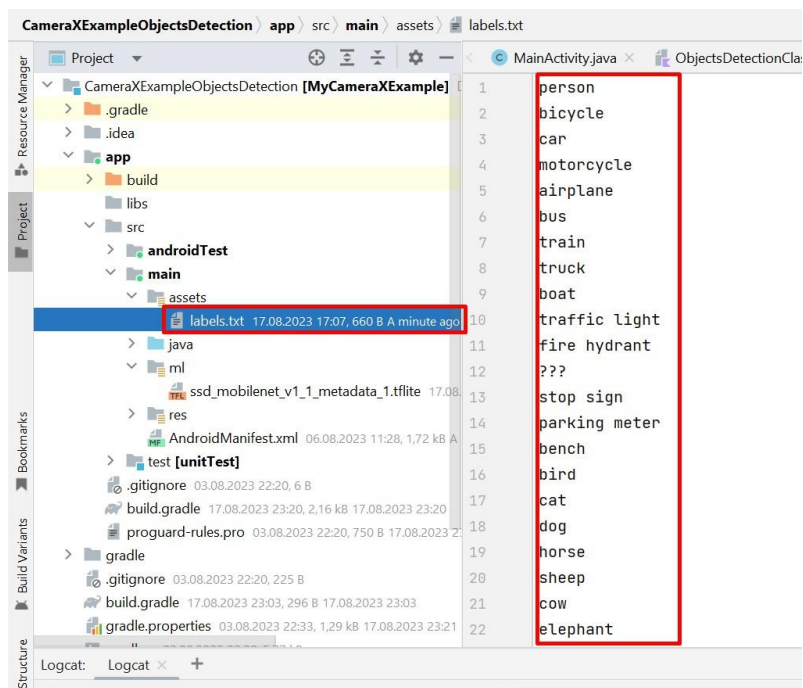


Рисунок 4 – Опис об'єктів, які може розпізнавати модель машинного навчання

ВИКОРИСТАННЯ МОДЕЛІ МАШИННОГО НАВЧАННЯ ДЛЯ РОЗПІЗНАВАННЯ ОБ'ЄКТІВ

Для підключення моделі використовується наступний фрагмент програмного коду (рисунок 5)

```

Labels = FileUtil.loadLabels(context, filePath: "Labels.txt")
model = SsdMobilenetV1Metadata1.newInstance(context)

```

Рисунок 5 – Підключення моделі

Для розпізнавання об'єктів на зображеннях та малювання на них bounding boxes (виділення розпізнаних об'єктів прямокутниками) використовується функція `get_predictions` (рисунок 6).

```

copy(Bitmap.Config.ARGB_8888, isMutable: true);
preview.setRotation(image.getImageInfo().getRotationDegrees());
ObjUtils.getPredictions(getApplicationContext(), bitmap, preview);
image.close();

```

Рисунок 6 – Використання `get_predictions` для розпізнавання об'єктів

Функція `get_predictions` має наступний вигляд (рисунок 7).

```

@JvmStatic
fun get_predictions(context: Context, bitmap: Bitmap, imageView: ImageView) {
    val paint = Paint()
    var colors = listOf<Int>(
        Color.BLUE, Color.GREEN, Color.RED, Color.CYAN, Color.GRAY, Color.BLACK,
        Color.DKGRAY, Color.MAGENTA, Color.YELLOW, Color.RED
    )
    lateinit var labels: List<String>
    lateinit var model: SsdMobilenetV1Metadata1
    val imageProcessor =
        ImageProcessor.Builder().add(ResizeOp( targetHeight: 300, targetWidth: 300, ResizeOp.ResizeMethod.BILINEAR))

    labels = FileUtil.loadLabels(context, filePath: "Labels.txt")
    model = SsdMobilenetV1Metadata1.newInstance(context)

    var image = TensorImage.fromBitmap(bitmap)
    image = imageProcessor.process(image)
    val outputs = model.process(image)
    val locations = outputs.locationsAsTensorBuffer.floatArray
    val classes = outputs.classesAsTensorBuffer.floatArray
    val scores = outputs.scoresAsTensorBuffer.floatArray

    var h = mutable.height
    var w = mutable.width

    paint.setTextSize = h / 15f
    paint.setStrokeWidth = h / 85f
    scores.forEachIndexed { index, fl ->
        if (fl > 0.7) {
            var x = index
            x += 4
            paint.setColor(colors.get(index))
            paint.style = Paint.Style.STROKE
            canvas.drawRect(
                RectF(
                    left: locations.get(x + 1) * w,
                    top: locations.get(x) * h,
                    right: locations.get(x + 3) * w,
                    bottom: locations.get(x + 2) * h
                ), paint
            )
            paint.style = Paint.Style.FILL
            canvas.drawText(

```

Рисунок 7 – Фрагмент програмного коду функції `get_predictions`

Також для коректої роботи мобільного додатку необхідно додати дозвіл на використання камери в `AndroidManifest.xml` (рисунок 8)

```

1 <?xml version="1.0" encoding="utf-8"?>
2 <manifest xmlns:android="http://schemas.android.com/apk/res/android"
3   xmlns:tools="http://schemas.android.com/tools"
4   <uses-permission android:name="android.permission.CAMERA"/>

```

Рисунок 8 – Дозвіл для використання камери

Необхідно також запросити дозвіл на використання камери у користувача у runtime режимі (рисунок 9).

```

if (ContextCompat.checkSelfPermission(context: this, android.Manifest.permission.CAMERA)
    != PackageManager.PERMISSION_GRANTED) {
    ActivityCompat.requestPermissions(activity: this, new String[] {android.Manifest.permission.CAMERA},
        PERMISSION_REQUEST_CAMERA);
} else {

```

Рисунок 9 – Запит дозволу на використання камери

Результати розпізнавання об'єктів з камери мобільного пристрою у реальному часі наведені на рисунку 10.

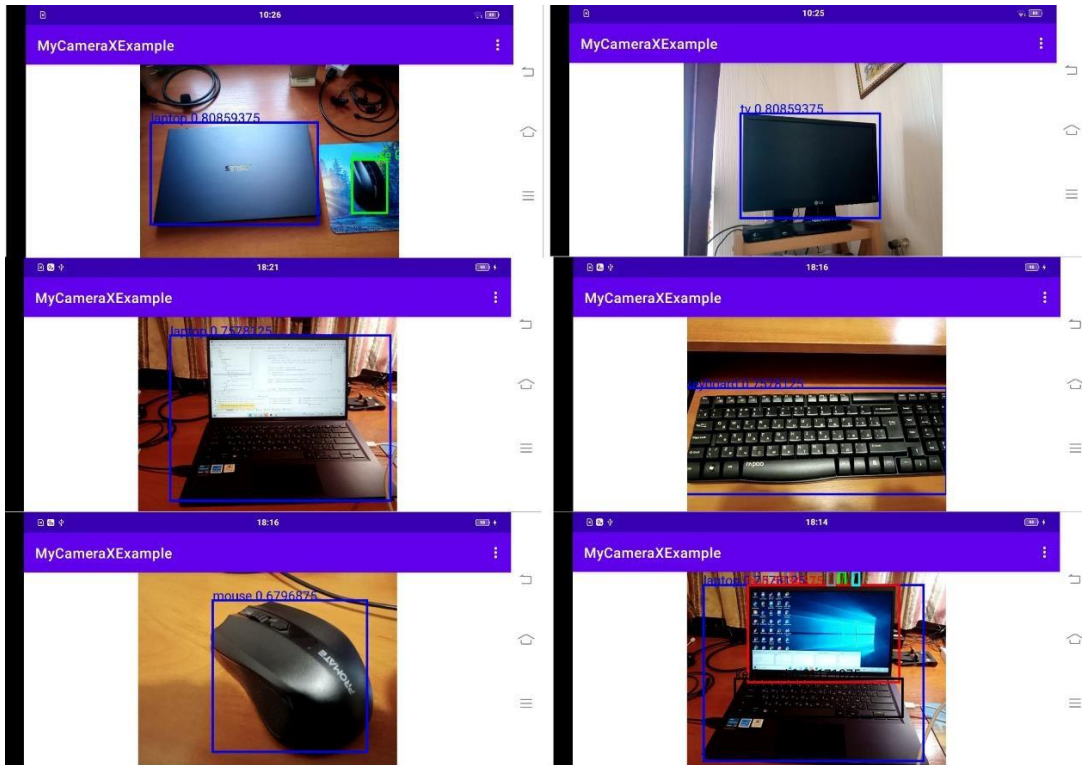


Рисунок 10 – Результати розпізнавання об'єктів з отриманих з камери мобільного пристрою у реальному часі

ВИСНОВКИ

В результаті виконаної роботи було розроблено мобільний (Android) додаток. Цей додаток здатен захоплювати та обробляти зображення з камери у реальному часі.

Для захоплення зображень був використаний клас CameraX. Для одночасного розпізнавання об'єктів використовувалася модель машинного навчання `ssd_mobilenet_v1_1_metadata_1.tflite`.

Також наведено результати розпізнавання об'єктів на зображеннях з камери у реальному часі.

СПИСОК ПОСИЛАНЬ

1. *TFLiteDetection [Electronic resource]* – Availble at: <https://github.com/joonb14/TFLiteDetection>
2. *CameraX Overview [Electronic resource]* – Availble at: <https://developer.android.com/training/camerax>
3. *IntelliJ IDEA – the Leading Java and Kotlin IDE [Electronic resource]* – Availble at: <https://www.jetbrains.com/idea/>

Бодяньський Є.В., Костюк С.О. (Харківський національний університет радіоелектроніки, Харків, Україна)

ПАРАМЕТРИЧНА ІДЕНТИФІКАЦІЯ АДАПТИВНИХ АКТИВАЦІЙНИХ ФУНКЦІЙ В ПОПЕРЕДНЬО НАВЧЕНИХ МОДЕЛЯХ ШТУЧНИХ НЕЙРОННИХ МЕРЕЖ

Анотація: В роботі розглядається питання автоматичного визначення параметрів адаптивних активаційних функцій в готових моделях, що мають сталу структуру та синаптичні параметри. У випадку багатошарової моделі штучної нейронної мережі, зміна одного з шарів викликає необхідність адаптації всіх подальших шарів, а отже введення адаптивних активаційних функцій в існуючу модель потребує обережної ініціалізації параметрів активаційних функцій для збереження апроксимаційної здатності попередньо навчених моделей. Таким чином, пропонується підхід градієнтної оптимізації для обрання значень параметрів активаційних функцій. Результативність та потенційні обмеження підходу досліджені експериментально на прикладі моделі KerasNet для класифікації зображень. З метою оцінки здатності функцій до адаптації, початкова форма адаптивних функцій обрана так, що вона не повторює форму фіксованих функцій в оригінальній моделі. Ключові слова: штучна нейронна мережа, адаптивна активаційна функція, доведення моделі.

Abstract: This paper investigates the topic of automatic parameter identification for adaptive activation functions in pre-trained models, i.e., models with pre-determined structure and synaptic weight values. In the case of a multi-layer neural network model, a change in one of the layers requires adapting all the following layers, so introducing adaptive activation functions into an existing model requires careful initialization of the activation function parameters to keep the approximation qualities of the pre-trained models. The gradient optimization approach is proposed as a solution for selecting the activation function parameter values. The experiment studies the performance and potential limitations of the proposed method using the KerasNet image classification model. The experiment sets the initial activation form and parameters to differ from the form of fixed functions in the original model to evaluate the adaptive properties of activation functions. Keywords: artificial neural network, adaptive activation function, model fine-tuning.

Вступ

Активаційні функції є одним з ключових блоків в моделях штучних нейронних мереж (ШНМ). Класичні активаційні функції, зокрема – сігма-функція, у своїй суперпозиції надають моделям властивостей універсального апроксиматора [1]. В той же час, різноманітні кусково-лінійні функції, як-от ReLU [2], PReLU [3] та інші, широко застосовуються в глибоких ШНМ через їх обчислювальну простоту, легке проходження інформації без ефектів зникаючого градієнту, а також їх фактичну достатність в прикладних задачах обробки даних [4].

Адаптивні активаційні функції дозволяють покращити апроксимаційні здатності мережі [5] шляхом додавання адаптивних параметрів, що підлаштовуються в процесі навчання на цільовому наборі даних. Адаптивні активаційні функції відрізняються за своєю природою та базовими блоками, з яких вони складаються. Наприклад, АНАФ [6] має в своїй основі сігма-функцію, PReLU – комбінацію лінійних функцій, а функція активації F-нейрону [7] – суперпозицію нечітких функцій належності, помножених на адаптивний коефіцієнт.

Практичне застосування адаптивних активаційних функцій ускладнене відсутністю попередньо навчених (pre-trained) моделей з такими функціями, що обмежує використання методів доведення параметрів (fine-tuning) та навчання з перенесенням інформації (transfer learning) [8], а отже призводить до потреби в додатковому часі та ресурсах на навчання таких моделей з нуля. Одним із методів подолання таких проблем є заміна фіксованих активаційних функцій на їх еквівалентні заміни [9], як-от заміна ReLU на PReLU, або SiLU [10] на АНАФ, однак існуючі методи передбачають ручний підбір значень параметрів для еквівалентної заміни та збереження решти синаптичних ваг та шарів у незмінному стані.

Метою даної роботи є вирішення задачі параметричної ідентифікації адаптивних активаційних функцій для існуючих мереж в автоматичному режимі в процесі fine-tuning.

1. Особливості заміни активаційних функцій

Штучні нейронні мережі можна описати як сукупність шарів нейронів, з'єднаних собою. Вихід кожного наступного шару L_i залежить від виходів попереднього шару L_{i-1} . Кожен шар L_i містить в собі набір синаптичних ваг w_i , що визначають вихід лінійної частини нейронів z_i , а на основі лінійного сигналу внутрішньої активації z_i формується вихід \hat{y}_i нейронів шару L_i . В мережах з прямим поширенням інформації вихід \hat{y}_n нейронів останнього шару L_n формує набір сигналів на виході всієї мережі \hat{y}_n .

В загальному випадку, зміна будь якого з параметрів w_1, w_2, \dots, w_i впливає на зміну вихідних значень $\hat{y}_1, \hat{y}_2, \hat{y}_3, \dots, \hat{y}_i$ відповідних шарів L_1, L_2, \dots, L_i , змінюючи розподілення значень на входах $x_2, x_3, x_4, \dots, x_{i+1}, \dots, x_n$ наступних шарів $L_2, L_3, \dots, L_{i+1}, \dots, L_n$. Аналогічно, заміна активаційної функції $f_i(z_i)$ в шарі L_i на альтернативну функцію $g_i(z_i)$ може призвести до зміни розподілення значень на входах наступних шарів, що веде до порушення проходження інформації в мережі, а в окремих випадках – до повної втрати мережею бажаних властивостей. У такому разі, зміна синаптичних ваг та активаційних функцій в попередніх шарах може потребувати корекції синаптичних ваг у всіх наступних шарах відповідно до оновленого розподілення вхідних значень таких шарів.

Таким чином, в задачах доведення параметрів мережі та навчання з перенесенням інформації практикують відкидання всіх шарів мережі, починаючи з шару L_i , замінюючи їх на нові шари, ініціалізовані з нуля. При цьому попередні шари L_1, L_2, \dots, L_{i-1} залишаються незмінними, а їх синаптичні ваги – фіксованими.

В задачах заміни активаційних функцій на адаптивні підбираються такі заміни, щоб оригінальна функція $f_i(z_i)$ являла собою окремий випадок функції-заміни $g_i(z_i)$, тобто $f_i(z_i) = g_i(z_i, p_i)$, де p_i – налаштовні параметри адаптивної функції [9]. Таким чином, розподілення вхідних значень x_i, \dots, x_{n-1}, x_n шарів L_i, \dots, L_{n-1}, L_n залишаються незмінними, що дозволяє повторно використати вже існуючі синаптичні параметри мережі w_i, \dots, w_{n-1}, w_n без додаткової адаптації та без викривлення інформації на виході мережі $\hat{y} = \hat{y}_n$.

Окремою задачею при заміні активаційних функцій є визначення набору параметрів p_i функції $g_i(z_i, p_i)$, які б дозволяли зберегти апроксимаційні здатності моделі на заданому наборі даних. Задача обрання таких параметрів ускладнюється, коли функції $f_i(z)$ та $g_i(z_i, p_i)$ відрізняються за своєю природою, тобто коли рівняння $f_i(z) = g_i(z_i, p_i)$ не може бути вирішене аналітичним чином. В такому разі можна звернутись до процедур оптимізації для вирішення задачі $f_i(z_i) \approx g_i(z_i, p_i)$, зокрема – до градієнтних інтервал, але для ряду функцій $g_i(z_i, p_i)$, зокрема – для нечіткої активації [7], умови $f_i(z_i) \approx g_i(z_i, p_i)$ можуть бути досягнені лише на визначеному інтервалі значень $z_i \in [z_{i,min}; z_{i,max}]$, де $z_{i,min} \neq -\infty$ та $z_{i,max} \neq +\infty$. Таким чином, критично важливим є визначення інтервалу значень $z_i \in [z_{i,min}; z_{i,max}]$, що поступають на вхід нейронів шару L_i мережі, для яких повинно виконуватись $f_i(z_i) \approx g_i(z_i, p_i)$. Для решти ж значень $z_i \notin [z_{i,min}; z_{i,max}]$ умовою $f_i(z_i) \approx g_i(z_i, p_i)$ можна знехтувати, оскільки вони не впливають на похибку E .

Одною з основних властивостей адаптивних активаційних функцій є їх здатність змінювати власні параметри форми та амплітуди в процесі навчання мережі, адаптуючись до особливостей поточної моделі, даних та вирішуваної задачі обробки інформації. Обрання неточної заміни $f_i(z_i) \neq g_i(z_i, p_i)$ для попередньо навченої моделі призводить до збільшення значення похибки E між очікуваними значенням на виході мережі у та фактичними виходами \hat{y} . За таких умов логічно припустити, що при фіксації синаптичних ваг $w = w_1, w_2, \dots, w_n$ моделі з використанням градієнтної процедури система буде змушена адаптувати параметри p_j, \dots, p_m $g_j(z_j, p_j), \dots, g_m(z_m, p_m)$ для зменшення значення похибки E та оптимізації цільової функції.

Таким чином, пропонуємо сформулювати задачу параметричної ідентифікації параметрів p_j, \dots, p_m адаптивних функцій $g_j(z_j, p_j), \dots, g_m(z_m, p_m)$ в попередньо навчених моделях у вигляді градієнтної процедури оптимізації цільової функції E . При цьому синаптичні ваги $w = w_1, w_2, \dots, w_n$ зберігаються фіксованими, а вільними залишаються лише параметри адаптивних активаційних функцій p_j, \dots, p_m . Для оптимізації використовується оригінальний набір даних, на якому була натренована вихідна модель штучної нейронної мережі. Використовуючи оригінальний набір даних, система отримує змогу налаштувати p_j, \dots, p_m безпосередньо на тому інтервалі даних $z_i \in [z_{i,min}; z_{i,max}]$, що має значення для визначення виходів \hat{y} .

Необхідною умовою для успішного налаштування параметрів p_j, \dots, p_m з використанням градієнтної процедури є підбір таких початкових значень $p_{j,init}, \dots, p_{m,init}$, щоб забезпечити пряме проходження інформації від входів x до виходів мережі $\hat{y} = \hat{y}_n$, тобто для часткових за параметрами p_j, \dots, p_m повинна виконуватись умова $dg_j(z_j, p_j)/dp_j \neq 0$.

2. Планування експерименту

Для експериментального підтвердження запропонованого методу розглядається проблема параметричної ідентифікації адаптивних активаційних функцій на прикладі моделі ШНМ KerasNet [11], навченої на наборі даних CIFAR-10 [12]. На відміну від досліджень в попередніх роботах [9], початкові значення параметрів $p_{j,init}, \dots, p_{m,init}$ функцій $g_j(z_j, p_j), \dots, g_m(z_m, p_m)$ навмисно обираються таким чином, щоб $f_i(z_i) \neq g_i(z_i, p_{i,start})$, тобто початкові значення параметрів не дозволяють задовільнити умови еквівалентності функцій. У якості адаптивних функцій використовуються АНАФ [6], LEAF [13], а також функція активації F-нейрону [7].

Загалом, експеримент демонструє здатність адаптивних активаційних функцій до зміни параметрів відповідно до вирішуваної задачі та архітектури моделі ШНМ. Одним з очікуваних результатів експерименту є поступова адаптація адаптивних активаційних функцій таким чином, щоб їх форма та амплітуда нагадували форму та амплітуду фіксованих активаційних функцій в оригінальній моделі ШНМ.

Для градієнтної процедури використовується оптимізатор Adam [14] з фіксованою швидкістю навчання $1e-3$. Оригінальні моделі з фіксованими активаційними функціями навчаються протягом 100 епох. Після 100 епох навчання синаптичні ваги оригінальної моделі фіксуються, а активаційні функції замінюються на адаптивні. Після заміни, навчання отриманої похідної моделі продовжується протягом додаткових 400 епох.

Заміни, використані в експерименті, описані в Таблиці 1.

Таблиця 1

Досліджувані похідні варіанти KerasNet з адаптивними активаційними функціями

№ п.п.	Оригінальна мережа		Похідна мережа	
	Згортк. шари	Повнозв. шари	Згортк. шари	Повнозв. шари
1	ReLU fixed	ReLU fixed	ReLU fixed	Fuzzy-as-Tanh
2	ReLU fixed	ReLU fixed	LEAF-as-ReLU	LEAF-as-Tanh
3	ReLU fixed	ReLU fixed	ReLU fixed	Fuzzy-as-Random
4	SiLU fixed	SiLU fixed	SiLU fixed	Fuzzy-as-Tanh
5	SiLU fixed	SiLU fixed	LEAF-as-SiLU	LEAF-as-Tanh
6	SiLU fixed	SiLU fixed	SiLU fixed	Fuzzy-as-Random
7	SiLU fixed	SiLU fixed	АНАФ-as-ReLU	АНАФ-as-ReLU
8	SiLU fixed	SiLU fixed	LEAF-as-ReLU	LEAF-as-ReLU
9	ReLU fixed	ReLU fixed	АНАФ-as-SiLU	АНАФ-as-SiLU
10	ReLU fixed	ReLU fixed	LEAF-as-SiLU	LEAF-as-SiLU

У підсумку:

- варіанти 1-6 досліджують адаптацію активаційних функцій для не споріднених активаційних функцій, для яких $f_i(z_i) \neq g_i(z_i, p_{i,start})$;
- варіанти 7-10 досліджують адаптацію відносно близьких функцій, для яких умова $f_i(z_i) \approx g_i(z_i, p_i)$ виконується за екстремальних значень $z_i = \pm\infty$.

В варіантах АНАF-as-ReLU, LEAF-as-ReLU параметри активаційних функцій обрані таким чином, щоб максимально близько повторювати форму фіксованої ReLU. В варіантах АНАF-as-SiLU, LEAF-as-SiLU – як точне повторення SiLU. У варіанті LEAF-as-Tanh – як точне повторення гіперболічного тангенсу (Tanh). У варіанті Fuzzy-as-Random використовується функція активації F-нейрону, параметри якої ініціалізовані рівномірним випадковим розподілом в діапазоні $[-1,0; +1,0]$. У варіанті Fuzzy-as-Tanh функція ініціалізована як кусково-лінійна апроксимація Tanh на інтервалі вхідних значень $[-4,0; +4,0]$.

В усіх експериментах параметри активаційних функцій є спільними для всіх нейронів кожного з шарів L_i . Таким чином, активаційні функції в процесі доведення змушені адаптуватись з урахуванням виходів всіх нейронів шару L_i одночасно.

Експеримент виконаний на комп'ютері з Nvidia RTX A4000, PyTorch 2.0.0 [15] та Python 3.10. Реалізація експерименту доступна на GitHub: github.com/s-kostyuk/aaf_identification, інтерактивний звіт – за посиланням: bit.ly/bk-learned-aaf-report.

3. Аналіз результатів експерименту

Рисунок 1 демонструє процес навчання похідних варіантів KerasNet під 1-3 порядковими номерами 1-3. Очікувано, заміна функції активації ReLU на Tanh та Random призводить до значного погіршення результатів роботи мережі на початкових епохах навчання. З кожною епохою, параметри активаційних функцій адаптуються, що дозволяє відновити точність класифікації. Треба відмітити, що LEAF-as-Tanh адаптується до нових умов значно повільніше за кусково-лінійні Fuzzy функції. Можливою причиною є початкове значення параметра $p_{1,start} = 0$, що обмежує проходження інформації та градієнтів в мережі.

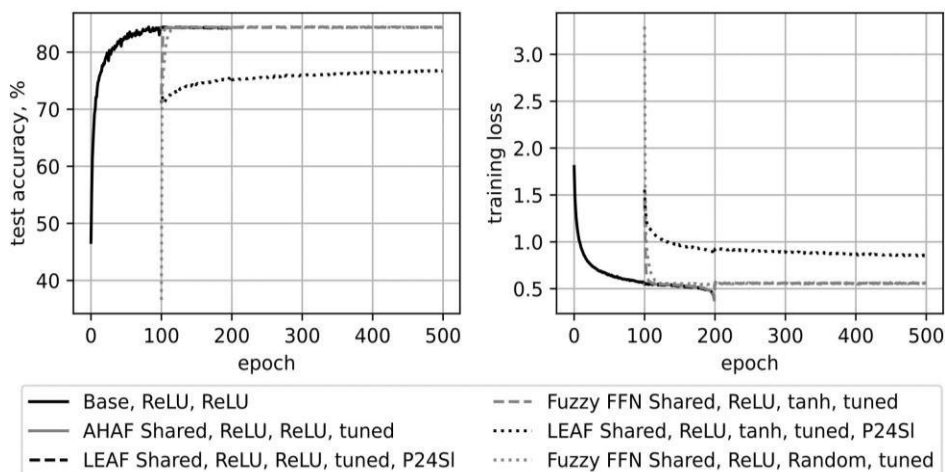


Рисунок 1 – Процес навчання похідних варіантів 1-3 моделі KerasNet

Цікаво, що за результатами доведення функції Fuzzy-as-Tanh та Fuzzy-as-Random набувають форм, подібних до функцій ReLU, тобто повторюють форму активаційних функцій в оригінальних моделях мережі. Рисунок 2 демонструє форму активаційних функцій Fuzzy-as-Tanh (а), LEAF-as-Tanh (б) та Fuzzy-as-Random (в), відповідно, за результатами навчання.

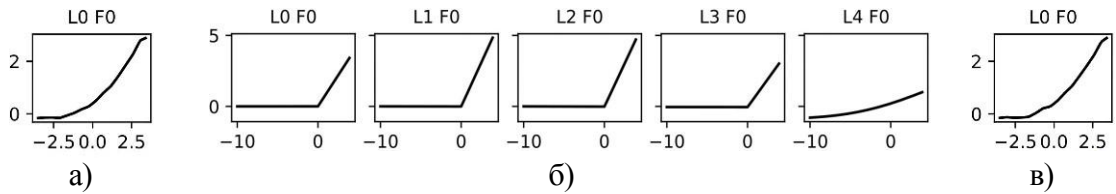


Рисунок 2 – Форма функцій Fuzzy-as-Tanh, LEAF-as-Tanh та Fuzzy-as-Random в моделях 1-3

При заміні SiLU на Tanh в варіантах 4-6 моделі KerasNet спостерігається значне погіршення якості моделі, що не компенсується додатковими епохами навчання. Як і у випадку з варіантами 1-3, LEAF-as-Tanh показує найгіршу здатність до зміни форми та відновлення характеристик моделі. Рисунок 3 відображає процес навчання, а рисунок 4 – форми функцій.

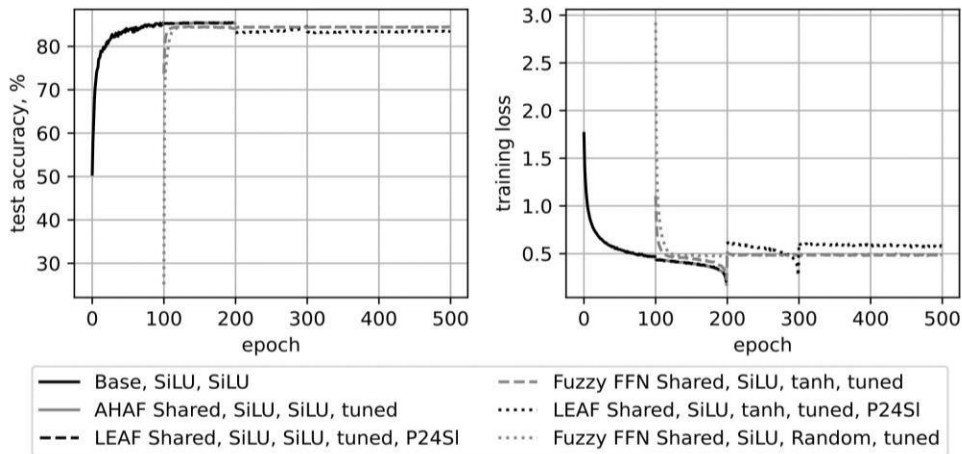


Рисунок 3 – Процес навчання похідних варіантів 4-6 моделі KerasNet

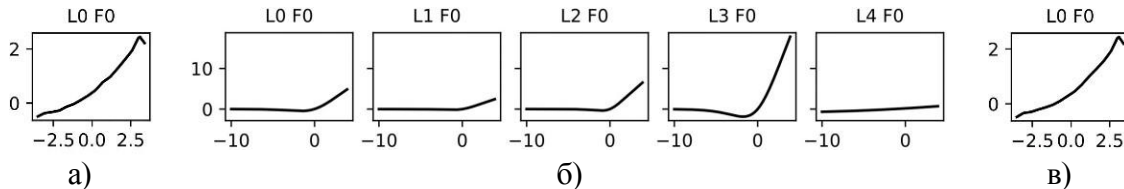


Рисунок 4 – Форма функцій Fuzzy-as-Tanh, LEAF-as-Tanh та Fuzzy-as-Random в моделях 4-6

В варіантах моделі 7-10 виконується заміна активаційних функцій на споріднені, а саме – з ReLU на SiLU та навпаки. При заміні SiLU на AHAF-as-ReLU та LEAF-as-ReLU активаційні функції не можуть успішно адаптувати власну форму, що зумовлено великим значенням та слабким градієнтом одного з початкових параметрів активаційних функцій, що відповідає за форму (γ для AHAF та p_3 для LEAF). При заміні ReLU на AHAF-as-SiLU та LEAF-as-SiLU спостерігається успішна адаптація активаційних функцій до форми ReLU.

Рисунки 5 та 6 зображують процес навчання похідних моделей 7-8 та 9-10, відповідно.

