

## ЗМІСТ

## ІНФОРМАЦІЙНІ ТЕХНОЛОГІЇ ТА ТЕОРІЯ КОДУВАННЯ

**О. А. Поплавський**

Інформаційна технологія обробки графічних даних на основі гібридних нейронних мереж з використанням геометричних особливостей об'єктів зображень 4

**Ю. В. Барішев, В. С. Ланова**

Інформаційна технологія захищеного зберігання результатів академічної успішності 17

**А.М. Тарновський, С.М. Захарченко, М.Г. Тарновський**

Проблеми сучасних методів тривимірної фотограметрії 31

**Д. Ю. Петрина, В. А. Корнута, О. В. Корнута**

Використання інструментів нейромереж для пришвидшення розробки веб-інтерфейсів 42

**Т.І. Коробейнікова, Л.А. Савицька, Л.В. Крупельницький**

Процеси асемблювання та реасемблювання документів у системі документообігу 51

## КОМП'ЮТЕРНІ СИСТЕМИ ТА КОМПОНЕНТИ

**V. M. Dubovoi**

Model-based learning of coordinators of the decentralized multi-zone objects control systems 66

**M. S. Forkaliuk, G. Franzl, O.V. Bisikalo**

Evaluating fast charging of electric vehicles along motorways using finite multi-server queueing system simulation 77

**Т. І. Коробейнікова, А.Б. Ямнич**

Багатовимірна матриця класифікування інформації для оцінки ризиків інформаційної безпеки 91

## БІОЛОГІЧНІ ТА МЕДИЧНІ ПРИЛАДИ І СИСТЕМИ

**Li Jinqiong, Sergii Pavlov, Anatolii Poplavskyu, Oleksandr Kaduk**

Construction Guidelines for Optical-Electronic Expert Systems in Blood Rheology 107

## МАТЕМАТИЧНЕ МОДЕЛЮВАННЯ ТА ОБЧИСЛЮВАЛЬНІ МЕТОДИ

**Р. В. Слободян, І. В. Богач, М. В. Барабан**

Застосування методів статистичного аналізу для вибору виконавця задачі у розподіленій обчислювальній системі 122

**До відома авторів**

133

## ІНФОРМАЦІЙНІ ТЕХНОЛОГІЇ ТА ТЕОРІЯ КОДУВАННЯ

УДК 681.5:004.67

### Інформаційна технологія обробки графічних даних на основі гібридних нейронних мереж з використанням геометричних особливостей об'єктів зображень

Олександр Анатолійович Поплавський

к.т.н, доцент кафедри інформаційних технологій,  
Київський національний університет будівництва і архітектури.  
03680, м. Київ, пр. Повітрофлотський, 31  
ORCID: 0000-0003-0465-6843

**Анотація.** Завдяки прогресу в галузі обчислювальної техніки спостерігається стабільне зростання обчислювальних потужностей, що призводить до експоненційного збільшення обсягів даних, які потребують обробки. Зокрема, підвищення продуктивності автоматизованих систем забезпечує можливість зберігання та аналізу великих масивів медичних даних з високою швидкістю та точністю. Сучасна медицина характеризується значним збільшенням інформаційного навантаження, що потребує складної обробки та глибокого аналізу для підтримки прийняття клінічних рішень. Інформаційні технології відіграють ключову роль у забезпеченні ефективної обробки цих великих обсягів даних, сприяючи підвищенню точності та швидкості діагностики, а також ефективності подальшого лікування пацієнтів. Метою даної статті є розробка та дослідження інформаційної технології обробки графічних даних на основі гібридних нейронних мереж з використанням геометричних особливостей об'єктів зображень. В роботі запропоновано передові методи машинного навчання, архітектури глибоких нейронних мереж, а також спеціалізовані інструменти для обробки графічних даних, такі як OpenCV, TensorFlow та інші. Процес обробки даних під час валідації запропонованих методів та архітектур включав декілька етапів: попередню обробку даних, навчання моделей та ретельне тестування отриманих результатів. Розроблена інформаційна технологія демонструє значне підвищення точності класифікації графічних даних. Експериментальні дослідження показали, що запропонований підхід забезпечує ефективну обробку великих обсягів біомедичних даних, що підтверджується високою точністю та швидкістю аналізу. Зокрема, точність класифікації патологій за допомогою гібридних нейронних мереж підвищено більш ніж на 11% у порівнянні з результатами отриманими за допомогою класичних методів. Практична цінність розробленої технології полягає в її високому потенціалі для застосування в області машинного зору, в т.ч для підвищення ефективності діагностики та лікування пацієнтів в медичній сфері. Вона може бути інтегрована у сучасні системи підтримки прийняття рішень, забезпечуючи більш точну та швидку обробку медичних зображень.

**Ключові слова:** обробка медичних даних, біоінженерія, інформаційні технології, машинне навчання, класифікація патологій, штучний інтелект.

**DOI:** <https://doi.org/10.31649/1999-9941-2024-60-2-4-16>

## Information technology for image data processing based on hybrid neural networks using geometric features

Oleksandr A. Poplavskiy

Ph.D., Associate Professor of the Department of Information Technologies,  
Kyiv National University of Construction and Architecture.  
03680, Kyiv, Povitrflotskyi Avenue, 31  
ORCID: 0000-0003-0465-6843

**Abstract.** Progress in computing technology has led to a steady increase in computing power, resulting in an exponential growth in the amount of data that needs to be processed. In particular, the enhanced performance of automated systems enables the storage and analysis of large volumes of medical data with high speed and accuracy. Modern medicine is characterized by a significant increase in the information load, necessitating complex processing and in-depth analysis to support clinical decision-making. Information technology plays a pivotal role in ensuring efficient processing of these large datasets, contributing to the accuracy and speed of diagnosis, as well as the effectiveness of subsequent patient treatment. The purpose of this article is to develop and study information technology for processing graphic data based on hybrid neural networks using geometric features of image objects. The paper proposes advanced machine learning methods, deep neural network architectures, and specialized tools for processing graphic data, such as OpenCV, TensorFlow, and others. The data processing workflow during the validation of the proposed methods and architectures included several stages: data pre-processing, model training, and thorough testing of the results. The developed information technology demonstrates a significant improvement in the accuracy of graphic data classification. Experimental studies have shown that the proposed approach ensures efficient processing of large volumes of biomedical data, as evidenced by the high accuracy and speed of analysis. In particular, the accuracy of pathology classification using hybrid neural networks increased by more than 11% compared to the results obtained using classical methods. The practical value of the developed technology lies in its high potential for use in the field of machine vision, including enhancing the efficiency of diagnosis and treatment of patients in the medical field. It can be integrated into modern decision support systems, providing more accurate and faster processing of medical images.

**Keywords:** medical data processing, bioengineering, information technology, machine learning, pathology classification, artificial intelligence.

**Вступ.** Завдяки сучасному розвитку обчислювальної техніки спостерігається постійне збільшення обчислювальних потужностей, що призводить до експоненційного зростання обсягів даних, які мають бути (A. Smith et al., 2023). Зокрема, підвищення продуктивності автоматизованих систем дозволяє зберігати та аналізувати великі масиви медичних даних з великою швидкістю та точністю. Це змушує науковців постійно шукати нові методи та моделі обробки зростаючої кількості інформації, адаптуючи інформаційні технології до нових стандартів якості та швидкості.

Останнім часом значного поширення набули нейронні мережі, які є особливо ефективними для обробки візуальної інформації, завдяки своїй здатності до узагальнення складних структур даних, та стали ключовим інструментом у сфері обробки даних (G. Davis et al., 2023). В цей час, більшість науковців та фахівців з обробки даних зосереджуються на модифікації архітектур нейронних мереж, налаштуванні гіперпараметрів та підвищенні

якості вхідних даних. Такі підходи значно підвищують точність, швидкість та якість обробки даних у порівнянні з застосуванням традиційних методів.

Ця стаття присвячена розробці інформаційної технології, що базується на поєднанні класичних методів обробки інформації та сучасних технологій, зокрема гібридних нейронних мереж, для підвищення точності аналізу даних у комп'ютерних системах підтримки рішень. Застосування такого підходу дозволяє поєднувати переваги методів, забезпечуючи ефективнішу обробку інформації та підвищуючи точність класифікації. Результати дослідження підтверджують доцільність таких підходів для покращення ефективності в інтелектуальних системах підтримки прийняття рішень.

**Актуальність.** Актуальність проблеми полягає в необхідності ефективної обробки великих масивів даних, які сьогодні генеруються з високою швидкістю та в значних обсягах. Сучасні інформаційні системи часто стикаються з труднощами в обробці та аналізі таких даних, що ускладнює прийняття точних і своєчасних рішень. Тому адаптація існуючих систем до динамічно змінюваних умов експлуатації, інтеграція новітніх досягнень в області машинного навчання та штучного інтелекту для забезпечення надійності та достовірності результатів обробки є актуальною і відповідає сучасним викликам.

**Мета.** Мета статті полягає в розробці та дослідженні ефективних методів обробки великих масивів даних, які базуються на використанні машинного навчання та нейронних мережі, ва відповідно нової інформаційної технології.

**Інформаційні технології обробки біомедичних даних.** У сфері біомедичних даних, графіка відіграє важливу роль в аналізі та візуалізації результатів досліджень. Комп'ютерний зір охоплює теорію та технологію створення штучних систем, які здатні обробляти вхідну інформацію у вигляді зображень. Візуальні дані можуть набувати різних форм, таких як відеоряди, зображення з різних камер або тривимірні дані, отримані за допомогою медичних сканерів (G. Litjens et al., 2017). Як технологічна дисципліна, комп'ютерний зір має на меті використовувати теорії та моделі комп'ютерного зору для створення систем, які можуть ефективно опрацьовувати графічну інформацію в режимі реального часу. Це особливо важливо у таких галузях, як системи управління промисловими процесами, автономні транспортні засоби, відеоспостереження, організація та індексація баз даних зображень, а також аналіз медичних зображень.

Сучасні інструменти комп'ютерного зору включають у себе бібліотеки для програмування, засоби для роботи з нейронними мережами та хмарні рішення для обробки зображень у реальному часі. Серед найпопулярніших інструментів варто відзначити OpenCV, TensorFlow, CUDA, Theano, Keras, Scikit-learn, YOLO та різноманітні API для обробки зображень у хмарі, такі як Google Cloud Vision API, Amazon Rekognition та Microsoft Azure Computer Vision API (Y. LeCun et al., 2017).

OpenCV (Open Source Computer Vision Library) є бібліотекою з відкритим вихідним кодом, що надає широкий набір функцій та алгоритмів для комп'ютерного зору, обробки зображень та числових алгоритмів. Вона дозволяє обробляти та аналізувати зміст зображень, розпізнавати об'єкти, відстежувати рух, конвертувати зображення, використовувати методи машинного навчання та ідентифікувати загальні елементи у різних зображеннях. TensorFlow, розроблена компанією Google, є найпопулярнішою бібліотекою для машинного та глибокого навчання, написаною на мові Python з підтримкою портів на JavaScript (tensorflow.js), що робить її доступною для широкого кола розробників (M. Abadi et al., 2016). TensorFlow надає інструменти для обробки даних та диференційованого програмування, включаючи підтримку нейронних мереж. CUDA (Compute Unified Device Architecture) від Nvidia дозволяє значно підвищити продуктивність обчислень за допомогою графічних процесорів, досягаючи інтерактивної швидкості обробки відео та інших графічних завдань, що є критичним для аналізу медичних зображень у реальному часі (J. Nickolls et al., 2008). Theano, розроблена командою LISA у Університеті Монреалю, є швидкою числовою бібліотекою для Python, яка підтримує обчислення на центральних та графічних процесорах, дозволяючи створювати

нові операції з графами у режимі роботи та виконувати численні операції з конвертацією графів (J. Bergstra et al., 2010). Keras є бібліотекою глибокого навчання для Python, що поєднує функції інших бібліотек, таких як TensorFlow, Theano та CNTK, акцентуючи на зручності, модульності та розширюваності, що дозволяє швидко експериментувати з глибокими нейронними мережами (F. Chollet, 2015). Scikit-learn, побудована на NumPy, SciPy та matplotlib, є популярною бібліотекою для машинного навчання, яка надає прості та ефективні інструменти для виявлення та аналізу даних, включаючи алгоритми навчання з учителем та без нього, лінійну та логістичну регресію, метод опорних векторів (SVM), градієнтний бустінг та кластеризацію (F. Pedregosa et al., 2011). YOLO (You Only Look Once) є системою виявлення об'єктів у реальному часі (M. Abadi et al., 2016), яка застосовує нейронну мережу до всього зображення, розбиваючи його на сітку та маркуючи регіони з виявленими об'єктами, що дозволяє досягти високої швидкості та точності виявлення об'єктів на зображеннях (J. Redmon та A. Farhadi, 2018).

Хмарні системи обробки зображень, такі як Google Cloud Vision API, Amazon Rekognition та Microsoft Azure Computer Vision API, надають потужні інструменти для аналізу зображень та відео (D. Kang et al., 2017; Y. Liu et al., 2017). Вони дозволяють розробникам використовувати попередньо навчені моделі машинного навчання для класифікації зображень, виявлення об'єктів, розпізнавання тексту та облич, а також аналізу відео у реальному часі.

Загалом, сучасні інформаційні технології обробки біомедичних даних включають широкий спектр інструментів та методів, які забезпечують ефективну та точну обробку візуальної інформації. Використання цих технологій у медичній практиці дозволяє підвищити якість діагностики та лікування, забезпечуючи кращі результати для пацієнтів (E. Shortliffe та J. Cimino, 2013).

Рис.1 ілюструє основні компоненти класичних експертних систем (ЕС), що забезпечує обробку біомедичних даних для підтримки прийняття клінічних рішень. Інтерфейс користувача призначений для введення інформації користувачем (менеджером, фахівцем) в експертну систему за допомогою меню, команд, природної мови або власного інтерфейсу. Команди містять параметри, що спрямовують процес обробки знань. Інтерфейс також забезпечує отримання вихідної інформації у формі значень, присвоєних певним змінним (S. Dilsizian та E. Siegel, 2014). Такі експертні системи, як правило, надають не тільки рішення, але й пояснення, які можуть бути запитані користувачем або стосуватися кроків вирішення поставлених задач.



**Рисунок 1.** Основні компоненти інформаційної технології класичних біомедичних експертних систем

База знань у сучасних медичних експертних системах являє собою великі масиви біомедичних даних, що містять детальну інформацію про різноманітні медичні аналізи, зображення, діагнози, лікування та інші аспекти клінічної практики (F. Jiang et al., 2017).. Ці дані включають як структуровану, так і неструктуровану інформацію, яка описує проблемну галузь і забезпечує логічний взаємозв'язок між різними медичними фактами. Завдяки обробці великих масивів біомедичних даних експертна система може надавати більш точні та обґрунтовані клінічні рішення, враховуючи всі можливі варіанти і підходи до лікування пацієнтів.

Модуль створення системи служить для формування набору правил, використовуючи або алгоритмічні мови програмування, або оболонки експертних систем. Останні представляють собою готову програмну середу, яка може бути пристосована до вирішення певної проблеми шляхом створення відповідної бази знань. Експерт і спеціаліст по знанням забезпечують визначення правил та алгоритмів, що використовуються для аналізу даних і прийняття рішень (X. Liu et al., 2019).

Як бачимо, переважна більшість сучасних технологій обробки біомедичних даних призначена для завантаження великих обсягів даних та їх навчання на основі відомих результатів. Це забезпечує можливість адаптації систем до нових умов та вдосконалення моделей прогнозування (G. Litjens et al., 2017).

Важливим аспектом використання великих даних у біомедичних дослідженнях є необхідність забезпечення високої точності та швидкості обробки інформації. Машинне навчання, зокрема глибокі нейронні мережі, дозволяють ефективно аналізувати великі обсяги даних, знаходити приховані закономірності та робити точні прогнози на основі історичних даних (A. Esteva et al., 2017). Такий підхід значно покращує якість діагностики та лікування, оскільки системи можуть швидко і точно визначати наявність патологій, прогнозувати розвиток захворювань та рекомендувати оптимальні стратегії лікування.

Одним із перспективних напрямів розвитку є інтеграція різних джерел біомедичних даних для створення більш повних та багатшарових моделей. Наприклад, об'єднання даних медичної візуалізації, геномних досліджень, електронних медичних записів, результатів обробки класичними методами та інших джерел може забезпечити більш глибоке розуміння складних біологічних процесів та підвищити точність діагностичних і прогностичних моделей (E. Topol, 2019). Такі підходи вже показали свою ефективність у ряді досліджень, демонструючи значне покращення результатів у порівнянні з традиційними методами.

**Методологія обробки великих обсягів даних з використанням нейронних мереж.** Сучасні методи обробки великих обсягів медичних даних часто застосовують нейронні мережі, які виявляють високу ефективність у розпізнаванні складних зв'язків у даних. Одним із ключових етапів розробки таких систем є ретельний аналіз і підготовка вхідних даних. Важливо усунути аномалії та нерелевантні дані, оскільки вони можуть впливати на результати аналізу.

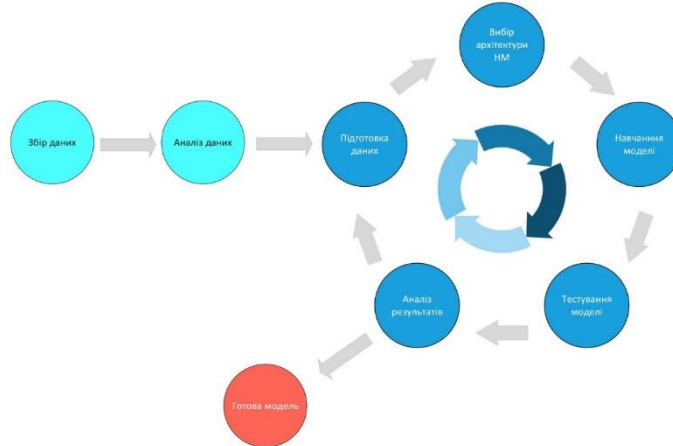
Для обробки графічних і статистичних даних застосовуються класичні методи попередньої обробки, які включають нормалізацію, вилучення текстурних характеристик та інші техніки (O. Ronneberger et al., 2015). Після попередньої обробки база знань, що містить великі масиви медичних даних, передається спроектованій нейронній мережі для подальшого навчання та аналізу.

Точність моделей машинного навчання, зокрема нейронних мереж, безпосередньо залежить від якості та обсягу тренувальних даних. Чим більша кількість якісних даних використовується для тренування моделі, тим більш точні та надійні результати можна отримати при класифікації. (V. Mnih et al., 2015).

Прикладом успішної реалізації може слугувати архітектура згорткової нейронної мережі, яка навчається на підготовлених великих даних (O. Poplavskiy et al., 2020). В цьому випадку навчання моделі включає оптимізацію вагових коефіцієнтів для мінімізації

розбіжності між передбаченими моделлю та фактичними мітками класів. Процес навчання завершується тестуванням моделі на нових даних для оцінки її точності та ефективності.

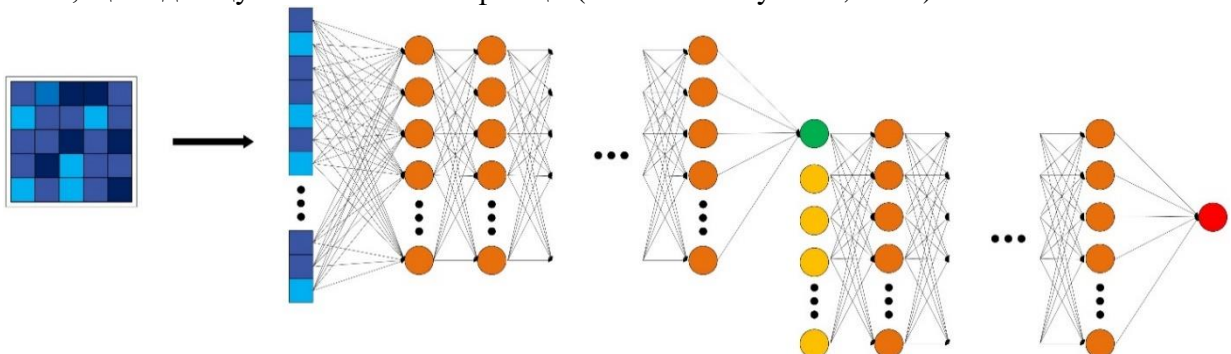
У даному алгоритмі ключову роль відіграють такі фактори, як вибір архітектури нейронної мережі, налаштування параметрів навчання та якість вхідних даних.