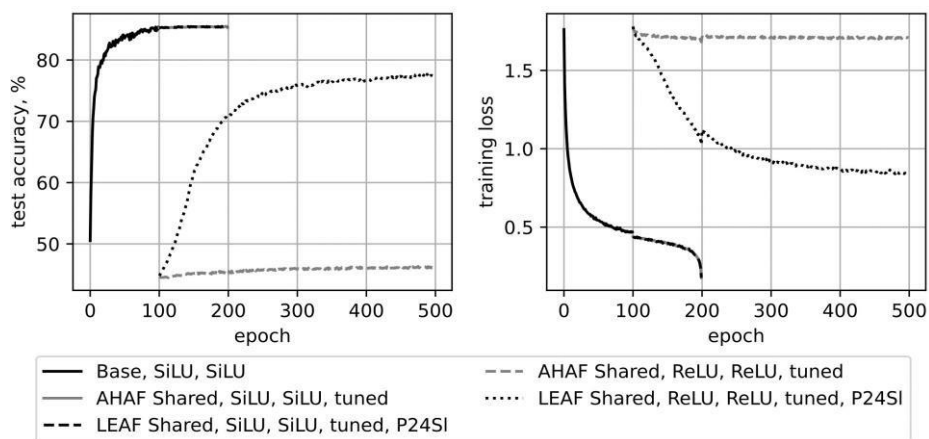


Рисунок 5 – Процес навчання похідних варіантів 7-8 моделі KerasNet

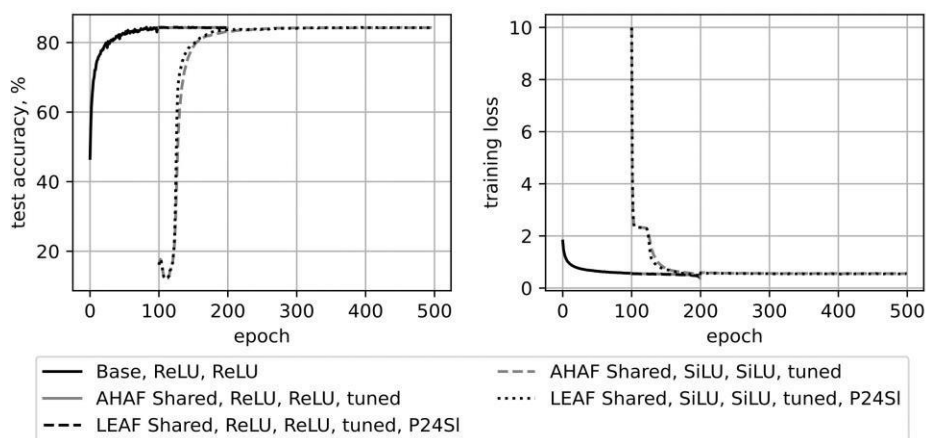


Рисунок 6 – Процес навчання похідних варіантів 9-10 моделі KerasNet

Рисунки 7 та 8 показують форми активаційних LEAF в моделях 8 та 10, відповідно. Варіанти з АНАФ схожі за формою, а тому їх ілюстрація опущена для стислості.

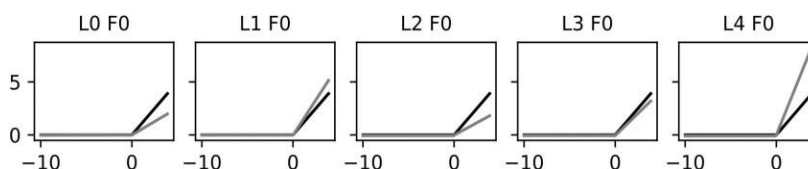


Рисунок 7 – Форма функції LEAF (сірим) в моделі 8 в порівнянні з ReLU (чорним)

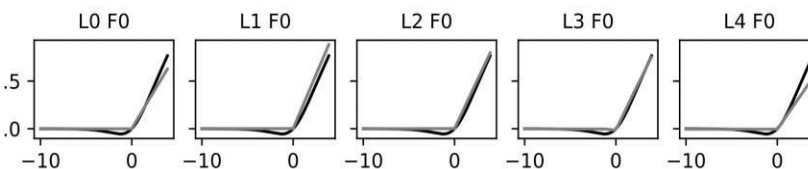


Рисунок 8 – Форма функції LEAF (сірим) в моделі 10 в порівнянні з SiLU (чорним)

ВИСНОВКИ

В роботі досліджено питання параметричної ідентифікації адаптивних активаційних функцій в автоматичному режимі. Запропоновано формулювання задачі параметричної

ідентифікації активаційних функцій як задачі оптимізації параметрів активаційних функцій з використанням градієнтної процедури на заданому наборі даних. Виділені особливості ідентифікації та заміни активаційних функцій в попередньо навчених моделях, зокрема – питання нееквівалентних замін та впливу змін в одному з шарів на всі наступні шари моделі.

Робота запропонованого методу досліджена на прикладі згорткової штучної нейронної мережі на наборі CIFAR-10. Експериментально визначено, що здатність адаптивної активаційної функції до адаптації власних параметрів значною мірою залежить від початкових значень параметрів активаційних функцій. За умов відповідної ініціалізації, активаційні функції відновлюють форму та амплітуду, що відповідають оригінальній моделі. Загалом, запропонований метод дозволяє виконувати ідентифікацію в автоматичному режимі.

Подальше покращення результатів та прискорення відновлення форми активаційної функції можливе за рахунок вдосконалення підбору початкових параметрів активаційних функцій, спрощення передачі інформації мережі, а також усунення перешкод на шляху зворотного поширення градієнтів в процесі навчання.

СПИСОК ПОСИЛАНЬ

1. Cybenko G. *Approximation by superpositions of a sigmoidal function*. *Mathematics of Control, Signals and Systems*. 1989. Vol. 2, no. 4. P. 303-314. URL: [10.1007/BF02551274](https://doi.org/10.1007/BF02551274).
2. Agarap A. F. *Deep Learning using Rectified Linear Units (ReLU)*. *CoRR*. 2018. Abs/1803.08375. URL: <http://arxiv.org/abs/1803.08375>.
3. *Delving Deep into Rectifiers: Surpassing Human-Level Performance on ImageNet Classification* / K. He et al. 2015 *IEEE International Conference on Computer Vision (ICCV), Santiago, Chile, 7–13 December 2015*. 2015. URL: <https://doi.org/10.1109/iccv.2015.123>.
4. Aggarwal C. C. *Neural Networks and Deep Learning*. Cham : Springer International Publishing, 2023. URL: <https://doi.org/10.1007/978-3-031-29642-0>.
5. *A survey on modern trainable activation functions* / A. Apicella et al. *Neural Networks*. 2021. Vol. 138. P. 14–32. URL: <https://doi.org/10.1016/j.neunet.2021.01.026>.
6. Bodyanskiy Y., Kostiuk S. *Adaptive hybrid activation function for deep neural networks*. *System research and information technologies*. 2022. No. 1. P. 87–96. URL: <https://doi.org/10.20535/srit.2308-8893.2022.1.07>.
7. Bodyanskiy Y., Kostiuk S. *Deep neural network based on F-neurons and its learning*. 2022. 11 p. (Preprint). URL: <https://doi.org/10.21203/rs.3.rs-2032768/v1>.
8. Paaß G., Giesselbach S. *Improving Pre-trained Language Models*. *Artificial Intelligence: Foundations, Theory, and Algorithms*. Cham, 2023. P. 79–159. URL: https://doi.org/10.1007/978-3-031-23190-2_3.
9. Bodyanskiy Y., Kostiuk S. *Using Adaptive Activation Functions in Pre-Trained Artificial Neural Network Models*. *Proceedings of the 11-th International Conference "Information Control Systems & Technologies"*, Odesa. 2023. P. 91–105. URL: <https://ceur-ws.org/Vol-3513/paper08.pdf>.
10. Elfving S., Uchibe E., Doya K. *Sigmoid-weighted linear units for neural network function approximation in reinforcement learning*. *Neural Networks*. 2018. Vol. 107. P. 3–11. URL: <https://doi.org/10.1016/j.neunet.2017.12.012>.
11. Chollet F., Others. *Train a simple deep CNN on the CIFAR10 small images dataset*. *Keras*. URL: https://github.com/keras-team/keras/blob/1.2.2/examples/cifar10_cnn.py.
12. Krizhevsky A. *Learning multiple layers of features from tiny images*. 2009.
13. Bodyanskiy Y., Kostiuk S. *Learnable Extended Activation Function for Deep Neural Networks*. *International Journal of Computing*. 2023. P. 311–318. URL: <https://doi.org/10.47839/ijc.22.3.3225>.
14. Kingma D. P., Ba J. *Adam: A Method for Stochastic Optimization*. 2017. (Preprint. arXiv:1412.6980). URL: <https://doi.org/10.48550/arXiv.1412.6980>.
15. *PyTorch: An Imperative Style, High-Performance Deep Learning Library* / A. Paszke et al. *Advances in Neural Information Processing Systems 32* / ed. by H. Wallach et al. 2019. P. 8024-8035. URL: <https://proceedings.neurips.cc/paper/2019/file/bdbca288fee7f92f2bfa9f7012727740-Paper.pdf>.

Воробель А. О. (Національний університет «Львівська політехніка», Україна)

ЗАСТОСУВАННЯ НЕЙРОМЕРЕЖЕВИХ ТЕХНОЛОГІЙ ДЛЯ ВИЯВЛЕННЯ ТА КЛАСИФІКАЦІЇ ПЕРЕЛОМІВ НА РЕНТГЕНІВСЬКИХ ЗНІМКАХ

Анотація: У статті досліджено проблему точності класифікації наявності перелому кістки на рентгенографічному зображенні. Актуальним підходом вважається застосування моделей сімейства YOLO для вирішення завдання виділення області перелому. Рентгенографічним зображенням властиве низьке значення контрасту в силу їх формування. З метою підвищення результатів класифікації розглянуто ряд методів попередньої обробки для підвищення контрасту таких зображень. В ході дослідження було проведено два експерименти: 1) мультикласове виявлення об'єктів, де фігурувало три класи тяжкості перелому, 2) однокласове виявлення, де модель тренувалась на виявленні єдиного класу "перелом". Завдяки попередній обробці зображень та звуженню множини класів вдалося значно збільшити показники точності Precision з 64.8% до 85.2%, Recall з 64.8% до 73.5%, mAP50 з 63.3% до 79.9% та mAP50-95 відповідно з 26.0% до 34.2%

Abstract: The article examines the problem of the accuracy of classification of the presence of a bone fracture on a radiographic image. The application of models of the YOLO family to solve the task of identifying the fracture area is considered a relevant approach. Radiographic images are characterized by a low contrast value due to their formation. In order to improve the classification results, a number of pre-processing methods were considered to increase the contrast of such images. In the course of the study, two experiments were conducted: 1) multi-class detection of objects, where three classes of fracture severity appeared, 2) single-class detection, where the model was trained to detect a single class "fracture". Due to pre-processing of images and narrowing of the set of classes, it was possible to significantly increase the accuracy indicators of Precision from 64.8% to 85.2%, Recall from 64.8% to 73.5%, mAP50 from 63.3% to 79.9%, and mAP50-95 from 26.0% to 34.2%, respectively.

Вступ

Сьогодні розробка моделей глибокого навчання для медичних зображень є одним із основних напрямків досліджень штучного інтелекту та машинного навчання. За допомогою цих моделей можна класифікувати різні типи зображень, наприклад рентгенівські, КТ та МРТ, а також виявляти патологію та інші зміни, які не завжди помітні людському оку.

Саме тому розробка моделей глибокого навчання для класифікації медичних зображень і виявлення патологій є важливим кроком на шляху до покращення діагностики та лікування захворювань. Використання цих моделей може допомогти покращити лікування та підвищити його ефективність, що може мати позитивний вплив на здоров'я та якість життя пацієнтів.

Рентгенівські знімки переломів є одними з видів медичних зображень. Їх виявлення залишається складним завданням для медичних фахівців і вимагає значних зусиль та часу. Іноді перелом може бути настільки непомітний, що пацієнт може навіть не підозрювати, що він є. Оптимізація процесу виявлення переломів є важливою задачею, оскільки швидка та точна діагностика є критично важливою для пацієнтів, які потребують невідкладної медичної допомоги.

Задача виявлення переломів на рентгенограмах займає своє місце в наукових роботах та широко досліджується, так у статті [1] розглядається проблема локалізації та ідентифікації захворювань на рентгенограмах хребта. Метою дослідження є розробка мережі для об'єктного виявлення, яка точно локалізує та класифікує аномальні ділянки на рентгенограмах хребта. Запропонована модель, AC-Faster R-CNN, використовує глибоке навчання та має структуру злиття ознак Deformable Convolution Feature Pyramid Network та структуру захоплення аномалій Abnormality Capture Head.

У дослідницькій роботі [4] зусилля були спрямовані на виявлення одного або кількох переломів в ліктьовій або променевої кістках за допомогою глибокого навчання, які були навчені на рентгенограмах верхніх кінцівок. Запропонований метод мав два основних етапи: препроцесінг, який включав видалення радіографічного фону та прибирання некісткової тканини, та використання глибоких моделей навчання, таких як RegNetX006, EfficientNet B0

та InceptionResNetV2, для обробки зображень, які містять тільки кісткову тканину, з метою виявлення одного або декількох переломів ліктьової або променевої кісток.

У роботі [10] досліджують використання глибокого навчання для виявлення та класифікації переломів ребер на рентгенограмах грудної клітки. У роботі автори представляють дві моделі для виявлення переломів, AB-YOLOv5 і PB-YOLOv5, які були навчені та оцінені на наборі даних EDARib-CXR. Ця структура демонструє покращені метрики на тестовому наборі даних, досягаючи значення AP30 на рівні 0.785.

У статті [17] дослідники розглядали можливість використання глибокого навчання для діагностики та оцінки важкості переломів нижніх кінцівок на звичайних рентгенограмах. Для розробки діагностичних моделей для виявлення та оцінки були використані мережа ResNet-50 та триплет-гілка. Результати показали, що моделі виявлення мають високу чутливість та добру специфічність для переломів тибіофібули та стопи та перевершують візуальну оцінку навіть від старших рентгенологів.

Методи дослідження

В даному дослідженні була використана модель YOLO 8-ї версії. YOLO, або “You Only Look Once”, — це популярна модель для виявлення об’єктів, відома своєю швидкістю та точністю. Ця модель є одноразовим детектором, що робить такі моделі обчислювально ефективними.

Модель YOLO пропонує використовувати наскрізну (end-to-end) нейронну мережу, яка робить передбачення рамок об’єктів (bounding boxes) і ймовірностей класів одночасно. Дотримуючись такого підходу до виявлення об’єктів, YOLO досягла дуже високих результатів, значно перевершивши інші алгоритми виявлення об’єктів у реальному часі.

З моменту першого випуску YOLO в 2015 році було розроблено кілька нових версій цієї моделі, кожна з яких базується та вдосконалює попередню. Остання її ітерація була випущена у 2023 році із номером версії 8. Ось хронологія, яка демонструє розвиток YOLO за останні роки.

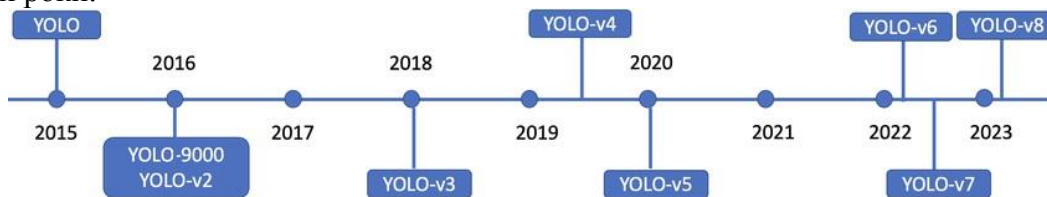


Рис.1. Хронологія розвитку YOLO

Тож можна побачити, що дана модель не просто популярна й користується успіхом, а й має великий потенціал для покращення. Саме тому для даної роботи була вибрана ця модель.

Архітектура YOLO складається із 3 основних частин:

- Backbone (хребет) - згортова нейронна мережа, яка об’єднує та формує характеристики зображення з різною деталізацією.
- Neck (шия) - серія шарів для змішування та комбінування функцій зображення, щоб передати їх у передбачення.
- Head (голова) - споживає функції з шиї та виконує кроки передбачення рамок та класу.

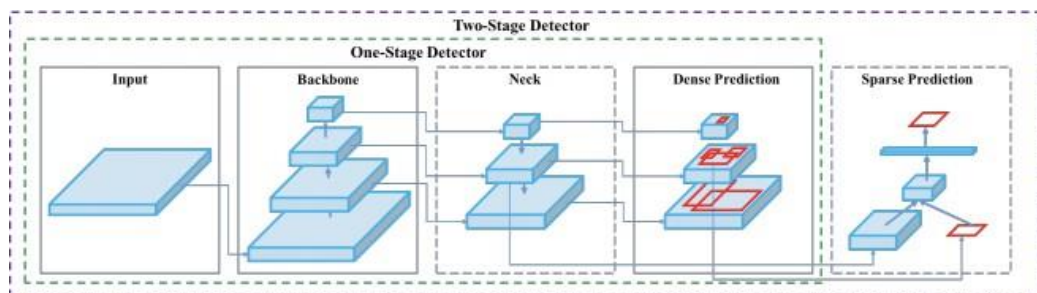


Рис.2. Архітектура YOLO

Алгоритм приймає зображення на вхід а потім використовує просту глибоку згорткову нейронну мережу для виявлення об'єктів на зображенні. Алгоритм YOLO ділить вхідне зображення на сітку $N \times N$. Якщо центр об'єкта потрапляє в клітинку сітки, ця клітинка відповідає за виявлення цього об'єкта. Кожна клітинка сітки передбачає прямокутники-рамки і оцінки достовірності для цих прямокутників, що позначають впевненість моделі в тому, що рамка містить об'єкт даного класу.

YOLO може передбачати кілька рамок на одну клітинку сітки. YOLO призначає один предиктор, який буде «відповідальним» за виявлення об'єкта на основі прогнозу, який має найвищий поточний IOU. IOU, або Intersection Over Union, — це метрика оцінки точності виявлення об'єкта, яка обраховує перекриття передбаченого регіону об'єкта із справжнім (вручну додана рамка, для навчання). Чим більше перекриття, тим краща оцінка.

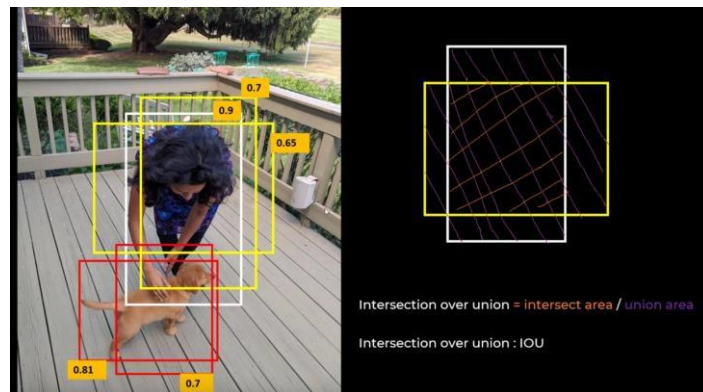


Рис.3. Візуалізація метрики IOU

Така робота алгоритму призводить до спеціалізації між предикторами рамки. Кожен предиктор стає кращим у прогнозуванні певних розмірів, пропорцій або класів об'єктів, покращуючи загальну оцінку запам'ятовування.

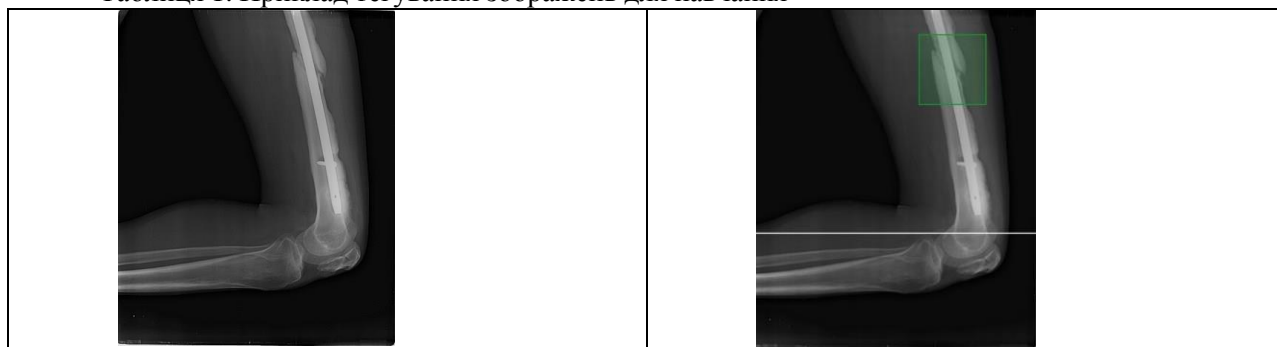
Набір даних та його обробка

Для цього дослідження було проаналізовано декілька датасетів, таких як “FracAtlas” [12], “bone fracture detection using x-rays” [9], “fracture types” [14], “long bones x rays” [16] та “long bones x rays” [16]. Кожен з них відрзнявся кількістю класів, розміров вибірки, наявністю чи відсутністю анотацій та якістю зображень. Для подальшої роботи було вирішено об'єднувати набори даних, задля збільшення навчальної вибірки та більшого різноманіття видів та місць переломів.

Також для самого навчання моделі потрібно вказати анотації (якщо їх немає), щоб модель “розуміла”, де знаходиться потрібний нам об'єкт. Для цього можна використати готові веб-застосунки [18], які набагато полегшують роботу з анотаціями (labeling) до зображень. Даний процес вимагає трохи ручної роботи.

Після завершення анотації зображень дані про встановлені рамки експортуються в спеціальному форматі, який очікує YOLO. Це є текстові (.txt) файли, що названі за ім'ям відповідного зображення, які містять 5 значень: порядковий номер класу та 4 координати рамки.

Таблиця 1. Приклад тегування зображень для навчання



Окрім цього було застосовано попередню обробку зображень для збільшення контрасту - CLAHE, або Contrast-Limited Adaptive Histogram Equalization. Цей метод застосовується для покращення контрастності зображень, що дає змогу моделі чіткіше виділити скелет на рентгенограмах. Цей метод відрізняється від звичайного вирівнювання гістограми тим, що адаптивний метод обчислює кілька гістограм, кожна з яких відповідає окремій ділянці зображення, і використовує їх для перерозподілу значень яскравості зображення. Таким чином, він може покращити локальний контраст і покращити чіткість країв у кожній області зображення.

Результати

Для мультикласового виявлення переломів були використана комбінація із двох датсетів — “fracture types” [14] та “long bones x rays” [16]. Цей набір має три класи які позначають складність перелому: А — 1 точка перелому, В — 2-3 точки, С — більше 3 точок. Оцінка результатів буде відбуватися за кількома метриками: Precision, Recall, mAP50, mAP50-95. Модель показала наступний результат по метриках: Precision — 64.8%, Recall — 64.8%, mAP50 — 63.3% та mAP50-95 — 26.0%. Ці значення можна інтерпретувати наступним чином: дана модель має посередню точність та може точно класифікувати більшу частину випадків, з яких найкраще класифікує перший клас.

Таблиця 2. Значення метрик для класів

Клас	Precision	Recall	mAP50	mAP50-95
всі	0.648	0.648	0.633	0.26
А	0.662	0.752	0.704	0.264
В	0.576	0.559	0.563	0.223
С	0.704	0.632	0.632	0.293

Причинами посередніх результатів мультикласового виявлення переломів можуть бути наступні моменти: 1) недостатньо якісні зображення — деякі зображення були низького розширення, 2) комбінація — набір був скомбінований із 2 інших наборів, які мали схожі класи, проте їх кількість була різною (в одного 3 класи, в іншого 9), 3) анотації — важко стверджувати, але є ймовірність, що набори даних у відкритому доступі можуть мати недостатньо коректно проставлені класи.

Нормалізована confusion-матриця (рис.5) показує, що доволі багато зображень без переломів (background) було передбачені як класи А та В, що підтверджує нахил до класу А.

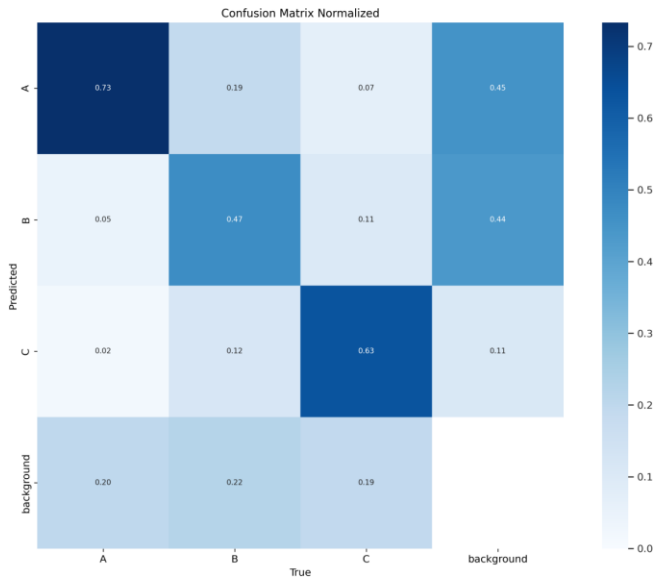


Рис.4. Нормалізована confusion-матриця

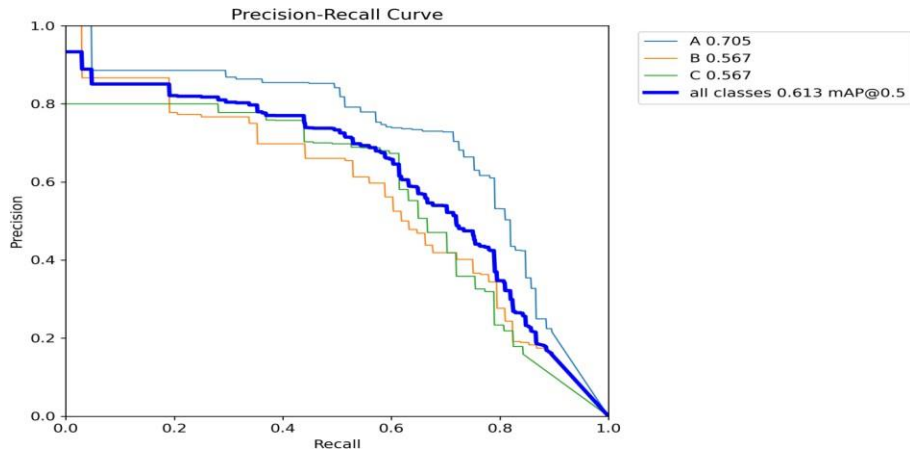


Рис.5. Графік кривої Precision-Recall

На тренувальній вибірці (рис.7) ми можемо спостерігати стабільне спадання функції для перебачення рамки, для передбачення класу можна помітити різке зниження на 75 епохах. Це можна пояснити тим, що на 75 епісі включається мозаїчна аугментація, що дозволяє краще навчитись моделі розрізняти класи.

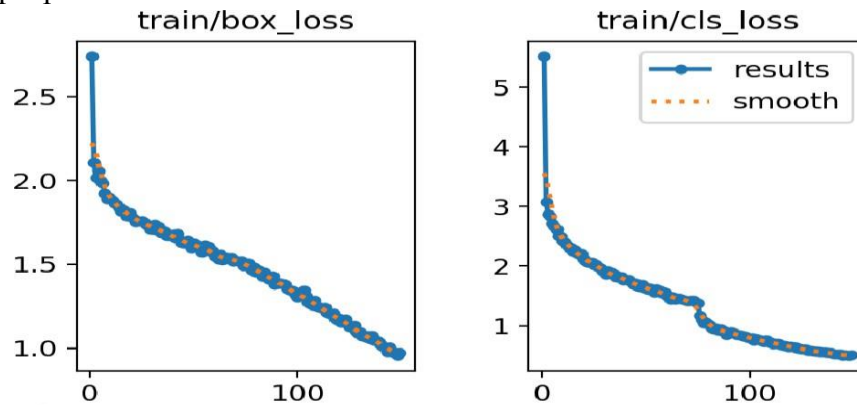



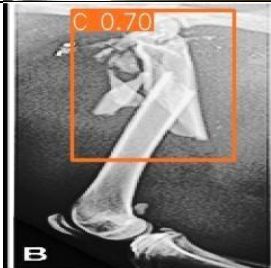


Рис.6. Графік функції втрат на тренувальній вибірці

Таблиця 3. Приклади роботи моделі

Оригінал	Передбачення
	
	

Також окрім мультикласового тренування було проведено експеримент з одним класом — “перелом”, який мав на меті перевірити точність оприділення місця перелому, без класифікації тяжкості. Тренування відбувалося на датасеті “Bone fracture detection” [7]. В результаті модель показала наступний результат по метриках: Precision — 85.2%, Recall — 73.5%, mAP50 — 79.9% та mAP50-95 — 34.2%.

Даний набір даних, мав зображення кращої якості, що також допомогло цій моделі перевищити результати першого експерименту, що показує важливість вибору якісного датасету.

На кривій Precision-Recall можна спостерігати доволі довгий проміжок з високим Precision, що означає високий шанс правильного передбачення класу. Проте із збільшенням значення Recall є суттєве зменшення точності Precision.

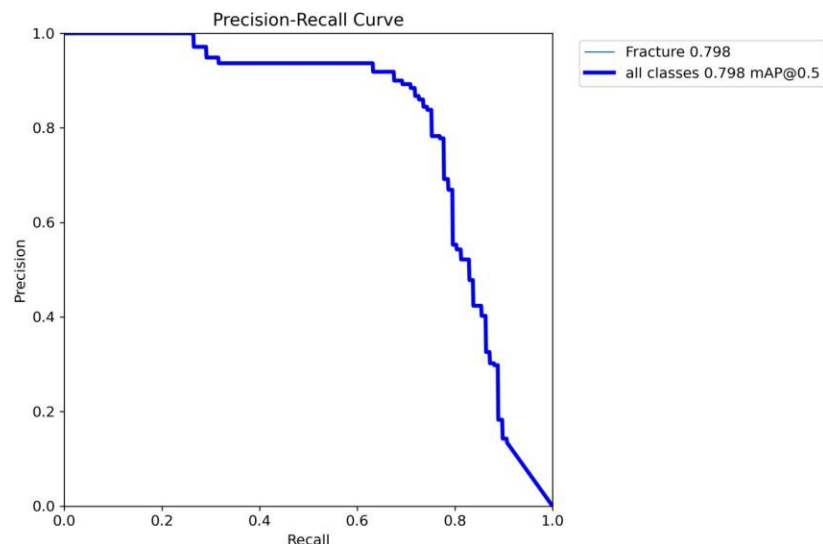






Рис. 7. Графік кривої Precision-Recall

Таблиця 4. Приклади передбачень однокласової моделі

Оригінал	Передбачення
	
	

Висновок

Якість медичних послуг є дуже важливим аспектом, оскільки чим точніше буде визначений діагноз, тим краще пройде лікування, а чим швидше буде визначений діагноз, тим більше шансів вчасно надати допомогу в лікуванні, чи навіть запобігти хворобі.

У даному дослідженні для поставленої задачі була використана модель YOLO останньої версії. Мультикласова стратегія, яка передбачала тренування моделі на датасеті із трьома класами, продемонструвала посередні результати. Ця модель виявилася менш точною, з Precision та Recall на рівні 0.648, та mAP50 у 0.633. Помітно, що складність датасету та, ймовірно недостатня якість, вплинула на загальну ефективність моделі. Натомість, однокласова стратегія, орієнтована на простіше завдання детекції переломів, продемонструвала кращі результати з Precision 0.852, Recall 0.735, та mAP50 0.799. Важливо відзначити, що цей експеримент проводився на іншому датасеті, що може пояснити поліпшення в ефективності.

У підсумку, модель YOLOv8 проявила свій потенціал у виявленні переломів на рентгенівських знімках, але вагомий вплив якості та репрезентації даних на результати підкреслює необхідність подальших досліджень та вдосконалення як самої моделі, так і якості наборів даних.

СПИСОК ПОСИЛАНЬ

1. Zhong B. AC-faster r-cnn: an improved detection architecture with high precision and sensitivity for abnormality in spine x-ray images / B. Zhong, J. Yi, Z. Jin // *Physics in medicine and biology*. — 2023. — Vol. 68, No. 19.
2. Kim M. W. Application of convolutional neural networks for distal radio-ulnar fracture detection on plain radiographs in the emergency room / M. W. Kim, J. Jung, S. J. Park, [et al.] // *Clinical and Experimental Emergency Medicine*. — 2021. — Vol. 8, No. 2. — P. 120–127.
3. Singh A. Automated detection of scaphoid fractures using deep neural networks in radiographs / A. Singh, A. A. Ardakani, H. W. Loh, [et al.] // *Engineering Applications of Artificial Intelligence*. — 2023. — Vol. 122.
4. Erdaş Ç. B. Automated fracture detection in the ulna and radius using deep learning on upper extremity radiographs / Ç. B. Erdaş // *Joint Diseases and Related Surgery*. — 2023. — Vol. 34, No. 3. — P. 598–604.
5. Lu S. Automated universal fractures detection in x-ray images based on deep learning approach / S. Lu, S. Wang, G. Wang // *Multimedia Tools and Applications*. — 2022. — Vol. 81, No. 30. — P. 44487–44503.

6. Guan B. Automatic detection and localization of thighbone fractures in x-ray based on improved deep learning method / B. Guan, J. Yao, S. Wang, [et al.] // *Computer Vision and Image Understanding*. — 2022. — Vol. 216.
7. Bone fracture detection / URL: <https://universe.roboflow.com/fracture-uofxm/bone-fracture-detection-ivsy6>.
8. Meena T. Bone fracture detection using deep supervised learning from radiological images: a paradigm shift / T. Meena, S. Roy // *Diagnostics (Basel, Switzerland)*. — 2022. — Vol. 12, No. 10. — P. 2420.
9. Bone fracture detection using x-rays / URL: <https://www.kaggle.com/datasets/vuppalaadithyasairam/bone-fracture-detection-using-xrays/data>.
10. Tsai H.-C. Cascading ab-yolov5 and pb-yolov5 for rib fracture detection in frontal and oblique chest x-ray images / H.-C. Tsai, N.-H. Lu, K.-Y. Liu, [et al.] // *IET Computer Vision*. — 2023. — Vol. 17, No. 7. — P. 750–762.
11. Cheng C.-T. Evaluation of ensemble strategy on the development of multiple view ankle fracture detection algorithm / C.-T. Cheng, C.-P. Hsu, C.-H. Ooyang, [et al.] // *The British Journal of Radiology*. — 2023. — Vol. 96, No. 1145. — P. 20220924.
12. FracAtlas: a dataset for fracture classification, localization and segmentation of musculoskeletal radiographs / figshare, 2023 / URL: https://figshare.com/articles/dataset/The_dataset/22363012.
13. Hardalaç F. Fracture detection in wrist x-ray images using deep learning-based object detection models / F. Hardalaç, F. Uysal, O. Peker, [et al.] // *Sensors*. — 2022. — Vol. 22, No. 3.
14. Fracture types object detection dataset and pre-trained model by modern academy for engineering and technology / URL: <https://universe.roboflow.com/modern-academy-for-engineering-and-technology/fracture-types>.
15. Yadav D. P. Hybrid sfnet model for bone fracture detection and classification using ml/dl / D. P. Yadav, A. Sharma, S. Athithan, [et al.] // *Sensors (Basel, Switzerland)*. — 2022. — Vol. 22, No. 15. — P. 5823.
16. Long bones x rays classification dataset and pre-trained model by modern academy for engineering and technology / URL: <https://universe.roboflow.com/modern-academy-for-engineering-and-technology/long-bones-x-rays>.
17. Wang Y. Lower-extremity fatigue fracture detection and grading based on deep learning models of radiographs / Y. Wang, Y. Li, G. Lin, [et al.] // *European Radiology*. — 2023. — Vol. 33, No. 1. — P. 555–565.
18. Make sense / URL: <https://www.makesense.ai/>.
19. Kim S. Multiclass datasets expand neural network utility: an example on ankle radiographs / S. Kim, P. Rebmann, P. H. Tran, [et al.] // *International Journal of Computer Assisted Radiology and Surgery*. — 2023. — Vol. 18, No. 5. — P. 819–826.
20. Kim T. Transfer learning-based ensemble convolutional neural network for accelerated diagnosis of foot fractures / T. Kim, T. S. Goh, J. S. Lee, [et al.] // *Physical and Engineering Sciences in Medicine*. — 2023. — Vol. 46, No. 1. — P. 265–277.
21. Shen L. Using artificial intelligence to diagnose osteoporotic vertebral fractures on plain radiographs / L. Shen, C. Gao, S. Hu, [et al.] // *Journal of Bone and Mineral Research*. — 2023. — Vol. 38, No. 9. — P. 1278–1287.

Ізонін І.В., Бляхар Р. С., Ткаченко Р. О. (Національний університет «Львівська політехніка», Львів, Україна)

НЕЙРОМЕРЕЖЕВИЙ МЕТОД ПОСЛІДОВНОЇ ГЛОБАЛЬНО-ЛОКАЛЬНОЇ АПРОКСИМАЦІЇ КОРОТКИХ НАБОРІВ МЕДИЧНИХ ДАНИХ

Анотація: В роботі представлено модифікацію методу глобально-локальної апроксимації коротких вибірок медичних даних за рахунок використання нейронної мережі узагальненої регресії замість RBF нейронної мережі для виділення локальної компоненти поверхні відгуку, що дозволило підвищити точність апроксимації на 7% за метрикою RMSE та зменшити тривалість часу виконання на 10% у порівнянні з базовим методом. Автори оптимізували продуктивність методу та дослідили його ефективність в порівнянні з низкою існуючих методів під час розв'язання задачі прогнозування вмісту підшкірного жиру в організмі людини.

Abstract: This paper presents a modification of the global-local approximation method by using a generalized regression neural network instead of an RBF neural network to extract the local component of the response surface, which allowed to increase the approximation accuracy by 7% in terms of RMSE and reduce the runtime by 10% compared to the basic method. The authors optimized the performance of the method and studied its efficiency in comparison with a number of existing methods in solving the problem of predicting the BMI.

Інтелектуальний аналіз та його інтерпретація у випадку опрацювання медичних даних мають надзвичайно високу актуальність у сучасній науці та медицині. Завдяки зростанню обсягу доступних медичних даних, вчені та медичні фахівці отримують безпрецедентну можливість вивчати хвороби, виявляти нові тенденції та залежності, а також удосконалювати методи діагностики та лікування.

Однак, саме завдяки розмаїттю та складності медичних даних виникають виклики, які потребують нових підходів та методів. Глобальна апроксимація даних, що описує усереднену поведінку моделі, недостатня для повного розуміння всіх аспектів медичних даних. Наприклад, вона може не враховувати індивідуальні особливості пацієнтів, відмінності між підгрупами хворих, або важливі локальні варіації. Тому локальні методи апроксимації набувають все більшої ваги. Вони дозволяють враховувати різні функції та аспекти даних в окремих точках або регіонах. Цей підхід особливо корисний для аналізу коротких наборів медичних даних, де різні регіони можуть виконувати різні функції, а також для детального вивчення специфічних медичних сценаріїв.

Зрозуміло, що глобальні та локальні методи мають свої обмеження. Глобальна модель може не відображати всіх нюансів та особливостей даних, а локальні моделі можуть бути недостатньо узагальненими для нових даних або не враховувати загальних тенденцій. Тому поєднання глобально-локальної апроксимації стає ключовим підходом.

Глобально-локальна апроксимація медичних даних надає можливість отримувати комплексне розуміння даних. Вона поєднує в собі гнучкість та широкий погляд глобальної моделі з точністю та адаптивністю локальних моделей. Цей підхід дозволяє виявляти складні залежності, розуміти індивідуальні реакції пацієнтів та враховувати контекстуальні фактори, що є критичними для ефективного прийняття рішень у медицині.

Однак існуючі методи глобально-локальної апроксимації можуть зазнавати обмежень, коли застосовуються на коротких наборах медичних даних. Це може бути пов'язано з недостатньою кількістю зразків, що ускладнює здатність моделей виявляти локальні особливості та деталі індивідуальних випадків. Такі обмеження вимагають уважного дослідження та розробки спеціалізованих підходів для ефективного застосування глобально-локальних методів на коротких медичних даних. Подолання цієї проблеми має важливе значення для забезпечення точності та надійності аналізу навіть у випадках з обмеженими

обсягами даних, що дозволить докладніше досліджувати клінічні сценарії та персоналізувати підходи до діагностики та лікування.

Метою роботи є підвищення ефективності прогнозування коротких наборів медичних даних на основі модифікації нейромережевого методу глобально-локальної апроксимації. Основною задачею, яку вирішує ця робота, є забезпечення більш точного аналізу медичних даних, зокрема у випадку їх обмежених обсягів, що може допомогти виявити складні залежності та зробити більш точні прогнози для важливих медичних показників.

У статті представлено покращений підхід до аналізу коротких наборів медичних даних, який базується на модифікації існуючого методу глобально-локальної апроксимації. Базовий метод використовує дві штучні нейронні мережі: нейроподібна структура МПГП для глобального аналізу та RBF для локального аналізу [1]. Застосування методу передбачає послідовну обробку даних з використанням нейромережевого аналогу PCA [2] для отримання глобальної компоненти перед використанням лінеаризованої нейроподібної структури МПГП та RBF нейронної мережі для аналізу локальних компонент. Значним недоліком базового методу є обмежена адаптивність RBF при роботі з малими наборами даних. RBF мережа вимагає великої кількості прикладів для навчання, і призводять до перенавчання при обмеженій кількості даних. Запропонована модифікація вирішує цю проблему, замінюючи RBF на GRNN (General Regression Neural Network) для аналізу локальної компоненти. GRNN володіє високими генерелізаційними властивостями і не потребує значної кількості прикладів для ефективного навчання. Таким чином, модифікований метод дозволяє отримувати точні та надійні результати аналізу медичних даних, особливо при роботі з невеликими обсягами інформації.

Структурно-функціональну схему запропонованого підходу наведено на рисунку 1.

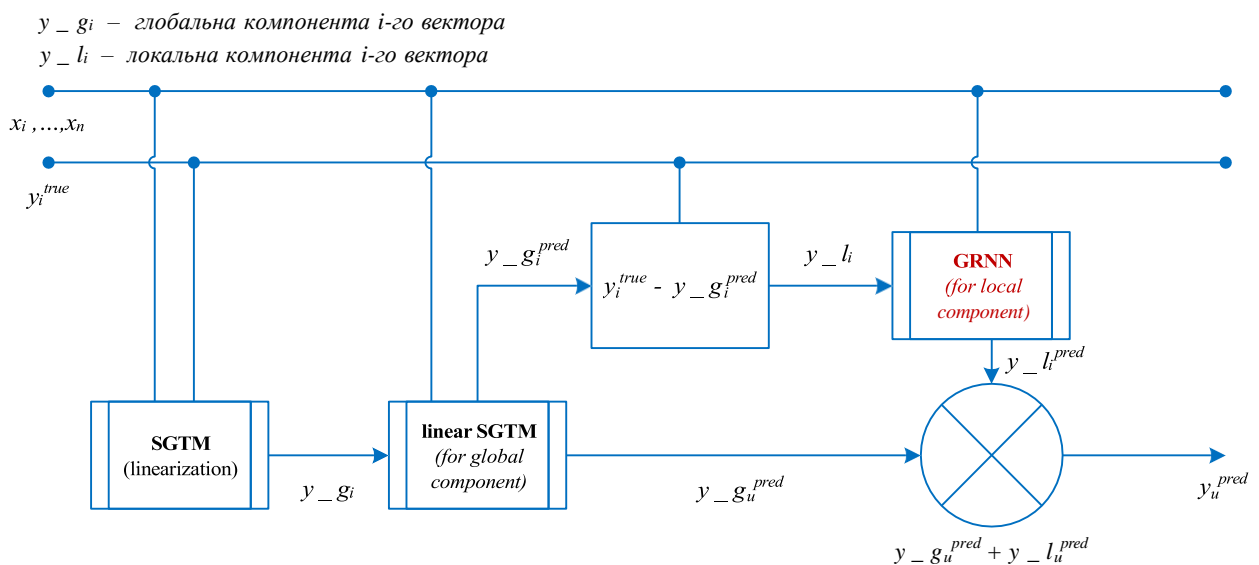


Рис. 1 Структурно-функціональна схема модифікації методу глобально-локальної апроксимації коротких наборів медичних даних

Для перевірки та порівняння ефективності модифікованого методу проведено серію експериментів на наборі даних невеликого розміру [3]. Усі експерименти були здійснені з використанням K-fold крос-валідації з N повтореннями, де $K = 5$ і $N = 20$. Порівняння запропонованого методу відбувалося із рядом існуючих методів. Зокрема дві ШНМ, які лежать в його основі, базовий метод та чотири існуючих методи які використовувалися авторами статті [5] для розв'язання поставленої задачі. Експерименти оцінювались на основі таких показників, як середньоквадратична помилка (RMSE), середня абсолютна помилка (MAE) та заміри часу виконання. Результати проведених експериментів зображені на рисунку 2.

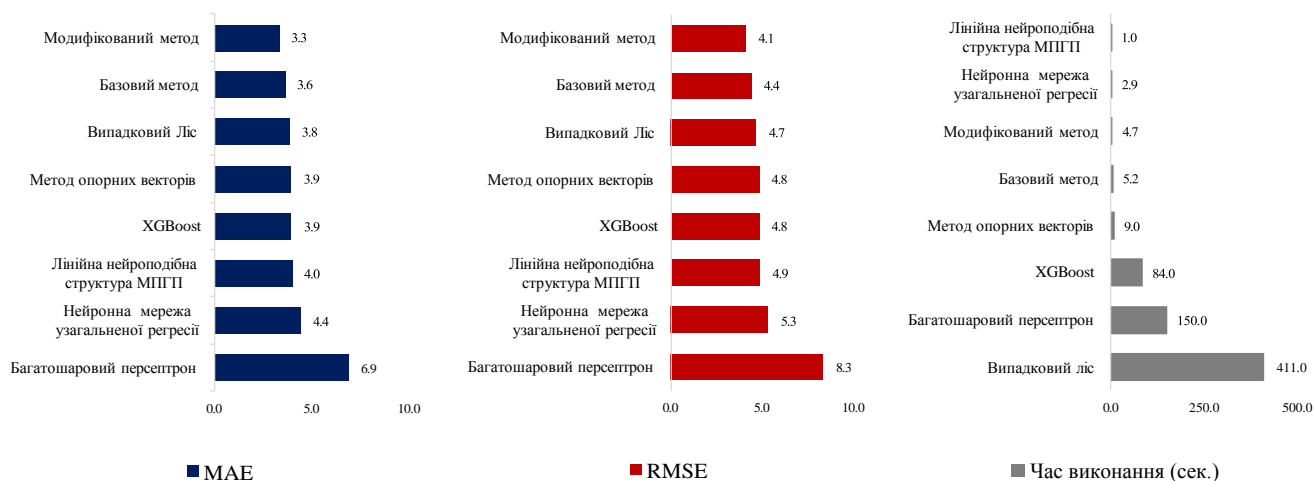


Рис. 2 Результати проведених експериментів

Як видно з наведеного рисунку, базовий та модифікований методи показали вищі показники точності, ніж методи, що досліджувалися у згаданій статті [5]. Окрім цього, модифікований метод демонструє найвищу точність серед усіх досліджуваних. Зокрема, він показав на 7% кращий результат за метрикою RMSE у порівнянні з базовим. Також, як видно з гістограми часу виконання – модифікований метод працює на 10% швидше ніж базовий. Це пояснюється, власне, використанням GRNN у запропонованій модифікації.

ВИСНОВКИ

У роботі представлено покращений підхід до інтелектуального аналізу коротких медичних даних, який має перспективи впровадження у медичну практику для покращення точності та ефективності аналізу даних. Проведені експерименти підтверджують високу ефективність та точність цього підходу, що робить його перспективним інструментом у сфері медичного аналізу даних.

СПИСОК ПОСИЛАНЬ

1. Tkachenko R, Doroshenko A, Izonin I, Tsymbal Y, Havrysh B. Imbalance Data Classification via Neural-Like Structures of Geometric Transformations Model: Local and Global Approaches. In: Hu Z, Petoukhov S, Dychka I, He M, editors. *Advances in Computer Science for Engineering and Education [Internet]*. Cham: Springer International Publishing; 2019. p. 112–22.
2. Tkachenko R. An Integral Software Solution of the SGTM Neural-Like Structures Implementation for Solving Different Data Mining Tasks. In: Babichev S, Lytvynenko V, editors. *Lecture Notes in Computational Intelligence and Decision Making*.
3. Ткаченко Р. О., Ткаченко П. Р., Ізонін І. В. *Нейромережеві засоби штучного інтелекту*. Львів, Львівська обл. : Львів. політехніка, 2017. 206 с.
4. Roger Johnson. *Body fat dataset [Internet]*. StatLi.. Available from: <http://lib.stat.cmu.edu/datasets/bodyfat>
5. *Body fat prediction through feature extraction based on anthropometric and laboratory measurements*. PLOS ONE. 2022;17(2):e0263333.

Кириченко М. Г. (Національний університет «Львівська політехніка», Україна)

ІНТЕЛЕКТУАЛЬНІ МЕТОДИ ОПТИМІЗАЦІЇ РОБОТИ ІНТЕРНЕТ-МАГАЗИНІВ

Анотація: У даній статті розглядається проблематика оптимізації роботи інтернет-магазинів за допомогою інтелектуальних методів. Особливу увагу приділено використанню чат-ботів та рекомендаційних систем, заснованих на алгоритмах машинного навчання та штучного інтелекту. Проаналізовано різні підходи до підвищення ефективності взаємодії між покупцями та торговельними платформами. В рамках дослідження було проведено експерименти з різними стратегіями оптимізації, використовуючи реальні дані з існуючих інтернет-магазинів. Розроблено чат-бот, який ефективно взаємодіє з повною базою знань інтернет-магазину, виконуючи широкий спектр завдань. Це включає надання відповідей на запитання користувачів у ролі асистента, а також використання зовнішніх інструментів, таких як рекомендаційні системи. Цей підхід дозволяє не тільки забезпечувати персоналізовану взаємодію, але й використовувати передові методи для оптимізації робочих процесів. Результати показали, що впровадження інтелектуальних методів, включаючи чат-боти та рекомендаційні системи, сприяє підвищенню задоволення клієнтів та оптимізації покупкової активності. Отримані дані свідчать про значне підвищення ефективності роботи інтернет-магазинів, в тому числі зростання конверсійних показників та поліпшення користувацького досвіду.

Abstract: This article examines the issue of optimizing the operation of online stores through intelligent methods. Special attention is given to the use of chatbots and recommendation systems based on machine learning algorithms and artificial intelligence. Various approaches to enhancing the interaction between customers and trading platforms have been analyzed. In the course of the study, experiments were conducted with various optimization strategies using real data from existing online stores. A chatbot has been developed that effectively interacts with the entire knowledge base of the online store, performing a wide range of tasks. This includes providing answers to customer queries in an assistant role, as well as utilizing external tools, such as recommendation systems. This approach not only ensures personalized interaction but also employs advanced methods for optimizing work processes. The results showed that the implementation of intelligent methods, including chatbots and recommendation systems, contributes to increasing customer satisfaction and optimizing shopping activity. The data indicate a significant increase in the efficiency of online store operations, including growth in conversion metrics and improvement of the user experience.

Вступ

Сучасний світ динамічно розвивається, а разом з ним і сфера електронної комерції, яка стала невід'ємною частиною життя суспільства. Інтернет-магазини, як основний елемент електронної торгівлі, за останні десятиліття перетворились з новинки у звичну реальність для більшості споживачів. Враховуючи цей факт, тема "Інтелектуальні методи оптимізації роботи інтернет-магазинів" є надзвичайно актуальною та вимагає детального дослідження та аналізу.

Імплементація інтелектуальних чат-ботів-асистентів відкриває нові горизонти для оптимізації різних аспектів інтернет-комерції. Ці інноваційні технології дозволяють значно покращити взаємодію з клієнтами, пропонуючи швидкі, точні та персоналізовані відповіді на їхні запитання. Використання чат-ботів допомагає зменшити час обробки запитів, підвищуючи задоволеність клієнтів та покращуючи їхній загальний досвід користування платформою.

Крім того, чат-боти здатні збирати та аналізувати великі об'єми даних, що дозволяє власникам інтернет-магазинів глибше зрозуміти потреби та уподобання своїх клієнтів. Це, в свою чергу, відкриває можливості для більш цілеспрямованих маркетингових кампаній та ефективнішого управління запасами.

В джерелах [2] та [31] особливу увагу приділено аналізу існуючих методів оптимізації та їхньої ефективності в контексті електронної комерції. Дослідження включає детальний розгляд практик великих компаній, таких як Amazon та Alibaba, з акцентом на їх підходи до оптимізації.

В іншому важливому джерелі [18], розглядаються сучасні техніки імплементації інтелектуальних чат-ботів та їх роль у підвищенні ефективності взаємодії з клієнтами. Автори аналізують, як розумні чат-боти можуть виконувати різноманітні функції, від відповідей на

запитання до інтеграції додаткових інструментів, та як це впливає на загальну динаміку роботи інтернет-магазинів.[16]

Окрім того, в джерелі [11] висвітлюється вплив глобалізації на стратегії маркетингу в епоху цифрових технологій. Аналізується, як передові технології, такі як GenAI, відкривають нові можливості для електронної комерції. Додатково, в роботі [12] розглядаються реалістичні вимоги та виклики, пов'язані з впровадженням GenAI, особливо для середнього бізнесу, в той час як в джерелі [23] наводиться аналіз потенційних викликів і можливостей, що виникають зі швидким розвитком цієї технології.

Кожне з цих джерел вносить значний вклад в розуміння широкого спектра аспектів, пов'язаних з оптимізацією роботи інтернет-магазинів, дозволяючи формувати комплексний погляд на цю актуальну тему.

Методи дослідження

Аналізуючи сучасні методи оптимізації роботи інтернет-магазинів, виявляється, що ключові стратегії охоплюють використання пошукової оптимізації (SEO), машинного навчання для персоналізації, соціальних мереж, аналітики даних та автоматизації маркетингу. Серед цих, чат-боти виступають як особливо важливий інструмент, що інтегрує багато з цих підходів, пропонуючи комплексне рішення для оптимізації.

SEO дозволяє підвищити видимість магазину в пошукових системах, залучаючи більше трафіку. Персоналізація через машинне навчання може збільшити продажі на 10-30%, адаптуючи пропозиції до потреб користувачів. Соціальні мережі допомагають у залученні клієнтів та підтримці бренду, тоді як аналітика даних дозволяє бізнесам більш точно налаштовувати свої стратегії та виробничі процеси. [24]

Однак, впровадження чат-ботів виявляється найбільш ефективним методом оптимізації, оскільки вони автоматизують обслуговування клієнтів, знижуючи витрати та підвищуючи задоволеність користувачів. Чат-боти відповідають на запитання клієнтів, допомагають у навігації по сайту, обробляють замовлення та інтегруються з рекомендаційними системами, надаючи персоналізовану взаємодію. Це сприяє зростанню продажів та підвищенню ефективності обслуговування, що робить чат-боти ключовим фактором успіху в динамічному ринку електронної комерції.[21]

Отже, враховуючи виклики сучасного ринку інтернет-торгівлі та швидкий розвиток технологій, чат-боти є найбільш оптимальним методом оптимізації, який забезпечує комплексне вирішення проблем та підвищення ефективності бізнес-процесів інтернет-магазинів.

Застосування чат-ботів є основною тенденцією останніх років і їхня популярність постійно зростає. Статистичні дані свідчать що все більше клієнтів надає перевагу користуватись чат-ботом, якщо це дозволить заощадити час.

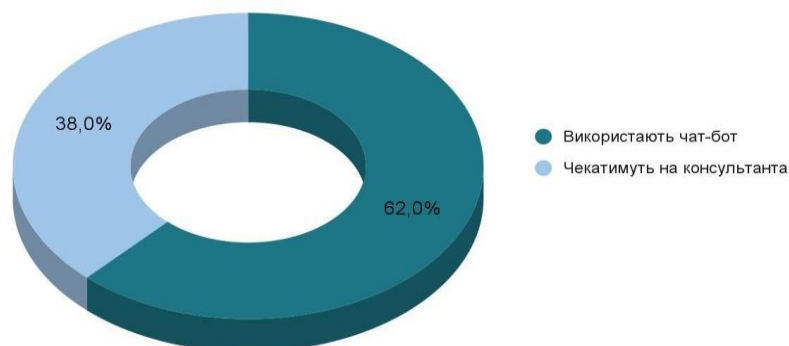


Рис. 1 Користування чат-ботом / консультантом

За минулий рік близько 9 з 10 користувачів користувались чат-ботом хоча б один раз. Більше того, 7 із 10 вважають цей досвід позитивним. Багато клієнтів висловлює думку, що компаніям слід більше використовувати чат-ботів, адже надання підтримку 24/7 та оперативність відповідей є ключовими для користувачів.

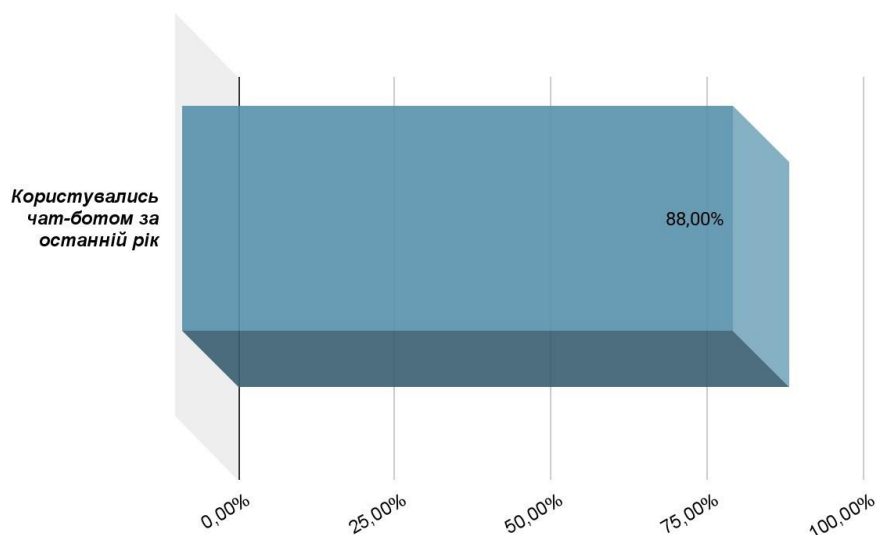


Рис. 2 Перший досвід користуванням

Основні проблеми, пов'язані з традиційними чат-ботами, включають труднощі з адаптацією до змінюваного контексту бесіди, високу вартість створення, обмеження в мультизадачності та персоналізації, а також проблеми з оновленням даних. Впровадження чат-ботів на основі LLM, таких як GPT-3.5, вносить значний прогрес у розв'язанні цих проблем, пропонуючи покращене розуміння контексту, генерацію більш природного тексту, і, як наслідок, більш гладку взаємодію з користувачами. [4] Ці чат-боти також відкривають нові можливості для бізнесу, покращуючи автоматизацію клієнтського обслуговування, точність відповідей на складні запити, персоналізацію обслуговування та ефективність маркетингових стратегій. У підсумку, використання LLM чат-ботів може суттєво вдосконалити досвід спілкування з клієнтами, знижуючи витрати компанії та підвищуючи задоволеність клієнтів.

GenAI не є стандартизованим поняттям, але з стрімким розвитком AI технологій за останній рік, він став загальновідомим. Генеративний штучний інтелект зазвичай відноситься до систем або алгоритмів AI, які здатні генерувати новий контент або дані, що не були явно закодовані в їх тренувальні дані. Це може включати, наприклад, створення нових зображень, текстів, музики або інших форм медіа на основі навчання на великій кількості даних. Ці системи часто базуються на складних моделях глибокого навчання, як-от генеративно-змагальні мережі (GAN) або трансформерні моделі, як GPT-3.5 від OpenAI. GenAI сприяє розвитку штучного інтелекту, що може генерувати нові ідеї, вивчати нові області та самовдосконалюватися. Використання GenAI в маркетингу може революціонізувати створення контенту, забезпечуючи персоналізоване спілкування та ефективніше вирішення завдань. [11] Ця технологія дозволяє створювати ботів, які покращують автоматизовані взаємодії та можуть ефективно спілкуватися з клієнтами, імітуючи людські розмови. GenAI також допомагає консультантам, надаючи дані клієнтів і пропонуючи рішення, що покращує час реакції та ефективність обслуговування. Однак, впровадження GenAI включає в себе технічні та організаційні виклики, які потребують обережного підходу для забезпечення ефективної інтеграції в існуючі системи.

Результати

Результатом даної роботи є імplementований чат-бот асистент інтернет магазину. Даний чат-бот працює з всією внутрішнією базою даних інтернет-магазину. Найбільша особливість даного чат-боту це його універсальність та легка інтеграція в будь який інтернет-магазин. Модель, яка виступає ядром чат-бота не тренується на даних інтернет-магазину, а швидше адаптовується під нього. В даній роботі використовується модель gpt-3.5-turbo та gpt-4.

Модель gpt-3.5-turbo може обробляти до 4096 токенів, але також доступний варіант gpt-3.5-turbo-16k, який може обробляти до 16384 токенів.

Запуск моделі відбувається за допомогою API запиту в якому один з параметрів є персональний ключ. Ці моделі натреновані на величезних об'ємах даних, що у вільному доступі, та вже навчені відповідати на питання користувача, а вот на які самі – це токенизований контекст, який повинен вміститись в модель. Алгоритм роботи чат боту можна розбити наступним чином

1. Створення внутрішньої бази знань для чат-бота.

Основний пріоритет чат-бота полягає у забезпеченні точної і актуалізованої відповіді користувачеві. Для цього бот оперує рядом джерел даних, які мають специфічну і строго визначену структуру, наприклад, наведу деякі з них:

а) Документація про інтернет магазин: цей ресурс зберігається в форматі .md і містить ключову інформацію про інтернет-магазин. Це може бути інформація про терміни доставки, гарантійні умови, основні засади роботи та інші загальні питання, які часто цікавлять користувачів.

б) База даних продуктів: зберігається в форматі .csv і містить деталізовані специфікації, ціни, наявність та інші характеристики товарів, представлених у магазині.

в) База даних користувачів: також зберігається у форматі .csv. Важливою особливістю є те, що ця інформація є строго персоналізованою, де кожен користувач з певним ID має окремий запис. Це робиться для забезпечення конфіденційності особистих даних користувачів.

г) Відгуки користувачів: файли, які зберігають відгуки та коментарі користувачів про товари або сервіс магазину, можуть бути представлені в форматі .json. Така структура дозволяє зберігати відгуки у вигляді структурованих записів, де кожен відгук може мати додаткові метадані (дата публікації, рейтинг тощо).

Ми розбиваємо документ на параграфи певного розміру – головна вимога, або кількість символів вміщались в вхідний шар моделі. Процес спліттингу буде відрізнятися в залежності до формату найбільш релевантного файлу.

Кожне з цих джерел даних створено таким чином, щоб чат-бот міг швидко та ефективно доступатися до потрібної інформації, реагуючи на запити користувачів. Всі джерела загружаються у векторну базу даних.