**Рисунок 2.** Класична схема створення автоматизованої системи класифікації, що базується на нейронних мережах

Для отримання зображень, як правило, дотримуються певного протоколу для зниження небажаних ефектів. Це включає належну підготовку пацієнта, точне налаштування обладнання та дотримання стандартних процедур зйомки (J. Bushberg et al., 2011). Після отримання зображень проводиться попередня оцінка для визначення їх якості та відповідності вимогам дослідження. На цьому етапі визначається, чи достатньо зібраних даних для проведення подальших аналізів та забезпечення достовірності результатів.

Для класифікації зображень можна використати згорткову нейронну мережу (CNN), яка складається з кількох типів шарів. Згортковий шар застосовує фільтри для виконання операцій згортки, що дозволяє сканувати вхідне зображення та генерувати карти активації, які містять всі виявлені ознаки. Шар пулінгу використовується для ущільнення ознак після згорткового шару, застосовуючи максимальне або середнє значення ознак. Повнозв'язаний шар який обробляє, плоский вхід, де кожен вхід пов'язаний з усіма нейронами, що дозволяє з'єднати приховані шари з вихідним шаром і оптимізувати оцінки класів. Комбінація цих шарів у CNN забезпечує ефективне витягнення, ущільнення та класифікацію ознак зображень, що підвищує точність класифікації (A. Krizhevsky et al., 2012).



**Рисунок 3.** Модифікація повнозв'язного шару на виході згорткової нейронної мережі

Створення архітектури нейронної мережі, в даному випадку включає кілька конволюційних та повнозв'язних шарів, що дозволяють аналізувати зображення та виявляти візуальні ознаки. Архітектура починається з конволюційного шару для сканування вхідних зображень та виявлення простих ознак, далі слідує шар активності та максимального

згортання для зменшення розмірності даних. Після серії конволюційних шарів використовується повнозв'язний шар для класифікації зображень.

Навчання моделі здійснюється на основі попереднь визначеного датасету, що включає зображення з відповідними мітками, за допомогою ітеративного підходу. Використання оптимізатора Adam та функції втрат CrossEntropyLoss дозволяє ефективно оновлювати ваги моделі, а процес тестування включає обробку зображень тестового набору даних для отримання прогнозів щодо стану здоров'я пацієнтів, що дозволяє оцінити точність моделі.

В контексті даного дослідження, особлива увага була приділена виокремленню текстурних характеристик (A. Poplavska et al., 2021) зображень, які містить датасет та обчисленню часу виконання окремих кроків методології. Такий підхід включав збір великої кількості текстурних ознак, що сприяло всебічному опису геометричних характеристик зображень. Вилучення таких характеристик як коефіцієнт асиметрії та площа асиметрії, та інших (V. Vassilenko et al., 2020) забезпечувало можливість точно класифікувати та аналізувати зображення на наявність патологічних змін.

**Результати експериментальних досліджень обробки графічних даних на основі гібридних нейронних мереж з використанням геометричних особливостей.** Методика проведення експериментальних досліджень має важливе значення для будь-якого наукового дослідження, оскільки вона лежить в основі валідності та надійності отриманих результатів. У даному дослідженні використовувалась методика проведення експериментальних досліджень, спрямованих на оцінку ефективності та надійності розроблених методів та інструментів, а також кінцевої моделі запропонованого методу. Апробація розроблених методів на принципово різних наборах великих даних, графічних та статистичних, застосовувалась для підтвердження їх універсальності та достовірності результатів. В цьому контексті, застосування технології до графічних даних, які охоплюють серії зображень вимагає аналізу складних візуальних взаємозв'язків та патернів. Це включає розпізнавання об'єктів, обробку та аналіз текстур, кольорів та форм. З іншого боку, статистичні дані, які в даному дослідженні складаються з числових записів-таблиць з 24 попередньо виокремлених для кожного зображення кількісних даних, таких як автокореляція, контраст, ентропія, коефіцієнт асиметрії (A. Poplavska et al., 2021) та інші.

Початковий етап дослідження включав обробку даних за допомогою згорткової нейронної мережі (CNN), що мала на меті опрацювання важливих характеристик медичних зображень для класифікації патологій. Цей процес передбачав кілька ключових кроків, кожен з яких сприяв підготовці та обробці великих даних для подальшого глибокого навчання.

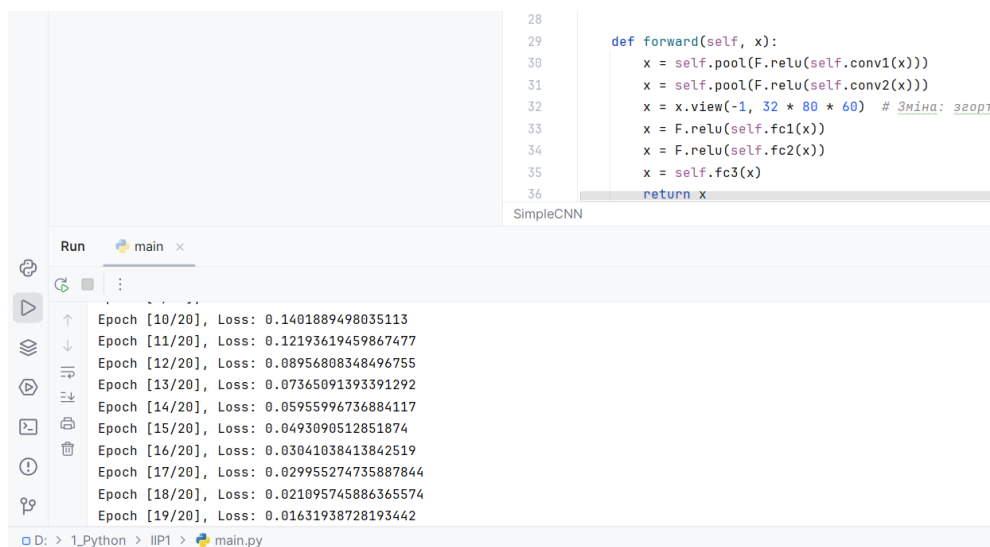
*Завантаження та попередня обробка даних* використовувалась з застосуванням спеціалізованих програмних бібліотек та інструментів, які дозволяють автоматизувати та оптимізувати обробку великих обсягів даних. Для завантаження та маніпуляції даними у дослідженні використовувалась бібліотека Pandas, яка забезпечила широкий спектр функціоналу для читання даних з різноманітних файлових форматів, в т.ч. для імпорту датасетів. Для підготовки зображень до подальшого аналізу використовувалась бібліотека Pillow (PIL), що дозволило відкривати, маніпулювати та зберігати зображення.

*Процес розділення даних* на тренувальні та тестові вибірки включало застосування функцій `train_test_split` з бібліотеки Scikit-learn для поділу датасету, де перші 90% даних були віднесені до тренувальної вибірки, а решта 10% – до тестової. Такий підхід є певним стандартом який забезпечує збалансоване тренування моделі та її адекватну оцінку на нових даних.

Застосування трансформацій до зображень перед тренуванням мережі допомогло підвищити ефективність навчання, зменшуючи ризик перенавчання та підвищити стійкість моделі до варіацій вхідних даних. У дослідженні використовувалися стандартні трансформації бібліотеки Torchvision, включаючи конвертацію зображень у тензори та їх нормалізацію, що є стандартною практикою при роботі з глибоким навчанням.

Розроблена модель CNN, SimpleCNN, яка є класичним прикладом згорткової нейронної мережі, включала декілька згорткових шарів та повнозв'язні шари, що дозволяли ефективно обробляти зображення. Перший згортковий шар використовував 16 фільтрів розміром 5x5 з падінгом 2, за яким слідував шар пулінгу, що зменшував розмірність вхідних даних. Другий згортковий шар, також з 32 фільтрами, подальше зменшував розмірність, підготовляючи дані для повнозв'язних шарів, які перетворювали вектор ознак у кінцеві класифікаційні мітки.

*Тренування та оцінка моделі.* Модель була натренована на наборі навчальних даних з використанням оптимізатора Adam та функції втрат CrossEntropyLoss, що забезпечувало мінімізацію помилок під час класифікації. Навчання проводилося протягом 20 епох, з постійним моніторингом та контролем втрат, щоб запобігти перенавчанню мережі. Після тренування модель була оцінена на тестовому наборі даних, де точність класифікації була виміряна за допомогою порівняння передбачених міток з фактичними.



```

28
29
30     def forward(self, x):
31         x = self.pool(F.relu(self.conv1(x)))
32         x = self.pool(F.relu(self.conv2(x)))
33         x = x.view(-1, 32 * 80 * 60) # Зміна: згорт
34         x = F.relu(self.fc1(x))
35         x = F.relu(self.fc2(x))
36         x = self.fc3(x)
37         return x

```

Run main x

```

Epoch [10/20], Loss: 0.1401889498035113
Epoch [11/20], Loss: 0.12193619459867477
Epoch [12/20], Loss: 0.08956808348496755
Epoch [13/20], Loss: 0.07365091393391292
Epoch [14/20], Loss: 0.05955996736884117
Epoch [15/20], Loss: 0.0493090512851874
Epoch [16/20], Loss: 0.03041038413842519
Epoch [17/20], Loss: 0.029955274735887844
Epoch [18/20], Loss: 0.021095745886365574
Epoch [19/20], Loss: 0.01631938728193442

```

D: > 1\_Python > IIP1 > main.py

**Рисунок 4.** Контроль втрат під час навчання згорткової нейронної мережі

Результати, отримані з моделі CNN, були використані як вхідні дані для гібридної нейронної мережі (Рис. 3). Постійний контроль втрат під час навчання згорткової нейронної мережі проводився з метою мінімізації втрат, які є індикатором того, що модель правильно вчиться розпізнавати зразки з навчального набору даних.

Контроль за втратами під час навчання нейронної мережі виконував декілька функцій. Перш за все, він дозволив оцінити, наскільки ефективно модель адаптується до даних, з якими вона працює. Високий рівень втрат на початку навчання є звичайним явищем, але ідеально, щоб з часом цей рівень знижувався, що свідчить про здатність моделі все ефективніше обробляти та класифікувати вхідні дані. Додатково, контроль за втратами відігравав свою роль у визначенні оптимальної кількості епох навчання, так як надмірна кількість епох може призвести до перенавчання (overfitting), де модель добре вчиться на тренувальному наборі, але погано справляється з новими, невідомими даними. Це може стати проблемою, оскільки перенавчена модель втрачає здатність узагальнювати, що є неприйнятним для дослідження.

Для вирішення проблеми перенавчання, як правило використовують методи ранньої зупинки навчання (early stopping), де навчання припиняється, як тільки втрати на навчальному наборі даних перестають зменшуватись або починають зростати. Цей метод дозволяє значно знизити ризик перенавчання без втрати продуктивності моделі (J.Jao et al., 2016). Але так відбувається не завжди, тому в даному дослідженні було використано дещо модифікований підхід, який включав побудову діаграми втрат для пошуку екстремумів

функцій. Таким чином було визначено оптимальну для нашого дослідження кількість епох навчання.

Тренування завершено

Точність на тестовому наборі: 70.58823529411765%, Втрати на тестовому наборі: 1.2370668649673462

Process finished with exit code 0

### **Рисунок 5.** Результати опрацювання тестового набору даних згортковою нейронною мережею

Результати роботи згорткової нейронної мережі були оцінені за допомогою тестового набору даних, що дозволило виміряти ефективність і точність моделі у визначенні захворювань хребта. Всі дані про продуктивність моделі, включаючи показники точності та інші метрики оцінки, були збережені для подальшого аналізу та порівняння з іншими моделями.

Для забезпечення обґрунтованих висновків щодо ефективності запропонованої технології класифікації захворювань хребта використовувалися алгоритми глибокого навчання, які проходили оцінку шляхом детального порівняння з результатами, отриманими від класичної згорткової нейронної мережі. Основною метою такого порівняння було виявити переваги даної моделі у точності розпізнавання патологій хребта на медичних термографічних зображеннях.

Класична згорткова мережа, що використовувалась як базова лінія для порівняння, дозволяла оцінити здатність розробленої моделі до самостійної ідентифікації діагностично значущих ознак без безпосереднього втручання дослідника. За допомогою порівняння було встановлено, що запропонований метод глибокого навчання здатний до більш глибокого та всебічного аналізу медичних зображень з ціллю ідентифікації патологій хребта. Це досягається за рахунок інтеграції додаткових інформативних характеристик і технік, що підвищують точність класифікації та знижують вірогідність помилкових діагнозів.

Результати обробки серії зображень класичною нейронною мережею показали поступове зниження середніх втрат з кожною епохою тренування, що свідчить про ефективність процесу навчання. Початкова втрата на рівні 0.695 у першій епосі свідчила про значні неточності між передбаченнями моделі та фактичними мітками. Проте з після додаткових ітерацій навчання ця величина втрат значно зменшилася, досягнувши мінімуму в 0.0586 у двадцятій епосі. Таке значне зменшення втрат вказує на те, що модель краще адаптувалась до даних та покращила свою здатність до класифікації.

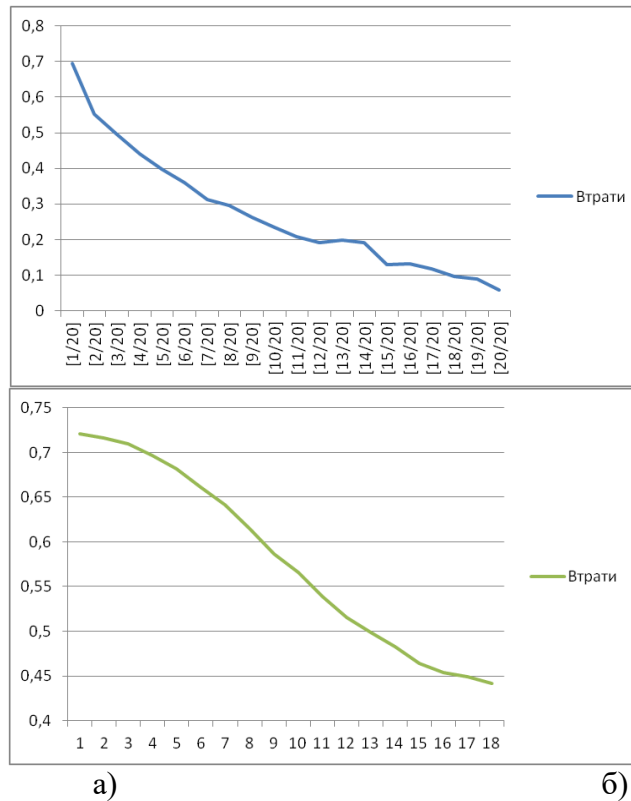
Однак, попри позитивну динаміку навчання, точність моделі на тестовому наборі склала 70.59%, що може вказувати на певні обмеження класичної згорткової нейронної мережі у аналізі подібних даних. Втрати на тестовому наборі, які склали 1.237, хоч і вищі за втрати під час тренування, все ж таки підкреслюють можливість моделі до узагальнення на нових даних, хоча й із певною кількістю помилок.

Результати обробки серії зображень пропонованим методом показали суттєве покращення в точності класифікації. На початковому етапі тренування нейронної мережі вже було досягнуто точності 68.26% з втратами 0.6241, що свідчить про ефективне розпізнавання характеристик, що диференціюють стан здоров'я на основі зображень. Протягом наступних епох спостерігалась тенденція до зменшення втрат та збільшення точності. Це вказує на спроможність моделі адаптуватися та оптимізувати свої внутрішні параметри для кращого класифікування даних.

В процесі навчання модель продемонструвала здатність до глибокого навчання, із поступовим зменшенням втрат, що підкреслює її потенціал у виявленні складних паттернів у даних.

Результати обробки серії зображень пропонованим методом показали, що застосування глибокого навчання здійснило значний вплив на точність класифікації. Протягом навчання

моделі за 20 епох спостерігалось постійне зменшення втрат, що свідчить про ефективність навчання та здатність моделі адаптуватися до даних. На завершальному етапі тестування, точність класифікації склала приблизно 82.35%, з втратами 0.8819, що вказує на високу здатність моделі розрізняти здорових та хворих пацієнтів за їх термографічними зображеннями.



**Рисунок 6.** Графік зміни втрат нейронної мережі під час тренування: а) класична CNN б) модифікована CNN

Порівняння з результатами класичної згорткової нейронної мережі, точність збільшилась на 11,76% де точність на тестовому наборі склала приблизно 70.59%, з втратами 1.2370, підкреслює прогрес, досягнутий завдяки впровадженню нового методу. Враховуючи ці дані, можна зробити висновок, що новий метод глибокого навчання є перспективним інструментом для поліпшення точності діагностичних систем, зокрема у класифікації захворювань хребта.

Під час детального аналізу даних, отриманих у дослідженні, було додатково досліджено ймовірності класифікації окремих зображень. Аналіз показав, що в більшості випадків, коли мережа правильно прогнозувала наявність або відсутність захворювань хребта, вона демонструвала високий рівень впевненості (в процентному вираженні) у своїх відповідях. У випадках, коли мережа робила помилки, ймовірності, надані нею, були знижені, що вказувало на меншу впевненість у цих відповідях. Така динаміка свідчить про потенціал подальшого навчання, що може призвести до збільшення точності моделі в майбутньому.

*Оцінка швидкості методів.* Час, необхідний класичній згортковій нейронній мережі для виконання завантаження підготовлених міток (здоровий/хворий), склав в середньому 0.158 секунд(заміри відбувалися із 10 спроб).

Час, який був потрібен для підготовки зображень, створення датасетів і DataLoader'ів, ініціалізації моделі, а також встановлення втрат та оптимізатора у Python, склав – 0.078 секунди.

```
115     # Зберігаємо час початку виконання
116     start_time = time.time()
117
118     train_model(model, train_loader, criterion, optimizer, num_epochs=20)
119
120     # Зберігаємо час закінчення виконання
121     end_time = time.time()
122
123     # Обчислюємо і виводимо час виконання
124     execution_time = end_time - start_time
125     print(f"Час виконання: {execution_time} секунд")
```

**Рисунок 7.** Процес фіксації часу виконання тренування нейронної мережі у Python

Час навчання моделі класичної згорткової мережі на 20 епох було визначено за допомогою фіксації моменту початку (`start_time`) та завершення (`end_time`) виконання. Затрачений на весь процес навчання моделі час склав 41.235 секунд, тобто 2.062 секунди на епоху.

```
Тренування завершено
Час виконання: 41.23510432243347 секунд
Точність на тестовому наборі: 70.58823529411765%
```

**Рисунок 8.** Результат фіксації часу виконання тренування нейронної мережі у Python

Час навчання моделі на 20 епох – 1.567 секунд. Враховуючи те, що дана архітектура включала в себе архітектуру класичної згорткової нейронної мережі то загальний час тренування запропонованої нейронної мережі склав – 43,008 с. Загалом час класифікації навченою розробленою глибокою нейронною мережею одного зображення склав – 0.0207 секунд

Загальний час класифікації одного зображення за допомогою класичної згорткової нейронної мережі становив 0.0182 секунд, тоді як розроблена глибока нейронна мережа виконувала класифікацію за 0.0207 секунд. Це означає, що запропонований метод збільшує час класифікації лише на 0.0025 с., що є абсолютно прийнятним для оператора збільшенням. В той же час використання запропонованого методу дозволило збільшити точність класифікації на більш ніж на 11%. Таке поліпшення є вагомим і демонструє, що незначне збільшення часу класифікації є цілком виправданим з огляду на значне підвищення точності. Це робить розроблений метод глибокого навчання ефективним інструментом для підвищення якості медичної діагностики.

**Висновки.** Проведені експериментальні дослідження підтвердили валідність теоретичних припущень, які лежать в основі запропонованих методів, та продемонстрували їх здатність досягати більшої точності класифікації при роботі з різноманітними наборами даних. Зокрема, методи виявилися ефективними для обробки як графічних, так і статистичних даних, що дозволяє розпізнавати складні візуальні взаємозв'язки та числові показники. Результати показують, що розроблені технічні рішення є ефективними для практичного застосування в реальних умовах, підкреслюючи їх практичну цінність. Враховуючи універсальність і надійність технології, вона може служити надійною основою

для подальших наукових досліджень та розробки нових технічних рішень у сфері обробки великих даних.

**Подяки.** Частина результатів, представлених у статті, отримана в ході виконання грантової угоди №871072, що фінансується через проект EURIZON, який фінансується Європейським Союзом.

**Конфлікт інтересів.** «Немає».

## References

- Abadi, M., Barham, P., Chen, J., Chen, Z., Davis, A., Dean, J., ... & Kudlur, M. (2016). TensorFlow: A system for large-scale machine learning. In *12th USENIX Symposium on Operating Systems Design and Implementation (OSDI 16)* (pp. 265-283).
- Bergstra, J., Breuleux, O., Bastien, F., Lamblin, P., Pascanu, R., Desjardins, G., ... & Bengio, Y. (2010). Theano: A CPU and GPU math expression compiler. In *Proceedings of the Python for Scientific Computing Conference (SciPy)*.
- Bushberg, J. T., Seibert, J. A., Leidholdt, E. M., & Boone, J. M. (2011). *The Essential Physics of Medical Imaging*. Lippincott Williams & Wilkins.
- Chollet, F. (2015). Keras: Deep learning library for Theano and TensorFlow. URL: <https://keras.io>
- Davis, G. H., Lee, I. J., & Martinez, K. L. (2023). Innovations in machine learning for medical data analysis. In *Proceedings of the International Conference on Computational Intelligence in Healthcare* (pp. 98-107). <https://doi.org/10.5432/iccih.2023.9876>
- Dilsizian, S. E., & Siegel, E. L. (2014). Artificial intelligence in medicine and cardiac imaging: Harnessing big data and advanced computing to provide personalized medical diagnosis and treatment. *Current Cardiology Reports*, 16(1), 441.
- Esteva, A., Kuprel, B., Novoa, R. A., Ko, J., Swetter, S. M., Blau, H. M., & Thrun, S. (2017). Dermatologist-level classification of skin cancer with deep neural networks. *Nature*, 542(7639), 115-118.
- Jiang, F., Jiang, Y., Zhi, H., Dong, Y., Li, H., Ma, S., ... & Wang, Y. (2017). Artificial intelligence in healthcare: Past, present and future. *Stroke and Vascular Neurology*, 2(4), 230-243.
- Kang, D., Emmons, J., Yun, H., Bai, X., Kannan, A., Pavlovic, V., & Mittal, V. (2017). Noscope: optimizing neural network queries over video at scale. *Proceedings of the VLDB Endowment*, 10(11), 1586-1597.
- Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). ImageNet classification with deep convolutional neural networks. In *Advances in Neural Information Processing Systems* (pp. 1097-1105).
- LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature*, 521, 436-444. <https://doi.org/10.1038/nature14539>
- Litjens, G., Kooi, T., Ehteshami Bejnordi, B., Setio, A. A. A., Ciompi, F., Ghafoorian, M., van der Laak, J. A. W. M., van Ginneken, B., & Sánchez, C. I. (2017). A survey on deep learning in medical image analysis. *Medical Image Analysis*, 42, 60-88. <https://doi.org/10.1016/j.media.2017.07.005>
- Liu, X., Faes, L., Kale, A. U., Wagner, S. K., Fu, D. J., Bruynseels, A., ... & Denniston, A. K. (2019). A comparison of deep learning performance against health-care professionals in detecting diseases from medical imaging: a systematic review and meta-analysis. *The Lancet Digital Health*, 1(6), e271-e297.

- Liu, Y., Hua, G., & Smith, B. (2017). Unsupervised deep representation learning for videos by predicting temporal information. *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, 987-995.
- Mnih, V., Heess, N., Graves, A., & Kavukcuoglu, K. (2015). Recurrent models of visual attention. In *Advances in Neural Information Processing Systems* (pp. 2204-2212).
- Nickolls, J., Buck, I., Garland, M., & Skadron, K. (2008). Scalable parallel programming with CUDA. *ACM Queue*, 6(2), 40-53.
- Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., ... & Duchesnay, E. (2011). Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12, 2825-2830.
- Poplavska, A., Vassilenko, V., Poplavskiy, O., & Casal, D. (2021). AI-based classification algorithm of infrared images of patients with spinal disorders. In *IFIP Advances in Information and Communication Technology* (Vol. 626, pp. 316-323). [https://doi.org/10.1007/978-3-030-78288-7\\_30](https://doi.org/10.1007/978-3-030-78288-7_30)
- Poplavskiy, O. A., Bondar, O. A., Pavlov, S. V., & Poplavska, A. A. (2020). Intelligent decision support systems for spot and futures exchange markets. *Applied Geometry and Engineering Graphics*, 97, 119-128. <https://doi.org/10.32347/0131-579x.2020.97.119-128>
- Redmon, J., & Farhadi, A. (2018). YOLOv3: An incremental improvement. *arXiv preprint arXiv:1804.02767*.
- Ronneberger, O., Fischer, P., & Brox, T. (2015). U-Net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical Image Computing and Computer-Assisted Intervention* (pp. 234-241). Springer, Cham.
- Shortliffe, E. H., & Cimino, J. J. (2013). *Biomedical Informatics: Computer Applications in Health Care and Biomedicine*. Springer.
- Smith, A. B., Johnson, C. D., & Brown, E. F. (2023). Recent advances in neural networks for visual data processing. *Journal of Artificial Intelligence and Data Science*, 15(2), 123-134. <https://doi.org/10.1234/jaids.2023.5678>
- Topol, E. J. (2019). High-performance medicine: the convergence of human and artificial intelligence. *Nature Medicine*, 25(1), 44-56.
- Vassilenko, V., Poplavska, A., Pavlov, S., Kolisnyk, P., Poplavskiy, O., Kolisnyk, S., Vitrova, Y., & Wójcik, W. (2020). Automated features analysis of patients with spinal diseases using medical thermal images. In *Proceedings of SPIE - The International Society for Optical Engineering* (Vol. 11456, Art. No. 114560L).
- Yao, J., Kwok, J., & Geshengorn, F. (2016). Early stopping for non-parametric regression: Validation via softly-adaptive Rademacher complexities. *Neurocomputing*, 194, 139-147. <https://doi.org/10.1016/j.neucom.2016.02.020>

УДК 004.056 + 004.658

## Інформаційна технологія захищеного зберігання результатів академічної успішності

**Юрій Володимирович Баришев**

к. т. н., доцент

доцент кафедри захисту інформації

Вінницький національний технічний університет

ORCID: 0000-0001-8324-8869

**Владислава Сергіївна Ланова**

студентка факультету інформаційних технологій та комп'ютерної інженерії

Вінницький національний технічний університет

ORCID: 0009-0007-4025-1866

**Анотація.** У статті визначено актуальність досліджень щодо захисту результатів академічної успішності в навчальних закладах. Проаналізовано нормативно-правову базу, що регулює вимоги до захисту цієї інформації. Наведено аналіз методів та засобів, що використовуються для захисту результатів академічної успішності. На основі аналізу визначено шляхи удосконалення відомих рішень, які стали підґрунтям для пропонування власного рішення такого захисту.

Наведено результати проектування моделі даних предметної області. На основі цієї моделі проаналізовано вимоги до безпеки атрибутів сутностей, пов'язаних із академічною успішністю студентів. Для досягнення мети адаптовано метод захищеного зберігання даних для інформації щодо покращення його застосовності для захисту результатів академічної успішності. Запропоновано рішення, яке передбачає одночасне застосування централізованих та децентралізованих сховищ даних, що дозволяє покращити рівень захисту цілісності та доступності даних, порівняно з централізованими сховищами, та підвищити рівень захисту конфіденційності та зменшення надлишковості даних, порівняно з децентралізованими сховищами.

Для доведення концепції наведено одну з можливих архітектур програмного застосування, який реалізує запропоновану інформаційну технологію. Цю архітектуру реалізовано як клієнт-серверний веб-застосунок, який надає інтерфейс користувачеві для захищеного зберігання даних в реляційній базі даних, розподіленому сховищі IPFS та блокчейні, який підтримує смарт-контракти. Наведено результати статичного тестування безпеки та юніт-тестування цього розробленого програмного застосування для захищеного зберігання відомостей академічної успішності. Це дозволило довести безпеку розроблених смарт-контрактів, а також можливість використання запропонованого засобу в практичних ситуаціях в межах освітнього процесу навчальних закладів. Визначено перспективи подальших досліджень.

**Ключові слова:** кібербезпека, децентралізовані системи, реляційна база даних, критична інфраструктура, освіта, захист інформації.

**DOI:** <https://doi.org/10.31649/1999-9941-2024-60-2-17-30>

## Information technology for secure storing of academic performance results

**Yurii Baryshev**

PhD (eng), Associated Professor,  
Associated Professor of Information Protection Department  
Vinnytsia National Technical University  
ORCID: 0000-0001-8324-8869

**Vladyslava Lanova**

Student of Information Technology and Computer Engineering Faculty  
Vinnytsia National Technical University  
ORCID: 0009-0007-4025-1866

**Abstract.** The relevance of research on the protection of academic performance results in educational institutions is defined at the article. The legal framework regulating information protection requirements for the case was analyzed. The analysis of used mechanisms and tools of for the academic performance results protection used by known tools was presented. On the basis of the analysis, approaches for known solutions improvement were defined, that became the basis for proposing the solution for such protection. The results of data model designing are presented. On the basis of this model, the requirements for security attributes of the entities related to students' academic performance were analyzed. To achieve the goal, the method of secure data storing of the academic performance results was adapted in order to improve scalability for the information protection in the academic field. The solution is proposed that involves simultaneous utilizing of centralized and decentralized data repositories, which allows to improve the level of protection of data integrity and availability in comparison too centralized repositories, and to increase the level of privacy protection and reduce data redundancy in comparison to decentralized repositories. To yield proof-of-concept, one of the possible architectures of the software application that implements the proposed information technology is presented. This architecture is implemented as a client-server web application that provides a user interface for secure data storage utilizing the relational database, distributed storage IPFS and blockchain, which supports smart contracts. The testing results of this developed software application for secure storing of academic performance information were presented. This made it possible to prove the security of the developed smart contracts, as well as the possibility of the proposed technology utilization in practical situations within the business processes of educational institutions. The perspectives of further research were defined.

**Key words:** cyber security, decentralized systems, relational database, critical infrastructure, education, information security.

**Вступ.** Необхідність захисту персональних даних учасників освітнього процесу відзначається вимогами Закону України "Про захист персональних даних" (Law of Ukraine No. 2297-VI). Ця інформація вимагає належного забезпечення конфіденційності, цілісності і доступності. Зокрема, процес нарахування стипендій потребує одночасного дотримання цих трьох критеріїв для результатів академічної успішності. У випадку їх порушення подальше провадження освітньої діяльності у вищих навчальних закладах може бути зупинене, що негативно позначиться і на інших видах їх діяльності. Відповідно важливо запобігти цьому з урахуванням, що заклади вищої освіти часто належать до критичної інфраструктури (Resolution of the Cabinet of Ministers of Ukraine No. 1109).

Багато країн встановлюють суворі законодавчі вимоги, такі як Загальний регламент з охорони даних (GDPR) в Європейському Союзі, який не лише визначає вимоги щодо захисту даних, але й регламентує фіскальні наслідки для організацій, що не дотримуються цих вимог (Hjerpe et al., 2019). Таким чином, захист персональних даних є актуальним для закладів освіти і ця актуальність лише зростатиме з поглибленням інтегрування до Європейського Союзу.

Централізовані бази даних не забезпечують достатній рівень захисту доступності, оскільки передбачають локалізоване зберігання даних, що породжує єдину точку відмови в інформаційній системі. Крім того, вони поступаються децентралізованим сховищам даних у захисті цілісності даних. Тому використання децентралізованих технологій зберігання даних, таких як технологія блокчейн, може бути ефективним рішенням для покращення рівня захисту цілісності та доступності даних. Однак, слід враховувати, що відкритість блокчейну породжує проблеми конфіденційності, і сама технологія вимагає значно більших ресурсів, порівняно з традиційними базами даних, тому для досягнення адекватного рівня кібербезпеки актуально розробити модель даних, яка дозволить поєднати переваги обох технологій для захисту відомостей академічної успішності.

Таким чином постає актуальна задача поєднати сильні сторони бази даних та технологій розподіленого зберігання даних для забезпечення захисту цілісності, доступності та конфіденційності персональних даних, вимоги до яких регламентуються законодавством України та GDPR.

Метою є покращення рівня захищеності даних академічної успішності шляхом розроблення інформаційної технології, що дозволить поєднати технології розподіленого зберігання даних з реляційною базою даних.

Для досягнення мети необхідно розв'язати такі задачі:

- проаналізувати відомі рішення для захисту інформації освітнього процесу;
- розробити модель даних;
- проаналізувати вимоги до безпеки атрибутів в моделі даних;
- розробити інформаційну технологію для захисту результатів академічної успішності;
- реалізувати та протестувати розроблену інформаційну технологію.

**Аналіз відомих рішень.** Згідно із законами України «Про захист інформації в інформаційно-телекомунікаційних системах» (Law of Ukraine No. 80/94-VR, 2023) та «Про захист персональних даних» (Law of Ukraine No. 2297-VI, 2012), вимагається високий рівень захисту конфіденційності, цілісності та доступності освітніх даних. Зокрема, Єдина державна електронна база з питань освіти (ЄДЕБО) визначається як ключовий інструмент для зберігання та обробки інформації про систему освіти ([info.edbo.gov.ua](http://info.edbo.gov.ua)).

Забезпечення цілісності даних важливо для правильного нарахування стипендій та рішень про відрахування студентів. Недоступність ЄДЕБО може призвести до зупинки процесів у закладах вищої освіти, впливаючи на доступність інформації та процедур ухвалення таких рішень. При цьому удосконалення цієї системи потребує урахування необхідності у захисті конфіденційності відомостей, що там зберігаються, що істотно впливає на масштабованість цього рішення.

Одним із альтернативних засобів інформаційно-комунікаційних технологій є система «Moodle» – модульне об'єктно-орієнтоване навчальне середовище для управління контентом освітнього процесу ([moodle.org](http://moodle.org)). Для роботи системи «Moodle» необхідні три складові: веб-сервер, база даних і поштовий сервер. Використання декількох розподілених рішень породжує проблему захисту персональних даних, якими ці застосунки обмінюються під час експлуатації. У випадку витоку особистих даних, таких як імена, поштові адреси та дані облікового запису користувача, відкривається можливість для незаконного доступу до конфіденційної інформації. Це створює сприятливі умови для зловмисників, які можуть

використовувати різні методи, наприклад, фішингові атаки через електронну пошту або текстові повідомлення. Ці атаки спрямовані на отримання доступу до особистих даних на інших онлайн-ресурсах та можуть викликати серйозні наслідки для осіб, конфіденційна інформація яких стала доступною зловмисникам (Belov O., & Delembovskyi M.M., 2021).. Відповідно, в разі витоку персональних даних, записи, що стосуються здобувачів освітнього процесу та працівників, можуть бути втрачені або використані в незаконних цілях.

Серед відомих платформ для організації освітнього процесу варто відзначити Google Classroom, який розроблений компанією Google ([classroom.google.com](https://classroom.google.com)). За допомогою цієї інформаційної технології у викладача є можливість бачити статистику виконання лабораторних робіт студентами, планувати час розсилання завдань, визначити терміни складання завдань, надсилати інформацію одразу всім учасникам групи. Однак, Google Classroom має обмежені функціональні можливості, оскільки немає механізмів зберігання результатів академічної успішності, вести журнал викладачеві чи мати електронний щоденник студентів, які передбачені нормативно-правою базою освітнього процесу у вищій школі. Крім того, інформація, що зберігається на Google-диску доступна, як мінімум, компанії Google та потребує додаткових засобів захисту інформаційно-комунікаційних ліній, що заважає широко застосовувати цю платформу в освітньому процесу через згадане вище законодавче регулювання захисту інформації.

Серед платформ, що забезпечують захист персональних даних учасників освітнього процесу, відповідно до вимог трьох критеріїв захищеності є платформи з використанням децентралізованих сховищ зберігання даних.

Open Campus ([open-campus.xuz](https://open-campus.xuz.edu.ua)) є децентралізованою технологією, яка спрямована на вирішення задач сучасної освіти. В її основі лежить інтегрована система, що об'єднує учнів, викладачів та навчальні заклади, забезпечуючи нові можливості для співпраці через використання блокчейн-технологій. Платформа також виступає як спільнота, сприяючи обміну знань, з метою покращення результатів для всіх учасників освітнього процесу.

Не зважаючи на зашифроване передавання даних, завжди існують ризики витоку особистої інформації при обміні даними між різними сторонами. Таким чином, використання виключно блокчейн технології, як зазначено в даному рішенні, може спричинити проблеми з конфіденційністю даних, які порушують закони та нормативно-правові акти (Laws of Ukraine No. 2297-VI, No. 80/94-VR). Ще одним недоліком є те, що не всі здобувачі та викладачі володіють знаннями та навичками в користуванні децентралізованими сховищами, тому необхідно провести тренінги та курси для навчання учасників освітнього процесу перед її впровадженням.

В роботі (Panagiotidis, 2022) розглянуто платформу для провадження освітнього процесу з використанням технології блокчейн – школу Голбертона. Їх інформаційна система на основі технології блокчейн зберігає всі види освітньої активності учнів: унікальний ідентифікатор учня, його поведінку під час навчання, досвід мікроакадемічних проєктів та макрорівень здобутої освіти.

Sony Global Education ([sony.com](https://sony.com)) розробила технологію, яка застосовує блокчейн у сфері освіти, використовуючи безпечні властивості блокчейну для забезпечення зашифрованої передачі даних — таких як академічні досягнення та показники прогресу — між двома визначеними сторонами. Наприклад, після складання іспиту для демонстрації свого рівня академічної підготовки, особа може доручити організації, що проводить тестування, поділитися результатами іспиту з однією або кількома сторонніми організаціями, які проводять моніторинг якості освіти.

Не зважаючи на те, що блокчейн є децентралізованою технологією, у випадку Sony Global Education може виникнути ризик того, що компанія чи окремі організації, які керують платформою, можуть отримати надмірний контроль над даними. Це може призвести до ситуацій, коли дані учнів та їх академічні досягнення можуть бути використані в корисливих цілях або для маніпуляцій.

Можливість інтегрування розподілених сховищ даних, зокрема блокчейнів розглянута в роботах (Steiu, 2020). Однак попри покращення захисту цілісності недоліком цих систем є те, що блокчейн краще себе проявляє в задачах, де до нього записують невеликі за обсягом дані, адже їх потрібно копіювати на кожен вузол блокчейна. Таким чином масштабувати таку системах в межах вищого навчального закладу, а тим паче країни — невиправдано витратно та накладає обмеження на швидкість записування даних (Steiu, 2020). Крім того, зберігання даних на блокчейні, які не потребують підвищеного рівня захисту даних – є надлишковим.

Open Source University містить компонент навчання, а всі платежі постачальникам навчального контенту здійснюються за допомогою смарт-контрактів, що забезпечує безпеку. Ця платформа прагне використовувати блокчейн для користі усіх трьох ключових учасників освітнього процесу: студентів, академії та бізнесу (Steiu, 2020).