2. Визначення релевантності файлів до запиту користувача.

Після отримання запиту від користувача, алгоритм чат-бота аналізує його та з'ясовує, до якого типу інформації відноситься запит. Це може бути загальна інформація про магазин, деталі товару або особисті дані користувача, зокрема його історія покупок. На даному кроці ми визначаємо, в якому файлі з внутрішніх знань чат бота знаходиться інформація, на основі якої можна дати точну відповідь користувачу. Далше ми визначаємо контексти (найбільш релевантні параграфи до вхідного запитання з обраних файлів) на основі яких і буде будуватись відповідь користувача. Цей процес також називається embeddings creation. Створюємо дані ембедінги ми теж за допомогою моделі. Пошук релевантних документів в ве відбувається на основі вхідного запиту. Запит перетворюється на вектор за допомогою моделі Sentence Transformers, яка використовується для кодування тексту в векторному просторі. Sentence Transformers - це модель, заснована на трансформаторах, яка була навчена на задачі знаходження подібних речень. Вона використовує BERT, RoBERTa, DistilBERT або XLM моделі для генерації векторних представлень речень.

В даній роботі використовується OpenAIEmbeddings a same text-embedding-ada-002.

Навідміну від вищезгаданих моделей, де input token limit становить 256 токенів, даної він становить більш ніж 8 тисяч.

3. Надання відповіді на основі знайденого контексту.

На цьому кроці чат-бот формує відповідь на запит користувача. Залежно від того, який контекст було визначено на попередньому кроці, відповідь може братися з різних джерел інформації, як і з внутрішнього опису продукту наприклад, так і доповнюючи опис загальновідомими фактами, якими може орудувати AI модель.

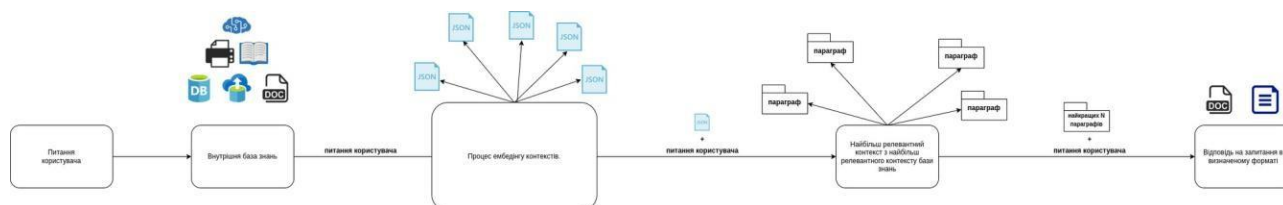


Рис. 3 Архітектура чат-бота

Одним з інструментів є рекомендаційна система. Її зміст для чат-бота це допомогти йому з вибіркою товарів по певному запиту. У своїй роботі я застосовую алгоритм колаборативної фільтрації (Collaborative Filtering, CF) для створення рекомендаційних систем. CF базується на припущенні, що користувачі, які згодні в одному випадку, ймовірно, будуть згодні й в інших. Мій підхід використовує user-based CF, де рекомендації генеруються на основі подібності між користувачами. По запиту користувача агент чат-боту визначає коли використовувати рекомендаційну систему.

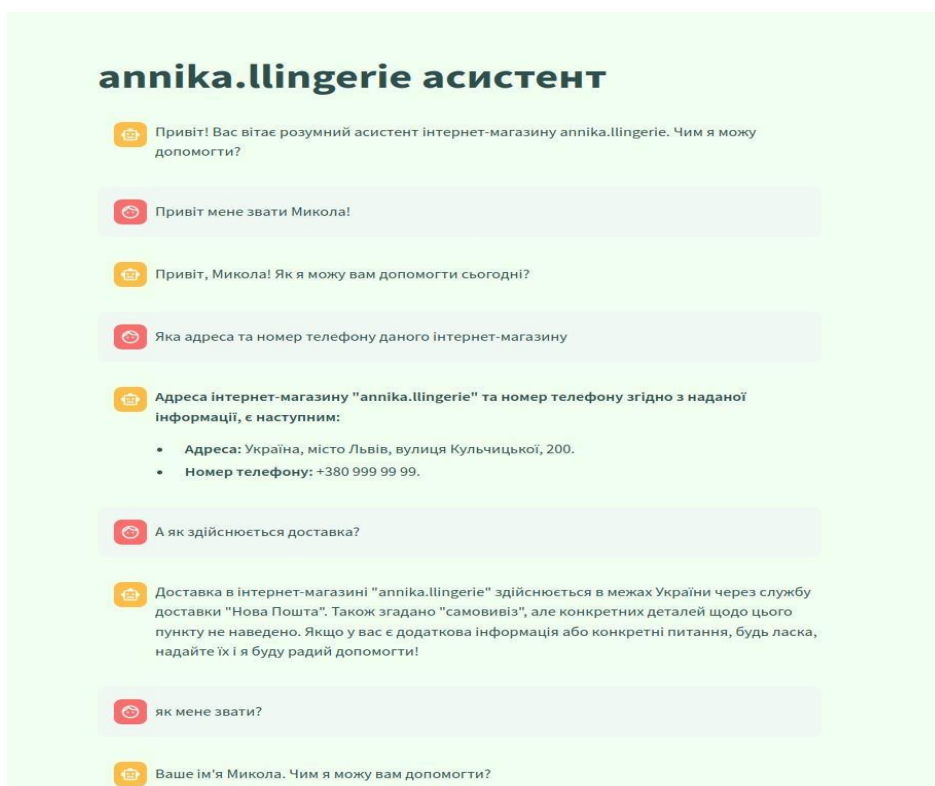


Рис. 4 Приклад демонстрації роботи пам'яті та доступу до бази знань

На даному прикладі продемонстроване запитання до внутрішньої документації інтернет-магазину ‘annika.lingerie’. Був знайдений відповідний корпус тексту, та збережений в пам’яті чату. На наступне запитання чат-бот знав де знайти відповідь без специфікації назви інтернет-магазину, тобто він пам’ятає попередній контекст. Також як видно чат-бот запам’ятав персональні дані наведені користувачем. Чат-бот може відповідати як і на питання з внутрішньої бази даних, так і вести просту бесіду. Тобто чат бот є універсальним і може допомогти як і з наданням базовних відповідей на основі даних, на який він був натренований (CommonCrawl, WebText, Вікіпедії та корпусу книг) так і з внутрішніми “вшитими йому” персональними даними інтернет-магазину.

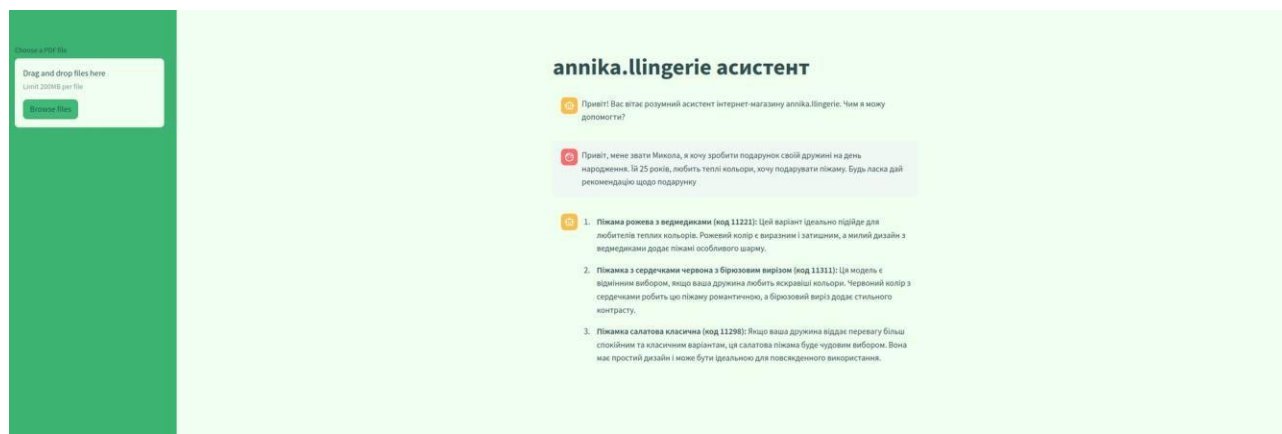


Рис. 5 Приклад інтеграції рекомендаційної системи з чат-ботом

В даному прикладі користувач попросив рекомендацію, щодо вибору товару. Тригернувся інструмент “Recomendation system” так як цей інструмент підходив найкраще. LLM за допомогою Pydantic визначила параметри для рекомендаційної системи(якщо користувача немає в базі, то буде зарекомендований продукт по загальним параметрам) та запустила функцію рекомендаційної системи. Як результати ми отримали товари, які нам повернула рекомендаційна система, яку в свою чергу чат-бот нам вивів в зручному для читання форматі.

Висновки

Ця робота акцентує увагу на важливості та перспективах розвитку інтелектуальних систем, особливо чат-ботів та рекомендаційних систем, у контексті оптимізації роботи інтернет-магазинів. Вона підкреслює роль інтернет-магазинів у сучасній електронній комерції, їх адаптацію до змінюваних умов ринку та потреб споживачів. Особлива увага приділяється розробці інтелектуального чат-бота, який спрощує взаємодію між покупцем та платформою, а також впровадженню рекомендаційних систем, які працюють у синергії з чат-ботом для надання персоналізованих пропозицій користувачам. Використання методів машинного навчання, LLM, забезпечує точність та ефективність запропонованих рішень. Практична імплементація та валідація розроблених методів і інструментів підтвердила їх ефективність та придатність для вирішення реальних задач інтернет-магазинів. Результати цієї роботи відкривають нові перспективи для подальшого розвитку інтелектуальних систем у сфері електронної комерції.

СПИСОК ПОСИЛАНЬ

1. Abdul Rahman M. S. Rashwan, Eitedal M. S. Alhelou: (2020) *The Impact of Using Artificial Intelligence on the Accounting and Auditing Profession in Light of the Corona Pandemic*
2. AliAyci, Shivam Tyagi: *The Role of Big Data Analysis and Machine Learning in Marketing Communications: A Case Study of Amazon*
3. Ali Borji QuinticAI (2023) *A Categorical Archive of ChatGPT Failures* 4. Amos Azaria (2022). *ChatGPT Usage and Limitations*
5. Anica-Popa, I., Anica-Popa, L., Rădulescu, C., & Vrîncianu, M. (2020). *The Integration of Artificial Intelligence in Retail: Benefits, Challenges and a Dedicated Conceptual Framework*
6. Bennett, M. (2023). *How generative AI fuels customer experience programmes* ComputerWeekly.
7. Blaise Agüera y Arcas(2022). *Do Large Language Models Understand Us?* *Daedalus*, 151(2):183–197
8. Bughin, J., Hazan, E., Ramaswamy, S., Chui, M., Allas, T., Dahlström, P., Henke, N., & Trench, M. (2017). *Artificial intelligence: the next digital frontier?* McKinsey & Company.
9. Capgemini. (2011). *Digital transformation: a roadmap for billion-dollar organization*
10. Ching-Yao Chuang, Varun Jampani, Yuanzhen Li, Antonio Torralba, and Stefanie Jegelka (2023). *Debiasing Vision-Language Models via Biased Prompts*
11. Chui, M., Hazan, E., Roberts, R., Singla, A., Smaje, K., Sukharevsky, A., Yee, L., & Zimmel, R. (2023). *The economic potential of generative AI*
12. Davenport, T. H., & Ronanki, R. (2023). *Artificial intelligence (AI) for the real world*
13. Debarag Banerjee, Pooja Singh, Arjun Avadhanam, Saksham Srivastava (2023) *Benchmarking LLM powered Chatbots: Methods and Metrics*
14. Dr. Shuchi Gautam and Dr. Shailender Kale: *Cost Optimization in E-Commerce Industry: An Analytical Study [THINK INDIA JOURNAL]*
15. Ernest Davis (2023). *Benchmarks for Automated Commonsense Reasoning: A Survey* 16. Gartner. *Chatbots Will Appeal to Modern Workers*
17. Gozalo-Brizuela, R., & Garrido-Mercha n, E. C. (2023). *A survey of Generative AI Applications*
18. Guendalina Caldarini , Sardar Jaf and Kenneth McGarry(2022) *A Literature Survey of Recent Advances in Chatbots*
19. Harreis, H., Koullias, T., Roberts, R., & Te, K. (2023). *Generative AI: Unlocking the future of fashion*
20. Jing Wei, Sungdong Kim, Hyunhoon Jung, Young-Ho Kim (2023) *Leveraging Large Language Models to Power Chatbots for collecting User Self-Reported Data*
21. Lihua Huang, Ganglan Hu, Xianghua Lu: *Research on E-business Ecosystem and Its Coordination Mechanism: A Case Study of Alibaba Group*
22. Likang Wu, Zhi Zheng, Zhaopeng Qiu, Hao Wang, Hongchao Gu, Tingjia Shen, Chuan Qin, Chen Zhu, Hengshu Zhu, Qi Liu, Hui Xiong, Enhong Chen(2023). *A Survey on Large Language Models for Recommendation*
23. Luan, H., Geczy, P., Lai, H., Gobert, J., Yang, S. J. H., Ogata, H., Baltes, J., Guerra, R., & Li, P. (2020). *Challenges and Future Directions of Big Data and Artificial Intelligence in Education*
24. Mohammad Waliul Hasanat, Ashikul Hoque, A. Bakar Abdul Hamid: *E-commerce optimization with the implementation of social media and SEO techniques to boost sales in retail*
25. Payal Dhar(2020). *The carbon impact of artificial intelligence* *Nat. Mach. Intell.*, 2(8):423–425
26. Rashad Yazdanifard & Merveen Tan Hunn Li: *The Review of Alibaba's Online Business Marketing Strategies Which Navigate them to Present Success*

27. *SoftServe. (6/ð). Generative AI: The race is on.*
28. *Som Biswas(2023) . ChatGPT and the Future of Medical Writing*
29. *Tom B. Brown, Benjamin Mann, Nick Ryder, Melanie Subbiah ma ihui (2020). Language Models are Few-Shot Learners*
30. *Vinu Sankar Sadasivan, Aounon Kumar, Sriram Balasubramanian, Wenxiao Wang, and Soheil Feizi(2023). Can AI-Generated Text be Reliably Detected?*
31. *Zana Majed Sadq, Hawre Nuraddin Sabir:Analyzing the Amazon success strategies (2018)*
32. *Zon, Maryia Fokina (2023): The Future of Chatbots*

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

ОСВІТНЬО-КОНСУЛЬТАЦІЙНИЙ ЦЕНТР ШТУЧНОГО ІНТЕЛЕКТУ ДЛЯ ВСІХ

Анотація: В роботі надані перспективи розвитку та соціального впливу. Зазначено ключові аспекти "Освітньо-консультаційного центру штучного інтелекту", його місію в розвитку навичок та розуміння у сфері штучного інтелекту серед всіх верств населення, сприяючи соціальному та економічному поліпшенню міста. Запропоновані рекомендації для залучення підтримки від власних структур та крупного бізнесу.

Abstract: Perspectives of development and social influence are provided in the paper. The key aspects of the "Artificial Intelligence Educational and Consulting Center", its mission to develop skills and understanding in the field of artificial intelligence among all segments of the population, contributing to the social and economic improvement of the city, are indicated. Suggested recommendations for attracting support from own structures and large businesses.

В сучасному світі освіта стає стратегічним фактором для забезпечення стабільного економічного розвитку та успішної інтеграції в глобальну спільноту. Ключове значення в цьому контексті має створення "Освітньо-консультаційного центру штучного інтелекту", який спрямований на розвиток розуміння та освоєння основних навичок в галузі застосування штучного інтелекту на нейронних мережах.

Широкий діапазон програм.

Центр відкритий для всіх громадян міста, незалежно від віку, освіти чи професії. Пропонуючи інноваційні програми та консультації, Центр має на меті відповідати потребам різних соціальних груп.

Кардинальне поліпшення соціального та економічного простору.

Створення Центру спрямоване на кардинальне поліпшення соціального та економічного розвитку міста, маючи на увазі:

- Зменшення соціальної нерівності, надаючи доступ до освіти та професійного розвитку усім верствам населення.
- Підвищення якості життя, забезпечуючи громадянам конкурентоспроможні знання та навички, необхідні для успіху в ері нових технологій та інновацій.
- Сприяння економічному розвитку через залучення інвестицій та створення нових робочих місць.

Центр як каталізатор змін

Очікується, що Центр стане потужним каталізатором позитивних змін у місті, сприяючи створенню більш інклюзивного та справедливого суспільства.

Рекомендації для підтримки центру

Для успішної реалізації проекту, необхідно:

- Привертати увагу власних структур і крупного бізнесу. Залучення сильних партнерів забезпечить ресурси та експертну підтримку.
- Розробляти чітку стратегію, яка враховує потреби суспільства та партнерів. Це дозволить Центру ефективно взаємодіяти з різними зацікавленими сторонами.

За допомогою цих заходів "Освітньо-консультаційний центр штучного інтелекту" може стати джерелом подальших позитивних змін для нашого міста.

Конкретні приклади того, як Центр може допомогти покращити соціальний і економічний простір міста:

- Центр може надавати безкоштовні або доступні за ціною курси з штучного інтелекту для людей, які не мають можливості отримати освіту в традиційному університеті. Це допоможе зменшити соціальну нерівність і надати можливість для професійного розвитку людям з усіх верств суспільства.

- Центр може розробляти програми навчання, які адаптовані до потреб різних груп населення. Наприклад, для людей похилого віку Центр може запропонувати курси з штучного інтелекту, які допоможуть їм використовувати нові технології в повсякденному житті.
- Центр може співпрацювати з місцевими підприємствами, щоб розробити програми навчання, які відповідають потребам їхніх працівників. Це допоможе підвищити продуктивність праці та заохотити інновації в місцевому бізнесі.

Особливості застосування можливостей штучного інтелекту

Ідея Освітньо-консультаційного центру штучного інтелекту має особливі акценти саме на використанні особливостей застосування можливостей штучного інтелекту. Це означає, що в Центрі не просто навчатимуть людей основам штучного інтелекту, а й допомагатимуть їм розвинути навички та знання, необхідні для ефективного використання штучного інтелекту в реальному світі.

Деякі з особливостей застосування можливостей штучного інтелекту, на які ідея Центру робить особливий акцент, включають:

- Інтеграція штучного інтелекту в різні сфери життя. Штучний інтелект може бути використаний в різних сферах життя, таких як бізнес, освіта, медицина, транспорт та інші. Центр допомагає людям зрозуміти, як штучний інтелект може бути використаний для вирішення конкретних проблем і покращення якості життя.
- Етика використання штучного інтелекту. Використання штучного інтелекту може мати як позитивні, так і негативні наслідки. Центр навчатиме людей про етичні аспекти використання штучного інтелекту та про те, як використовувати штучний інтелект відповідально.
- Безпека та безпека використання штучного інтелекту. Штучний інтелект може бути використаний для створення нових загроз безпеці та безпеці. Центр навчає людей про потенційні загрози, пов'язані з використанням штучного інтелекту, та про те, як захиститися від цих загроз.

Центр також допомагатиме людям розвивати навички критичного мислення та аналізу, необхідні для ефективного використання штучного інтелекту. Ці навички допомагають людям оцінювати достовірність інформації, яка генерується штучним інтелектом, і приймати обґрунтовані рішення на основі цієї інформації.

За допомогою цих зусиль Центр допомагатиме людям зрозуміти, як штучний інтелект може бути використаний для вирішення конкретних проблем і покращення якості життя, планувати будь які заходи і втілювати їх в життя на підставі оптимізації використання всіх наявних ресурсів.

Майстренко О.С., Ключин Д.А. (Київський національний університет імені Тараса Шевченка м. Київ, Україна)

ДІАГНОСТИКА РАКУ МОЛОЧНОЇ ЗАЛОЗИ ЗА ДОПОМОГОЮ ЗГОРТКОВОЇ НЕЙРОМЕРЕЖІ І ФРАКТАЛЬНОГО АНАЛІЗУ ЯДЕР БУКАЛЬНОГО ЕПІТЕЛІЮ

Анотація: Мета публікації — описати новий ефективний метод скринінгу раку молочної залози, заснований на дослідженні фрактальних властивостей хроматину в пофарбованих за Фельгеном ядрах буккального епітелію за допомогою машинного навчання. Спочатку йде короткий огляд документів про зміни, пов'язані зі злякисними новоутвореннями. Далі описано морфометричне дослідження та аналіз зображення зміни, пов'язані зі злякисними новоутвореннями в буккальному епітелії. Після цього описується фрактальний аналіз хроматину. Маючи фрактальну інтерпретацію вхідних зображень, було побудовано згорткова нейронна мережа з описом архітектури.

Abstract: The purpose of the publication is to describe a novel effective method for screening of breast cancer based on investigation of fractal properties of chromatin in Feulgen-stained nuclei of buccal epithelium using machine learning. It starts with a short survey of papers on the malignancy-associated changes. Then the morphometric research and image analysis of malignancy-associated changes in buccal epithelium are described. After that it describes the fractal analysis of the chromatin. Having fractal interpretation of input images, convolutional neural network is built, proceeded with architecture description.

Сучасний стандарт для діагностики раку молочної залози включає проведення клінічного обстеження, мамографії та аспіраційної біопсії. Цей підхід забезпечує високу точність діагностики, проте мамографія включає в себе вплив радіації, а процедура аспіраційної біопсії може призводити до ушкодження пухлини. Це протирічить вимогам безпеки при скринінгу, тому великою необхідністю є розробка ефективного методу скринінгу, який був би неінвазивним і безпечним [1]. Таким чином, ми пропонуємо використовувати неінвазивні методи дослідження змін, пов'язаних із злякисними новоутвореннями в інтерфазних ядрах буккального епітелію.

Перші повідомлення про злякисні зміни з'явилися в 1960-х роках, коли було широко вивчено вміст Х-хроматину в соматичних клітинах і виявлено його лабільність при різних функціональних змінах організму і загальносоматичної патології. При наявності в організмі пухлини спостерігаються значні зміни вмісту Х-хроматину в буккальному епітелії та нейтрофілах периферичної крові. Також було показано, що зміни кількості клітин з Х-хроматином зумовлені порушеннями функціонального стану гетероциклічної Х-хромосоми.

Особливий інтерес представляють роботи, що показують зміни в епітеліоцитах буккального епітелію у хворих на пухлини. Так, у 1960-х роках Н. Нібургз і його співавтори у [2] повідомили про характерний перерозподіл мас хроматину в соматичних клітинах у 77% хворих на рак і назвали ці зміни пухлиноасоційованими змінами. Останні характеризувалися збільшенням розмірів ядер епітеліоцитів, збільшенням розмірів зон «обмеженого» хроматину, які були оточені світлими зонами. Такі ж зміни спостерігалися в клітинах печінки, нирок та інших органів. Також було повідомлено, що зміни, пов'язані зі злякисними новоутвореннями спостерігалися в буккальному епітелії 74% пацієнтів зі злякисними пухлинами. Показано збільшення вмісту ДНК у ядрах епітеліоцитів у хворих на злякисну меланому порівняно з практично здоровими жінками. Водночас виявлено зменшення кількості хроматинпозитивних клітин (Х-хроматин) у хворих на злякисну меланому порівняно з такою у пацієнтів із доброякісними невусами та в контрольній групі. У хворих на рак молочної залози виявлено

збільшення вмісту ДНК та розміру інтерфазних ядер буккального епітелію. Але деякі автори при цитоспектрофотометричному визначенні кількості ДНК в епітеліоцитах буккального епітелію у чоловіків з епітеліомою бронхів не виявили істотної різниці між цим показником у хворих і майже здорових чоловіків [3].

Пізніше була зроблена спроба використати зміни буккального епітелію для характеристики впливу пухлини на її стан. Також були спроби охарактеризувати та обґрунтувати можливість впливу пухлини на функціональний стан буккального епітелію з метою використання отриманих даних для характеристики перебігу процесів, що відбувалися в органах, віддалених від пухлини, та виявлення закономірностей, що характеризують перебіг цих процесів. У 77% пацієнтів з пухлинами різної локалізації (карциноми, лімфоми, семіноми) спостерігаються порушення, що виражаються змінами ядерного матеріалу, неоднорідністю речовин хроматину та змінами ядерних мембран. Критеріями оцінки пухлинно-асоційованих змін були цитофотометричні дослідження вмісту ДНК, розмірів ядра і цитоплазми пухлинних клітин, характеру розподілу хроматину в ядрі. Але не вдалося виявити чіткі закономірності, властиві пухлинному процесу, крім збільшення розмірів ядер пухлинних клітин і зміни ядерно-цитоплазматичних відносин. Однак не можна було заперечувати, що виявлені порушення пов'язані з впливом пухлин на функціональний стан слизової оболонки порожнини рота.

У [4] було виявлено, що упаковка ДНК у клітинних ядрах має фрактальні властивості, тобто ДНК закручена як тривимірна крива Пеано. Областю інтенсивних досліджень став фрактальний аналіз клітин. Фрактальну розмірність вважають ефективним показником гетерогенності клітин комплексної гіперплазії ендометрія та добре диференційованої ендометріюїдної карциноми, а також як прогностичний фактор для виживання при меланомі, лейкемії та інших захворювань. Також були дослідження, які показали значний потенціал фрактальної розмірності для оцінки морфологічних даних. Однак вони були зосереджені на пухлинних клітинах, а не на клітинах буккального епітелію. Таким чином, можна припустити, що фрактальні властивості відображаються на розподілі хроматину в ядрах буккального епітелію і пухлина може впливати на цей розподіл, викликаючи зміни, пов'язані зі злоякісним новоутворенням.

У роботі [3] було досліджено контрольну групу (29 осіб), групу хворих на рак молочної залози II стадії (68 хворих) та групу хворих на фіброаденоматоз (33 хворих). Усі діагнози перевірені гістологічно. Морфологічний набір даних складався з 20256 зображень інтерфазних ядер буккального епітелію (6752 ядра, сканованих у трьох варіантах: без фільтра, через жовтий фільтр і через фіолетовий фільтр).

Морфологічними матеріалами є мазки епітеліоцитів слизової оболонки порожнини рота середньої глибини остистого шару, висушені при кімнатній температурі, фіксовані сумішшю Нікіфорова та забарвлені за Фельгеном холодним гідролізом у HCl протягом 15 хв при $t = 21-22$ C. Хроматин, пофарбований Фельгеном, аналізували за допомогою аналізатора Olympus, що складається з мікроскопа Olympus BX, камери Camedia C-5050 з цифровим зумом і комп'ютера. У середньому кожен препарат складається з 52 клітин у кожному препараті. Вміст ДНК-фуксину в ядрах епітеліоцитів розраховували як добуток оптичної густини на площу. У результаті першого етапу аналізу було отримано зображення розподілу хроматину у вигляді матриці розміром 128×128 пікселів.

Намагаючись відобразити фрактальний характер розподілу хроматину та забезпечити інваріантність щодо повороту зображення, було побудовано криву заповнення простору, що проходить через кожен піксель зображення та послідовно зчитано значення RGB кольорів пікселів зображення, а не рядок за рядком. У результаті можна відобразити матрицю пікселів на три вектори, що відповідають трьом каналам колірної моделі RGB. В якості кривої заповнення простору було використано криву Серпінського.

Перед застосуванням методів аналізу фрактального зображення його необхідно попередньо обробити. Для цього у роботі [3] було застосовано метод Отсу. Цей метод використовується для виконання порогової бінаризації напівтонових зображень. Алгоритм передбачає наявність двох класів пікселів (основних і фонових) на зображенні і шукає оптимальний поріг, який розділяє їх на два класи, щоб їх внутрішньокласова дисперсія була мінімальною.

Існує кілька методів обчислення фрактальної розмірності зображення. У [3] було обрано показник Херста, тому що він дуже підходить для послідовного аналізу. Показник Херста пов'язаний з фрактальною розмірністю D формулою $H = 2 - D$.

Відповідно до показника Херста можна класифікувати послідовність даних за їх хаотичними властивостями. Якщо $0 < H < 0,5$, послідовність вважається ергодичною, тобто якщо послідовність збільшилася в попередньому сегменті, велика ймовірність того, що вона зменшиться в наступному сегменті, і навпаки. Якщо $H = 0,5$, то послідовність є хаотичною, тобто їх значення не впливають на наступні значення. Якщо $0,5 < H < 1,0$, то послідовність є трендово-стабільною. Якщо послідовність у попередньому сегменті збільшується або зменшується в попередньому сегменті, вона дуже ймовірно збереже цю тенденцію в наступному сегменті. Якщо $H > 1$, то послідовність є фрактальним випадковим процесом із незалежними стрибками амплітуди, використовуючи розподіл Леві.

В результаті, для кожного пацієнта дані представляли з себе трьохканальні (RGB) вибірки фрактальних розмірностей ядер. Вибірki істотно відрізнялися за кількістю елементів. Тому при підготовці даних перед навчанням нейронної мережі для кожної вибірки було обчислено n квантилів, де n – кількість елементів у найменшій з вибірок (рис. 1).

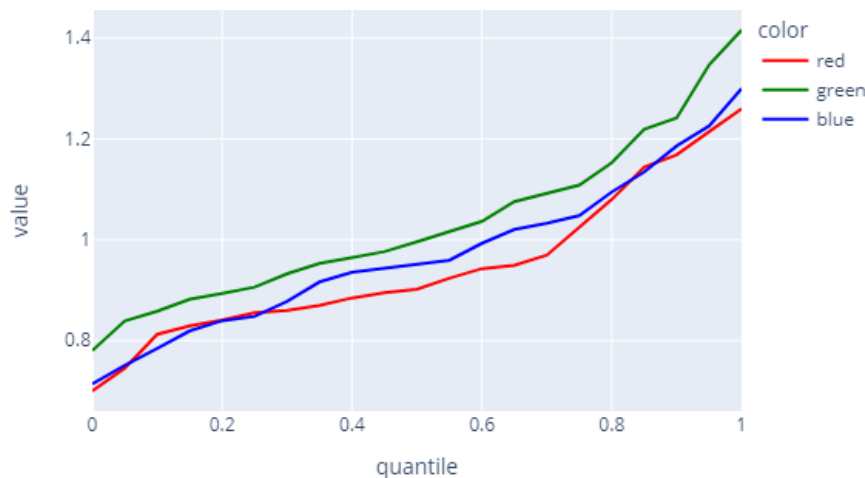


Рис.1. Приклад узагальнення навчальних вибірок через квантилі

Внаслідок проведених експериментів була розроблена згорткова нейронна мережа для класифікації даних, яка включає ряд компонентів: згорткові шари Conv1d, шари пулінгу, повністю з'єднані шари та шари нормалізації та регуляризації. У прямому проході вхідні дані пройшовши згортковий шар, проходять через шари пулінгу та повністю з'єднані шари, забезпечуючи комплексну обробку і адаптацію до особливостей даних. Conv1d є ключовим компонентом архітектури мережі і призначений для виявлення локальних патернів у вхідних даних. Згортковий шар Conv1d дозволяє моделі ефективно аналізувати просторові особливості

у вхідних даних, використовуючи ядро згортки та враховуючи обсяг інформації завдяки параметру padding.

Для виділення прихованих ознак в навчальних даних було проведено декілька експериментів з початковими шарами нейронної мережі. Найкращі результати вдалося отримати з одномірним згортковим шаром на вході моделі. За допомогою згортки було виділено k прихованих послідовностей менших по довжині за вхідну, які далі передавалися підмережі, що складається з декількох fully connected шарів.