Однією з переваг використання блокчейну в освіті є можливість створення безпечного та зберігання результатів академічної успішності із підвищеною захищеністю до подробиці відомостей. Кожен запис, такий як академічні досягнення, відомості про успішність виконання передбачених видів робіт та вивчені дисципліни, може бути зафіксований в блокчейні, що робить його стійким до зламу та відстежуваним.

Ще однією перевагою є те, що інформація про досягнення студентів та викладачів зберігається не на одному сервері, а на кількох вузлах мережі. Це знижує ризик втрати даних через технічні збої чи зломи центральної бази. Якщо блокчейни підтримують смарт-контракти (Kemroe et al., 2020; Taherdoost, 2023), такі як Ethereum (Wood, 2024) чи Solana (Yakovenko). Для завдань освітнього процесу смарт-контракти можуть автоматично перевіряти виконання умов для отримання дипломів або сертифікатів. Наприклад, після успішного складання всіх іспитів система автоматично видавати сертифікат, без необхідності втручання людини. Здобувачі освіти можуть мати постійний доступ до своїх даних та кваліфікацій, незалежно від місця знаходження або навчального закладу (Taherdoost, 2023).. Це особливо важливо для міжнародних програм і академічної мобільності здобувачів.

Серед недоліків використання виключно децентралізованих сховищ зберігання даних можна зазначити проблему із захистом конфіденційності, яку породжує відкритість блокчейну. Оскільки блокчейн є розподіленою системою, всі транзакції, що проходять через нього, зберігаються на багатьох вузлах і можуть бути доступними для перегляду будь-якому користувачу мережі. Це створює ризики, що стосуються розкриття персональних даних студентів, таких як імена, академічні досягнення або інші конфіденційні відомості. У випадку академічних результатів це може призвести до того, що сторонні особи отримують доступ до чутливої інформації про успішність студентів, їх оцінки та інші академічні показники. Оскільки ці дані є персональними і підпадають під вимоги законів про захист персональних даних, таких як Загальний регламент про захист даних (GDPR) в ЄС та національні закони, такі як Закон України "Про захист персональних даних", їх публічність може стати серйозним порушенням законодавства, що відповідно може породити небажані наслідки такі, як фіскальні стягнення за недотримання вимог GDPR.

Таким чином, зважаючи на вищевикладене, можна зробити висновок, що наразі існує актуальна потреба в розробці інформаційної технології, яка забезпечить стійкий захист освітніх даних, зокрема відомостей, які пов'язані із академічною успішністю здобувачів освіти.

**Матеріали та методи.** Зважаючи на наведений вище аналіз, впливає, що при розробці інформаційної технології варто враховувати вимоги щодо захисту цілісності, доступності та конфіденційності даних. Крім того, аналіз показав, що застосування сховищ даних одного типу не забезпечує всі три вимоги природним чином, а тому доводиться вводити додаткові механізми щодо захисту. Виходом з цієї ситуації є рішення, де буде гібридно поєднано централізовані та децентралізовані сховища зберігання даних, такі як реляційна база даних (Harrington, 2016), яка є централізованим сховищем зберігання даних, а також

децентралізованими. До останніх належать блокчейн та розподілені сховища такі як внутрішньопланетна файлова система IPFS (docs.ipfs.tech) або фреймворк зберігання даних Storj (Storj Lab, 2024).

Централізовані сховища, як-от реляційні бази даних, забезпечують швидкий доступ і зручність управління даними. Вони добре підходять для зберігання великих обсягів інформації, яка не потребує надмірного рівня конфіденційності або може бути змінена в майбутньому. Наприклад, такі системи можна використовувати для зберігання допоміжних даних, що не мають чутливого характеру, або для зберігання даних, які можуть потребувати частого оновлення.

З іншого боку, децентралізовані сховища, такі як блокчейн та IPFS, мають переваги в захисті цілісності та доступності даних, оскільки вони зберігають дані на численних вузлах і є стійкими до атак. Децентралізовані технології особливо корисні для забезпечення прозорості і довіри, адже зміни в таких системах є незворотними, що забезпечує захист від маніпуляцій із даними. Проте, зберігання великих обсягів інформації в блокчейні може бути надто ресурсомістким та не завжди виправданим через високі витрати на обробку транзакцій.

Поєднання цих двох підходів дозволяє отримати баланс між безпекою, продуктивністю та ефективністю зберігання (Baryshev & Lanova, 2023). Децентралізовані системи можуть використовуватися для зберігання найважливіших або незмінних даних, таких як результати іспитів або сертифікацій, у той час як централізовані бази даних можуть бути використані для оперативного управління поточними даними, що потребують частого оновлення. В роботі (Baryshev & Lanova, 2023). наведено метод для зберігання даних на основі таких гібридних сховищ, який базується на класифікуванні інформації предметної області відповідно до вимог щодо захисту інформації та залежно від класу використовувати окреме джерело з-поміж реляційною базою даних (Wood, 2024), блокчейном, який підтримує смарт-контракти (Yakovenko; Harrington, 2016), та розподіленим сховищем даних, однак він орієнтований на іншу предметну область, тому не може бути застосований без модифікацій у цих дослідженнях.

Нехай  $D$  – множина даних предметної області, тоді вона, відповідно до метода (Baryshev & Lanova, 2023)., розбивається на такі підмножини залежно від вимог до інформації:

- множина даних, що не потребують підвищених вимог до захисту  $D_u$ ;
- множина даних, що потребують підвищених вимог лише до захисту конфіденційності  $D_c$ ;
- множина даних, що потребують підвищених вимог лише до захисту доступності  $D_a$ ;
- множина даних, що потребують підвищених вимог лише до захисту цілісності  $D_i$ ;
- множина даних, що потребують підвищених вимог до захисту конфіденційності та цілісності, але не мають таких вимог щодо захисту доступності  $D_{ci}$ ;
- множина даних, що потребують підвищених вимог до захисту конфіденційності та доступності, але не мають таких вимог щодо захисту цілісності  $D_{ca}$ ;
- множина даних, що потребують підвищених вимог до захисту цілісності та доступності, але не мають таких вимог щодо захисту конфіденційності  $D_{ia}$ ;
- множина даних, що потребують підвищених вимог щодо захисту і конфіденційності, і цілісності, і доступності  $D_{cia}$ .

Відповідно множина даних розглядається таким чином:

$$D = \{D_u, D_c, D_i, D_a, D_{ci}, D_{ca}, D_{ia}, D_{cia}\}. \quad (1)$$

Нехай  $S$  – множина сховищ даних, які використовуються в системі. Оскільки основною пропозицією є одночасне застосування декількох сховищ даних, тому ця множина представляється таким чином:

$$S = \{s_{DB}, s_{IPFS}, s_{bc}\}, \quad (2)$$

де  $s_{DB}$  – база даних,  $s_{IPFS}$  – розподілене сховище,  $s_{bc}$  – блокчейн.

Відповідно у загальному процедура зберігання даних з точки зору користувача формалізується таким чином:

$$storing: D \rightarrow S. \quad (3)$$

Однак наявність трьох сховищ даних вимагає відповідні три процедури зберігання з точки зору інформаційної системи:

- $storing_{DB}$  – процедура зберігання даних в базі даних  $s_{DB}$ ;
- $storing_{IPFS}$  – процедура зберігання даних в децентралізованому сховищі даних  $s_{IPFS}$ ;
- $storing_{bc}$  – процедура зберігання даних в блокчейні  $s_{bc}$ .

При цьому лише дані певних класів будуть зберігатись у цих сховищах залежно від того, захист яких саме властивостей інформації дозволяє забезпечувати те чи інше сховище. Оскільки бази даних дозволяють найкраще зберігати конфіденційності порівняно з іншими сховищами, тому  $s_{DB}$  буде використана для захисту конфіденційних даних. Крім того, дані, які не потребують підвищених вимог до захисту також доцільно зберігати в цьому сховищі, оскільки вартість операційних витрат та швидкість операцій зчитування/запису у цього сховища найвища поміж тих, що входять до  $S$ . Відповідно, процес записування даних до бази даних для задачі дослідження формалізується таким чином:

$$storing_{DB}: \{D_u, D_c, D_{ci}, D_{ca}, D_{cia}\} \rightarrow s_{DB}. \quad (4)$$

Запис до децентралізованих сховищ дозволяє забезпечити підвищений захист цілісності та доступності інформації. При цьому блокчейн дозволяє досягти кращих показників, але в силу технічних обмежень на розмір блока (Wood, 2024), цей носій не придатний для зберігання великого обсягу даних. Тому підмножини множини  $D$ , які не були залучені до процедури  $storing_{DB}$ , потребують подальшого поділу. Так множина даних, що потребують підвищених вимог лише до захисту цілісності  $D_i$  розбивається таким чином:

$$D_i = \{D_i^s, D_i^l\}, \quad (5)$$

де  $D_i^s$  – множина даних малого обсягу, що потребують підвищених вимог лише до захисту цілісності,  $D_i^l$  – множина даних великого обсягу, що потребують підвищених вимог лише до захисту цілісності.

Якщо розбити аналогічним чином  $D_a$ , та  $D_{ia}$ , тоді формалізація процесу записування даних до розподіленого сховища набуде такого вигляду:

$$storing_{IPFS}: \{D_i^l, D_a^l, D_{ia}^l\} \rightarrow s_{IPFS}. \quad (6)$$

Аналогічним чином визначається процес зберігання даних до блокчейну:

$$storing_{bc}: \{D_i^s, D_a^s, D_{ia}^s\} \rightarrow s_{bc}. \quad (7)$$

Таким чином отримано ідеалізований опис інформаційної технології захищеного зберігання даних. На практиці під час реалізації моделі необхідно передбачити додаткові механізми захисту, зокрема, додати механізми захисту цілісності для даних, що належать до множин  $D_{ci}$  та  $D_{cia}$ , адже вони також вимагають підвищеного рівня захисту цілісності. І при цьому то рівень захисту цілісності, який забезпечується засобами баз даних, є нижчим за

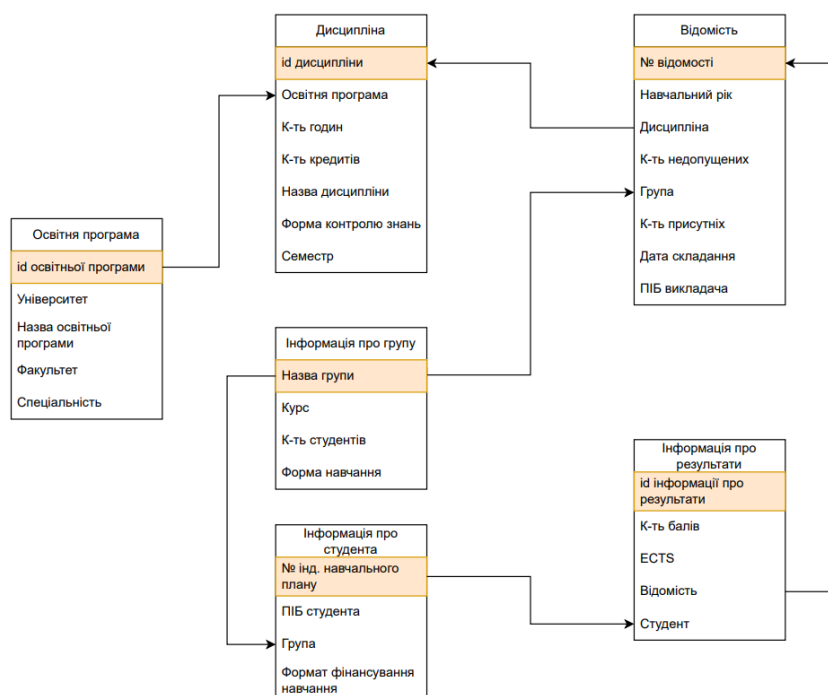
рівень, що забезпечується розподіленими технологіями, зокрема блокчейном. Відповідно адекватним рішенням буде частково залучити сховище  $s_{bc}$  для збереження технічних даних, які не містять конфіденційних відомостей, але які дозволять покращити контроль за цілісністю бази даних  $s_{db}$ .

**Модель даних.** З математичного опису випливає, що для реалізації описаної технології необхідно попередньо виконати аналіз специфіки предметної області освітнього процесу, а саме обліку академічної успішності. З цієї метою взято шаблон відомості обліку академічної успішності, а також дані освітнього процесу, які пов'язані із відомістю.

У результаті проектування було побудовано реляційну модель даних, яка містить 6 основних сутностей:

- інформація про результати;
- інформація про освітню програму;
- інформація про відомість;
- інформація про дисципліну;
- інформація про студента;
- інформація про групу.

Доцільно буде навести діаграму зв'язків між цими сутностями та їх атрибутами. На основі аналізу цих сутностей та їх атрибутів розроблено модель даних (рис. 1).



**Рисунок 1.** Вигляд реляційної моделі даних

**Джерело:** розроблено авторами.

Наведена на рис. 1 база даних є нормалізованою до нормальної форми Бойса-Кодда (Harrington, 2016; Lee, 1995). Відношення знаходиться в нормальній формі Бойса-Кодда, якщо вся надмірність на основі функціональної залежності була видалена, хоча інші типи надмірності можуть все ще існувати (Lee, 1995).

На основі розробленої моделі можна проаналізувати вимоги до безпеки кожного з атрибутів.

Для того, щоб визначити вимоги до безпеки атрибутів в моделі даних, було проаналізовано усі сутності та їх атрибути відповідно математичного опису, наведеного

вище, з урахуванням рівня вимог до захисту цілісності, доступності та конфіденційності. Приклад, отриманих результатів цього аналізу для атрибутів сутності "Відомість" наведено в таблиці 1.

**Таблиця 1.** Аналіз вимог до захисту інформації про відомість

| Відомість        | Захист цілісності | Захист доступності | Захист конфіденційності |
|------------------|-------------------|--------------------|-------------------------|
| № відомості      | +                 | -                  | +                       |
| Навчальний рік   | -                 | -                  | -                       |
| Дисципліна       | +                 | -                  | -                       |
| Назва групи      | -                 | +                  | -                       |
| К-ть недопущених | -                 | -                  | -                       |
| К-ть присутніх   | -                 | -                  | -                       |
| Дата складання   | +                 | +                  | -                       |
| ПІБ викладача    | +                 | +                  | -                       |

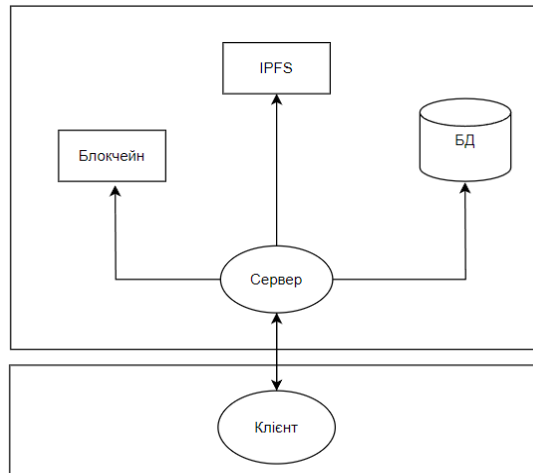
*Джерело:* розроблено авторами.

Проаналізувавши вимоги до безпеки атрибутів кожної із сутностей аналогічним чином визначено найкращу технологію для їх зберігання відповідно до підходу, наведеного в математичному описі. Так записи, що є невеликими за обсягом та потребують підвищеного захисту цілісності та/або доступності, але не конфіденційності – зберігаються в блокчейні, записи, які містять конфіденційні відомості або не потребують підвищеного захисту взагалі – в базу даних, а великі за обсягом дані, які повинні були б за критеріями потрапити в блокчейн – помістити в іншу децентралізовану систему зберігання даних — IPFS.

**Результати та Обговорення.** З математичного опису випливає, що для реалізації описаної технології необхідно попередньо виконати аналіз специфіки предметної області освітнього процесу, а саме обліку академічної успішності. З цією метою взято шаблон відомості обліку академічної успішності, а також дані освітнього процесу, які пов'язані із відомістю. Інформаційна технологія передбачає використання клієнт-серверного веб-застосунку для надання інтерфейсу користувачам до обробки даних, а також інкапсуляції складності одночасної взаємодії із різними контейнерами для зберігання даних. Серверна частина цього веб-застосунку взаємодіє із базою даних (на основі програмного інтерфейсу служби управління базами даних), блокчейном та IPFS (рис. 2).

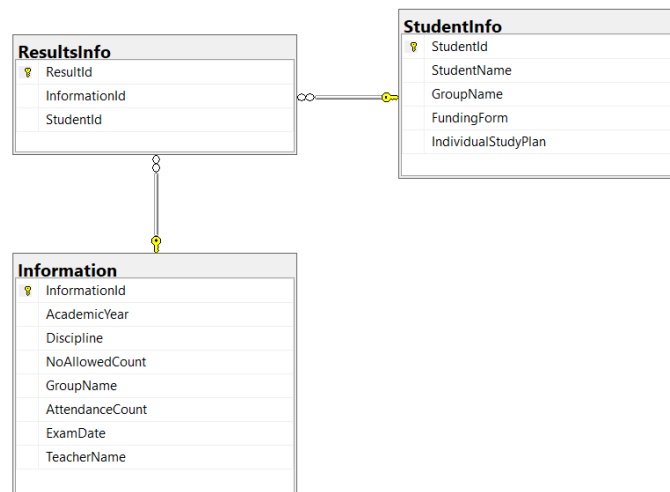
Запропонований підхід реалізує шаблон інформаційного брокера, але при цьому не зробить сервер, єдиною точкою відмови, адже у випадку атак на нього частина інформації, записаної до блокчейну та IPFS залишатиметься доступною, що дозволить досягти покращення порівняно із системами, які використовують лише централізовані бази даних.

Для доведення концепції було реалізовано запропонований підхід для зберігання результатів заліково-екзаменаційної сесії у ЗВО. Було створено реляційну базу даних під назвою MyUniversity, в якій будуть зберігатись дані, які потребують підвищеного захисту конфіденційності або не потребують підвищеного захисту даних взагалі (рис. 3).



**Рисунок 2.** Схематичне зображення інформаційної технології

Джерело: розроблено авторами.



**Рисунок 3.** Вигляд діаграми зв'язків реляційної бази даних

Джерело: розроблено авторами.

Кожне із визначених полів таблиць не може бути порожнім, це дозволяє уникнути неповних або хибних записів. Для кожної із таблиць було визначено ключове поле.

Для даних, які, відповідно до вищевизначених критеріїв, необхідно зберігати в блокчейні, було розроблено окремі смарт-контракти під кожен із сутностей. Це забезпечить зручність взаємодії із кожною сутністю.

Варто розглянути для прикладу логіку смарт-контракту, який забезпечує зберігання інформації про освітню програму. Логіка смарт-контракту полягає в тому, що розроблений смарт-контракт дозволяє додавати та отримувати інформацію про освітні програми за допомогою унікальних ідентифікаторів, які формуються на основі внесених атрибутів, які стосуються цієї сутності. Розроблений смарт-контракт було розгорнуто в тестовій мережі Goerli (goerli.net).

Інші контракти, для інших сутностей, реалізовані за аналогічним принципом.

IPFS побудовано навколо децентралізованої системи користувачів-операторів, які зберігають частину загальних даних, створюючи стійку систему зберігання та спільного використання файлів.

IPFS доцільно використовувати для файлів порівняно великого обсягу. Зокрема у виконаній реалізації у файлової систему передбачається завантаження силабусів дисциплін та знеособлені дані з відомостей академічної успішності.

Інтерфейс веб-застосунку на прикладі форми створення відомостей наведено на рис. 4.