Моделі з одновимірними згортковими шарами виявляється дуже ефективною у класифікації даних завдяки здатності Conv1d виділяти локальні особливості. Повністю з'єднані шари та шари нормалізації додають гнучкості та стабільності моделі. (рис. 2.)

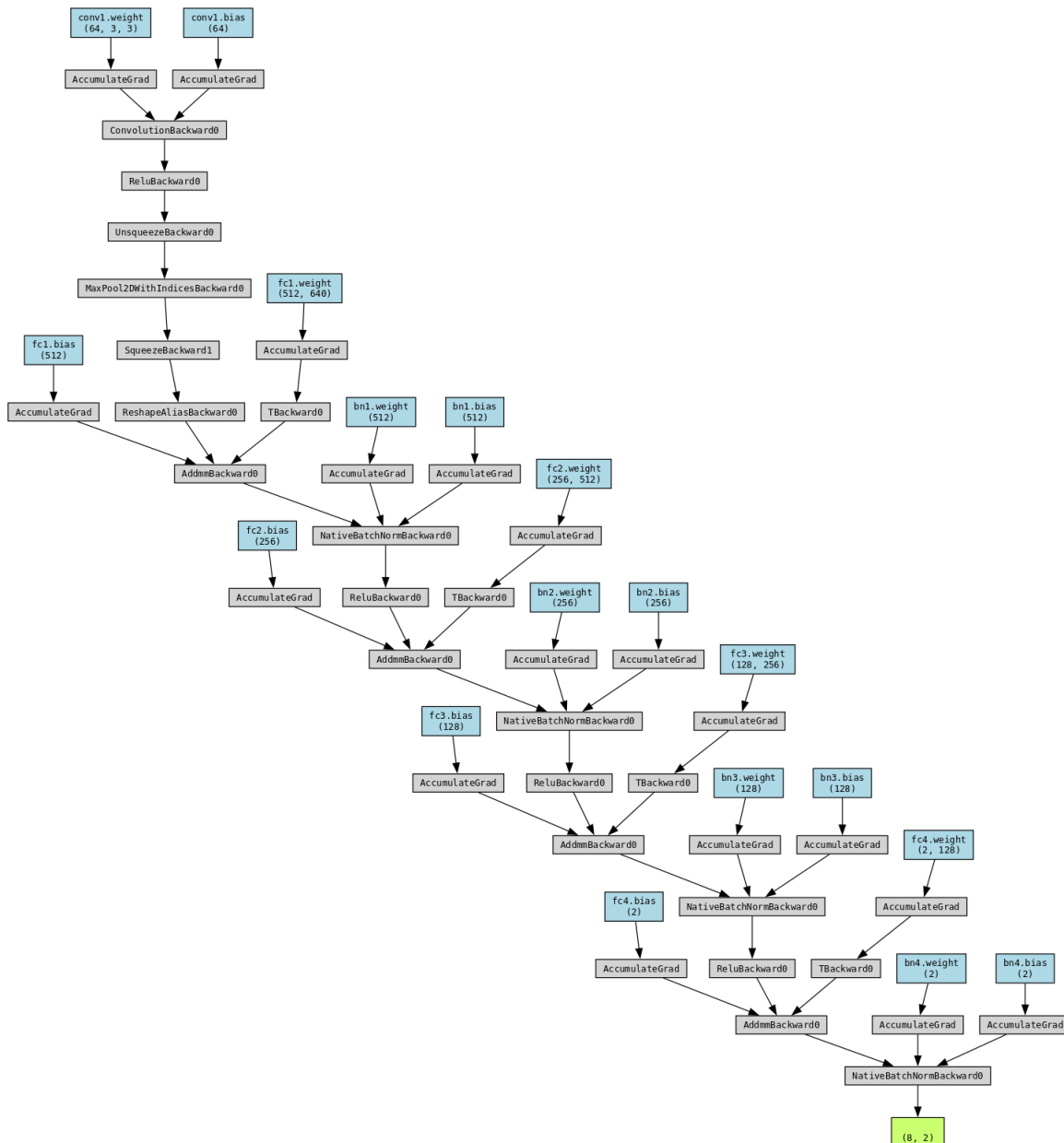


Рис. 2. Архітектура моделі

Для моніторингу навчання мережі було обрано ROC AUC – міру точності для задач класифікації при різних порогових значеннях. Можна побачити, що з більшою кількістю епох, модель досягає впевнених результатів на тестовому датасеті. (рис. 3.)

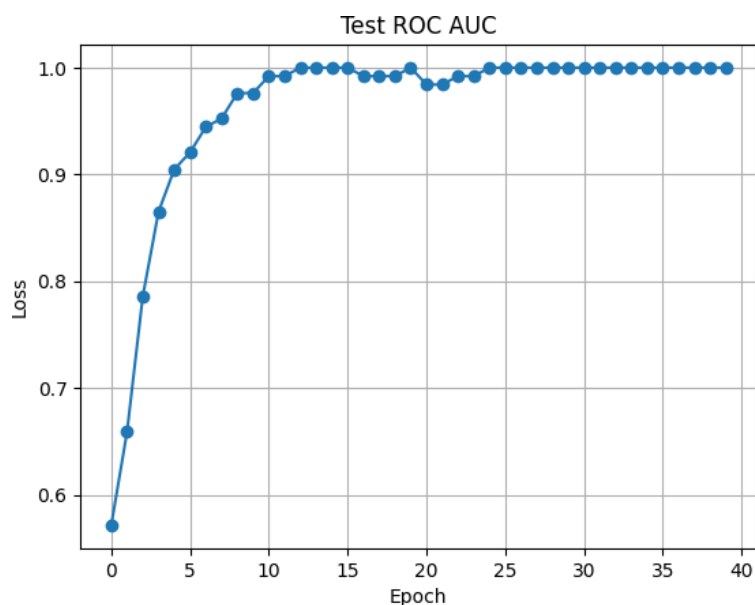


Рис. 3. Графік ROC AUC на тестових даних під час навчання моделі.

За рахунок специфічності даних та задачі, потребувалися додаткові підходи для забезпечення стабільної роботи моделі. Цієї цілі було досягнуто за рахунок сучасних підходів в оптимізації нейромереж, зокрема batch normalization та dropout.

Для евалюації результатів моделі було проведено декілька експериментів. Тренувальний датасет не був рівномірним – кількість елементів в одному з класів переважала – тому за основні метрики було обрано чутливість (precision) та специфічність (recall). Щоб встановити базовий рівень для метрик, спочатку було використано випадкову модель. Далі було порівняно дві нейромережі, остання з яких мала згортковий шар, це дало приріст в значеннях метрик (табл. 1.)

	Випадковий генератор	FCN	FCN+Conv1d
Чутливість (precision)	0.56	0.83	0.94
Специфічність (recall)	0.49	0.8	0.91

Табл. 1. Порівняння результатів різних моделей

ВИСНОВКИ

У цьому дослідженні описано успішну розробку згорткової нейронної мережі для прогнозування раку молочної залози. Вхідні дані включають трьохканальні вибірки, які представляють кольорову інформацію з трьох основних каналів: червоного, зеленого та синього. Щоб стандартизувати ці дані, вони розбиті на N квантилі.

Архітектура моделі включає в себе перший шар у вигляді нетипової одновимірної згорткової мережі. Це рішення сприяє виявленню ознак, пов'язаних із раком молочної залози. За результатами експериментів модель продемонструвала велику влучність, перевищуючи 94%.

Порівняння з існуючими методами діагностики підкреслило переваги запропонованого підходу. Висока точність та здатність адаптуватися до різноманітних форм даних роблять цю

модель потенційно корисною у клінічній практиці. Детальний аналіз показав, що модель ефективно реагує на різні характеристики даних, що дозволяє зробити висновок про її гнучкість та універсальність.

Потенційні напрямки подальших досліджень включають розширення датасету для поліпшення загальної надійності моделі, а також удосконалення архітектури мережі. Додаткові аспекти, такі як врахування інших медичних параметрів, можуть також покращити прогностичні можливості системи.

У заключенні, високі результати цього дослідження свідчать про значущий внесок згорткових нейронних мереж у сферу медичних досліджень, зокрема в прогнозуванні раку молочної залози.

Автори висловлюють вдячність Голубевої К. М. та Бородай Н. В. за надані дані для навчання моделей.

СПИСОК ПОСИЛАНЬ

1. Radak M., Haider Y., Lafta, Fallahi H.. 2023. *Machine learning and deep learning techniques for breast cancer diagnosis and classification: a comprehensive review of medical imaging studies*. 49(12):10473-10491. doi: 10.1007/s00432-023-04956-z.
2. Nieburgs H.E. 1968. *Recent progress in the interpretation of malignancy associated changes (MAC)*. *Acta Cytologica*. 12, pp. 445–453. doi: 10.1155/238921
3. Ogden G.R., Cowpe J.G., Green M.W. 1990 *The effect of distant malignancy upon quantitative cytologic assessment of normal oral mucosa*. *Cancer*. 65, pp. 477–480. doi:10.1002/1097-0142(19900201)65:3<477::AID-CNCR2820650317>3.0.CO;2-G.
4. Lieberman-Aiden E. et al. 2009. *Comprehensive mapping of long-range interactions reveals folding principles of the human Genome*. *Science*. 326, 5959, pp. 289–193. doi: 10.1126/science.1181369.
5. Klyushin D., Golubeva K., Boroday N., Shervarly D.. 2021. *Breast Cancer Diagnosis Using Machine Learning and Fractal Analysis of Malignancy-Associated Changes in Buccal Epithelium*. pp. 2–15. doi: 10.1201/9781003153405-1
6. Boroday N., Chekhun V., Golubeva E. and Klyushin D. 2016. *In vitro and in vivo densitometric analysis of DNA content and chromatin texture in nuclei of tumor cells under the influence of a nano composite and magnetic field*. *Advances in Cancer Research & Treatment*. 2016, 706183. pp. 1–11, doi: 10.5171/2016.706183.
7. *Breast Cancer Surveillance Consortium (BCSC)*. 2017. *Sensitivity, specificity, and false negative rate for 1,682,504 screening mammography examinations from 2007–2013*. www.bcsc-research.org/statistics/screening-performance-benchmarks/screening-sens-spec-false-negative.
8. Adam R., Silva R., Pereira F. et al. 2006. *The fractal dimension of nuclear chromatin as a prognostic factor in acute precursor B lymphoblastic leukemia*. *Cellular Oncology*. 28, pp. 55–59. doi: 10.1155/2006/409593.
9. Adhraryn S.G., Dave B.J., Trivedi A.H. 1991. *Cytogenetic surveillance of tobacco-areca nut (mava) chewers, including patient with oral cancers and premalignant conditions*. *Mutation Research*. 261, 1, pp. 41–49. doi: 10.1016/0165-1218(91)90096-5.
10. Andrushkiw R.I., Boroday N.V., Klyushin D.A., Petunin Y.A. *Computer-aided cytogenetic method of cancer diagnosis*. New York: Nova Publishers, 2007. ISBN 10: 1-59454-882-X. Bedin V. et al. 2010. *Fractal dimension of chromatin is an independent prognostic factor for survival in melanoma*. *BMC Cancer*. 10, 260. doi:10.1186/1471-2407-10-260.
11. Bikou O. et al. 2016. *Fractal dimension as a diagnostic tool of complex endometrial hyperplasia and well-differentiated endometrioid carcinoma*. *In Vivo*. 30, pp. 681–690
12. Butakov V., Grakovskiy A. 2005. *Evaluation of arbitrary time series stochastic level by Hurst parameter*. *Computer Modelling and New Technologies*. 9, 2, pp. 27–32.

Мельников О.Ю., Капелешук А.О. (Донбаська державна машинобудівна академія, м. Краматорськ, Україна)

ЗАСТОСУВАННЯ ЗГОРТКОВОЇ НЕЙРОННОЇ МЕРЕЖІ ДЛЯ КЛАСИФІКАЦІЇ МОНЕТ АНТИЧНОГО СВІТУ

***Анотація:** Робота присвячена можливості застосування методів штучного інтелекту для класифікації античних монет. Запропоновано архітектуру нейронної мережі, проведено її тестування, проведено аналіз результатів, та описано шляхи подальшого розвитку.*

***Abstract:** The work is dedicated to exploring the application of artificial intelligence methods for the classification of ancient coins. An architecture for a neural network is proposed, and its testing is conducted. The results are analyzed, and avenues for further development are described.*

Гроші – це невіддільна частина будь-якої високорозвиненої цивілізації та інколи майже єдиний слід існування цієї цивілізації. Через те, що у давнину гроші робили з коштовних металів, археологи знаходять досить велику кількість монет на місцях існування великих розвинених цивілізацій. Інколи тільки завдяки зображенням на монетах ми маємо уявлення про те, як виглядали античні царі, де проходили торгові шляхи та іншу цікаву і корисну для істориків інформацію.

Зазвичай на античних монетах зображали портрети царів та/або богів, яких шанували у тій місцевості. На ранніх монетах (VI – V ст. до н.е.) переважно зображали тварин та геометричні фігури, а з IV ст. до н.е. зображення тварин витісняють зображення людей. Номіналом монети слугував її матеріал та вага, тому його на монеті не вказували.

На території України в античні часи знаходилось 7 великих античних міст-полісів, а також значна частина Боспорського царства зі столицею у місті Пантікапей. За період існування міст-полісів (з середини VI ст. до н.е. до першої половини V ст. н.е.) було сумарно випущено десятки тисяч монет з золота, срібла, міді, свинцю, заліза та інших металів.

Наразі сумарна кількість знайдених античних монет, випущених в полісах, що знаходилися на території України, значно перевищує відмітку у 30 тисяч одиниць, і з кожним роком поповнюється новими знахідками, знайденими як у землі, так, на жаль, і на чорному ринку. Україна в цьому не унікальна держава, бо схожа ситуація, тільки ще в більшому масштабі, є в усіх країнах Середземномор'я.

Однак автоматична класифікація монет є задачею, яку до кінця неможливо вирішити, через те що в античні часи кожне місто та кожний цар карбували свої монети. Але є невелика (порівняно з усією різноманітністю видів монет) кількість типів монет, які зустрічаються у десятки разів частіше, ніж інші. Нас зараз не цікавить, чому так осталося, що одних монет карбували багато, а інших мало (багато монет дійшло до нас в єдиному екземплярі або в надто малому тиражі), а нас цікавить те, що ймовірність відшукати монету одного типу значно вища, ніж відшукати монету іншого типу. Тому якщо перевіряти монету, чи не належить вона до одного з популярних типів, можна автоматично класифікувати більшість знахідок.

Іншою проблемою є те, що кількість монет кожного типу величезна та може стрімко зростати. Наприклад, якщо сьогодні ми знаємо три монети умовного типу «А» та за її рідкість її не включають до класифікації, ніхто не дасть гарантії, що через рік, а може і раніше не відшукають скарб в якому буде ще 20 – 30 таких монет.

Тому при розробці системи класифікації монет були передбачені наступні особливості такої знахідки як монета:

1. Невідома кількість категорій класифікації: тільки в Україні нараховують майже дві з половиною тисячі типів античних монет, не кажучи про інші держави, що знаходяться на території колишньої Римської імперії, та не враховуючи те, що антична монета може бути сплутана з монетою середньовічною. Також треба врахувати міграцію людей в античні часи, тому монету, накарбовану в Борисфені (Миколаївська область), можуть відшукати десь на околицях Риму чи Константинополю, а в Борисфені можуть знайти монети, накарбовані зовсім в інших місцях імперії.

2. Основною інформацією для системи класифікації є не одне, а два зображення однієї монети (з обох боків).

3. Монета може бути частково ушкодженою, покритою шаром оксидів (особливо актуально для мідних та залізних монет) або бруду.

4. Монета може бути деформованою (особливо актуально для свинцю) або доробленою механічно після штампування, тобто її малюнок трошки відрізняється від більшості монет (рис. 1).

5. Монети одного типу можуть мати абсолютно різні, але схожі малюнки (рис. 2 та рис. 3). Більшість монет кожного типу мають такі незначні відмінності, а на малюнках показані лише деякі з них. Також, наприклад, на рис. 3 дві монети у нижньому ряду з ліворуч мають не опуклий, а вдвлений малюнок. Тому добитися високої точності розпізнавання для античних монет є складною задачею.



Рисунок 1 – Порівняння монети з механічною доробкою (зліва) з монетою без доробки (справа)



Рисунок 2 – Порівняння реверсів монет типу 783-7831



Рисунок 3 – Порівняння реверсів монет типу 109-2075

Тому виходячи з описаних вище вимог та з урахуванням описаних проблем було запропоновано універсальну архітектуру нейромережі, яка буде навчена класифікувати монети одного виду. На вхід такої мережі буде подаватися зображення монети у градаціях сірого (щоб знизити залежність від кольору знахідки, тому, що через окислення він може значно змінюватись), з одного боку, і на виході отримувати значення ймовірності, з якою ця монета може належати до цієї групи. Таким чином, у майбутньому бібліотеку монет, які вмітиме класифікувати система, можна буде збільшувати з додаванням нових навчених моделей. Кожну сторону монети перед завантаженням потрібно помістити у центр квадратного зображення з білим фоном (розмір 150 на 150 пікселів). Також для більш якісного навчання мережі було зроблено аргументацію даних – генерування нових фотографій шляхом їх випадкових трансформацій. Особливо актуальною трансформацією для монет є повороти на декілька градусів (8), масштабування (0,08), невеличкі зсуви пікселів (0,08).

Але зі зростанням бібліотеки у майбутньому описаний вище підхід може призвести до того, що на аналіз однієї монети буде йти досить багато машинного часу, тому потрібен алгоритм який виключав ті моделі зі спуску аналізу, негативний результат яких можна передбачити заздалегідь. Через те алгоритм роботи можна описати наступним чином:

1. Підготовлене зображення обох сторін монети по черзі завантажують до програми. Також разом з монетою передаємо контекстуальні дані, які допоможуть провести первинну оцінку монети.

2. Кожну зі сторін (по черзі) подаємо до алгоритму, який підбере моделі для її класифікації з загальної бібліотеки моделей на основі контекстуальної інформації.

3. Кожна сторона монети по черзі подається набору одібраних моделей, кожна з яких на виході видає ймовірність того, що ця монета належить до її категорії. (Критерієм для відбору моделей для другої сторони монети є те, що якщо при аналізі першої сторони якась з моделей покаже велике значення ймовірності, то для другої сторони потрібно першою поставити модель для другої сторони монети того ж виду.)

4. Обраховуються складні ймовірності для моделей обох сторін монет одного виду. Виводиться звіт користувачу.

Зрозуміло, що основною або центральною частиною цієї системи є нейронна мережа, яку можна навчити для будь-яких монет. У ході ряду експериментів з різними архітекторами згорткових нейронних мереж, бо саме ця архітектура найкраще себе показує у роботі з графічними образами, було виявлено, що оптимальною архітектурою є мережа, яка складається з чотирьох шарів згортки з ядром згортки 2×2 та чотирьох шарів пулінгу, кожен з яких зменшує вхідний вектор в 4 рази. Для контролю перенавчання використовуються шари Dropout які розташовані після першої та другої групи згортка-пулінг з коефіцієнтом 0,2. Більш детально структуру показано у таблиці 1. Функція активації на всіх шарах крім вихідного ReLU, функція активації вихідного шару Softmax, який видає ймовірності того, що монета належить або не належить до заданого типу, що може бути корисним при використанні системи експертом: наразі система обирає більше значення для класифікації монети, але у ситуації, коли показники ймовірності досить схожі, такий вибір може призвести до неправильного результату. Функція втрат – категоріальна кросентропія, оптимізатор – adam. Усього модель має 2 мільйони параметрів, для роботи яких знадобиться 8,66 МВ оперативної пам'яті. Для всіх тестів використовувався фреймворк Tensorflow, та його надбудова для Python – Keras.

Також тестувалися варіанти мережі, які використовують не монохромні, а кольорові RGB зображення, і хоча така модель трохи краще відрізняє золоті монети, вона показує абсолютну не ефективність для будь-яких інших металів, через те, що вони мають схильність до окису. Задача порівняння кольорів і відсортування золотих монет від не золотих вирішується за допомогою звичайного алгоритму, тому ідеї використовувати модель для трьох каналів були відкинуті.

Таблиця 1. Структура моделі

Layer (type)	Output Shape	Param #
conv2d (Conv2D)	(None, 149, 149, 32)	160
max_pooling2d (MaxPooling2D)	(None, 74, 74, 32)	0
dropout (Dropout)	(None, 74, 74, 32)	0
conv2d_1 (Conv2D)	(None, 73, 73, 64)	8256
max_pooling2d_1 (MaxPooling2D)	(None, 36, 36, 64)	0
dropout_1 (Dropout)	(None, 36, 36, 64)	0
conv2d_2 (Conv2D)	(None, 35, 35, 128)	32896
max_pooling2d_2 (MaxPooling2D)	(None, 17, 17, 128)	0
conv2d_3 (Conv2D)	(None, 16, 16, 256)	131328
max_pooling2d_3 (MaxPooling2D)	(None, 8, 8, 256)	0
flatten (Flatten)	(None, 16384)	0
dense (Dense)	(None, 128)	2097280
dense_1 (Dense)	(None, 2)	258

Для навчання моделей використовувалися створені власноруч авторами вибірки даних, які складаються з наборів кольорових зображень монет (роздільною здатністю 200 на 200) декількох типів та зображень монет, які відіграють роль усіх інших монет і повинні класифікуватися як не належні до заданого класу. Усі фотографії монет взяті авторами з відкритих інтернет-каталогів <https://bosporan-kingdom.com/> та <https://tauriscoins.ru/>, як набори даних використовувалися аверси та реверси (окремо) наступних монет: 109-2075 (78 шт.), 783-7831 (21 шт.), 111-3002 (100 шт., обраних с більше ніж 1000 монет каталогу, інші взяті всі). Вибірка інших монет складалась з ~170 знімків (аверси та реверси разом).

Для монет 109-2075, 111-3002 та аверсу 783-7831 вдалося досягти точності для валідаційної вибірки у 100%. Навчити модель для реверсів 783-7831 не вдалося через велику варіативність чеканки (рис. 2). Склад валідаційної вибірки приблизно 10% для всіх випадків. Навчання відбувалося протягом 20 епох. Детальніше дивіться на графіку рис 4.

Було також виявлено залежність між відношенням монет класу до інших монет. Для вибірки аверс 783-7831 точність 100% за десять епох вдалося досягти коли відношення між вибірками було 1:1, при відношенні 1:4 та 1:7 для досягнення точності 100% знадобилося навчання протягом 30 епох. Докладніше на графіку рис 5.

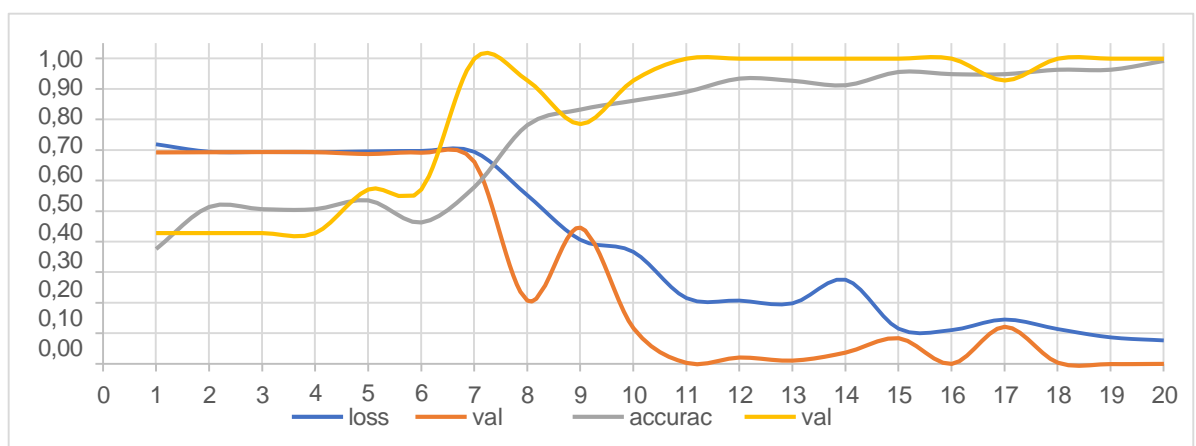


Рисунок 4 – Показники навчання для аверсу 109-2075

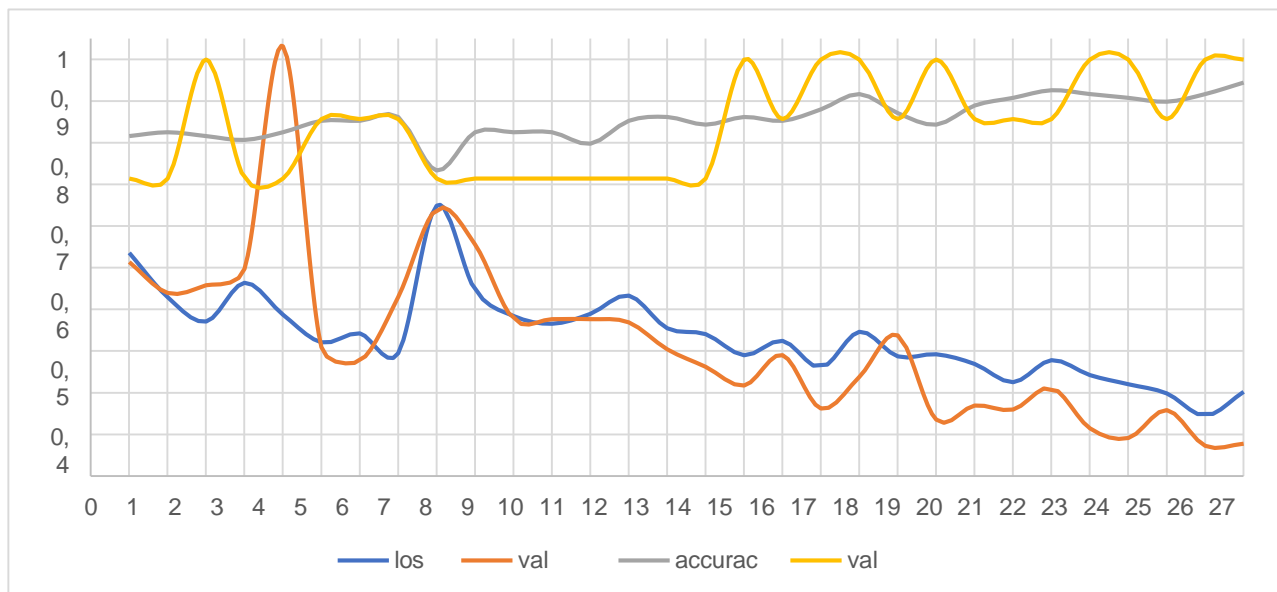


Рисунок 5 – Показники навчання для аверсу 783-7831 (30 епох)

ВИСНОВКИ

Проведені дослідження показали, що застосування нейронних мереж для автоматизації класифікації монет цілком можливе та потребує подальших досліджень. Було запропоновано архітектуру універсальної системи класифікації монет та архітектуру нейронної мережі яка досить добре справляється з класифікацією монет за приналежністю чи не приналежністю до певної категорії. Було проведення тестування запропонованої моделі на 6 вибірках, на 5-х моделях показала 100% результати, для одного набору навчити модель не вдалось.

ПЕРЕЛІК ПОСИЛАНЬ

1. Монети Боспора. [Електронний ресурс]. Режим доступу: <https://bosporan-kingdom.com/>
2. Гіміс В. Б. Нейромережні технології : навчальний посібник / В. Б. Гіміс. – Краматорськ : ДДМА, 2021. – 248 с.

АНАЛІЗ СТАНУ ЯКОСТІ ПОВІТРЯ ТА ВИЗНАЧЕННЯ ЧИННИКІВ, ЩО НА НЬОГО ВПЛИВАЮТЬ

Анотація: В даній роботі було піднято проблему якості стану повітря, яка є досить великою в сучасному світі. Для її дослідження було згенеровано набір даних за допомогою різних відкритих джерел. Основну увагу було звернуто на промислові об'єкти, які могли б забруднювати повітря. Для цього з ресурсу Google Maps було зібрано усі можливі промислові об'єкти на території України за ключовими словами, після чого були вручну відібрані необхідні нам типи об'єктів. На основі створеного набору даних, було побудовано нейронну мережу, гіперпараметри якої були підібрані методом перебору. За допомогою підходу Permutation Importance було запропоновано алгоритм зменшення розмірності даних з підвищенням якості моделі. В результаті, вдалося досягти показника RMSE для нашої моделі ~ 0.0022 . Побудована модель нейронної мережі, дозволила оцінити вплив різних чинників на показник Air Quality Index за допомогою вище вказаного алгоритму Permutation Importance. Результати дослідження виявили, що найбільшими забрудниками є об'єкти електричних підстанцій, хімічних заводів та сонячних фотоелектричних електростанцій.

Abstract: In this work, the problem of air quality, which is quite large in the modern world, was raised. For this research, a dataset was generated using various open sources. The main attention was paid to industrial facilities that could pollute the air. For this purpose, all possible industrial objects in the territory of Ukraine were collected from the Google Maps resource by keywords, after this the types of objects we needed were manually selected. For example, such objects as "Tzegelny Zavod Bus Stop" were rejected. Next, on the basis of the created data set, a neural network was built, the hyperparameters of which were selected by the selection method. Next, using the Permutation Importance algorithm, an algorithm to reduce the dimensionality of the data and improve the quality of the model was proposed. As a result, it was possible to achieve an RMSE of ~ 0.0022 for our model. After the neural network model has been built, we can evaluate the influence of various factors on the AQI indicator using the same Permutation Importance algorithm. As a result, the objects of electrical substations, chemical plants and solar photovoltaic power plants turned out to be the biggest polluters.

Вступ

Проблема якості повітря, яким ми дихаємо існувала завжди, тим більше у сучасному світі, де, за статистикою 2020 року, в Україні на 1000 осіб приходиться 245 автомобілів. Провівши нескладні обрахунки, є можливість дізнатися приблизну кількість автомобілів, що дорівнює більше 10 млн одиниць. Так, існує певна тенденція переходу на електромобілі, проте станом на 1 лютого 2023 року, в Україні було зареєстровано 48 тис електроавтомобілів, що складає лише 0.48% від загальної кількості автомобілів.

Ще одна причина забруднення повітря – це заводи, які є майже у кожному районному центрі України, та кількість яких зростає. За період з 2015 по 2019 роки було побудовано 207 нових заводів.

У цій роботі ми спробуємо дослідити це питання, порівнявши вплив різних типів промислових об'єктів на показник AQI (Air Quality Index).

Ця робота не спрямована на безпосереднє вирішення цієї проблеми, адже їх потрібно вирішувати явно не методами машинного навчання. Але, проведене дослідження дасть змогу виявляти об'єкти, що є найбільшими забрудниками повітря, тим самим дати поштовх для подальшого запуску процесів, що спричинять посилену увагу на вирішення проблеми забруднення навколишнього середовища.

Критичний аналіз літературних джерел

В сучасному світі дана проблема є надзвичайно актуальною, тому над її дослідженням працювало досить багато науковців. Також багато таких статей спрямовані й на дослідження впливу різних очевидних та другорядних забруднювачів повітря. Розглянемо декілька таких робіт.

У [1] більше спрямування на опис системи MoreAir (система моніторингу забруднення

повітря в містах). У статті описано методологія при розробці цієї системи. Також, проведено дослідження на основі Марокко, яке показало, що в деяких місцях вихлопи транспорту – це далеко не найбільший забруднювач.