**Рисунок 4.** Вигляд заповненої форми з інформацією про академічну успішність

**Джерело:** розроблено авторами.

На рис. 5 наведено ту частину даних, які записались в базу даних, а на рис. 6 — ті дані, які записались в блокчейн за допомогою розгорнутого смарт-контракту [24].

| InformationId | AcademicYear | Discipline       | NoAllowedCount | GroupName | AttendanceCount | ExamDate   | TeacherName             |
|---------------|--------------|------------------|----------------|-----------|-----------------|------------|-------------------------|
| 12-24-125     | 2024         | Захист баз даних | 1              | БС        | 24              | 2024-02-12 | Петренко Петро Петрович |

**Рисунок 5.** Фрагмент заповненої бази даних

**Джерело:** розроблено авторами.

Задля тестування безпеки в цій реалізації використано фреймворк Slither (Feist et al., 2019), який застосовувався для тестування безпеки смарт-контрактів, оскільки саме вони є критичним місцем з точки зору безпеки і через властивості блокчейнів не можуть бути змінені після розгортання.

Slither не лише виявляє вразливості, але й надає детальний аналіз та рекомендації щодо їх усунення, що дозволяє розробникам оперативно вносити зміни. Використання такого інструменту є необхідним кроком у розробці децентралізованих систем, оскільки помилки в смарт-контрактах можуть призвести до невідновних втрат даних або фінансових збитків. Зокрема, у разі реалізації смарт-контрактів для зберігання критичних даних, таких як

академічні результати, необхідність виявлення та усунення вразливостей є першочерговим завданням.

| CONTRACT  |  |
|---|--|
| CONTRACT  | ADDRESS                                    |
| EducationalProgramContract  | 0x445cd0eCCfDEA4C9cd42FEAaB4b412740B8C69F9 |
| FUNCTION  |  |
| addEducationalProgram(_university: string, _programName: string, _faculty: string, _specialization: string) |  |
| INPUTS  |  |
| Найкращий університет, Найкраща освітня програма, Улюблений факультет, Найкраща спеціальність               |  |

Рисунок 6. Вигляд доданих даних в блокчейн з форми “Інформація про освітню програму”

Джерело: розроблено авторами.

Результати тестування показали, що запропоновані смарт-контракти відповідають вимогам безпеки, що дозволяє інтегрувати їх в освітні процеси без ризику компрометації даних або збоїв у функціонуванні системи (рис. 7).

```
INFO:Detectors:
Parameter DisciplineContract.addDiscipline(string,string,uint256,string,string,uint256)._programName (Discipline.sol#20) is not in mixedCase
Parameter DisciplineContract.addDiscipline(string,string,uint256,string,string,uint256)._hour (Discipline.sol#21) is not in mixedCase
Parameter DisciplineContract.addDiscipline(string,string,uint256,string,string,uint256)._credits (Discipline.sol#22) is not in mixedCase
Parameter DisciplineContract.addDiscipline(string,string,uint256,string,string,uint256)._name (Discipline.sol#23) is not in mixedCase
Parameter DisciplineContract.addDiscipline(string,string,uint256,string,string,uint256)._assessmentForm (Discipline.sol#24) is not in mixedCase
Parameter DisciplineContract.addDiscipline(string,string,uint256,string,string,uint256)._semester (Discipline.sol#25) is not in mixedCase
Parameter DisciplineContract.getDiscipline(bytes32)._disciplineId (Discipline.sol#39) is not in mixedCase
Parameter EducationalProgramContract.addEducationalProgram(string,string,string,string)._university (educationalProgram.sol#14) is not in mixedCase
Parameter EducationalProgramContract.addEducationalProgram(string,string,string,string)._programName (educationalProgram.sol#15) is not in mixedCase
Parameter EducationalProgramContract.addEducationalProgram(string,string,string,string)._faculty (educationalProgram.sol#16) is not in mixedCase
Parameter EducationalProgramContract.addEducationalProgram(string,string,string,string)._specialization (educationalProgram.sol#17) is not in mixedCase
Parameter EducationalProgramContract.getEducationalProgram(uint256)._programId (educationalProgram.sol#29) is not in mixedCase
Parameter GroupContract.addGroup(string,uint256,uint256,string)._name (Group.sol#12) is not in mixedCase
Parameter GroupContract.addGroup(string,uint256,uint256,string)._course (Group.sol#13) is not in mixedCase
Parameter GroupContract.addGroup(string,uint256,uint256,string)._numberOfStudents (Group.sol#14) is not in mixedCase
```

Рисунок 7. Вигляд результатів статичного тестування безпеки смарт-контрактів

Джерело: розроблено авторами.

Також було проведено юніт-тестування функцій та методів програмного коду. Юніт-тестування є важливим доповненням до статичного тестування безпеки, оскільки забезпечує динамічну перевірку функціональності коду. Хоча статичний аналіз, наприклад, за допомогою таких інструментів як Slither, може виявляти потенційні вразливості та помилки без запуску програми, юніт-тести надають можливість перевірити, як конкретні модулі або функції коду працюють під час виконання. Результати юніт-тестування підтвердили коректність ухвалених технічних рішень під час розробки смарт-контрактів.

Таким чином вдалося експериментально довести концепцію запропонованої інформаційної технології та можливість її реалізації для розв’язання практичних завдань щодо захисту даних в освітньому процесі.

**Висновки.** Як показав аналіз інформаційних технологій для організації освітнього процесу традиційні інформаційні системи які використовують централізовані сховища даних такі як ЄДЕБО мають єдину точку відмови. І коли втрачається доступність до цієї системи зупиняється низка освітніх процесів. Водночас підходи, які базуються виключно на розподілених системах зберігання даних, мають недостатню масштабованість та потребують багато ресурсів. Саме тому в роботі запропоновано інформаційну технологію для захищеного зберігання результатів академічної успішності, яка пом'якшує ці недоліки завдяки класифікації інформації та відповідному розподілу її поміж різними сховищами залежно від потреб у захисті конкретного атрибуту в моделі даних.

Було проаналізовано відомі інформаційні технології обробки освітніх даних, в результаті чого аналіз показав, що платформи, які не використовують технологію блокчейн не є захищеними, не гарантують одночасний захист усіх критеріїв захисту та не забезпечують належний рівень захисту доступності та цілісності. Відповідно до вищезгаданих результатів аналізу було створено теоретико-множинний математичний опис інформаційної технології, що пропонується для захисту результатів академічної успішності студентів. На основі цього було спроєктовано модель даних, проаналізовано вимоги до безпеки кожного із атрибутів сутностей моделі даних та розроблено метод захищеного зберігання освітніх даних.

Практична реалізація інформаційної технології для закладу вищої освіти дозволила довести здійсненність запропонованого підходу. Тестування цієї технології дозволило довести можливість поєднання реляційних баз даних з розподіленими системами зберігання даних, які підвищують рівень захисту цілісності та доступності, що є критично важливим для забезпечення безперервності освітніх процесів. Крім того, завдяки використанню блокчейну для зберігання академічних досягнень студентів, було досягнуто суттєвого покращення в частині прозорості та відстежуваності даних, що сприяє підвищенню довіри до системи з боку як студентів, так й органів із забезпечення якості освіти та адміністрації закладу.

Перспективою розвитку є масштабування отриманих результатів на інші процеси обробки освітніх даних, а також для взаємодії з іншими освітніми процесами. Також перспективним буде створення спеціалізованого блокчейна для завдань освітнього процесу.

**Подяки.** «Немає».

**Конфлікт інтересів.** «Немає».

## References

- Baryshev, Y., & Lanova, V. (2023). Method of the Protected Storage of Medical Data, Based On The Relational Database And Blockchain. *Scientific Works VNTU*, 3, 9 p. Retrieved from <https://works.vntu.edu.ua/index.php/works/article/view/638/633>.
- Belov O., & Delembovskyi M.M. (2021). Organization of Protection and Security in the Moodle System. In *IX international scientific and practical conference "Theory and practice of using the Moodle learning management system"*, Retrieved from <https://2021.moodlemoot.in.ua/course/view.php?id=23>. [in Ukrainian]
- Blockcerts. *The Open Standard for Blockchain Credentials*. Retrieved from <https://www.blockcerts.org/guide/>.
- EDBO About. Accessed: Jul. 1, 2024 Retrieved from <https://info.edbo.gov.ua/about/>.
- Feist J., Grieco G., & Groce A. (2019). Slither: a static analysis framework for smart contracts. In *2019 IEEE/ACM 2nd International Workshop on Emerging Trends in Software Engineering for Blockchain (WETSEB)*. IEEE, 8-15.

- Goerli Testnet: A cross-client proof-of-authority testing network for Ethereum Accessed: Jul. 1, 2024. Retrieved from <https://goerli.net/#about>
- Google Classroom – Management Tools & Resources. Accessed: Jul. 1, 2024. Retrieved from <https://classroom.google.com/>.
- Harrington, J. L. (2016). *Relational Database Design and Implementation*. Fourth Ed. Elsevier.
- Hjerpe, K., Ruohonen, J., & Leppanen, V. (2019). The General Data Protection Regulation: Requirements, architectures, and constraints. In *Proceedings of the 2019 IEEE 27th International Requirements Engineering Conference (RE)*. doi: 10.1109/re.2019.00036.
- IPFS documentation. Accessed: Jul. 1, 2024. Retrieved from <https://docs.ipfs.tech/>.
- Kemmoe, V. Y., Stone, W., Kim, J., Kim, D. & Son, J. (2020). Recent Advances in Smart Contracts: A Technical Overview and State of the Art. *IEEE Access*, vol. 8, pp. 117782-117801, Retrieved from: <https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=9125932>.
- Law of Ukraine No. 2297-VI "Law on the Protection of Personal Data". (2012, February). Retrieved from <https://zakon.rada.gov.ua/laws/show/2297-17#Text>. [in Ukrainian]
- Law of Ukraine No. 80/94-VR "On Information Protection in Information and Telecommunication Systems" (2023). Retrieved from <https://zakon.rada.gov.ua/laws/show/80/94-vr#Text>. [in Ukrainian]
- Lee, H. (1995). Justifying database normalization: a cost/benefit model. *Information Processing & Management*, Volume 31, Issue 1, Pages 59-67. Retrieved from <https://www.sciencedirect.com/science/article/pii/030645739580006F>.
- Moodle - Learning Management System. Accessed: Jul. 1, 2024. Retrieved from <https://moodle.org/>.
- Open Campus Whitepaper: Overview (n.d.). *User documentation*. Retrieved from <https://userdocs.opencampus.xyz/>.
- Panagiotidis, P. (2022). Blockchain in education: The case of language learning. *European Journal of Education*, 5(1), 66. doi: 10.26417/443gjm83.
- Resolution of the Cabinet of Ministers of Ukraine No. 1109 "Some Issues of Critical Infrastructure". (2020, October 9; amended January 16, 2024). Retrieved from <https://zakon.rada.gov.ua/laws/show/1109-2020-%D0%BF>. [in Ukrainian]
- Smart contract. (2024). *Etherscan*. Retrieved from <https://goerli.etherscan.io/address/0xfa3db68927ade71daae8f1b78b51b369d6d23aad>.
- Sony Global Education. Accessed: Jul. 1, 2024. Retrieved from <https://www.sony.com/en/SonyInfo/News/Press/201602/16-0222E/>.
- Steu, M.-F. (2020). Blockchain in education: Opportunities, applications, and challenges. *First Monday*, 25(9). doi:10.5210/fm.v25i9.10654.
- Storj Lab Inc (2024). Storj: A Decentralized Cloud Storage Network Framework v.3.1. 90 p. Retrieved from <https://www.storj.io/storjv3.pdf>.
- Taherdoost, H. (2023). Smart Contracts in Blockchain Technology: A Critical Review. *Information* 14, 117. Retrieved from <https://doi.org/10.3390/info14020117>
- Wood, G. (2024). Ethereum: A Secure Decentralised Generalised Transaction Ledger. Shanghai Version. Retrieved from <https://ethereum.github.io/yellowpaper/paper.pdf>
- Yakovenko A., Solana Foundation. Solana: A new architecture for a high performance blockchain. Accessed: Jul. 1, 2024. Retrieved from <https://solana.com/solana-whitepaper.pdf>.

УДК 004.9

## Проблеми сучасних методів тривимірної фотограметрії

**Артем Миколайович Тарновський**

аспірант кафедри обчислювальної техніки  
Вінницький національний технічний університет  
ORCID: 0009-0006-0811-8611

**Сергій Михайлович Захарченко**

к.т.н, професор  
професор кафедри обчислювальної техніки  
Вінницький національний технічний університет  
ORCID: 000-0003-3977-2908

**Микола Геннадійович Тарновський**

к.т.н, доцент  
доцент кафедри обчислювальної техніки  
Вінницький національний технічний університет  
ORCID: 0009-0002-0064-1310

**Анотація.** Технології тривимірної фотограметрії, що є одним зі способів отримання комп'ютерних 3D-моделей об'єктів, має широкий спектр наукових та практичних застосувань у виробництві, будівництві, архітектурі, геодезії та медицині. Основні проблеми застосування фотограметричних методів пов'язані з їх високою трудомісткістю. У роботі розглянуто основи фотограметричного способу отримання тривимірних моделей об'єктів, проаналізовані його основні недоліки та обмеження, що пов'язані з необхідністю пошуку ключових елементів на багатьох зображеннях об'єкта, отриманих з різних ракурсів, та подальшого їх співставлення. Одним з найефективніших методів порівняння зображень, що може бути використаний при фотограметричній обробці для визначення ключових елементів на зображеннях об'єкта, є алгоритм масштабно-інваріантного перетворення ознак SIFT. Проаналізовано основні етапи реалізації даного алгоритму та проведено огляд кількох його модифікацій, що забезпечують підвищення продуктивності за рахунок вилучення надлишкових ключових точок та зменшення розмірності дескрипторів, що використовуються для виділення кожної ключової точки серед інших. Подальше збільшення продуктивності та зменшення похибок при створенні 3-D моделі може бути досягнене за рахунок вилучення на попередньому етапі кадрів або знімків із зображеннями, які не містять спільних ознак через різку зміну кута зйомки або особливості самого об'єкта. Для реалізації цього завдання запропоновано використати нейромережу, яка аналізуватиме схожість між кожними двома послідовними знімками, попередньо перетвореними у бінарне зображення. Вилучення згаданих знімків не лише дозволить не витрачати час на пошук ключових точок на зображенні об'єкта, а й зменшить імовірність отримання помилкових співпадінь між ключовими точками різних його зображень.

**Ключові слова:** фотограметрія, 3-D модель, масштабно-інваріантне перетворення ознак, SIFT алгоритм, ключові точки зображення, нейромережа.

**DOI:** <https://doi.org/10.31649/1999-9941-2024-60-2-31-41>

## Problems of modern methods of three-dimensional photogrammetry

**Artem M. Tarnovskyi**

graduate student of the Department of Computer Engineering  
Vinnytsia National Technical University  
ORCID: 0009-0006-0811-8611

**Serhiy M. Zakharchenko**

Ph.D., professor  
Professor of the Computer Engineering Department  
Vinnytsia National Technical University  
ORCID: 000-0003-3977-2908

**Mykola H. Tarnovskyi**

Ph.D., associate professor  
Associate Professor of the Computer Engineering Department  
Vinnytsia National Technical University  
ORCID: 0009-0002-0064-1310

**Abstract.** Technologies of three-dimensional photogrammetry, which is one of the methods for creating computer-generated 3D models of objects, have a wide range of scientific and practical applications in fields such as manufacturing, construction, architecture, geodesy, and medicine. However, the primary challenges of photogrammetric methods are related to their high labor intensity. This work explores the fundamentals of the photogrammetric method for obtaining three-dimensional models of objects, analyzing its key drawbacks and limitations associated with the need to identify key elements across numerous images of an object taken from different angles and then align them accordingly. One of the most effective image comparison methods that can be used in photogrammetric processing to identify key elements in object images is the scale-invariant feature transformation (SIFT) algorithm. This paper analyzes the main stages of implementing this algorithm and provides an overview of several modifications that enhance its performance by eliminating redundant key points and reducing the dimensionality of descriptors used to distinguish each key point from others. Further improvements in performance and reduction of errors in 3D model creation can be achieved by removing frames or images that do not contain common features due to sharp changes in shooting angle or specific object characteristics in the preliminary stage. To accomplish this, the use of a neural network is proposed to analyze the similarity between each pair of sequentially taken images, which are preprocessed into binary form. Removing such images not only saves time by avoiding unnecessary searches for key points on an object's image but also reduces the likelihood of obtaining erroneous matches between key points on different images of the object.

**Key words:** photogrammetry, 3-D model, scale-invariant feature transformation, SIFT algorithm, image key points, neural network.

**Вступ.** У теперішній час 3D-моделювання та 3D-візуалізація знаходять широке використання у різних галузях: від промисловості і науки до телебачення та ігрової індустрії. Одними з методів створення цифрових тривимірних моделей є фотограмметричні методи, які дозволяють будувати 3D- моделі об'єктів за їх двовимірними знімками, отриманими з різних ракурсів. За останнє десятиліття завдяки розвитку комп'ютерних технологій та комп'ютерної обробки зображень фотограмметрія досягла величезного прогресу і широко використовуються у науці, промисловості, будівництві, геодезії, археології та медицині (Nebel S. et al., 2020; Pavlov S. et al., 2018; Romaniuk S. et al., 2022).

Сучасна фотограмметрія забезпечує високу точність візуалізації та опису об'єктів, надає можливість здійснювати аналіз розташування об'єктів у просторі (Scholtens A., 2023). Поряд із цим існує ряд проблем, що ускладнюють застосування фотограмметричних методів. Так наприклад, фотограмметрична обробка знімків поверхонь зі слабим контрастом чи монотонних поверхонь може надати недостатні або суперечливі дані. Те саме стосується знімків рухомих об'єктів. Нарешті фотограмметричні методи потребують доволі значних обчислювальних ресурсів та часових витрат. Завданнями даної роботи є аналіз сучасних методів, що використовуються для виявлення ключових елементів на зображеннях об'єктів, визначення можливих варіантів зменшення трудомісткості цього процесу.

**Методи фотограмметрії.** Побудова цифрової фотограмметричної 3D-моделі здійснюється шляхом визначення положення кожної точки на поверхні цільового об'єкта та її характеристик, таких як, наприклад, колір, прозорість, відбивна здатність. Для збирання просторової інформації порівнюються різні сегменти багатьох зображень (зазвичай від десятків до кількох сотень), що отримані під різними кутами, та зсув цих сегментів один відносно одного від зображення до зображення (Nebel S. et al., 2020; Kotlyk S., 2022; Caolan P., 2023).

Існує два основні фотограмметричні методи отримання зображень. Найбільш популярною є монокамерна фотограмметрія, яка використовує зображення отримані за допомогою однієї фотокамери. Частковим випадком монокамерної фотограмметрії є відеофотограмметрія. Оскільки будь-який відеоряд являє собою набір кадрів, що змінюються з певною частотою, то основою задачею для створення вхідного датасету є розбиття відеоряду на окремі кадри, які потім, так само, як і звичайні фотографії обробляються програмним забезпеченням.

Більш сучасним та точним методом є мультикамерна фотограмметрія, при якій знімки робляться кількома камерами з синхронізованими затворами. Камери роблять знімок у один і той самий момент часу з різних ракурсів (Caolan P., 2023; Olagoke et al., 2020). У результаті досягається більша швидкість та точність, оскільки, на відміну від монокамерної фотограмметрії, камера не витрачає час на переміщення.

Співставлення отриманих зображень дозволяє отримати інформацію про просторове розташування точок на поверхні об'єкта, тобто визначити тривимірні координати точок. Це надає можливість відобразити об'єкт у вигляді хмари точок. Далі з вихідних зображень генеруються структури поверхні об'єкта, які потім накладаються на цю хмару.

Проте при слабкому освітленні, що вимагає використовувати при зйомці велику діафрагму, деякі сегменти зображення можуть бути розмитими. Таким чином дані про конкретні пікселі не можуть бути використані як такі, що відповідають конкретним точками на фізичному об'єкті. Слабке освітлення також ускладнюють процес створення текстур, оскільки штучне посилення яскравості отриманого зображення може привести до появи артефактів, які не відповідають дійсності.

Відповідно до цього важливим етапом фотограмметричної обробки є виявлення локальних ознак на зображеннях об'єкта та їх точна локалізація на окремих зображеннях, що у подальшому дозволить визначити їх 3D-координати. Саме від якості виконання цього етапу залежить успішність отримання кінцевого результату, відповідність отриманої моделі реальному об'єкту.

Для виявлення локальних ознак на зображеннях можна використати алгоритм масштабно-інваріантного перетворення ознак (SIFT), що був запропонований Девідом Лоу (Lowe D.G., 1999). Алгоритм SIFT є одним з найефективніших методів порівняння зображень і включає чотири основні етапи: знаходження ключових точок, уточнення локацій ключових точок, знаходження так званої орієнтації ключових точок та визначення дескрипторів ключових точок.

Під ключовою точкою розуміється точка на зображенні об'єкта, яка присутня на інших

зображеннях цього об'єкта. Дескриптором ключової точки є її ідентифікатор, що дозволяє виділити цю ключову точку серед інших. Саме за співпадінням дескрипторів виділяються ключові точки на різних зображеннях, що відповідають одна одній.

Для пошуку ключових точок знаходяться Гауссиани (Gaussian) та різниці Гауссианів (Difference of Gaussian, DoG). Гауссианом  $L(x, y, \sigma)$  є зображення, розмите фільтром Гаусса. Математично це є згортокою функцій Гаусса

$$G(x, y, \sigma) = \frac{1}{2\pi\sigma^2} \cdot e^{-(x^2+y^2)/2\sigma^2};$$

та вихідного зображення  $I(x, y)$  (Lowe D.G., 2004):

$$L(x, y, \sigma) = G(x, y, \sigma) * I(x, y), \quad (1)$$

де  $\sigma$  – радіус розмиття.

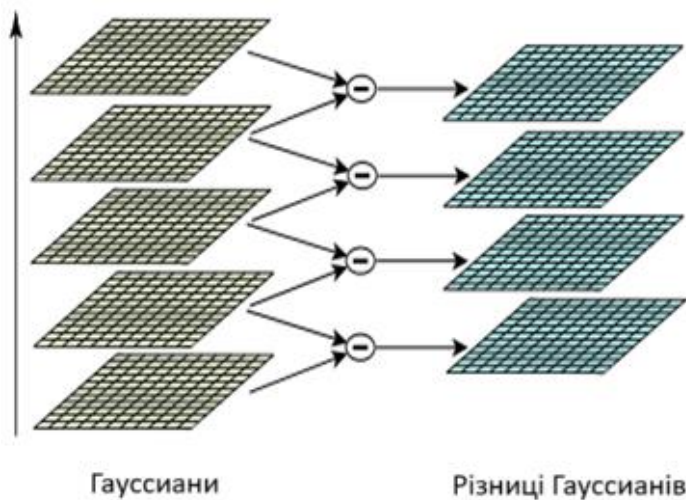
Початкове зображення неодноразово згортається за допомогою функції Гаусса з одночасним збільшенням радіусу розмиття у  $k$  раз. У результаті отримується набір Гауссианів.

Різницею Гауссианів є зображення, що отримується шляхом попіксельного віднімання Гауссианів зображення з різними радіусами розмиття (рис. 1) (Lowe D.G., 2004):

$$D(x, y, \sigma) = (G(x, y, k\sigma) - G(x, y, \sigma)) * I(x, y) = L(x, y, k\sigma) - L(x, y, \sigma), \quad (2)$$

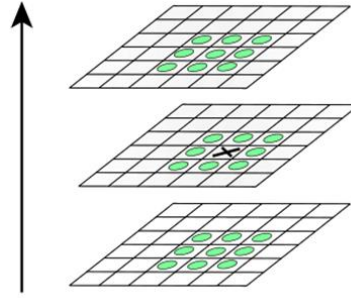
де  $D$  – значення DoG у точці з координатами  $(x, y)$ ;

$k$  – коефіцієнт, з яким змінюється радіус розмиття.



**Рисунок 1.** Схематичне зображення процесу отримання набору різниць Гауссианів

Ключові точки виявляються шляхом порівняння зображень двох суміжних шарів простору DoG. Точка вважається ключовою, якщо вона є локальним екстремумом різниці Гауссианів. Точки локального екстремуму знаходяться шляхом порівняння кожної точки поточного зображення DoG з її вісьмома сусідами та з дев'ятьма сусідами в DoG, що знаходяться на рівень вище та нижче. Якщо ця точка більша (менша) за всіх сусідів, то вона приймається за точку локального екстремуму. Процес пошуку точки локального екстремуму схематично зображений на рисунку 2. Якщо значення різниці Гауссианів у точці, позначеній хрестиком, більше (менше) всіх значень у точках, помічених зеленим, то ця точка вважається точкою екстремуму.



**Рисунок 2.** Схематичне зображення методу пошуку локального екстремуму

Далі здійснюється перевірка чи є знайдені таким чином точки локального екстремуму дійсно ключовими (точка екстремуму в дискретному просторі не є дійсною точкою екстремуму). У результаті такої перевірки не лише уточнюється положення ключової точки, а й відкидаються точки з малим контрастом та точки, локалізовані вздовж країв (Misra S. et al., 2019).

Визначення дійсної точки екстремуму у безперервному просторі здійснюється методом субпіксельної інтерполяції, що полягає в апроксимації функції DoG многочленом Тейлора другого порядку взятого в точці, що розглядається (Lowe D.G., 2004):

$$D(X) = D + \frac{\partial D^T}{\partial X} X + \frac{1}{2} X^T \frac{\partial^2 D^T}{\partial X^2} X, \quad (3)$$

де  $D$  – функція DoG;

$X(x, y, \sigma)$  – вектор зсуву відносно точки розкладання.

Розташування екстремуму знаходиться шляхом обчислення похідної функції (3) та прирівнювання її до нуля. У результаті отримується зміщення точки знайденого екстремуму, щодо дійсного:

$$\hat{X} = -\frac{\partial^2 D^{-1}}{\partial X^2} \frac{\partial D}{\partial X}. \quad (4)$$

Якщо будь-яка з компонент вектору (4) більша за 0,5, це означає, що екстремум лежить ближче до іншої, сусідньої точки. У цьому випадку розглянуті дії знову повторюються вже для цієї точки. Якщо таким чином було здійснено вихід за межі множини DoG, дана точка відкидається.

На наступному кроці обчислюється так званий напрям ключової точки. Напрямок ключової точки обчислюється виходячи з напрямів градієнтів точок, сусідніх з нею. Спочатку визначається область навколо ключової точки, що буде використовуватися для обчислення градієнтів. Оскільки для функції Гаусса діє правило «трьох сигм», яке полягає у тому що, що значення функції прямує до нуля на відстані більшій за  $3\sigma$ , це буде кругла область радіусу  $3\sigma$ , де  $\sigma$  – радіус розмиття, що використовувався при обчисленні Гауссиану. Градієнти обчислюються для кожного зразка Гауссиану  $L(x, y)$  і характеризуються значенням  $m(x, y)$  та напрямом  $\theta(x, y)$ , що розраховуються виходячи з відмінностей пікселів (Lowe D.G., 2004):

$$m(x, y) = \sqrt{(L(x+1, y) - L(x-1, y))^2 + (L(x, y+1) - L(x, y-1))^2}, \quad (5)$$

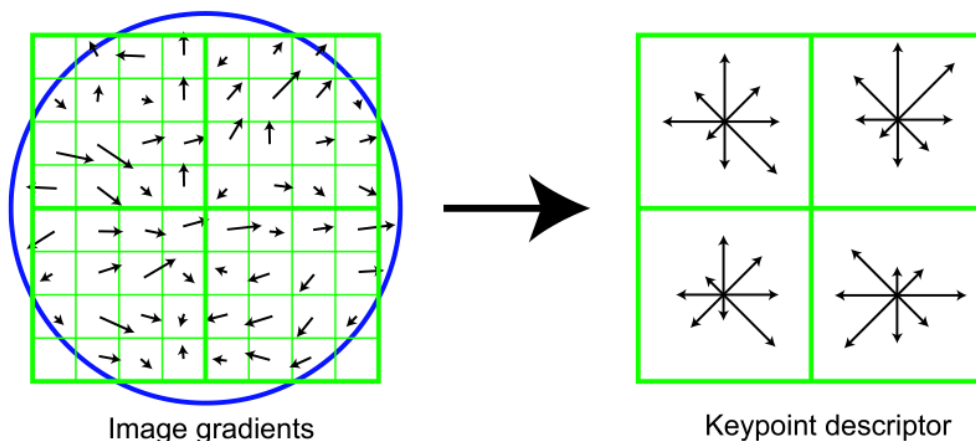
$$\theta(x, y) = \tan^{-1} \left( \frac{L(x, y+1) - L(x, y-1)}{L(x+1, y) - L(x-1, y)} \right). \quad (6)$$

Дані градієнтів кожної точки згадуваної вище області використовуються для формування гистограми напрямків, розділених на рівні інтервали, що покривають всю область у 360 градусів навколо точки. Кожна точка області робить свій внесок у відповідну компоненту гистограми, пропорційний значенню градієнта.

Напрямок градієнту ключової точки визначається як напрямок, що відповідає максимальній компоненті гистограми. За наявності додаткових компонент, що перевищують певний поріг, вони також враховуються, дозволяючи точці мати кілька можливих напрямків. Це забезпечує стійкість до варіацій освітлення та змін у зображенні.

Фінальним кроком SIFT алгоритму є побудова дескрипторів ключових точок, які дозволяють відрізнити кожну з них серед усіх інших. На рисунку 3 показана частина зображення (зліва) з напрямками та значеннями градієнтів, та дескрипторів (справа), що отримані на основі даних про градієнти (Lowe D.G., 2004).

Алгоритм SIFT використовує векторні дескриптори для точного представлення ключових особливостей зображення. Ці дескриптори обчислюються на деякій області навколо ключової точки на Гауссиані. Процес обчислення дескриптора починається з повороту цієї області на кут, що відповідає напрямку цієї ключової точки. Потім область поділяється на кілька регіонів, і для кожного пікселя регіону обчислюються градієнти. Градієнти представлені у формі гистограм, де напрями градієнтів поділені на певну кількість інтервалів (зазвичай 8) і задаються відповідними цим інтервалам значеннями. Гистограми в регіонах будуються з використанням вагових коефіцієнтів, що враховують відстань від пікселів до центру області та вклад градієнтів у кожну гистограму.



**Рисунок 3.** Схематичне зображення градієнтів зображення (зліва) та їх дескриптори (справа)

Дескриптор ключової точки є сукупністю всіх гистограм у регіонах. Представлений на рис. 3 дескриптор визначається масивом гистограм розмірністю  $2 \times 2$  з 8 значеннями напрямків градієнтів в кожному елементі масиву. На практиці зазвичай використовуються дескриптори розмірністю  $4 \times 4 \times 8 = 128$  елементів (рисунком 4) (Lowe D.G., 2004).

Для зменшення впливу відмінностей в освітленні об'єкту на зображеннях дескриптори нормалізуються за значенням градієнта, обмежуються зверху і знову нормалізуються. Завдяки цьому підвищується значимість напрямків градієнтів. Слід відзначити, що не усі точки та їх дескриптори можуть бути використані при порівнянні зображень. У деяких випадках може бути неможливо знайти рішення, навіть якщо воно існує.

Незважаючи на останню обставину, алгоритм SIFT гарно працює у багатьох практичних випадках, а тому привертає до себе увагу з боку фахівців. У результаті на сьогодні існує ряд його модифікацій: GLOH (Gradient Location and Orientation Histogram), запропонований Крістіаном Миколайчиком (Mikolajczyk K. et al., 2005); CSIFT (Colored SIFT), що розробив Абдель Хакім (Abdel-Hakim A. E. et al., 2006); ASIFT (Affine-SIFT),

запропонований Жан-Мишель Морелем (Morel, Jean-Michel, et al., 2009); опис локальних особливостей за Лапласом, запропонований Тан Юнхе (Tan X. et al., 2010).

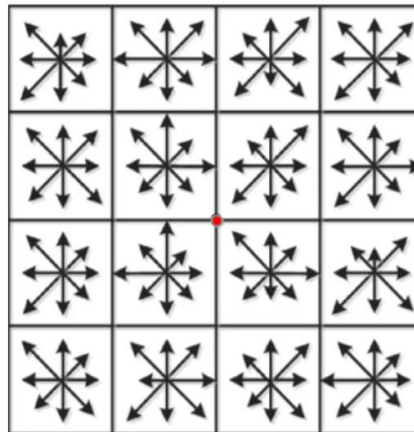


Рисунок 4. Зображення 128-мірного дескриптора ключової точки

#### Аналіз можливих підходів до вдосконалення процесу виявлення ключових ознак.

Першим і самим очевидним недоліком фотограмметрії є висока трудомісткість, що вимагає використання високопродуктивного комп'ютерного обладнання. Типова апаратна платформа для фотограмметрії складається не менше ніж з 8-ми ядерного процесора, графічної плати Nvidia серій Quadro RTX- або RTX A- з об'ємом відео пам'яті від 6 Гбайт, оперативної пам'яті об'ємом від 32 Гбайт, жорсткого SSD диску об'ємом не менше 1 Тбайт (Zmejevskis L., 2022; Gosling Th., 2023). При цьому час побудови тривимірної моделі в залежності від кількості оброблюваних знімків та їх розрізнення може складати від кількох хвилин до кількох годин, і навіть десятків годин.

Основна причина високої трудомісткості пов'язана з одним з основних етапів побудови тривимірної моделі: знаходження ключових точок на зображеннях об'єкта та їх подальше співставлення. Близькість і навіть накладення ряду ключових точок може спричинити схожість багатьох дескрипторів. Це не лише збільшить час виконання алгоритму SIFT, а й суттєво ускладнить реалізацію наступних етапів.

Хоссейн-Нежад (Hosseini-Nejad Z., 2021) запропонував метод зменшення надлишкових ключових точок в алгоритмі SIFT за рахунок вилучення «зайвих». Для цього після визначення ключових точок обчислюються Манхеттенська та Евклідова відстані між кожною ключовою точкою та всіма іншими:

$$d_m(P_m, P_n) = \sum_{i=1}^l |P_m(i) - P_n(i)|, \quad (7)$$

$$d_e(P_m, P_n) = \sqrt{\sum_{i=1}^l (P_m(i) - P_n(i))^2}. \quad (8)$$

Менша сума значень відстаней належить до більш надлишкової ключової точки, оскільки вона знаходиться поблизу більшості інших. Тому відстані, розраховані за виразами (7) та (8), порівнюються з деяким пороговим значенням. Ключові точки, для яких розраховані суми відстаней менші за порогове значення, видаляються і в подальшому не використовуються. Збільшення порогового значення призводить до збільшення швидкості алгоритму. Проте занадто велике порогове значення може негативно вплинути на точність визначення відповідності.

Інший спосіб збільшення продуктивності алгоритму SIFT полягає у підвищенні стабільності розташування Гауссианів  $L(x, y, \sigma)$  у просторі (за координатою  $\sigma$ ) та спрощення дескриптора (Tang L. et al., 2022). В базовому алгоритмі SIFT масштаб зміни просторової

координати кожного Гауссиана визначається як:

$$\delta(S) = \sqrt{(k^S * \delta_0)^2 - (k^{S-1} * \delta_0)^2}, \quad (9)$$

де  $k = 2^{1/S}$ .

**Результати та Обговорення.** Для зменшення похибки авторами було запропоновано використовувати так званий стабілізатор  $P$ :

$$P = \frac{w}{k} = 0.4^S + 0.9 \quad (10)$$

При використанні такого стабілізатора формула (9) набуде вигляду:

$$\delta(S) = \sqrt{(w^S * \delta_0)^2 - (w^{S-1} * \delta_0)^2}. \quad (11)$$

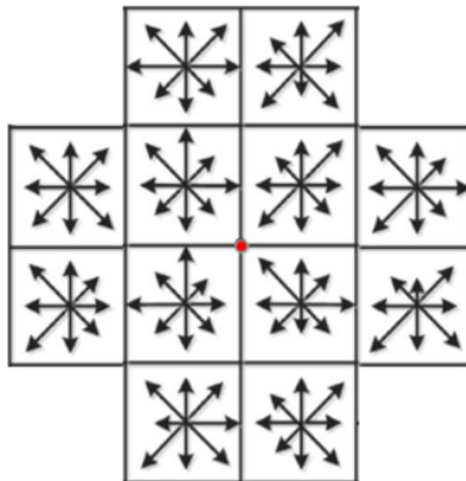
Авторами було доведено, що використовуючи саме таке значення стабілізатора  $P$  можна зменшити кількість хибних точок в середньому на 15%, що збільшило швидкодію алгоритму на 26%, а похибку зменшило на 20%.

Для отримання ще більшої продуктивності запропоновано зменшити розмірність дескриптора. У базовому алгоритмі SIFT обчислювальна складність дескрипторів становить

$$O(\text{кількістьключовихточок} \times d_1 \times d_2 \times (n^2 + \text{scale}^2)) \quad (12)$$

де  $d_1$  та  $d_2$  — розмірність дескриптора (4 та 4 у базовому алгоритмі відповідно);  
 $n$  — кількість ітерацій алгоритму;  
 $\text{scale}$  — розмір шкали напрямів градієнтів (8 у базовому алгоритмі).

З виразу (12) видно, що чим більші розміри дескриптора, тим вища складність обчислень. Авторами було запропоновано ігнорувати в обчисленнях кутові сектори матриці дескрипторів. Ідея ґрунтується на тому, що чим далі піксель знаходиться від ключової точки, тим меншим є його вплив на взаємозв'язок між ключовими точками. Таким чином можна зменшити розмір дескриптора до 96 (рис. 5), що дозволяє скоротити час обчислень.

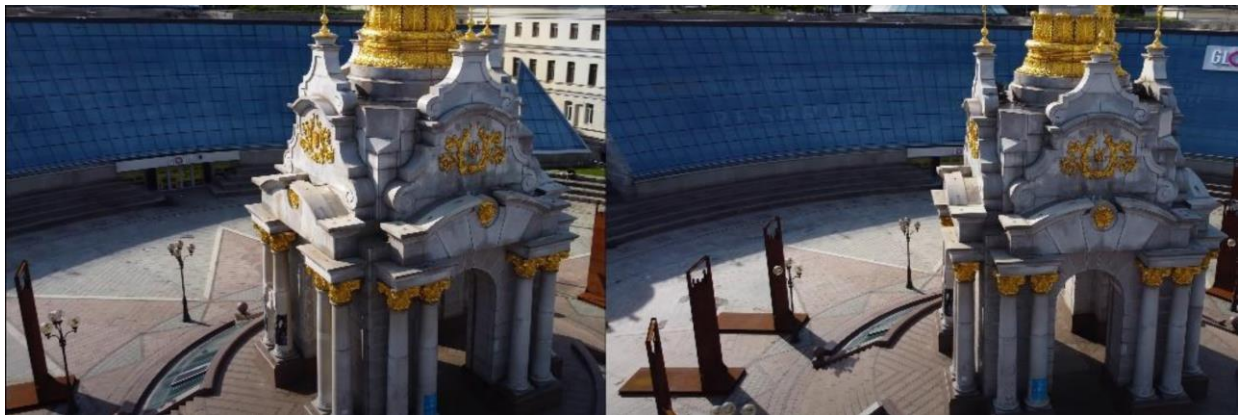


**Рисунок 5.** Зображення 96-мірного дескриптора

Іншим підходом до підвищення продуктивності та покращення відповідності моделі реальному об'єкту може стати вилученням кадрів або знімків із зображеннями, які апріорі не можуть мати спільних точок. Наприклад, знімки об'єкту, виконані з протилежних ракурсів не можуть мати спільних точок, хоча і відносяться до того самого об'єкта. При пошуку

відповідності між такими знімками витрати часу та обчислювальних ресурсів будуть недоцільними, оскільки не дадуть очікуваного результату. Вилученню можуть підлягати і знімки з однотипними зображеннями, що може пов'язане з невірно підібраними ракурсами при зйомці об'єктів, які мають симетрію відносно деякої осі.