У [2] було завдання створити погодинний прогноз концентрації PM_{2.5} в повітрі. За основу було взято 2 великих міста в Південній Кореї. За основний алгоритм використовувався LSTM та порівнювався з іншими відомими алгоритмами, де показав себе найкращим чином.

Досить цікава стаття [3], у якій досліджений вплив страйків на якість повітря. Здавалося б, що це абсолютно не очевидний забруднювач, але він є, і досить немалий. Звичайно, страйки бувають різні, тому, так, це може мати значний вплив на якість повітря. Для цього аналізу використанні різні алгоритми машинного навчання, проте найкраще себе показав GBM (gradient boosting machine).

У [4] досліджується залежність між станом повітря та пандемією COVID-19. Тобто, пандемія знизила активність на вулицях, менше транспорту їздило, відповідно – це однозначно мало свій вплив.

На оцінку впливу COVID-19 була спрямована й робота [5]. Зазначається, що під час пандемії спостерігалось явне зниження вмісту шкідливих речовин у повітрі, що пояснювалося значним зниженням мобільності людей. Проте автор статті запевняє, що немалий вплив на це мали й метеорологічні зміни. За основу дослідження було взято Сінгапур.

У роботі [6], на основі різних забруднювачів повітря, таких як наявність великої кількості автомобілів, викидів з промислових виробництв, спалювання нафтопродуктів і виробництво електроенергії, складається прогноз забруднення повітря. Для цього використовуються методи машинного навчання, такі як LSTM, CNN тощо.

Генерація набору даних

Спочатку було зібрано необхідні дані для подальшого дослідження. З ресурсу [10] було взято міста для яких був доступний показник AQI (Air Quality Index). В таблиці 1 наведено звідки було взято потрібні дані для проведення дослідження.

Таблиця 1. Ресурси, на основі даних з яких був згенерований набір даних

Ресурс	Зібрані дані	Пояснення
www.coordinates_converter.com [11]	Координати міст	
Вікіпедія	Площа, населення та густота населення міст	Єдиний ресурс, який може надати нам цю інформацію про всі потрібні нам міста
Google Maps	Всі виробництва-потенційні забруднювачі на території України	Пошук за ключовими словами (виробництво, фабрика, АЕС, ТЕС, ГЕС, АЕС). Після чого вручну було відібрано 26 типів, що можуть забруднювати повітря.
www.meteopost.com [8]	Погодні показники наявних на ресурсі міст з інтервалом 30 хв	

www.sinoptik.ua [9]	Погодні показники усіх інших міст з інтервалом 3 години	
--	---	--

В результаті отримано наступний набір даних (рис. 1):

merged_data	
area	real
state	text
population	real
date	datetime
temperature	integer
wind_speed	real
wind_dir	integer
pressure	integer
humidity	integer
density	real

Рис. 1. Об'єднані дані без даних з Google Maps

Також до цих даних було додано кількість об'єктів знайдених на Google Maps з різними радіусами до міст. Під радіусом мається на увазі різниця площ більшого та меншого кіл з центром в одній точці (рис. 2).

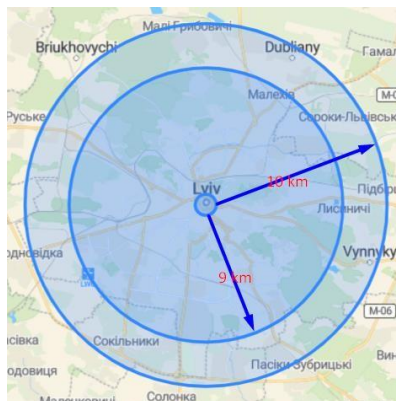


Рис. 2. Візуалізація пошуку об'єктів у частині кола радіусом між 9 км і 10 км

Як вказано в таблиці 1, було виділено 26 типів, по 20 кожного, так як для кожного типу ми брали 20 радіусів (від 1 до 20). Отже отримаємо $26 \cdot 20 = 520$ колонок. Для кращого визначення окремого впливу кожного параметра на забруднення повітря, зменшування розміру вхідних даних проводитись не буде. Також перетворимо колонку state за допомогою One-Hot encoding, так як в нашому наборі ця колонка у текстовому форматі, а нам потрібно перетворити її в числовий.

Підбір параметрів нейронної мережі

Підбір гіперпараметрів нейронної мережі буде здійснено методом перебору. Підбираються наступні параметри: функція активації, оптимізатор, кількість шарів, розмір виходу в першому шарі, а вихід кожного наступного шару буде вдвічі менший.

Найкраще себе показав наступний набір параметрів:

- функція активації – AdaMax;
- оптимізатор – ReLU;
- кількість шарів – 5;

- розмір виходу в першому шарі – 600.

На рис. 3. наведено графік порівняння даного набору параметрів з іншими на прикладі метрики RMSE.

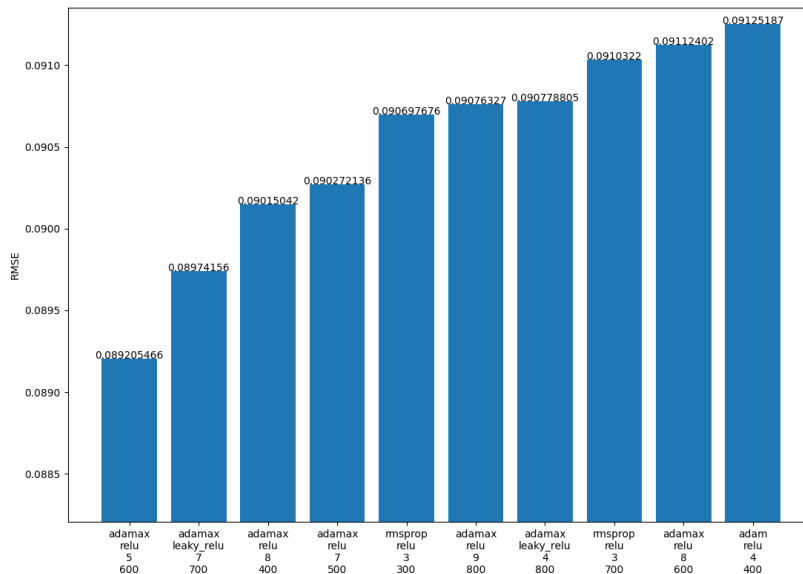


Рис. 3. Стовпчикова діаграма показника RMSE для 10 найкращих дослідів

Зменшення розмірності набору даних

Для зменшення розмірності набору даних недоцільно застосувати PCA, адже в такому випадку буде неможливо дослідити вплив окремих ознак на AQI. Тому для цього використано алгоритм Permutation Importance. Суть роботи алгоритму полягає в наступних діях: береться натренована модель та набір даних, кожна колонка перемішується певну кількість разів та алгоритм дивиться на те, наскільки змінилася похибка алгоритму. Результат – це коефіцієнт, який позначає вплив кожної колонки на результуючу змінну.

Зменшування розмірності за допомогою вище вказаного алгоритму буде проводитись наступним чином:

- 1) беремо натреновану модель;
- 2) запускаємо роботу алгоритму Permutation Importance;
- 3) видаляємо колонки, які отримали коефіцієнт менше або дорівнює нуля;
- 4) тренуємо модель на новому наборі даних та порівнюємо її похибку з попередньою;
- 5) повторюємо кроки 1-4 доки похибка буде зменшуватися.

На даному етапі розмірність набору даних складає 331042 x 444. На рис. 4 та 5 зображено графіки зменшення розмірності набору даних та зміни похибки.

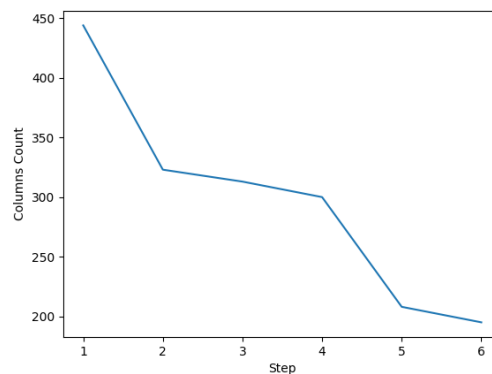


Рис. 4. Кількість колонок при поступовому зменшенні розмірності даних

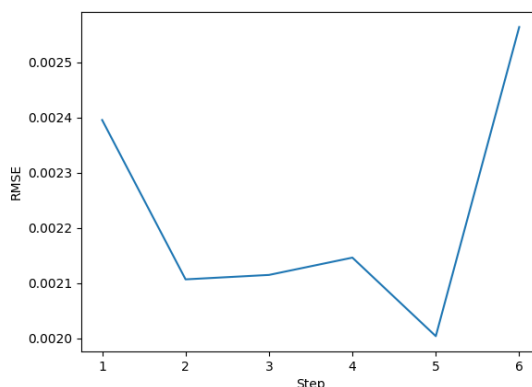


Рис. 5. Показник RMSE при поступовому зменшенні розмірності даних

Виходячи з графіку на рис. 5, найкращого результату похибки було досягнуто на 5 кроці. На 6 кроці похибка стала різко збільшуватись. При цьому, на 5 кроці кількість ознак обраного набору даних дорівнює 210. Отже, тепер розмірність даного набору даних дорівнює 331042×210 .

Цей алгоритм кращий за PCA в тому плані, що при зменшенні розмірності покращується сама модель. Проте за цим методом не вийде виконати зменшення розмірності масштабу PCA, адже Permutation Importance – це алгоритм, який працює досить довго.

Оцінка моделі нейронної мережі

Для подальшого дослідження буде використано набір даних, що було отримано на 5 кроці алгоритму зменшення розмірності даних. Нижче наведемо графік тренування даної моделі (рис. 6).

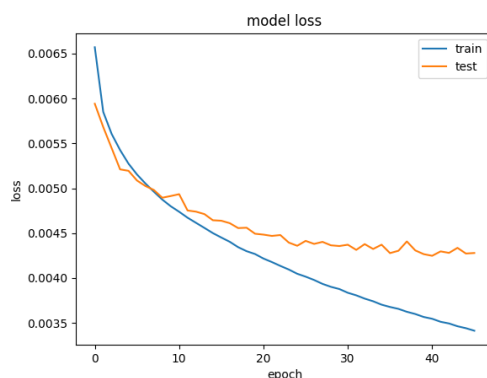


Рис. 6. Графік втрат нейронної мережі

Показник RMSE отриманої моделі вийшов 0.0022528581129933702.

Аналіз впливу різних типів промислових об'єктів на стан якості повітря

Виведемо 5 найбільших забруднювачів повітря згідно з нашою моделлю. Коефіцієнт впливовості рахується як середній для усіх радіусів одного типу об'єктів.

Таблиця 2. 5 найбільш впливових промислових об'єктів

Показник	Коефіцієнт впливовості
Електрична підстанція	0.00013180925648873578
Хімічний завод	0.00013038753893535756
Показник	Коефіцієнт впливовості
Сонячна фотоелектрична електростанція	8.770038767039458e-05
Косметична індустрія	8.650569002812188e-05
Виробництво	8.011425893648674e-05

Як бачимо, найбільшими забруднювачами є електричні підстанції, які найчастіше будується біля теплових електростанцій. Якщо вірити виданню DW та їхній статті від 25 травня 2021 року [7], то теплові електростанції дійсно є найбільшими забруднювачами повітря в Україні, що може вказувати на достовірність нашого дослідження. Також пояснити те, що самі електростанції менше впливають на стан якості повітря, можна тим, що їх значно менше на території України, ніж підстанцій. Отже, після проведеного дослідження, було складено перелік типів об'єктів, що нам доступні з 6 по 15 місця, які найбільше впливають на забруднення повітря:

- 6) Фармацевтична компанія (7.330393704337653e-05);
- 7) Маслоробня (7.168745519625006e-05);
- 8) Сільськогосподарський кооператив (5.9996562520565663e-05);
- 9) Кар'єр (5.951612837496133e-05);
- 10) Електростанція (5.8587293709198875e-05);
- 11) Пивоварня (5.436671631549445e-05);
- 12) Водно-комунальна компанія (2.7024660464011773e-05);
- 13) Гідроелектростанція (2.135397761805352e-05);
- 14) Виробник косметичної продукції (1.6045247643868207e-05);
- 15) Заправка (8.229845310878288e-06)

З отриманих даних виходить, що найбільшими забруднювачами після електричних підстанцій є хімічні заводи, що цілком очікуваний результат. Викликає здивування косметична індустрія на четвертому місці, адже це абсолютно не очевидний глобальний забруднювач повітря, але потрібно брати до уваги, що виробництва – це потенційні забруднювачі повітря. Тому на 5 місці об'єкт з досить загальною назвою, просто виробництва, вони можуть виробляти різні товари, проте в загальному всі вони є шкідливими для атмосфери. Також на 6 та 7 місцях фармацевтична компанія та маслоробня відповідно. Усі ці об'єкти можуть викидати в атмосферу шкідливі відходи та пари.

ВИСНОВКИ

Однією з найбільших проблем сучасності є стан якості повітря. У цій роботі вона була досліджена. Для цього був згенерований набір даних на основі різних відкритих джерел, таких як saveecobot, sinoptik, Google Maps, і навіть Вікіпедія. Для останнього ресурсу було аргументовано доцільність використання саме цього ресурсу в даній ситуації. На основі моделі нейронної мережі, гіперпараметри для якої були підібрані методом перебору, та алгоритму Permutation Importance було спочатку зменшено розмірність набору даних з покращенням якості самої моделі, а після чого й визначено найбільші забруднювачі повітря серед промислових об'єктів (електрична підстанція, хімічний завод, сонячна фотоелектрична електростанція).

СПИСОК ПОСИЛАНЬ

1. Gryech I. *Moreair: a low-cost urban air pollution monitoring system* / I. Gryech, Y. Ben-Aboud, B. Guermah, [et al.] // *Sensors (Switzerland)*. — 2020. — Vol. 20, No. 4.
2. Qadeer K. *A long short-term memory (lstm) network for hourly estimation of pm2.5 concentration in two cities of south korea* / K. Qadeer, W. U. Rehman, A. M. Sheri, [et al.] // *Applied Sciences (Switzerland)*. — 2020. — Vol. 10, No. 11.
3. Zalakeviciute R. *Gradient boosting machine to assess the public protest impact on urban air quality* / R. Zalakeviciute, Y. Rybarczyk, K. Alexandrino, [et al.] // *Applied Sciences (Switzerland)*. — 2021. — Vol. 11, No. 24.
4. Lin G.-Y. *Chang impact analysis of level 3 covid-19 alert on air pollution indicators using artificial neural network* / G.-Y. Lin, W.-Y. Chen, S.-H. Chieh, Y.-T. Yang // *Ecological Informatics*. — 2022. — Vol. 69.

5. Li Y. *The impact of covid-19 on no2 and pm2.5 levels and their associations with human mobility patterns in singapore* / Y. Li, Y. Zhu, J. Y. K. Tan, [et al.] // *Annals of GIS*. — 2022. — Vol. 28, No. 4. — P. 515–531.
6. Ghufuran Isam Drewil, Dr. Riyadh Jabbar Al-Bahadili *Forecast air pollution in smart city using deep learning techniques: a review ghufuran isam drewil , dr. riyadh jabbar al-bahadili* / Ghufuran Isam Drewil, Dr. Riyadh Jabbar Al-Bahadili // 2021.
7. «Вугільні ТЕС України забруднюють повітря найбільше в Європі – DW – 25.05.2021». Дата звернення: 4, Листопад 2023. [Online]. Доступний у: <https://www.dw.com/uk/vuhilni-tes-ukrainy-zabrudniuiut-povitria-naibilshe-v-yevropi-doslidzhennia/a-57653232>
8. Р. by Метеопост, «Метеопост - Архів погоди». Дата звернення: 23, Вересень 2023. [Online]. Доступний у: <https://meteopost.com/weather/archive/>
9. «Погода у Києві», SINOPTIK.UA. Дата звернення: 25, Вересень 2023. [Online]. Доступний у: [//ua.sinoptik.ua/ногода-київ](http://ua.sinoptik.ua/ногода-київ)
10. «Якість повітря в Україні онлайн: карта моніторингу якості повітря - SaveEcoBot». Дата звернення: 18, Вересень 2023. [Online]. Доступний у: <https://www.saveecobot.com/maps>
11. *coordinates-converter.com*, «Online converter to all coordinate systems | UTM, WGS.. | with map», *Online-Umrechner in alle gängigen Koordinatensysteme wie UTM, WGS, WGS84 und CH1903, Gauss-Krueger mit grosser Karte*. Дата звернення: 20, Вересень 2023. [Online]. Доступний у: <https://coordinates-converter.com/en/decimal/51.000000,10.000000?karte=OpenStreetMap&zoom=8>

ЗАСТОСУВАННЯ МЕТОДІВ МАШИННОГО НАВЧАННЯ ДЛЯ ДІАГНОСТИКИ ПНЕВМОНІЇ

Анотація: В роботі розглядається можливий спосіб виявлення ознак захворювання пневмонії на рентгенологічних зображеннях за допомогою методів машинного навчання. Окрім виявлення ознак захворювання також враховується їх класифікація відповідно до 3-рьох видів пневмонії (вірусна, бактеріальна, covid-19). Для розпізнавання ознак на зображеннях було використано згорткову нейронну мережу. Проаналізовано результати класифікації та шляхи підвищення точності системи.

Abstract: The paper considers a possible way to detect signs of pneumonia in X-ray images using machine learning methods. In addition to detecting the signs of the disease, their classification according to 3 types of pneumonia (viral, bacterial, covid-19) is also taken into account. A convolutional neural networks was used to recognize the features in the images. The results of classification and ways to improve the accuracy of the system are analyzed.

1. Вступ

Актуальність роботи. З розвитком технологій, зокрема з появою методів машинного навчання та їх вдосконаленням стало можливим покращувати продуктивність багатьох сфер сучасного життя. Серед найбільш актуальних шляхів застосування методів глибокого навчання на сьогоднішній день залишається впровадження рішень в медичній сфері. Швидкість, точність та якість є критичними параметрами, які враховуються при розробці рішення завдань пов'язаних з діагностикою та лікуванням. З появою пандемії covid-19 людство зіштовхнулось з такими проблемами як великий потім хворих та нестачанеобхідних ресурсів, що в свою чергу спровокувало високу смертність даного захворювання. Людський фактор також не можна виключити, трапляються випадки неправильно встановленого діагнозу, що значно подовжує процес лікування або ж навіть може зашкодити здоров'ю людини. Застосування методів машинного навчання для розпізнавання ознак пневмонії на рентгенологічних зображеннях може значно пришвидшити процес встановлення правильного діагнозу. Окрім того такий інструмент можна використовуватись як додаткове джерело знань, оскільки трапляються випадки коли складно точно визначити вид захворювання просто подивившись на рентген знімок.

Мета й завдання роботи. Метою роботи є розробка методу діагностики різних видів пневмонії за рентгенологічними зображеннями.

2. Аналіз існуючих підходів

На сьогоднішній день існують деякі дослідження, пов'язані з розпізнаванням ознак захворювання легень на рентгенологічних зображеннях. Зокрема у дослідженні [1] описано метод виявлення covid-19 за допомогою нейронної мережі контекстної агрегації (CANN), яка складається з трьох модулів: context fuse model (CFM), attention mix module (AMM), residual convolutional module (RCM). Суть методу даного дослідження полягає у виявленні двох основних областей, пов'язаних з covid-19: симптом «матового скла» (туманне зниження прозорості легеневої тканини) та зону ущільнення легеневої тканини, внаслідок заповнення повітряних альвеол. В даному випадку визначається одне захворювання, проте також є дослідження з визначення різних захворювань легень, як у дослідженні [2], наприклад, класифікація двох схожих за ознаками на рентген-знімках захворювань: пневмонії та

туберкульозу. Запропонована в даному дослідженні система працює на основі згорткових нейронних мереж в поєднанні з такими методами машинного навчання як SVM та PCA.

Дослідження [3] присвячено класифікації більше ніж двох захворювань легень, а саме: пневмонії, раку легень, туберкульозу, непрозорості легень та covid-19. Методологія даного дослідження полягає у поєднанні VGG19 та CNN. Також, класифікація більше ніж двох захворювань представлена у дослідженні [4]. Створена система у вигляді поєднання моделей EfficientNetB0, EfficientNetB1, EfficientNetB2 для видобування ознак та класифікатора у вигляді Random Forest + SVM + логістична регресія здатна розпізнавати пневмонію, туберкульоз та covid-19.

У дослідженні [5] на основі згорткової нейронної мережі було побудовано систему, здатну розпізнавати пневмонію, covid-19, кардіомегалію, помутніння легень та плевру на цифрових рентгенівських знімках грудної клітки. Окрім того, в даному дослідженні було проведено експеримент з порівняння багатокласової та бінарної класифікації за допомогою розробленої системи. В результаті бінарна класифікація показала значно вищі результати точності. Для покращення результатів класифікації захворювань за рентгенологічними знімками у дослідженні [6] було також запропоновано метод попередньої обробки зображень на основі фільтра Гауса та логарифмічного оператора. Такий метод виявився економічно ефективним та простим у використанні.

Високої точності, приблизно 97%, було досягнуто при класифікації пневмонії та covid-19 в дослідженні [7]. В основі системи, розробленої в даному дослідженні, згорткова нейронна мережа VGG та попередня обробка зображень за допомогою методу обмеженого контрасту адаптивного вирівнювання гістограми (CLAHE) в поєднанні з фільтром гомоморфної трансформації.

В дослідженні [8] запропоновано архітектуру CNN, яка працює на зменшеній кількості даних, що суттєво пришвидшує швидкість її навчання. Метою даного дослідження було досягнути максимально можливої точності при максимально можливому зменшенні навчальних даних. При побудові моделі було застосовано методіку, що імітує когнітивну увагу. В результаті отримана система визначала пневмонію з точністю 95.47%. Згорткову нейронну мережу також було використано у дослідженні [9] для класифікації рентгенологічних зображень на два класи: «норма» та «пневмонія».

Постановка проблеми. На даний момент основними проблемами при вирішенні задачі розпізнавання захворювання за рентгенологічними зображеннями є підвищення точності та ефективності системи, а також наявність необхідних даних. Як відомо медичні дані часто є недоступними або їх немає в достатній кількості, рішенням цього є штучне доповнення датасету за допомогою методів аугментації або ж шляхом підвищення ефективності класифікаційної моделі, як запропоновано у дослідженні [8].

3. Опис запропонованого методу

В даному дослідженні запропоновано використання згорткової нейронної мережі для класифікації рентгенологічних зображень з метою виявлення бактеріальної пневмонії, вірусної пневмонії та спричиненої covid-19. Різні види пневмонії мають свої характерні ознаки, представлені на *Рис. 1* та Таблиці 1.

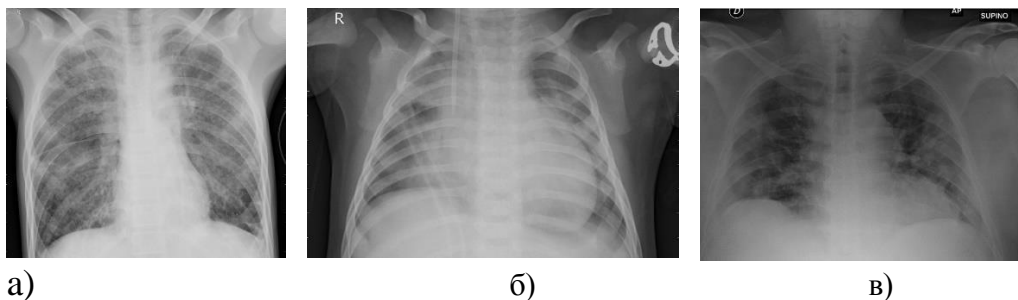


Рис.1 Рентгенологічні зображення легень з ознаками наявності пневмонії різни видів: а) вірусна, б) бактеріальна, в) covid-19

Таблиця 1.1. Характерні ознаки різних видів пневмонії

Вид пневмонії	Характерні візуальні ознаки
Вірусна	Характеризується м'якими, непрозорими затемненнями, які дозволяють бачити тканину легень. Менш видимі консолідації (згущення тканини легень) (приклад на <i>Рис. 1.2.</i>).
Бактеріальна	Характеризується більш вираженими консолідаціями (згущеннями легень), має локалізований характер (приклад на <i>Рис. 1.3.</i>).
Covid-19	Затемнення у вигляді мозаїки, є асиметричними (приклад на <i>Рис. 1.4.</i>).

Для збільшення кількості тренувальних даних та стійкості до змін застосуємо аугментацію до тренувального датасету. Аугментація даних – це процес створення нових прикладів даних шляхом застосування різних операцій трансформації до зображень з датасету. Серед операцій, які були застосовані до зображень в даному дослідженні наступні:

- зміна розміру зображення, шляхом обрізки інформаційно неважливих країв зображення, адже для вирішення задачі важлива лише область легень, що розташована ближче до центру;
- поворот зображення на випадково обраний кут та у довільно обрану сторону дозволяє створити варіації зображень під різними кутами;
- зміна контрастності та яскравості зображення.

Аргументація даних покращує навчання моделей, готуючи їх до нових умов та ситуацій. У випадку з рентгенологічними зображеннями, в залежності від пристрою на якому було створено зображення, можливі певні варіації якості таких зображень. А також враховуючи, що зазвичай медичні дані є не доступними, або ж їх недостатньо, за допомогою аргументації можна штучним способом збільшити кількість даних. Окрім того, проведено попередню обробку зображень, шляхом стандартизації даних та підвищення контрастності. Отримані зображення передаються на вхід згорткової нейронної мережі, схема якої представлена на *Рис. 2.*

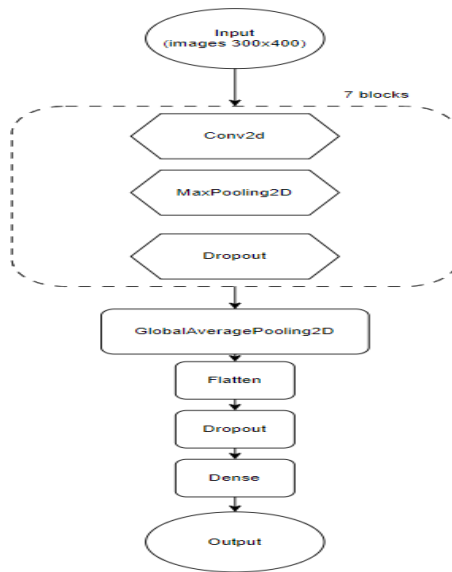


Рис. 2. Схема архітектури CNN

Згорткова нейронна мережа, використана для класифікації в даному дослідженні, складається з 7-ми згорткових блоків, кожен з яких включає згортковий шар, за яким слідує MaxPooling та Dropout. Згорткові шари використовуються для виявлення різних ознак на зображеннях, MaxPooling зменшує розмірність ображення, зберігаючи важливі ознаки. Dropout використовується для регуляризації мережі (запобігання перенавчання). Після останнього згорткового блоку використовується Global Average Pooling, для зменшення розмірності даних і створення вектора ознак, та 2 повноз'язні шари для класифікації зображень на 4 класи. Вихідний шар становить повністю зв'язаний шар з 4-рьох нейронів та функцією активації softmax, що генерує імовірність належності зображення до кожного з 4-рьох класів. Детальніший опис мережі представлено на Рис. 3 з урахуванням кількості застосованих фільтрів кожного шару.

Layer (Type)	Output Shape	Param #
input (InputLayer)	[(None, 300, 400, 1)]	0
conv2d_1 (Conv2D)	(None, 300, 400, 128)	1280
maxpool2d_1 (MaxPooling2D)	(None, 150, 200, 128)	0
dropout_1 (Dropout)	(None, 150, 200, 128)	0
conv2d_2 (Conv2D)	(None, 150, 200, 128)	147584
maxpool2d_2 (MaxPooling2D)	(None, 75, 100, 128)	0
dropout_2 (Dropout)	(None, 75, 100, 128)	0
conv2d_3 (Conv2D)	(None, 75, 100, 128)	147584
maxpool2d_3 (MaxPooling2D)	(None, 37, 50, 128)	0
dropout_3 (Dropout)	(None, 37, 50, 128)	0
conv2d_4 (Conv2D)	(None, 37, 50, 256)	295168
maxpool2d_4 (MaxPooling2D)	(None, 18, 25, 256)	0
dropout_4 (Dropout)	(None, 18, 25, 256)	0
conv2d_5 (Conv2D)	(None, 18, 25, 256)	590880
maxpool2d_5 (MaxPooling2D)	(None, 9, 12, 256)	0
dropout_5 (Dropout)	(None, 9, 12, 256)	0
conv2d_6 (Conv2D)	(None, 9, 12, 512)	1188160
maxpool2d_6 (MaxPooling2D)	(None, 4, 6, 512)	0
dropout_6 (Dropout)	(None, 4, 6, 512)	0
conv2d_7 (Conv2D)	(None, 4, 6, 512)	2358880
maxpool2d_7 (MaxPooling2D)	(None, 2, 3, 512)	0
dropout_7 (Dropout)	(None, 2, 3, 512)	0
global_average_pooling2d (GlobalAveragePooling2D)	(None, 512)	0
flatten (Flatten)	(None, 512)	0
dense (Dense)	(None, 1824)	525312
dropout_head_2 (Dropout)	(None, 1824)	0
dense_1 (Dense)	(None, 128)	131200
output (Dense)	(None, 4)	516

Total params: 5378692 (20.52 MB)
 Trainable params: 5378692 (20.52 MB)
 Non-trainable params: 0 (0.00 Byte)

Рис. 3. Архітектура розробленої CNN

Тренування мережі було проведено при поділі датасету на 70% тренувальних, 30 % тестувальних даних, а також при поділі – на 80% тренувальних, 20 % тестувальних даних. Вищої точності 88% було досягнуто притренуванні мережі на 80% даних. При тренуванні на 70% даних було досягнуто точності 86%. На Рис. 4 представлено графік функції втрат під час тренування мережі.

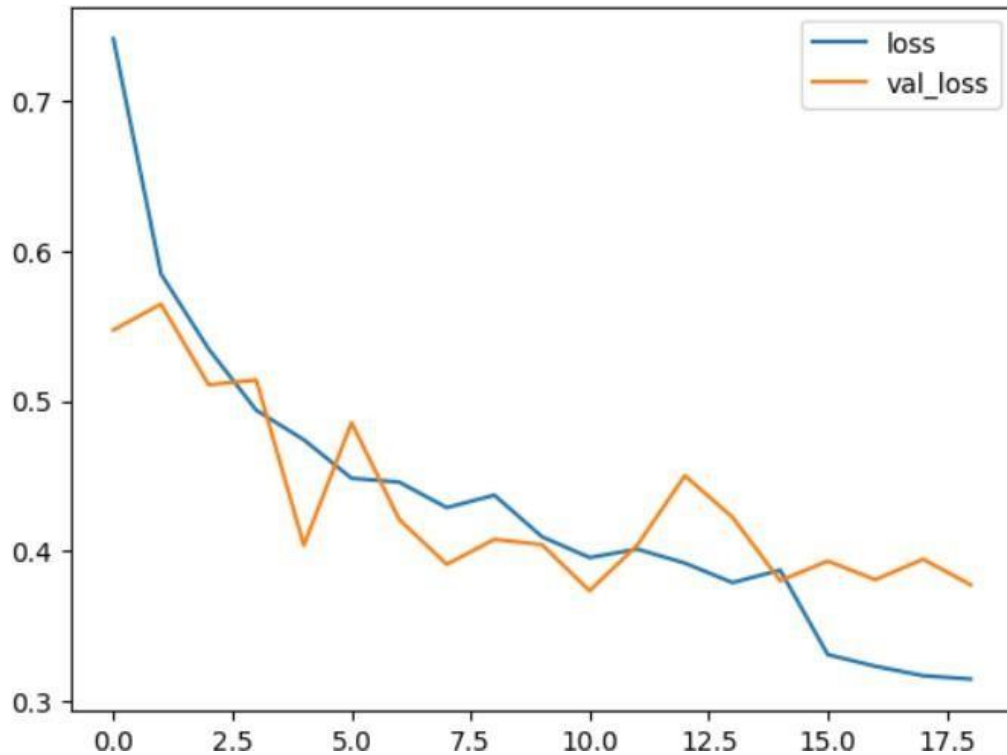


Рис. 4. Графік функції втрат

ВИСНОВКИ

Запропоновано архітектуру згорткової нейронної мережі для класифікації рентгенологічних зображень на 4 класи (вірусна пневмонія, бактеріальна пневмонія, covid-19, відсутність захворювання). Було досягнуто точності 88% при класифікації мережею, натренованою на 80% зображень використовуваного датасету. В якості попередньої обробки зображень було застосовано аргументацію, що дозволило розширити датасет.

Для подальшого вдосконалення мережі та підвищення точності класифікації можна спробувати ускладнити архітектуру мережі, а також застосувати ансамблі декількох таких мереж для більш ефективної роботи та кращих результатів.

СПИСОК ПОСИЛАНЬ

1. Afif M., Ayachi R., Said Y., Atri M. Deep learning-based technique for lesions segmentation in CT scan images for COVID-19 prediction. *Multimedia Tools and Applications*. 2023.
2. Ahmed I. A., Senan E. M., Shatnawi H. S. A., Alkhraisha Z. M., Al-Azzam M. M. A. Multi-Techniques for Analyzing X-ray Images for Early Detection and Differentiation of Pneumonia and Tuberculosis Based on Hybrid Features. *Diagnostics*. 2023. Bun. 13, № 4. С. 814.
3. Alshmrani G. M. M., Ni Q., Jiang R., Pervaiz H., Elshennawy N. M. A deep learning architecture for multi-class lung diseases classification using chest X-ray (CXR) images. *Alexandria Engineering Journal*.

2023. *Bun.* 64. C. 923–935.

4. Ravi V., Acharya V., Alazab M. A multichannel EfficientNet deep learning-based stacking ensemble approach for lung disease detection using chest X-ray images. *Cluster Computing*. 2023. *Bun.* 26, №2. C. 1181–1203.
5. Nahiduzzaman Md., Islam Md. R., Hassan R. ChestX-Ray6: Prediction of multiple diseases including COVID-19 from chest X-ray images using convolutional neural network. *Expert Systems with Applications*. 2023. *Bun.* 211. C. 118576.
6. Shaheed K., Szczuko P., Abbas Q., Hussain A., Albathan M. Computer-Aided Diagnosis of COVID-19 from Chest X-ray Images Using Hybrid-Features and Random Forest Classifier. *Healthcare*. 2023. *Bun.* 11, № 6. C. 837.
7. Shibu George G., Raj Mishra P., Sinha P., Ranjan Prusty M. COVID-19 detection on chest X-ray images using Homomorphic Transformation and VGG inspired deep convolutional neural network. *Biocybernetics and Biomedical Engineering*. 2023. *Bun.* 43, № 1. C. 1–16.
8. Singh S., Singh Rawat S., Gupta M., K. Tripathi B., Alanzi F., Majumdar A., Khuwuthyakorn P., Thinnukool O. Deep Attention Network for Pneumonia Detection Using Chest X-Ray Images. *Computers, Materials & Continua*. 2023. *Bun.* 74, № 1. C. 1673–1691.
9. Raheja K., Goel A., Mahajan M. Detecting Pneumonia Lung Infection From X-Ray Images with Deep Learning. *Journal of Physics: Conference Series*. 2021. *Bun.* 1950, № 1. C. 012052.

Нечволода Л.В., Крикуненко К.М., Макаров С.І. (Донбаська державна машинобудівна академія, м. Краматорськ-Тернопіль, Україна)

ВИКОРИСТАННЯ НЕЙРОННОЇ МЕРЕЖІ ДЛЯ ПРОГНОЗУВАННЯ ТРЕНДІВ ОНЛАЙН ІГОР

Анотація. У роботі розглядається математичний апарат для прогнозування трендів та загальної зацікавленості користувачів онлайн ігор з використанням технологій нейронних мереж. Для прогнозування трендів пропонується модель, у якій використовується двошарова нейронна мережа з нормалізацією даних і підготовкою навчальних пар для подальшого навчання мережі. В якості вхідних даних для такої моделі виконується виборка показників онлайн-активності з соціальної мережі користувачів онлайн ігор Steam. Відзначається системний підхід щодо використання нейронних мереж для прогнозування соціально-економічних процесів у мережі Інтернет.

Abstract. This paper considers a mathematical apparatus for forecasting trends and general interest of users of online games using neural network technologies. For forecasting trends, a model is proposed that uses a two-layer neural network with data normalization and training of training pairs for further training of the network. As input data for such a model, a selection of online activity indicators from the social network of Steam online game users is performed. A systematic approach to the use of neural networks for forecasting socio-economic processes on the Internet is noted.

ПОСТАНОВКА ПРОБЛЕМИ. У сфері ігрової промисловості успіх нерідко вимірюється сумарним доходом і кількістю завантажень. Ці показники можуть досягати позначок у мільйони доларів, що не завжди є доступним для більшості ігрових продуктів на ринку. Проте, навіть невеликі компанії можуть успішно розвивати свій бізнес, якщо оберуть правильний підхід. Для досягнення успіху у довгостроковій перспективі необхідно приймати виважені рішення для подальшого розвитку. Успіх на ринку залежить від того, наскільки ігри відповідають попиту на ринку та інтересам користувачів. Цей процес у сучасному світі має назву «трендвотчінг».

За цих умов можливим інструментом для прийняття рішень компаніями – розробниками програмного забезпечення може виступити застосування прогнозування трендів у комп'ютерних іграх. В якості обробки великих масивів та прогнозу трендів пропонується розглянути середній онлайн комп'ютерних ігор, використати статистичні методи [1] та штучні нейронні мережі.

АНАЛІЗ ОСТАННІХ ПУБЛІКАЦІЙ. Об'єктом соціально-економічного прогнозування є соціально-економічні процеси – тобто. сукупність економічних та соціальних процесів формування та функціонування соціально-економічної системи, що характеризують динаміку зміни її параметрів на певному рівні господарювання. Застосування методів соціально-економічного прогнозування для прийняття управлінських рішень стало об'єктом досліджень В. Г. Воронкової [2], Геєць В. М. [3], Б. Є. Грабовецького [4]. У роботі [5] І. Мункачій для розрахунку прогнозного значення показника пропонується використовувати метод екстраполяції трендів. У роботі [6] декомпозиція трендів виявила наявність сезонної компоненти для соціальної інженерії. У роботі [7] розглядається прогнозування тренду світових цін на базі СМІ-моделі економічних циклів.

Метою роботи є підвищення ефективності процесу прийняття рішень підприємствами з розробки та впровадження комп'ютерних ігор за рахунок створення методики прогнозування трендів та загальної зацікавленості користувачів онлайн ігор з використанням технологій нейронних мереж. Така методика дозволить розробити інформаційну систему підтримки прийняття рішень та автоматизувати процес отримання рекомендацій на базі трендвотчингу.

Авторами пропонується модель, у якій використовується двошарова нейронна мережа з нормалізацією даних і підготовкою навчальних пар для подальшого навчання мережі. Мережа має два входи і один вихід. Для підготовки даних перед навчанням проводиться нормалізація вхідних даних в діапазоні від 0 до 1 [8]. На практиці дана модель базуватиметься на використанні бібліотеки `brain.js` для прогнозування онлайн-активності в геймінговому сервісі Steam. Steam – онлайн-сервіс цифрового розповсюдження комп'ютерних ігор та програм, розроблений компанією Valve. Steam виконує роль платформи для багатокористувацьких ігор і потокового мовлення, а також соціальної мережі для гравців.

Під терміном «трендвотчінг», який має англомовне походження (від *trend watching*), розуміється здатність правильно відстежувати тренди на ринку, а також своєчасно реагувати на їхню появу в інформаційному середовищі. Під трендом варто розуміти явище, якого зберігається стійкий інтерес з боку користувачів. При цьому і саме це явище характеризується своєрідною довговічністю та стабільністю.

Така прогнозна інформація дасть можливість зрозуміти, яким чином змінюється кількість гравців від місяця до місяця та визначити майбутні тренди. Модель використовує нейронні мережі для прогнозування майбутнього онлайн у іграх. Її підготовка включає кілька етапів:

1. Збір даних – використання даних про середню кількість гравців, приріст гравців, приріст у відсотках та пік онлайн для кожного місяця з жовтня 2021 року по жовтень 2023 року.

2. Нормалізація даних – перетворення значень на шкалу від 0 до 1. Це допомагає моделі краще адаптуватися до вхідних даних та уникнути проблем, пов'язаних з різницею в масштабі значень різних атрибутів, для нормалізації даних буде використовуватися наступна формула:

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

де X_i – нормалізоване значення, X_I – оригінальне значення, X_{max} – максимальне значення, X_{min} – мінімальне значення.

3. Побудова навчальних пар. На основі нормалізованих даних формуються навчальні пари [9], в яких вхідні дані відповідають одному місяцю, а вихідні дані передбачають кількість середніх гравців наступного місяця.

4. Створення та навчання нейронної мережі. Модель використовує нейронну мережу з параметрами, які можна налаштовувати. Після підготовки даних модель навчається на цих навчальних парах.

5. Прогнозування майбутнього онлайн. Після навчання модель використовується для прогнозу середньої кількості гравців на наступний місяць на основі останніх доступних даних. Після прогнозу виконується перегляд результату для отримання прогнозованої кількості гравців у реальних одиницях, для цього дані повинні бути денормалізовані. Використовуючи формулу нормалізації виразимо X_I , внаслідок чого отримаємо наступну формулу:

$$X_I = X_i (X_{max} - X_{min}) + X_{min} \quad (2)$$

де X_I – денормалізоване значення, X_i – нормалізоване значення, X_{max} – максимальне значення, X_{min} – мінімальне значення.

6. Побудування графіку. На основі отриманих даних, будується графік, який відображає динаміку змін середньої кількості гравців, а також включає в себе навчальні дані,

тестові дані та прогнозовані значення. Побудований графік дозволить візуально порівняти реальні дані з прогнозними, це дасть змогу зрозуміти, наскільки добре модель вирішує задачу прогнозування.

ВИСНОВКИ

Таким чином, запропонована модель дозволяє краще розуміти тенденції та взаємозв'язки між різними атрибутами геймінгових даних і прогнозувати майбутні зміни у геймінговому середовищі. Вона може бути використана для покращення управління розробкою програмних продуктів у сфері комп'ютерних ігор, а також для прийняття інформованих рішень щодо розвитку геймінгової індустрії.

СПИСОК ПОСИЛАНЬ

1. *Максішко Н.К. (2006) Аналіз та прогнозування еволюції економічних систем/Н.К. Максішко, В.О. Перепелиця. Запоріжжя: Поліграф, 236 с.*
1. *Воронкова В. Г. (2004) Соціально-економічне прогнозування : навчальний посібник. Київ: Професіонал, 284 с.*
2. *Геєць В. М., Клебанова Т. С., Черняк О. І. (2005) Моделі і методи соціально-економічного прогнозування : підручник. Харків : ВД "ІНЖЕК", 396 с.*
3. *Грабовецький Б. Є. (2003) Економічне прогнозування та планування : навчальний посібник. Київ: Центр учбової л-ри, 188 с.*
4. *Мункачій Ірина, Грицишин Анна, Білоус Софія. (2023) Методика екстраполяції трендів як інструмент прогнозування туристичних потоків в Україні. International Science Journal of Management, Economics & Finance. Vol. 2, No. 4, 2023, pp. 65-76. doi: 10.46299/j.isjmef.20230204.08.*
5. *Яровенко, Г., Солярова, К. (2023). ПРОГНОЗУВАННЯ ІНФОРМАЦІЙНИХ ТРЕНДІВ КІБЕРАТАК ЯК ІНСТРУМЕНТ ПРОТИДІЇ ВРАЗЛИВОСТЕЙ В ЕКОНОМІЦІ. Економіка та суспільство, (51). <https://doi.org/10.32782/2524-0072/2023-51-27>.*
6. *Бандура О.В. (2018) Прогнозування тренду світових цін на нафту на базі СМІ-моделі економічних циклів. Екон. prognosuvannâ 2018; 2:91-110. <https://doi.org/10.15407/eip2018.02.091>*
7. *Субботін С. О. (2014) Нейронні мережі : навч. посіб. для студентів ВНЗ, які навчаються за напрямом підгот. "Програмна інженерія"/С. О. Субботін, А. О. Олійник ; під заг. ред. д-ра техн. наук, проф. С. О. Субботіна ; Запоріз. нац. техн. ун-т. Запоріжжя : ЗНТУ, 131 с.*
8. *Нейромережеві засоби прогнозування споживання енергоресурсів [Електронний ресурс]. Режим доступу: https://nv.nltu.edu.ua/Archive/2018/28_5/31.pdf*

Олійник С.Ю., Онищук С.Г., Тулупов В.І. (Донбаська державна машинобудівна академія, м. Краматорськ, Україна)

ІНТЕГРАЦІЯ ШТУЧНОГО ІНТЕЛЕКТУ ТА НЕЙРОННИХ МЕРЕЖ У КУРСІ «ОСНОВИ ТЕХНІЧНОЇ ТВОРЧОСТІ»

***Анотація.** У статті розглянуті проблеми освоєння дисципліни «Основи технічної творчості» в умовах дистанційного та очного навчання. Визначені основні перепони для досягнення результатів навчання.*

***Abstract:** The article deals with the problems of mastering the discipline "Fundamentals of Technical Creativity" in the conditions of distance and full-time education. The obstacles to achieving learning outcomes are identified.*

Дисципліна "Основи технічної творчості" в сучасних умовах визначається як ключовий елемент освіти на бакалаврському рівні на кафедрі інноваційних технологій та управління [1]. Освітня програма цієї дисципліни має глибокі корені та визначену структуру, яка розвиває важливі навички для студентів [2, 3]. Швидкі зміни в технічному світі вимагають вирішення нових завдань у техніці та технологіях, що робить цю дисципліну актуальною.

У сучасному освітньому середовищі важливо не лише отримувати професійні навички, але і розвивати творчі здібності майбутніх інженерів. Фахівці сьогодення розв'язують складні завдання, об'єднуючи різні наукові галузі та забезпечуючи конкурентоздатність своїх продуктів. Активізація технічної творчості сприяє виникненню нових і вдосконалених рішень.

Застосування штучного інтелекту та нейронних мереж може значно поліпшити процес викладання дисципліни "Основи технічної творчості". Нейромережеві технології можуть бути використані для аналізу індивідуальних особливостей студентів, а також для пропозицій індивідуальних завдань, спрямованих на їхні сильні сторони. Штучний інтелект може служити інструментом для автоматизації оцінювання творчих робіт студентів та розробки індивідуальних підходів до кожного з них.

Курс "Основи технічної творчості" може бути перетворений за допомогою штучного інтелекту і нейромереж. Наприклад, система може використовувати алгоритми машинного навчання для аналізу та класифікації ідей, які пропонують студенти, і надавати індивідуальний фідбек для покращення їхніх творчих навичок.

Застосування інноваційних технологій, таких як штучний інтелект та нейромережі, в контексті дисципліни "Основи технічної творчості" може стати перспективним напрямком розвитку, сприяючи більш ефективному вивченню та розвитку творчих здібностей студентів.

Використання штучного інтелекту та нейронних мереж в курсі "Основи технічної творчості" може бути вдосконаленою стратегією для підтримки студентів у їхньому навчальному процесі та розвитку творчих навичок. Ось деякі напрямки, які можна розглядати:

1. **Персоналізоване навчання:**

- Системи штучного інтелекту можуть аналізувати індивідуальні навички та слабкі сторони кожного студента.

- Нейромережі можуть створювати персоналізовані навчальні траєкторії, де студентам пропонуються завдання та матеріали, які відповідають їхньому рівню знань і темпу навчання.

2. **Автоматизована оцінка творчості:**

- Застосування нейронних мереж для оцінювання творчих робіт може допомогти об'єктивізувати процес оцінювання.

- Можливість визначення інноваційних елементів та оцінка їхнього впливу на рішення проблеми може бути автоматизованою.

3. **Генерація творчих ідей:**

- Застосування генеративних нейронних мереж для стимулювання творчого мислення та генерації ідей.
- Використання алгоритмів машинного навчання для аналізу ідей студентів та надання їм рекомендацій для поліпшення та розширення їхніх концепцій.
- 4. **Сприяння вирішенню творчих завдань:**
 - Створення віртуальних асистентів, які базуються на алгоритмах машинного навчання та можуть надавати студентам підтримку під час вирішення творчих завдань.
 - Використання нейромереж для аналізу підходів до розв'язання проблем та розробки стратегій для оптимального вирішення.
- 5. **Відстеження та аналіз прогресу:**
 - Застосування алгоритмів навчання для відстеження та аналізу прогресу студентів у розвитку творчих навичок.
 - Визначення індивідуальних досягнень та запропонування вдосконалень для подальшого зростання.

Загальна мета використання штучного інтелекту та нейронних мереж у курсі "Основи технічної творчості" полягає в тому, щоб створити ефективне та інтерактивне середовище для студентів, яке сприяє розвитку їх творчих здібностей та глибокому розумінню технічних концепцій.

СПИСОК ПОСИЛАНЬ

1. *Основи технічної творчості та наукових досліджень : конспект лекцій, методичні вказівки до практичних робіт для студентів спеціальності 131 Прикладна механіка / В. І. Тулупов, С. Ю. Олійник. – Краматорськ: ДДМА, 2017. – 124 с.*
2. *Основи технічної творчості. Частина 1: Навчальний посібник для студентів спеціальності 136 – металургія (бакалаврський рівень) / Укл.: Б.М. Бойченко, Л.С. Молчанов, Є.В. Синегін. – Дніпро: НМетАУ, 2019. – 53 с.*
3. *Прасолов Є.Я. Основи технічної творчості: навчальний посібник / Є.Я. Прасолов, С.А. Браженко, О.П. Новицький. – 2023, С.128*

Павлюк О.М., Медиковський М.О., Міщук М.В. (Національний університет «Львівська політехніка», Львів, Україна)

МЕТОД ДАТА-МАЙНІНГУ У НЕЙРОМЕРЕЖЕВОМУ ПРОГНОЗУВАННІ ЗАЛИШКОВОГО ЗАРЯДУ АКУМУЛЯТОРНОЇ БАТАРЕЇ КЕРОВАНОЇ РОБОТОТЕХНІЧНОЇ ПЛАТФОРМИ

Анотація: В роботі представлено метод на основі глибоких нейронних мереж та дата-майнінгу для прогнозування напруги в елементах акумуляторних батарей керованої робототехнічної платформи. Для виявлення відсутніх даних проведено їх попереднє опрацювання на основі аналізу варіації шляхом розрахунку: стандартного відхилення, варіантності, мінімальних та максимальних значень, діапазону, лінійного та стандартного відхилення для всіх дозволених значень сигма. Втраченими вважали дані, сигма яких перевищувала 1,5 та їх відновили ковзним середнім з вікном 12. Шлях який проходить керована робототехнічна платформа розбили на сегменти. Для кожного з них на основі дата-майнінгу розраховували додаткові параметри: кількість, тривалість, дельта напруги, кінцеву та початкову напругу під час проїзду сегмента. Дані розбили на статичні та динамічні категорії. Підготували тренувальні, валідаційні та тестові набори для глибоких нейронних мереж. Експерименти, проведені на даних КРП Formica 1 компанії AIUT Ltd., показали високу точність короткострокового (MAPE = 0,09%) та середньострокового (MAPE = 0,18%) прогнозування.

Abstract: This paper presents a method based on deep neural networks and data mining for predicting the voltage in the battery cells of a controlled robotic platform. To identify the missing data, they were pre-processed based on the analysis of variation by calculating: standard deviation, variability, minimum and maximum values, range, linear and standard deviation for all allowed sigma values. Data with a sigma of more than 1.5 were considered lost and were recovered by a moving average with a window of 12. The path of the controlled robotic platform was divided into segments. For each of them, additional parameters were calculated based on data mining: number, duration, voltage delta, final and initial voltage during the segment. The data was divided into static and dynamic categories. We prepared training, validation, and test sets for deep neural networks. Experiments conducted on the data of the Formica 1 switchgear of AIUT Ltd. showed high accuracy of short-term (MAPE = 0.09%) and medium-term (MAPE = 0.18%) forecasting.

Із розвитком концепції індустрії 4.0 керовані робототехнічні платформи (КРП) стали не лише джерелом різноманітних даних отриманих з їх сенсорів [1,2]. Виникла необхідність у інтелектуальному аналізі цих великих масивів даних з метою покращення управлінських рішень щодо логістики КРП. У сучасних внутрішніх транспортних системах, які відзначаються гнучкістю та ефективністю, важливо максимально ефективно використовувати акумуляторні батареї КРП. Оскільки КРП є енергозалежними, то від точного прогнозу залишкового заряду їх акумуляторних батарей залежить весь процес використання цих платформ на підприємствах. Тому проблема прогнозування розряду батареї та загального управління нею залишається актуальною, незалежно від конкретного типу КРП.

На сьогоднішній день одним із найперспективніших напрямків прогнозування залишкового заряду в елементах акумуляторних батарей КРП є методи машинного навчання [3,4]. Особливість системи нейромережевого моделювання як методу для побудови прогностичних моделей полягає в її здатності автоматично вивчати та адаптуватися до складних нелінійних залежностей у вхідних даних. В їх основу покладено штучні нейронні мережі (ШНМ). Вони дозволяють враховувати складні взаємозв'язки між різними характеристиками і враховувати їх динаміку в часі та адаптуватися до змінних умов. ШНМ можуть працювати з великим обсягом різнотипних даних. Точність навчання ШНМ залежить не лише від об'єму, але й від якості даних.

Основними недоліками які повинні бути усунуті в даних є [5, 6]:

- наявність аномалій, помилоки чи відсутність важливої інформації;
- нерепрезентативність вибірки з неадекватним відображенням розподілу змінних;
- малий обсяг даних;

- наявність шуму в даних;
- відсутність балансу між різними класами у вибірці;
- наявність викривлень з прихованими чи систематичними спотвореннями.

Збір даних КРП Formica 1 на основі який робили прогноз здійснювали на одному поверсі підприємства АІУТ, Глівіце Польща. На Рис 1. представлена мапа маршруту КРП Formica 1 побудована за допомогою спеціалізованого програмного забезпечення Navitrol v 6.4 [6, 7, 8].

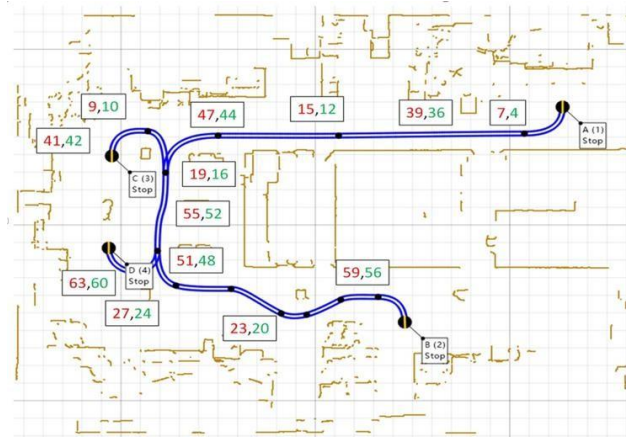


Рис 1. Мапа маршруту КРП Formica 1 на підприємстві АІУТ.

Світло-коричневим кольором позначені стаціонарні перешкоди (стіни, стаціонарні робочі місця, де працюють люди, палети та ін.), де рух КРП був заборонений. Синіми лініями позначено шлях, яким дозволений рух КРП. Навігаційні точки, між якими їздить КРП, позначені великими чорними крапками з коричневою вертикальною лінією. Маленькі чорні крапки - це відрізки, на які поділено весь маршрут. Сукупність навігаційних точок утворює агрегат, тобто логічну одиницю, в межах якої КРП повністю виконає одне поставлене перед ним завдання. Маршрут складається з пронумерованих сегментів червоними та зеленими кольорами. Червоним кольором, коли КРП рухається вперед, і зеленим, коли він рухається задом. Цим сегментам присвоюються унікальні номери при створенні карти маршруту і не змінюються до тих пір, поки карта маршруту не буде змінена.

На рис.2. показано фрейм структури збору даних КРП Formica 1.

Field	BEGINNING	ID	Status	TIMESTAMP	NUMBER	VER	LENGTH	DATA	ENDING			
Size (bytes)	3	2	2	12	4	2	2	various	3			
Byte number	1 2 3	4 5	6 7	8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23	24 25 26 27	28 29 30 31 32			40 41 42			
Format	ASCII	ASCII	INT	WORD	DTL	UINT	INT	INT	various	ASCII	ASCII	ASCII
Example	K	1	1							1	1	V

Рис. 2 Фрейм структури збору даних КРП Formica 1.

Дані щодо енергоспоживання розміщені від 83 байту до 98 байту. В них йде запис таких параметрів:

- миттєве споживання струму (mA);
- напруга елемента живлення (mV)
- миттєве споживання енергії (W)
- миттєве споживання енергії (Ws)
- акумулятивне споживання енергії (Wh)

Наступних 3 байти є зарезервованими під інші функції пов'язані з енергоспоживанням КРП Formica 1. Всі історичні дані зчитувалися з бази даних та надсилалися на порт 5000. За допомогою клієнтського протоколу OPC UA дані записувалися у файл з розширенням ".pk1" у

форматі CSV (comma-separated values). Для отримання необхідних параметрів КРП з файлу ".pkl" використовувалася програмна бібліотека Pandas мови Python.

Дата-майнінг здійснювали у кілька етапів. Згідно стандарту протоколу OPC UA з метою економії дискового простору та оптимізації пропускну здатності каналу зв'язку, дані записуються лише при зміні значення параметра. Спочатку їх відновили з щосекундною рівномірною дискретизацією по осі часу. Далі усереднили щохвилинно. Дані запис яких не здійснювався в задані часові проміжки – вилучили з датасету (наприклад, "ActualSpeed_L" та "ActualSpeed_R").

Наступним кроком - видалили спонтанні викиди для покращення якості навчання ШНМ, оскільки після усереднення викиди все ще залишилися в сигналах. Метод, який було використано для виявлення та видалення викидів, базувався на дисперсійному аналізі. Побудували для кожного сигналу гістограму розподілу та втраченими даними вважали спонтанні викиди, сигма яких перевищувала 1,5. Якщо значення шуму певного параметра перевищувало поріг сигми, його замінювали ковзним середнім значенням по 12 точках (оптимальне значення вибрали експериментально). Також розраховували середнє квадратичне відхилення, дисперсію, мінімальне та максимальне значення, розмах, лінійне відхилення та середнє квадратичне відхилення.

Далі провели аналіз даних шляхом обчислення наступних параметрів для кожного сегмента:

- сегмент - номер зазначеного сегмента;
- тривалість - середня тривалість присутності КРП у певному сегменті;
- кількість вибірок - середня кількість вибірок для даного сегмента;
- дельта підрахунку напруги - математичне сподівання падіння напруги батареї після одного проходження даного сегмента;
- дисперсія дельти напруги - дисперсія падіння напруги акумулятора після проходження певного сегмента;
- маса - маса, яку КРП перевозить на даному відрізку;
- напруга на початку сегмента - усереднена напруга акумулятора на початку руху на даному сегменті.

Під час аналізу виявили 25 унікальних сегментів на основі яких розраховували такі нові параметри: сегмент, кількість відліків, тривалість, дельта напруги, маса та початкова напруга сегмента. Отримані параметри були згруповані за значенням "сегмента". Для кожного сегмента розраховували "кількість вибірок" - середню кількість вимірювань для всіх досліджуваних параметрів. Оскільки час, проведений КРП у сегменті, може бути різним, то також розраховували "тривалість" часу проходження КРП через сегмент. Коли КРП входила у сегмент, напруга на елементах батареї записувалася як параметр "початкова напруга сегмента". Різниця між значеннями напруги на елементах батареї на вході і виході з сегмента зберігалася в параметрі "дельта-напруги". Параметр "маса" відображав масу яку перевозить на собі КРП. Усі розраховані параметри розділили на статичні та динамічні. До статичних параметрів належали сегмент, кількість сегментів, тривалість, дельта напруги та маса. Динамічним параметром була напруга на початковому сегменті.

Для навчання ШНМ всі дані окрім "сегмент" нормалізували за допомогою функції Python MinMaxScaler() в діапазоні від 0 до 1. Для виконання коротко- та середньострокових прогнозів побудували комітет ШНМ прямого поширення без зворотного зв'язку, в якій сигнал подавався на вхідний шар з 12 нейронами, прихований шар з різною кількістю нейронів в залежності від параметра та вихідний шар з одним нейроном. Кількість нейронів на прихованому шарі вибиралася експериментально в діапазоні від 2 до 11. Число ШНМ в комітеті дорівнює кількості розрахованих за допомогою дата-майнінгу параметрів (5- якщо КРП не перевозив на собі вантажу, та 6 – з вантажем). Об'єднання ШНМ в комітет здійснювали за допомогою `tf.keras.layers.concatenate`.

Комітет ШНМ навчався за допомогою методу зворотного поширення з градієнтним спуском який є доволі часозатратним. Навчання здійснювали на ПК: Intel(R) Core(TM) i7-5500U CPU 2.40 GHz; 16.0 GB of installed RAM; a 64-bit operating system, and an x64-based processor. 5204 опрацьованих за допомогою запропонованого методу дана-майнінгу використали для навчання та тестування ШНМ. Перші 5052 вибірок були використані для навчання моделі, а решта 152 - для оцінки точності передбачення побудованої моделі. Для навчання ШНМ було використано Keras - бібліотеку нейронних мереж з відкритим вихідним кодом, що працює на Ten-sorFlow. Для вхідного та вихідного шарів було використано лінійну функцію активації, а для прихованого шару - функцію ReLU. Алгоритм adam використовувався як оптимізатор з середньою абсолютною відсотковою похибкою (MAPE) в якості міри. Для оцінки моделі було використано функцію "evaluate". Комітет нейромереж навчали та тестували в двох режимах TW1 – з перенавчанням на кожному кроці та TW2 – прогнозуванням методом «часових вікон» (ширина вікна 12).

Час навчання комітету ШНМ тривав від 0,4 хв до 1,96 хв. Швидкість обчислень можна підвищити, використовуючи такі методи, як міні-пакування та одночасне обчислення градієнта на декількох навчальних прикладах, а також на декількох персональних комп'ютерах об'єднаних в кластер. На Рис.3. представлений результат прогнозу залишкового заряду акумуляторної батареї керованої робототехнічної платформи після послідовного проходження 21 сегменту.