Вилучення зазначених зображень може бути здійснено за допомогою нейромережі, що навчена на визначення схожості кадрів. Перетворюючи кожен вхідний кадр (рис. 6а) у чорно-білий формат та виділяючи пікселі з максимальним та мінімальним значеннями яскравості можна отримати контрастне чорно-біле зображення (рис. 6б). Аналізуючи розташування світлих та темних фрагментів зображення нейромережа зможе передбачити наявність спільних деталей на двох послідовних зображеннях ще до того, як вони будуть оброблені SIFT алгоритмом.



a)



b)

**Рисунок 6.** а) Вхідні зображення  
б) Вхідні зображення у контрастному чорно-білому форматі

Якщо нейромережа не виявить на зображеннях спільних елементів, програмі не треба буде витрачати час на пошук спільних точок між цими зображеннями. Завдяки цьому не лише підвищується швидкість отримання вихідної 3D моделі, а й збільшиться її точність, оскільки SIFT алгоритм не генеруватиме помилкових спільних точок.

**Висновки.** Фотограмметричні методи отримання тривимірних моделей об'єктів сьогодні широко використовуються у різних сферах. При цьому вони є єдино можливими для застосування у значній кількості випадків. Основними недоліками фотограмметрії є неможливість отримання тривимірних моделей об'єктів зі слабконтрастною структурою поверхні або рухомих об'єктів, а також висока трудомісткість, яка вимагає використання

високопродуктивного комп'ютерного обладнання. Останній недолік обумовлений необхідністю пошуку спільних елементів на багатьох зображеннях та їх співставлення між собою. Гарні результати з виявлення ключових елементів зображення дає алгоритм масштабно-інваріантного перетворення ознак SIFT, який має хороші перспективи для подальшого вдосконалення щодо зменшення часу роботи, про що свідчить значна кількість його покращених модифікацій. Ще одним підходом до зменшення часу побудови моделі може стати використання нейромережевих технологій для попередньої обробки зображень з метою вилучення тих, що не мають спільних елементів або повторюються.

**Подяки** «Немає».

**Конфлікт інтересів** «Немає».

## References

- Abdel-Hakim A. E. and Farag, A. A. (2006) CSIFT: A SIFT Descriptor with Color Invariant Characteristics, *IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'06)* (pp. 1978-1983), New York, NY, USA, doi: 10.1109/CVPR.2006.95.
- Caolan P., Farzad R., Diptangshu P., Hannah Th., Nigel Cl. A (2023) Framework for Realistic Virtual Representation for Immersive Training Environments. In *Proceedings Of The 23rd International Conference On Construction Applications Of Virtual Reality* (pp. 274-287). Florence, Italy: University of Florence.
- Gosling Th. Recommended Computer Workstation For Agisoft Metashape (2023). Retrieved from <https://www.workstationspecialist.com/recommended-computer-workstation-for-agisoft-metashape/>.
- Hossein-Nejad Z., Agahi H. and Mahmoodzadeh A. (2021) Image matching based on the adaptive redundant keypoint elimination method in the SIFT algorithm. *Theoretical Advances*, vol.24, 669–683.
- Kotlyk S., Romanyuk O., Sokolova O., Kotlyk D. (2022) Development of affordable technology for creating 3D computer models based on photogrammetry. Part I. *Automation of technological and business processes*, 14 (2), 37-50.
- Lowe D.G. (1999) Object recognition from local scale-invariant features. In *Proceedings of the Seventh IEEE International Conference on Computer Vision* (pp. 1150 - 1157). Kerkyra, Greece.
- Lowe D.G. (2004) Distinctive Image Features from Scale-Invariant Keypoints. *International Journal of Computer Vision* 60, 91–110.
- Mikolajczyk K. and Schmid, C. (2005) A performance evaluation of local descriptors, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 27 (10), 1615-1630, doi: 10.1109/TPAMI.2005.188.
- Misra S., Li H., He J. (2019). *Machine Learning for Subsurface Characterization*. Gulf Professional Publishing.
- Morel Jean-Michel and Yu Guoshen. (2009). ASIFT: a new framework for fully affine invariant image comparison. *SIAM Journal on Imaging Sciences*, 2 (2), 438-469.
- Nebel S, Beege M, Schneider S and Rey GD (2020) A Review of Photogrammetry and Photorealistic 3D Models in Education From a Psychological Perspective. *Front. Educ.* vol.5, 1–15.
- Olagoke A.S., Ibrahiman H., Teoh S.S. (2020) Literature survey on multi-camera system and its application. *IEEE Access*, 8, 172892-172922.

- Pavlov S. V., Romanyuk S. O., Nechiporuk M. L. (2018) Adaptive determination of diffuse and specular components of color for the rendering of face images when planning plastic surgery. *Scientific Journal "ScienceRise"*, 49 (8), 2018, 24-28.
- Romaniuk S. O., Pavlov S. V., Titova N. V. and Koval L. G. (2022) Using graphic 3D images of faces for express diagnosis and construction of biomedical devices. *Optoelectronic Information-Power Technologies*, 42 (2), 12–20.
- Scholtens A. (2023) *Capturing Reality in the Fascinating World of Photogrammetry*. Sas155
- Tan X. and Triggs B. (2010) Enhanced local texture feature sets for face recognition under difficult lighting conditions. *IEEE Trans Image Process*, 19 (6), 1635-1650.
- Tang L., Ma S., Ma X., You H. (2022) Research on Image Matching of Improved SIFT Algorithm Based on Stability Factor and Feature Descriptor Simplification. *Applied Sciences*, 12 (17), 8448 - 8466.
- Zmejevskis L. PC For Photogrammetry – What Hardware Do You Need? (2022) Retrieved from <https://www.pix-pro.com/blog/photogrammetry-pc>.

УДК 004.8

## Використання інструментів нейромереж для пришвидшення розробки веб-інтерфейсів

**Дмитро Юрійович Петрина,**

д.т.н., професор, професор кафедри технічної механіки,  
інженерної та комп'ютерної графіки,  
Івано-Франківський національний технічний університету нафти і газу  
ORCID: 0000-0003-0663-746X

**Володимир Андрійович Корнута,**

к. т. н., доцент, доцент кафедри інженерії програмного забезпечення,  
Івано-Франківський національний технічний університету нафти і газу  
ORCID: 0000-0002-3345-2578

**Олена Володимирівна Корнута,**

кандидат технічних наук, доцент, доцент кафедри технічної механіки,  
інженерної та комп'ютерної графіки,  
Івано-Франківський національний технічний університету нафти і газу  
ORCID: 0000-0002-0626-888X

**Анотація.** Стаття присвячена розгляду сучасних інструментів нейромереж, що дозволяють пришвидшити розробку веб-інтерфейсів та спростити роботу UI/UX дизайнерів. Одна з основних проблем сучасного дизайну – швидкий доступ до загальної інформації та можливого структурування сайту зі спеціалізованим контентом, а також отримання візуального його наповнення. Зараз нейромережі ще не можуть замінити дизайнерів, але значною мірою допомагають їм вирішувати поставлені завдання. Всі нейромережі, що можуть використовуватися у дизайні веб-інтерфейсів, можна поділити на чотири основні типи: конволюційні, рекурентні, мережі прямого розповсюдження та генеративна змагальна мережа. В своїй роботі дизайнер може використовувати переважно генеративні мережі, класифікацію їх можна здійснити за принципом «інформація на вході – інформація на виході». При роботі над проектом дизайнер може створити запит до нейромережі та отримати декілька варіантів, генерувати різні ідеї та створювати мудборди на їх основі, підбирати кольори, градієнти, текстуру, типографіку тощо. Нейромережа може створити різноманітні графічні елементи: іконки, кнопки, ілюстрації та фотографії з потрібною перспективою, стилем, кольорами. Також є перспективним використання нейромереж для покращення зображень, домальовування або видалення необхідних елементів. Розглянуто процес пришвидшення створення інтерфейсу лендингової сторінки за допомогою додатку Midjourney. Наведені приклади написання запитів (prompts), які впливатимуть на остаточну якість згенерованого зображення. Отримані результати є якісним візуальним контентом, який можна або розмістити в проєкті, або використати ідею, розташування елементів, композицію, кольорову гаму, фотографії, іконки тощо. Після створення графічних елементів дизайну за допомогою Chat GPT 3.5 був створений контент лендингової сторінки. Безпосередньо в середовищі Figma для швидкої генерації потрібного контенту можна використати плагін FIG GPT. Існуючі певні недоліки та неточності генерування, що виникають в роботі, можуть бути виправлені шляхом швидкого оновлення та створення нових версій нейромереж.

**Ключові слова:** нейромережа, веб-інтерфейс, UI/UX дизайн, Chat GPT, Midjourney.

## Using neural network tools to accelerate the development of Web interfaces

**Dmytro Y. Petryna,**

Doctor of Technical Sciences, Professor,  
Professor of the Department of Technical Mechanics, Engineering and Computer Graphics  
Ivano-Frankivsk National Technical University of Oil and Gas  
ORCID: 0000-0003-0663-746X

**Volodymyr A. Kornuta,**

Candidate of Technical Sciences, Associate Professor,  
Associate Professor of the software engineering department  
Ivano-Frankivsk National Technical University of Oil and Gas  
ORCID: 0000-0002-3345-2578

**Олена Володимирівна Корнута,**

Olena V. Kornuta,  
Candidate of Technical Sciences, Associate Professor,  
Associate Professor of the Department of Technical Mechanics, Engineering and Computer Graphic  
Ivano-Frankivsk National Technical University of Oil and Gas  
ORCID: 0000-0002-0626-888X

**Abstract.** The article is devoted to considering modern neural network tools that allow for speeding up the development of web interfaces and simplifying the work of UI/UX designers. One of the main problems of modern design is quick access to general information and possible structuring of a site with specialized content, as well as obtaining its visual content. Currently, neural networks cannot replace designers, but to a large extent help them solve tasks. All neural networks that can be used in the design of web interfaces can be divided into four main types: convolutional, recurrent, forward propagation, and generative adversarial networks. In his work, the designer can mainly use generative networks, they can be classified according to the principle of "information at the input - information at the output". When working on a project, the designer can create a request to the neural network and get several options, generate different ideas, and create mood boards based on them, selecting colors, gradients, texture, typography, etc. The neural network can create various graphic elements: icons, buttons, illustrations, and photos with the right perspective, style, and colors. Using neural networks to improve images and refine or remove necessary elements is also promising. The process of speeding up the creation of the landing page interface using the Midjourney application is considered. Examples of writing prompts (prompts) that will affect the final quality of the generated image are given. The results are high-quality visual content that can either be placed in a project or used as an idea, element placement, composition, color scheme, photos, icons, etc. After creating the graphic design elements using Chat GPT 3.5, the landing page's content was created. You can use the FIG GPT plugin directly in the Figma environment to quickly generate the required content. Existing shortcomings and generation inaccuracies that arise in the work can be corrected by quickly updating and creating new versions of neural networks.

**Key words:** neural network, web interface, UI/UX design, Chat GPT, Midjourney.

**DOI:** <https://doi.org/10.31649/1999-9941-2024-60-2-42-50>

**Вступ.** Стрімкий розвиток штучного інтелекту (ШІ) викликає еволюційні зміни в багатьох сферах та галузях економіки. Вже зараз нейромережі допомагають створювати креативні та унікальні проекти копірайтерам, SMM – фахівцям, маркетологам, та звичайно, веб дизайнерам. Одна з основних проблем сучасного дизайну – швидкий доступ до

загальної інформації та можливого структурування сайту зі спеціалізованим контентом, а також отримання візуального його наповнення. Зараз нейромережі ще не можуть замінити дизайнерів, але значною мірою допомагають їм вирішувати поставлені завдання. Наприклад, нейромережа може підібрати контент для макетів, створити фон або навіть фрагменти проектів, анімацію, логотипи, значки та іконки, підібрати елементи фірмового стилю тощо, тобто автоматизувати процес створення веб-сайту, починаючи від веб дизайну та завершуючи розробкою користувацького інтерфейсу. Використання нейромереж дозволяє створювати сайти, що є більш персоналізованими, інтуїтивними та ефективними, значно прискорює процес розробки та відкриває нові можливості для інновацій. Сучасні нейромережі постійно розвиваються, вдосконалюються у вигляді оновлених версій, а також створюються нові, враховуючи помилки попередників. Однією з перших нейромереж, що базується на взаємному діалозі, є Chat GPT.

Сьогодні доступна 4-та версія навіть у безкоштовному варіанті, проте з певними обмеженнями. Потужним конкурентом Chat GPT є нова розробка від Google – Gemini, яка майже необмежена у використанні і застосовує більші можливості у спілкуванні та виводі остаточного результату. Для створення інтернет сторінок застосовуються зображення, які генеруються лідером зі створень візуального контенту – лабораторією Midjourney з доступною версією V6.1. Ці програми є ключовими інструментами у розробці сучасних сайтів і значно прискорюють роботу дизайнера. Принципи використання цих інструментів та їх переваги і досліджувалися у даній роботі.

**Огляд літератури.** Аналіз сучасних досягнень у галузі штучних нейронних мереж, машинного навчання та обчислювального інтелекту розглянуто (Maltsev, A. 2022). Зараз з'явилася низка публікацій, присвячених використанню штучного інтелекту в дизайні та обговорення його ролі в процесі проектування. Автори (Томіс І. та ін., 2023) вважають, що первинною метою інструментів ШІ у графічному дизайні раніше була оптимізація та швидкість – заміна дизайнерів у роботі при виконанні повторюваних завдань або аналіз великої кількості даних користувача для створення кращих рішень. Сьогодні ж ШІ дозволяє дизайнерам зосередитися на творчій частині своєї роботи, але також створювати дизайни з нуля, дотримуючись вказівок користувачів. Інструменти ШІ пропонують більшу гнучкість та креативність пошуку найкращого рішення для різних завдань, таких як вибір шрифту та кольору, редагування зображення, створення найкращих макетів та композицій. Вони також забезпечують високий рівень автоматизації, переважно в редагуванні та дизайні UI/UX. Нові алгоритми ШІ, практична інтерпретація та ефективність для реальних користувачів розглядається в (J. Zhu et al.,). У (Slityuk O. et al., 2023) представлено результати дослідження сучасних методів впровадження нейронних мереж штучного інтелекту у веб-дизайні та визначенні принципів використання нейромереж для створення сайтів, їх вплив на взаємодію з дизайнером цифрового ресурсу. Інтелектуальна система, що генерує рекламні плакати з можливістю онлайн-редагування та зворотнім зв'язком наведена (Shunan G. et al., 2021). В праці (Gang Chen et al., 2019) розглянутий креативний процес при використанні ШІ. Узагальнення досліджень впливу ШІ на професійну діяльність дизайнера та перевірка можливостей та результатів проектних рішень на основі ШІ досліджено (Irbite A. et al., 2021). Відповідність нейронних мереж кожному з видів продукції графічного дизайну та ефективність мереж у генеруванні концептів персонажів і локацій для комп'ютерних ігор, ілюстрацій для друкованих та електронних видань, торгових марок і логотипів, носіїв фірмового та іміджевого стилю доведено (Bozhko T., & Arefiev V. 2023). Ноуе Farhana (2019) обговорює наслідки використання ШІ в роботі дизайнера, як технологія штучного інтелекту полегшить обов'язки дизайнерів та причини, через які вона не зможе замінити дизайнерів та архітекторів у найближчому майбутньому. У статтях (Verganti Roberto et al., 2020; Ying Du et al, 2023) розглядаються етичні міркування щодо використання ШІ в графічному дизайні та наслідки, які він має для галузі. Потенційні переваги та недоліки

штучного інтелекту в дизайні наведені в (Bahaa Mustafa, 2023). Singh K. (2023) досліджує навички, які знадобляться майбутнім дизайнерам для ефективної роботи зі ШІ.

**Актуальність.** Використання нейромереж для пришвидшення розробки веб-інтерфейсів є досить актуальним, оскільки дозволяє допомогти у генерації дизайнів веб-сторінок на основі вхідних даних та заданих критеріїв, що значно прискорює процес розробки, аналізувати поведінку користувачів на веб-сайті для виявлення паттернів та покращення веб-інтерфейсу з урахуванням цих даних, оптимізує завантаження веб-сторінок, попередження затримок і покращення роботи веб-сайтів на мобільних пристроях. Також нейромережі допомагають у створенні персоналізованого контенту для користувачів, що дозволяє підлаштовувати веб-інтерфейс під індивідуальні потреби та інтереси.

**Мета та задачі.** Метою дослідження є визначення принципів використання різних типів нейромереж як інструмента створення інтерфейсів у веб дизайні та визначення відповідності такого інструментарію як нейромережі, потребам створення кожного з складників продукції UI дизайну.






























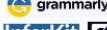






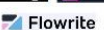



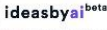
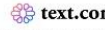




















Задачі дослідження містять наступні аспекти:

- 1) проведення огляду різних типів нейромереж, їх функціональних можливостей та сервісів;
- 2) аналіз методів автоматизації та пришвидшення розробки інтерфейсу лендингової сторінки з використанням нейромереж;
- 3) формулювання висновків та рекомендацій щодо використання нейромереж для пришвидшення розробки веб-інтерфейсів.

**Методологічне обґрунтування.** На даний момент існує багато нейромереж і кількість їх невинно зростає. Якщо брати до уваги нейромережі, що можуть використовуватися у дизайні веб-інтерфейсів, то їх можна поділити на чотири основні типи: конволюційні, рекурентні, мережі прямого розповсюдження та генеративна змагальна мережа. В своїй роботі дизайнер може використовувати переважно генеративні мережі, класифікацію їх можна здійснити за принципом «інформація на вході – інформація на виході» (рис.1).

За допомогою нейромереж дизайнер може знаходити референси та натхнення для нової роботи. На початку роботи над новим проектом дизайнер може створити запит до нейромережі та отримати декілька варіантів, які згодом можна використати в роботі. Також можна генерувати різні ідеї та створювати мудборди на їх основі, підбирати кольори, градієнти, текстуру, типографіку тощо. Нейромережа може створити різноманітні графічні елементи: іконки, кнопки, ілюстрації та фотографії з потрібною перспективою, стилем, кольорами. Також є перспективним використання нейромереж для покращення зображень, домальовування або видалення необхідних елементів. Для цього використовується функція Image-to-Image (I2I): завантажується зображення, пишеться промт з описом необхідних на зображенні змін, нейромережа обробляє запит та видає новий варіант зображення з необхідними змінами. Крім того, можна покращувати якість ілюстрацій, змінювати їх розмір, видаляти або додавати необхідне тло тощо.