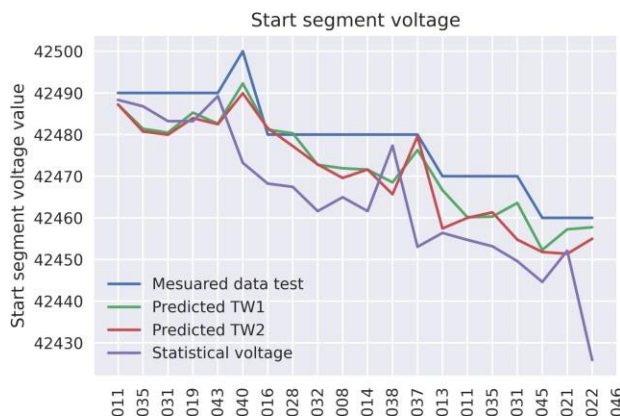


Рис.3. Результат прогнозу залишкового заряду акумуляторної батареї керованої робототехнічної платформи після послідовного проходження 21 сегменту

Точність прогнозу оцінювали за допомогою похибок MAPE (середню абсолютну похибку виражену у відсотках) та RMS (середню квадратичну похибку). Похибки розраховували для комітету ШНМ які працювали в режимах TW1 та TW2, а також stat (статистично розрахованих середніх значень в кожному сегменті). Результати довгострокового прогнозу для 21 сегменту представлені на рис 4.

MAPE (ANN:TW1) %	0.014653
RMS (ANN:TW1) %	0.003573
MAPE (ANN:TW2) %	0.01867
RMS (ANN:TW2) %	0.004514
MAPE (stat) %	0.032447
RMS (stat) %	0.003573

Рис.4. Точність довготерманового прогнозу заряду акумуляторної батареї КРП Formica 1 для 21 сегменту

Розроблений метод дана-майнінгу у нейромережевому прогнозуванні залишкового заряду акумуляторної батареї КРП Formica 1 показав ефективність і для іншої кількості сегментів. Короткостроковим прогнозом вважали прогноз для кількості сегментів від 1 до 7,

середньотерміновим від 8 до 14. Короткостроковий прогноз для семи сегментів і середньостроковий прогноз для дев'яти сегментів становив 0,09% і 0,18% відповідно, а найдовший час навчання - 1,96 хв. Таким чином відносно попередніх нейромережових прогнозів без розробленого методі дата-майнінгу [6, 7, 8] точність прогнозів зросла до 4 разів.

ВИСНОВКИ

У роботі представлено розроблений метод дата-майнінгу у нейромережевому прогнозуванні залишкового заряду акумуляторної батареї КРП. Основними кроками якого є: попереднє опрацювання даних; усунення втрачених даних; відновлення спонтанних викидів; розрахунок додаткових параметрів для кожного сегменту шляху. Метод протестовано на датасеті КРП Formica 1. Проведені експерименти показують зростання точності прогнозів до 5 разів, що підтверджує високу ефективність та точність цього методу.

СПИСОК ПОСИЛАНЬ

1. Abderrahim, Moussa, et al. 'Manufacturing 4.0 Operations Scheduling with AGV Battery Management Constraints'. *Energies*, vol. 13, no. 18, Sept. 2020, p. 4948. DOI.org (Crossref), <https://doi.org/10.3390/en13184948>.
2. Hu, Hao, et al. 'Self-Adaptive Traffic Control Model With Behavior Trees and Reinforcement Learning for AGV in Industry 4.0'. *IEEE Transactions on Industrial Informatics*, vol. 17, no. 12, Dec. 2021, pp. 7968–79. DOI.org (Crossref), <https://doi.org/10.1109/TII.2021.3059676>
3. Benecki, Pawel, et al. 'Forecasting of Energy Consumption for Anomaly Detection in Automated Guided Vehicles: Models and Feature Selection'. *2022 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*, IEEE, 2022, pp. 2073–79. DOI.org (Crossref), <https://doi.org/10.1109/SMC53654.2022.9945146>.
4. Medykovskvi, Mykola, et al. 'Use of Machine Learning Technologys for the Electric Consumption Forecast'. *2018 IEEE 13th International Scientific and Technical Conference on Computer Sciences and Information Technologies (CSIT)*, IEEE, 2018, pp. 432–35. DOI.org (Crossref), <https://doi.org/10.1109/STC-CSIT.2018.8526617>.
5. Izonin, Ivan, et al. 'An Approach towards Missing Data Management Using Improved GRNN-SGTM Ensemble Method'. *Engineering Science and Technology, an International Journal*, vol. 24, no. 3, June 2021, pp. 749–59. DOI.org (Crossref), <https://doi.org/10.1016/j.jestch.2020.10.005>.
6. Pavliuk, Olena, et al. 'The Forecast of the AGV Battery Discharging via the Machine Learning Methods'. *2022 IEEE International Conference on Big Data (Big Data)*, IEEE, 2022, pp. 6315–24. DOI.org (Crossref), <https://doi.org/10.1109/BigData55660.2022.10020968>.
7. Олена Павлюк, Микола Медиковський, Наталія Луса, Мирослав Міщук. Інтелектуальна система аналізу процесів споживання заряду акумуляторними батареями. *Information systems and networks Issue 13*, 2023
8. О.М. Павлюк, М.О. Медиковський, М.В. Міщук. Методологія машинного навчання для прогнозу спаду заряду акумуляторної батареї автоматизованого керованого транспортного засобу. *Computer technologies of printing 2023/1(49)*

Слюсар В.І. (Центральний науково-дослідний інститут озброєння та військової техніки ЗС України), Слюсар І.І. (Полтавський державний аграрний університет, Україна)

КЛАСИФІКАЦІЯ ЗИМОВОГО ТА ЛІТНЬОГО СЕЗОНІВ З БОРТА БПЛА

Анотація: Сезонні зміни в середовищі та ландшафті вимагають від систем Object Detection високого рівня гнучкості та адаптивності. Нейромережі, які не враховують ці зміни, можуть мати знижену точність та ефективність у різних сезонних умовах. Тому включення сезонних варіацій у тренувальні та тестувальні набори даних є критично важливим для розробки надійних та ефективних систем Object Detection. В роботі запропоновано підхід для врахування сезонно-специфічних умов виявлення об'єктів з борта БПЛА на основі згорткової нейронної мережі класифікації зображень. Робота такої нейромережі має передувати застосуванню нейромережі YOLO, в якій відповідно змінюються сезонно-залежні вагові коефіцієнти. Спеціально сформований датасет дозволив класифікувати зимові та літні сцени з точністю 97.6 %.

Abstract: Seasonal changes in the environment and landscape require a high level of flexibility and adaptability from Object Detection systems. Neural networks that do not take these changes into account may have reduced accuracy and effectiveness in different seasonal conditions. Therefore, including seasonal variations in training and testing data sets is critically important for developing reliable and efficient Object Detection systems. The work proposed an approach to account for season-specific conditions in object detection from UAVs based on a convolutional neural network for image classification. The operation of such a neural network must precede the application of the YOLO neural network, in which season-dependent weight coefficients are accordingly changed. A specially formed dataset allowed for the classification of winter and summer scenes with 97.6% accuracy.

Важливим аспектом при реалізації нейромереж виявлення об'єктів (Object Detection [1, 2]) є врахування пори року. Сезонні зміни можуть радикально змінювати зовнішній вигляд ландшафтів та середовищ. Наприклад, взимку снігопад може покривати та деформувати візуальні ознаки, за якими алгоритм нейромережі YOLO розпізнає об'єкти (рис. 1), в той час як восени листя може забарвлюватися та опадати, змінюючи зовнішній вигляд дерев та іншої рослинності як фону для візуальних сцен. Сезонність суттєво впливає на візуальні характеристики об'єктів і вносить варіації, які можуть заплутати нейромережі, якщо вони не були належним чином треновані. Наприклад, відображення снігу на вулицях може вплинути на розпізнавання на них машин, дорожніх знаків або інших об'єктів. Сезонні зміни неминуче впливають на кольори та текстури у навколишньому середовищі, що знижує здатність алгоритмів виявлення об'єктів відрізнити об'єкти від фону.

Отже, для ефективної роботи нейромережі в різних сезонних умовах, важливо включати диверсифіковані набори даних, які враховують сезонні варіації у візуальних характеристиках об'єктів та середовищ.

Метою роботи є врахування сезонної специфіки при виявленні об'єктів з борта БПЛА на основі попередньої класифікації сезонних змін за допомогою нейронних мереж загорткового типу [3].

Питання про те, чи краще мати окремі датасети для зими та літа, або один єдиний датасет, що включає сцени обох сезонів, залежить від конкретних цілей та обмежень того чи іншого проєкту з виявлення об'єктів. В контексті БПЛА відповідні міркування обмежуються необхідністю здешевлення та спрощення електронного сегменту бортового обладнання. З урахуванням жорстких вимог до вартості, габаритів та електроспоживання обчислювальних засобів окремі датасети для зими та літа дозволяють більш точно налаштувати нейромережу на специфічні умови кожного сезону. Це може бути корисно в період міжсезонних коливань погодних умов або при вирішенні завдань з великою дальністю польоту, коли маршрут БПЛА перетинає кілька різних кліматично-географічних зон.

Окремі датасети для різних сезонів можуть дозволити ефективніше використовувати доступні ресурси, зосереджуючись на найбільш релевантних сценах за умови обмежених ресурсів для тренування.



Рис. 1. Сніг на поверхні ґрунту

Для того, щоб нейромережі Object Detection могли надійно працювати в усіх сезонах, необхідно зробити так, щоб вони могли адаптуватися до цих змін. Створення сезонного набору даних є одним із ключових етапів у підвищенні ефективності систем Object Detection. Цей процес включає кілька кроків.

1) Розробка методології збору та анотації сезонних зображень для створення представницького набору даних. Перед початком збору даних необхідно визначити, які сезони і погодні умови представлятимуть найбільший інтерес для подальшого аналізу і використання. Наприклад, для аграрного сектору важливими будуть зображення посівів у різні періоди їх росту та збору врожаю, в той час як для міської інфраструктури може бути корисним аналіз стану доріг у зимовий період. Збір зображень може відбуватися через різноманітні джерела: супутникові знімки, аерофотознімання, знімки з беспілотних літальних апаратів або наземні фотографії. Для забезпечення різноманітності набору даних, зображення повинні бути отримані в різні часові періоди і за різних погодних умов. Кожне зібране зображення необхідно анотувати, що включає в себе відмітку об'єктів, які повинні бути розпізнані системою Object Detection. Анотація може бути ручною або напівавтоматичною з використанням попередньо навчених моделей для підвищення ефективності процесу. Після анотації проводиться перевірка на помилки і неточності. Цей крок важливий, оскільки від якості анотації безпосередньо залежить якість навчання нейронної мережі. Для підвищення узагальнюючої спроможності моделей розпізнавання об'єктів використовують розширення даних.

2) Використання розширених технік передобробки та аугментації зображень для підвищення їх різноманітності та якості. Це може включати зміну масштабу зображень, обертання, зміни освітлення та колірної гами, що допомагає моделі краще адаптуватися до різноманітних умов. Для розширення датасету та його збалансування по класах доцільно використовувати нейромережі генерації зображень типу DALL-E3, Midjourney, Stable Diffusion XL тощо. Перед використанням набору даних у навчанні моделей виконується його валідація. Вона може включати аналіз розподілу класів, перевірку на збалансованість даних та виявлення можливих викидів.

На основі проведених досліджень в роботі визначено структуру сезонно-специфічного сегменту багаторівневої архітектури нейронної мережі класифікації зображень місцевості. Вона містить два блоки. Перший блок проводить класифікацію сезону, представленому на

зображенні. В залежності від результату класифікації, у другому блоці активується відповідна до сезону модель нейронної мережі або завантажується специфічний для неї масив вагових коефіцієнтів за основним завданням, наприклад, реалізація автопілоту, Object Detection, Object Tracking та ін. Такий підхід дозволяє знизити вимоги до обчислювальних потужностей Edge Computing, зберігаючи продуктивність проєктів Edge Computing + AI Computer Vision. При цьому, необхідно вибрати архітектуру селективної нейромережі для реалізації першого блоку.

Формування датасету для навчання нейромережі класифікації сезонів здійснювалось на основі персонального фотоархіву, отриманого під час польотів авіарейсами різних європейських маршрутів. Датасет містить зображення двох класів: поверхня землі зі снігом (клас SNOW) – 154 шт. та без нього (клас NOT_SNOW) – 196 шт. Зображення робились з борту пасажирських літаків переважно на етапах злету та заходу на посадку на висотах до 5 км. При цьому на відповідних сценах представлені лісна та гориста місцевість, міські забудови, сільськогосподарські угіддя, об'єкти промислової та транспортної інфраструктури, річки та озера. Весь сукупний датасет був розділений у пропорції 70 % – на навчальну вибірку (рис. 2) та 30 % – на перевіірочну (SNOW – 47 шт. і NOT_SNOW – 59 шт.). Тестова вибірка містила 27 зображень, що не перетиналися з навчальною та перевіірочною вибірками.

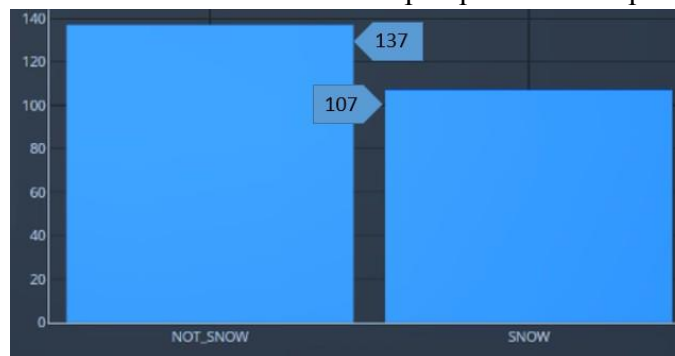


Рис. 2. Структура навчальної вибірки датасету

Для класифікації зображень на основі сформованого датасету було обрано архітектуру з використанням нейронної мережі MobileNetV2 [4], попередньо навченої на датасеті ImageNet [5 - 7]. В роботі розглядається архітектура, що наведена на рис. 3. Вона містить $\approx 2,4 \times 10^6$ параметрів (рис. 4).

Фрагмент шаблону коду на мові Python, що описує зазначену архітектуру, має вигляд:

```
from tensorflow.keras.layers import Input
from tensorflow.keras.layers import BatchNormalization
from tensorflow.keras.applications.mobilenet_v2 import MobileNetV2
from tensorflow.keras.layers import Dense
from tensorflow.keras.models import Model

input_1 = Input(shape=(180, 240, 3), name='1')
x_3 = BatchNormalization(axis=-1, momentum=0.99, epsilon=0.001, center=True,
scale=True, beta_initializer='zeros', gamma_initializer='ones',
moving_mean_initializer='zeros', moving_variance_initializer='ones',
beta_regularizer='l1_l2', gamma_regularizer='l1_l2', beta_constraint='min_max_norm',
gamma_constraint='min_max_norm', name='BatchNormalization_3')(input_1)

x_16 = MobileNetV2(include_top=False, weights='imagenet', pooling='avg', classes=1000,
classifier_activation='softmax', alpha=1.0)
for layer in x_16.layers:
    layer.trainable = False
x_16 = x_16(x_3)
```

Для даної моделі виконувалась підготовка зображень датасету шляхом зміни їх розмірів до величини 240x180 пікселів, а також оптимізація гіперпараметрів. Спочатку був вибраний розмір батчу 32 і крок навчання 0,001. Однак при вказаних значеннях параметрів

навчання спостерігався так званий ефект перенавчання після 40 епох. Тому для даної комбінації налаштувань процесу навчання довелося обмежитися саме такою кількістю епох.



Рис. 3. Архітектура нейронної мережі на основі MobileNetV2.

Model: "model_1"

Layer (type)	Output Shape	Param #
1 (InputLayer)	[(None, 180, 240, 3)]	0
BatchNormalization_3 (Batch Normalization)	(None, 180, 240, 3)	12
mobilenetv2_1.00_224 (Functional)	(None, 1280)	2257984
MobileNetV2_16 (Activation)	(None, 1280)	0
Dense_5 (Dense)	(None, 120)	153720
Dense_6 (Dense)	(None, 80)	9680
2 (Dense)	(None, 2)	162

=====
 Total params: 2,421,558
 Trainable params: 2,387,440
 Non-trainable params: 34,118

Рис. 4. Характеристика моделі нейромережі на основі MobileNetV2

В результаті навчання було зафіксовано точність класифікації літнього та зимового сезонів на навчальній вибірці 99,3 %, а на перевіірочній – 89,2 % (рис. 5). Графік помилки навчання наведений на рис. 6 (кращий результат на навчальній вибірці – 0,0115, на перевіірочній – 0,55). Фрагмент візуалізації результатів класифікації по тестовій вибірці наведений на рис. 7.

Наступним етапом була зміна розміру батча до 8 та кроку навчання – до величини 0,0001. Зменшення батчу дозволило знизити навантаження на апаратний сегмент (рис. 8) та отримати ефективність за показником точності *Balanced Recall* на рівні 100 % для навчальної і 93,2 % - на перевіірочній вибірках (рис. 9). Помилка навчання на перевіірочній вибірці становила 0,0227 (рис. 10).

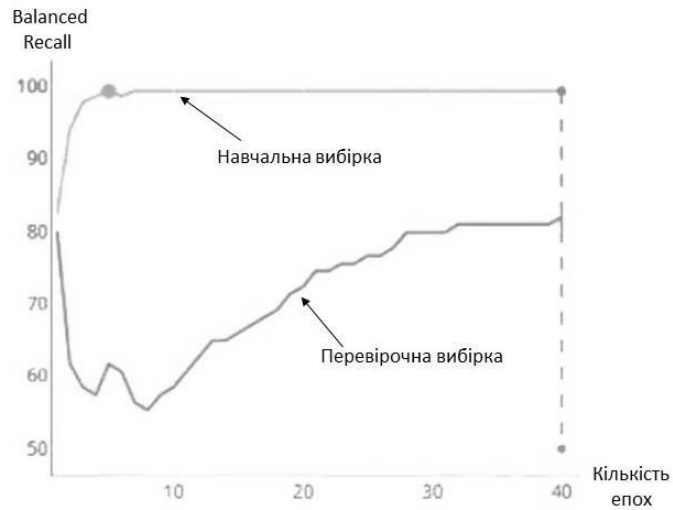


Рис.5. Точність неймережі на основі MobileNetV2

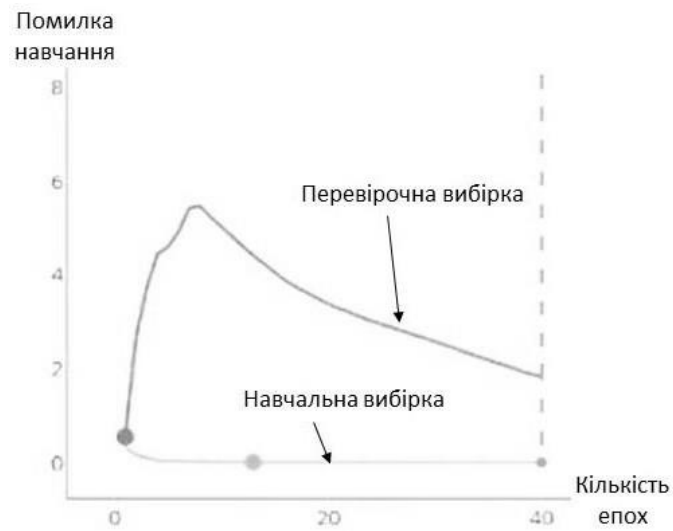


Рис.6. Помилка навчання







Шар	Вхідні дані	Вихідні дані	Справжнє значення	Статистика прикладів, %	
	Вхідний шар			Вихідний шар	
	Зображення			Клас	Клас
1		SNOW	SNOW	0	100
2		SNOW	SNOW	0	100
3		SNOW	SNOW	0	100
4		SNOW	SNOW	0	100
5		SNOW	NOT_SNOW	0	100
6		NOT_SNOW	SNOW	1,6	98,4

Рис. 7. Фрагмент візуалізації результатів класифікації



Рис. 8. Навантаження на обчислювальне середовище Google Colab при розмірі батчу 8.

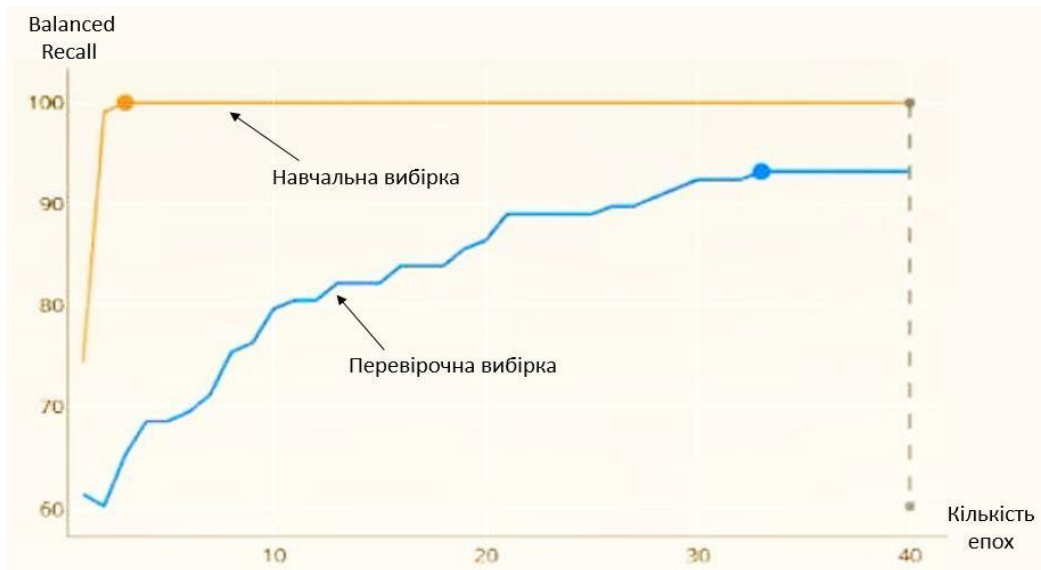


Рис. 9. Ефективність неймережі на основі MobileNetV2 (батч 8).

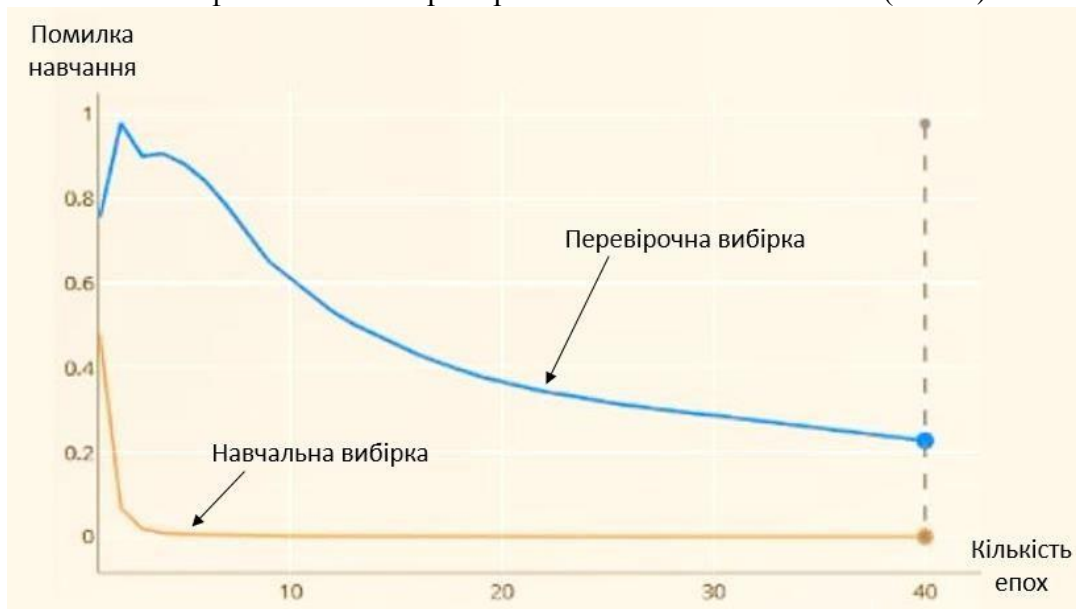


Рис. 10. Помилка навчання при батчі 8.

Матриця помилок наведена на рис. 11. Приклад роботи мережі наведено на рис. 12. Враховуючи поведінку *Balanced Recall*, можливо збільшити кількість епох навчання з метою підвищення точності класифікації. Подальші дослідження дозволили отримати *Balanced*

$Recall = 96,2\%$ при кількості епох 75 (рис. 13), помилка навчання на перевіірочній вибірці становила 0,1 (рис. 14).

Дійсне значення	NOT_SNOW	52 88.1%	7 11.9%
	SNOW	2 4.3%	45 95.7%
		NOT_SNOW	SNOW
		Прогноз	

Рис. 11. Матриця помилок





	NOT_SNOW	NOT_SNOW	100.0%	0.0%
	NOT_SNOW	NOT_SNOW	100.0%	0.0%
	NOT_SNOW	NOT_SNOW	100.0%	0.0%
	SNOW	SNOW	0.0%	100.0%

Рис. 12. Приклад роботи мережі

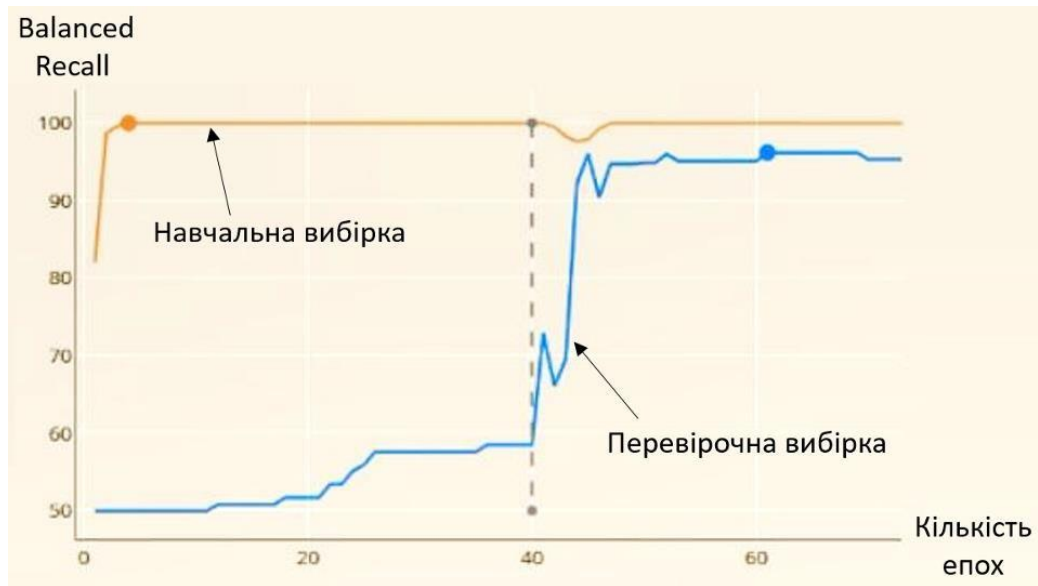


Рис. 13. Точність неймережі класифікації при 75 епохах навчання.

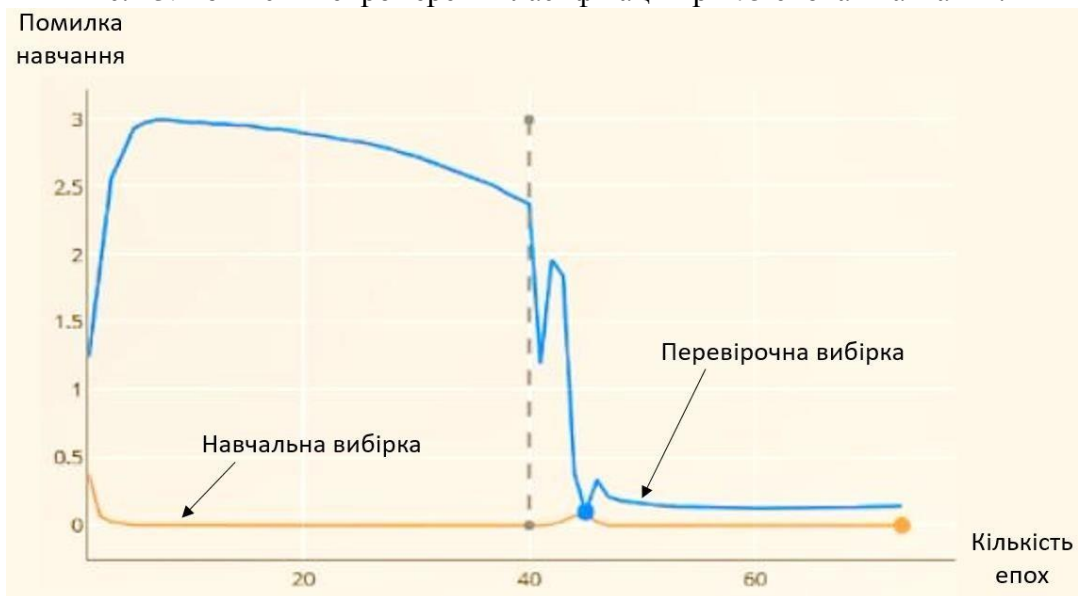


Рис.14. Помилка навчання

Проведені дослідження свідчать про можливість використання MobileNetV2 як основи для створення сезонно-специфічного сегменту багаторівневої архітектури нейронної мережі класифікації зображень місцевості.

Вибір між окремими сезонними датасетами та єдиним датасетом, що включає обидва сезони, залежить від специфіки задачі та доступних ресурсів. Якщо система має бути ефективною в різних сезонних умовах та має можливість обробляти великі об'єми даних, єдиний датасет з різноманітними сценами може бути кращим варіантом. Проте, якщо потрібна висока спеціалізація або обмежені ресурси, окремі датасети можуть бути більш доцільними. Подальші дослідження доцільно провести для інших попередньонавчених архітектур неймереж, зокрема Xception [7, 8], Inception [9] та ін.

СПИСОК ПОСИЛАНЬ

1. Vadym Slyusar, Mykhailo Protsenko, Anton Chernukha, Pavlo Kovalov, Pavlo Borodych, Serhii Shevchenko, Oleksandr Chernikov, Serhii Vazhynskiy, 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.
2. 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.
3. 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.
4. 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).
5. *ImageNet.* URL: <https://www.image-net.org>. (дата звернення: 27.10.2023).
6. Слюсар В.І. *Архітектурно-математичні основи удосконалення нейронних мереж з класифікації зображень. Штучний інтелект, 2022, № 1. С. 127-138. DOI: 10.15407/jai2022.01.127.*
7. *Xception.* URL: <https://stephan-osterburg.gitbook.io/coding/coding/ml-dl/tensorfow/ch3-xception>. (дата звернення: 27.10.2023).
8. *Inception.* URL: <https://en.wikipedia.org/wiki/Inception>. (дата звернення: 27.10.2023).

Наукове видання

**НЕЙРОМЕРЕЖНІ ТЕХНОЛОГІЇ
ТА ЇХ ЗАСТОСУВАННЯ
НМТіЗ-2023**

Збірник наукових праць

Загальною редакцією
д-ра техн. наук, проф. С. В. Ковалевського і
Hon.D.Sc., prof. Predrag Dašić

32/2016 Формат 60 x 84/16. Ум. друк. арк. 16,1
Обл.-вид. арк. 17,2. Тираж 100 прим. Зам. № 26

Видавець і виготівник
Донбаська державна машинобудівна академія
84313, м. Краматорськ, вул. Академічна, 72.
Свідоцтво суб'єкта видавничої справи
ДК № 1633 від 24.12.2003