При написанні запиту необхідно чітко та максимально докладно описати поставлене завдання (кольор, стиль, ...). Запит повинен бути максимально конкретним та деталізованим, необхідно вказувати всі, навіть найменші подробиці. Якщо вхідної інформації недостатньо, то нейромережа може додумати та самостійно згенерувати необхідні елементи. Існують сайти, що мають колекції вже підготовлених промтів, що допомагає користувачам конструювати їх власні запити, не лише для проектів, але і для портфоліо дизайнера.

|  <b>Generative AI</b><br><small>Curated by @aaronSIM</small> |  |
|---|--|
| Text-to-Image (T2I)   |    <br>   <br>     |
| Text-to-Video (T2V)   |         |
| Text-to-Audio (T2A)   |        |
| Text-to-Text (T2T)  |     <br>    <br>     <br>    <br>    |
| Text-to-Motion (T2M)  |  <small>MDM: Human Motion Diffusion Model</small>   |
| Text-to-Code (T2C)  |    |
| Text-to-NFT (T2N)   |   |
| Text-to-3D (T2D)  |     |
| Audio-to-Text (A2T)   |     |
| Audio-to-Audio (A2A)  |    |
| Brain-to-Text (B2T)   | <small>speech from brain</small>   |

**Рисунок 1.** Класифікація генеративних нейромереж, що використовуються в дизайні

Джерело: Sim Aaron, (2023)

Незважаючи на те, що на даний момент створення інтерфейсів веб-сайтів виключно за допомогою інструментів ШІ не є можливим, нейромережі можуть бути гарним орієнтиром та помічником при створенні реальних проєктів. Нашою основною задачею було пришвидшити та по можливості автоматизувати процес створення інтерфейсу лендингової сторінки. Оскільки існують різні налаштування, які можуть ввести певні відхилення в порівняльний аналіз різних версій додатків, то для процесу генерації використовували однакові налаштування, що діють за замовчуванням. Для дослідження обрали генерацію лендингової сторінки для навчання графічному дизайну. Спочатку для генерації зображення обрали додаток **Midjourney**. За допомогою цієї нейромережі можна отримати високоякісні результати, що є одними з найкращих, при цьому можна за результатами запитів або отримати візуальні матеріали, або покращити вже отримане зображення. Отримані зображення можна використовувати з комерційною метою або приватно. При роботі з **Midjourney** використовуються запити (пропти), що коротко описують тематику та характеристики зображення, яке потрібно згенерувати. Отримані результати є якісним візуальним контентом, який можна або розмістити в проєкті, або використати ідею, розташування елементів, композицію, кольорову гаму, фотографії, іконки тощо. Всі ці елементи є унікальними, що робить можливим їх використання в комерційних проєктах. Наприклад, інколи для візуалізації потрібні зображення людей (наприклад, лікарі для медичного сайту, будівельники для сайту з будівництва тощо). Нейромережа може створити зображення людей, які в дійсності не існують, але слід звертати увагу на деталі, наприклад, не завжди є п'ять пальців на руках тощо. При генеруванні зображень обов'язково вказуємо стиль (можна вказати прізвище художника або дизайнера), перспективу, колір, необхідні елементи.

Зазвичай нейромережа генерує чотири варіанти. Існує функціонал для збільшення найбільш вдалого варіанту та отримання його в гарній якості та у великому розмірі.

Зображення можна скопіювати, завантажити у Figma та здійснити його редагування. Також до обраного, найбільш вдалого зображення можна згенерувати ще чотири варіації. На перший погляд ці варіації є схожими, але існують відмінності у дрібних деталях, елементах та інтерфейсі. При генеруванні інтерфейсів до запиту обов'язово слід додавати слова «UI, UX, design, web, dribbble, behance, awwwards». При необхідності згенерувати фото, до запиту додаємо слова «фотореалізм, full HD тощо».

**Результати та обговорення.** Особливістю роботи з додатком Midjourney є можливість вказання пропорцій зображення, яке потрібно згенерувати (за замовчуванням генерується квадратне зображення). Розглянемо конструкцію запиту, що використовувався для генерації лендингу:

*luxury landing page for graphic design education, ui, ux, ui/ux, bright colors, watercolor style --s 500 --no mockup --ar 3:2 --v 6.0 --style raw.*

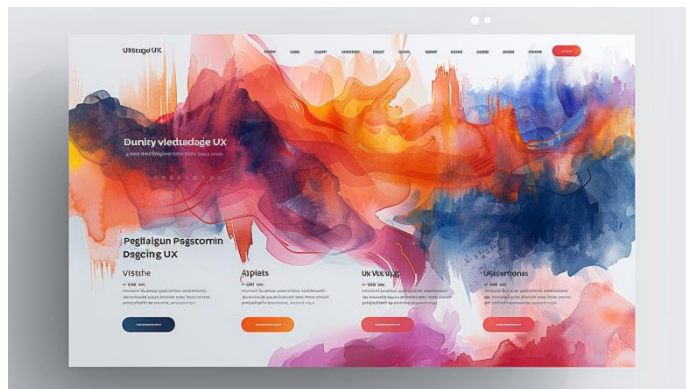
В результаті генерації одержали декілька зображень екрану, одне з них наведено на рисунку 2.

Для створення зображення графічного логотипу для курсів навчання дизайну знову використали неймережу Midjourney та написали відповідний промт:

*minimalistic logo for graphic design education company, flat pictogram, black and white, white background.*

В результаті одержали декілька варіантів чорно-білих логотипів на білому фоні, для подальшої роботи обрали логотип, зображений на рисунку 3.

Після створення графічних елементів дизайну за допомогою Chat GPT 3.5 створили контент лендингової сторінки. Контент за змістом можна створювати за допомогою текстогенераторів, наприклад, Chat GPT. Текст «Lorem ipsum», який на протязі довгого часу використовувався для створення дизайнів інтерфейсу на початкових стадіях (поки не було майбутнього контенту) зараз не використовується. Рекомендується відразу вставляти в сайт контент, який буде подібний до реального. Нажаль, пошук такого контенту займає багато часу, особливо якщо він стосується специфічних галузей (наприклад, медицина, енергетика, будівництво), тому дизайнер може витратити багато часу на пошук таких даних в інтернеті. Chat GPT допомагає швидко генерувати текст не лише для контенту, але й для елементів інтерфейсу (помилки, сповіщення, порожні сторінки тощо). Наприклад, при описі сторінки 404 написати «Page not found» буде не «User Friendly» по відношенню до користувача. Chat GPT може згенерувати «User Friendly» повідомлення, які будуть приємними для користувача та викличуть у нього позитивні емоції. Також Chat GPT може генерувати ідеї для дизайнера, наприклад, при брейнштурмі, для контенту в соцмережах, для портфолію.



**Рисунок 2.** Результати генерації зображення додатком Midjourney

Джерело: Згенеровано авторами додатком Midjourney



**Рисунок 3.** Варіант логотипу, згенерованого додатком Midjourney

Джерело: Згенеровано авторами додатком Midjourney

Також Chat GPT може допомагати генерувати інформацію, що пов'язана з UX search. Наприклад, скласти архітектуру продукту (перед цим треба детально описати продукт), написати User Flow, User Persons, Custom and journe Maps. Також можна згенерувати структуру сторінки чи лендингу на задану тематику, вказати кількість секцій, стиль, структури різних елементів, наприклад фільтрів. Ця мережа також може бути корисною при аналізі та порівнянні конкуруючих компаній, оскільки швидко та точно збирає необхідні дані. Слід пам'ятати, що при складанні запитів Chat GPT важливо, щоби першим словом було дієслово – порівняй, досліди, проаналізуй тощо.

Для швидкої генерації потрібного контенту безпосередньо в середовищі Figma може допомогти плагін FIG GPT, що працює в режимі редагування та створення і має контроль стилю письма та кількості слів.

Для створення контенту в нашому прикладі лендингової сторінки використовували Chat GPT 3.5. Після написання відповідного запиту нейромережа запропонувала структуру лендингової сторінки (заголовок, підзаголовок, вступне відео або зображення, опис курсу, переваги курсу, свідоцтва та відгуки, цінова пропозиція) та перелік графічних елементів інтерфейсу сторінки (пунктів меню або кнопок), що будуть знаходитися на головному екрані (заклик до дії, контактна інформація, FAQ, блок новин, посилання на соціальні мережі, адреса та мапа, підписка на розсилку, політика конфіденційності та умови використання). Приклад діалогу з Chat GPT 3.5 та одержаний контент наведені на рисунках 4, 5.

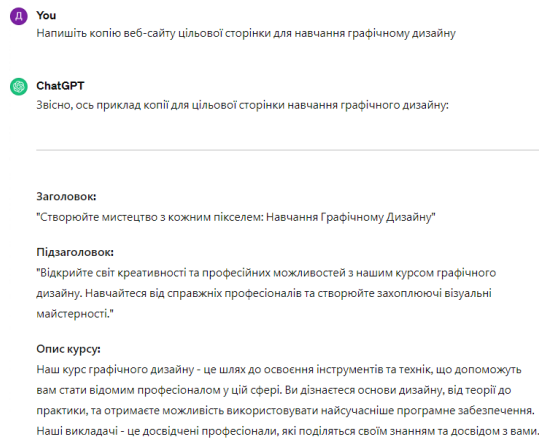
**You**  
Напиши структуру landing page для навчання графічного дизайну

**ChatGPT**  
Ось пропозиція структури для лендингу, який пропонує навчання графічного дизайну:

- Заголовок (Headline)**
  - Привабливий заголовок, який відображає основну перевагу вашого курсу.
- Підзаголовок (Subheadline)**
  - Коротке пояснення того, що користувач зможе досягти, навчаючись від вас.
- Вступне відео або зображення**
  - Коротке відео або захоплююче зображення, що демонструє важливість і захоплення графічним дизайном.
- Опис курсу**
  - Короткий огляд програми курсу, включаючи теми, які будуть вивчені, та цілі, які будуть досягнуті.
- Переваги курсу**
  - Перелік основних переваг вашого курсу, таких як доступність відмінних викладачів, практичні вправи, можливість взаємодії з іншими студентами тощо.
- Свідоцтва та відгуки**
  - Короткі відгуки попередніх учасників курсу або висловлення задоволення від попередніх студентів.
- Цінова пропозиція**
  - Пояснення вартості курсу та можливостей оплати, таких як розстрочка або знижки для ранніх реєстрацій.
- Виклик до дії (Call to Action)**
  - Чіткий та привабливий виклик до дії, який надихне відвідувачів лендингу зареєструватись або отримати додаткову інформацію.
- Контактна інформація**
  - Контактна інформація для зв'язку, така як електронна пошта, телефон або форма зворотного зв'язку.
- Часті питання (FAQ)**
  - Найбільш поширені запитання та відповіді, щоб допомогти відвідувачам знайти необхідну інформацію.
- Блок новин або оновлень**
  - Інформація про нові курси, події або оновлення в програмі навчання.
- Соціальні мережі**
  - Посилання на ваші сторінки у соціальних мережах, де відвідувачі можуть отримати додаткову інформацію та спілкуватись з іншими студентами.
- Адреса та мапа (якщо це офлайн курс)**
  - Якщо курс відбувається в офлайн форматі, включіть адресу навчального закладу та мапу для зручності учасників.
- Підписка на розсилку (Optional)**
  - Пропозиція підписатися на розсилку для отримання оновлень, акцій або цікавих статей з області графічного дизайну.
- Політика конфіденційності та умови використання**
  - Посилання на сторінку з політикою конфіденційності та умовами використання, які регулюють взаємодію з сайтом та надання інформації.

**Рисунок 4.** Діалог з Chat GPT 3.5 (запити щодо структури лендингової сторінки та переліку графічних елементів)

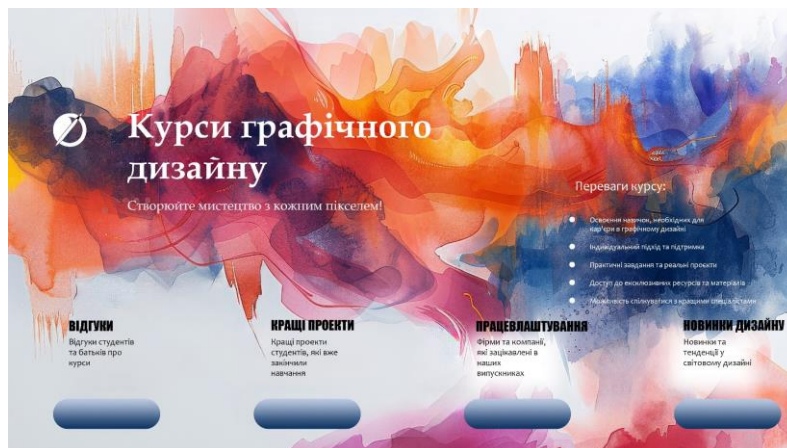
Джерело: Згенерований авторами діалог з Chat GPT



**Рисунок 5.** Одержаний за допомогою Chat GPT 3.5 контент

**Джерело:** Згенерований авторами контент з Chat GPT

Як було вказано вище, штучний інтелект не може поки що якісно створювати графічні інтерфейси, тому одержане зображення для остаточного редагування завантажили в програму Figma, де вставили логотип, видалили зайві написи та дописали потрібний контент, одержаний в чаті Chat GPT 3.5 (рис.6).



**Рисунок 6.** Остаточний варіант інтерфейсу

**Джерело:** розроблено авторами

**Висновки.** Для створення дизайну інтернет сторінок ми використали найбільш популярні нейромережі, а саме Chat GPT, який є текстогенеруючим та Midjourney для генерування візуальних зображень та фонів. Отримання швидкої інформації та цікавих ідей - є ключовим завданням для сучасного дизайнера WEB сайтів і саме Chat GPT допоміг у структуризації сторінки сайту та можливих варіантів його дизайну. Візуальна лабораторія Midjourney дозволила швидко отримати візуальне наповнення інтернет сторінки. Дальший розвиток нейромереж призводить до появи нових програм та додатків, що працюють із зображенням і навіть відео контентом. Знання і вміння використовувати ці інструменти буде ключовою задачею для сучасного дизайнера і значно спростить та вдосконалив всі аспекти такої роботи. Існуючі певні недоліки та неточності генерування, що виникають в роботі, можуть бути виправлені шляхом швидкого оновлення та створення нових версій нейромереж.

**Подяки.** «Немає».

**Конфлікт інтересів.** «Немає».

## References

- Bozhko, T., & Arefiev, V. (2023). Neural Networks as a Graphic Design Tool. *Bulletin of KNUKiM. Series in Arts*, (48), 125–135. <https://doi.org/10.31866/2410-1176.48.2023.282475>.
- Chen, G., Xie, P., Dong, J., & Wang, T. (2019). Understanding Programmatic Creative: The Role of AI. *Journal of Advertising*, 48(4), <https://doi.org/10.1080/00913367.2019.1654421>.
- Farhana Hoque (2024). Does Artificial Intelligence have the Possibility of Taking Over Designers' Jobs in the Future? *International Journal of Science and Business*, 31(1), <https://doi.org/10.58970/IJSB.2273>.
- Irbite, A., & Strode, A. (2021). Artificial intelligence vs designer: the impact of artificial intelligence on design practice. society. integration. education. In *Proceedings of the International Scientific Conference*, 4, (p.539-549). <https://doi.org/10.17770/sie2021vol4.6310>
- Maltsev, A. (2022). Analysis of modern achievements in the field of artificial neural networks, machine learning and computational intelligence. *Information Technology and Society*, 2 (4), <https://doi.org/10.32689/maup.it.2022.2.9>.
- Mustafa, Bahaa (2023). The Impact of Artificial Intelligence on the Graphic Design Industry. *Arts and Design Studies*. 104 <https://doi.org/10.7176/ADS/104-01>.
- Shunan Guo, Zhuochen Jin, Fuling Sun, Jingwen Li, Zhaorui Li, Yang Shi, and Nan Cao. 2021. Vinci: An Intelligent Graphic Design System for Generating Advertising Posters. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems (CHI '21)*. Association for Computing Machinery, New York, USA, Article 577, (p. 1–17). <https://doi.org/10.1145/3411764.3445117>.
- Sim Aaron, (2023). Retrieved from <https://x.com/aaronsiim/status/1595544909540458496>
- Singh, K.D., Duo, Y.X. (2023). Future Design: An Analysis of the Impact of AI on Designers' Workflow and Skill Sets. In: Vasant, P., et al. *Intelligent Computing and Optimization. ICO 2023. Lecture Notes in Networks and Systems*, vol 852. Springer, Cham. [https://doi.org/10.1007/978-3-031-50330-6\\_23](https://doi.org/10.1007/978-3-031-50330-6_23)
- Slityuk O., Struminska T., & Hlinska A. (2023). Application of neural networks to provide artistic expression in the filling of websites, In *V International Scientific Conference «Topical issues of modern design»*, Kyiv, KNUiM, 27.04.2023 (p.379-382).
- Tomić, Ivana & Jurič, Ivana & Dedijer, Sandra & Adamovic, Savka (2023). Artificial Intelligence in Graphic Design, In *54th Annual Scientific Conference of the International Circle of Educational Institutes of Graphic-Media Technology and Management*, 2023. [https://www.researchgate.net/publication/375423443\\_Artificial\\_Intelligence\\_in\\_Graphic\\_Design](https://www.researchgate.net/publication/375423443_Artificial_Intelligence_in_Graphic_Design).
- Verganti, R., Vendraminelli, L. and Iansiti, M. (2020). Innovation and Design in the Age of Artificial Intelligence. *Journal of Product Innovation Management*, 37. <https://doi.org/10.1111/jpim.12523>.
- Ying Du, Tianyu Li, Chang Gao (2023). Why do designers in various fields have different attitude and behavioral intention towards AI painting tools? an extended UTAUT model, *Procedia Computer Science*, Volume 221, <https://doi.org/10.1016/j.procs.2023.08.010>.
- Zhu J., A. Liapis, S. Risi, R. Bidarra and G. M. Youngblood, Explainable AI for Designers: A Human-Centered Perspective on Mixed-Initiative Co-Creation (2018), in *2018 IEEE Conference on Computational Intelligence and Games (CIG)*, Maastricht, Netherlands. 2018, (pp. 1-8). <https://doi.org/10.1109/CIG.2018.8490433>.

УДК 004.7

## Процеси асемблювання та реасемблювання документів у системі документообігу

**Тетяна Іванівна Коробейнікова**

к.т.н., доцент кафедри безпеки інформаційних технологій  
Національний університет «Львівська Політехніка»  
79013, м. Львів. вул. Степана Бандери, 12  
ORCID 0000-0003-2487-8742

**Людмила Анатоліївна Савицька**

к.т.н., доцент кафедри обчислювальної техніки  
Вінницький національний технічний університет  
21021, м. Вінниця. Хмельницьке шосе, 95  
ORCID 0000-0003-1130-2621

**Леонід Віталійович Крупельницький**

к.т.н., доцент кафедри обчислювальної техніки  
Вінницький національний технічний університет  
21021, м. Вінниця. Хмельницьке шосе, 95  
ORCID 0000-0001-7370-9772

**Анотація.** Дана робота присвячена аналізу та вдосконаленню методів обробки документів в середовищі електронних систем документообігу, зокрема методів збереження цілісності та автентичності документів і їх генерації в автоматизованому режимі. Основна увага приділяється процесам асемблювання та реасемблювання документів. Дослідження ґрунтуються на широкому досвіді використання електронних систем документообігу та використовують загальнодоступну інформацію про найновіші методи та практики обробки, захисту та генерації документів загального користування. Під час оглядового аналізу літературних джерел був виконаний аналіз сучасних систем документообігу та водночас розглянуто мануальний підхід обробки документів. Метою оглядової частини роботи стало ознайомлення з існуючими імплементаціями електронних систем документообігу та розробки їх характеристик для виконання порівняння їх між собою для подальшої роботи з виявленими перевагами та недоліками. Зокрема розглянуті такі ЕСД, як «Дія», «PandaDoc» та «GoogleDocs». В результаті аналізу сучасного стану питання в галузі автоматизованої та мануальної обробки документів було розроблено технологічний ланцюжок спеціалізованої автоматизованої системи документообігу, розроблені та описані механізми асемблювання та реасемблювання документів, водночас описано інші процеси, що супроводжують даний технологічний ланцюжок. Метою технічної частини даної роботи став детальний огляд необхідних механізмів спеціалізованої автоматизованої системи документообігу, їх загальна взаємодія на рівні клієнт-серверної архітектури. У підсумку наукова новизна полягає у використанні спеціалізованого ПЗ для асемблювання та реасемблювання документів з метою вдосконалення технологічного ланцюжка спеціалізованої автоматизованої системи документообігу. Під час дослідження запропоновано аналітичний опис запропонованого ПЗ для асемблювання та реасемблювання документів, яке оригінально враховує можливість автоматизованої генерації документів.

**Ключові слова:** документообіг, електронна система документообігу, автоматизована генерація, процеси асемблювання та реасемблювання документів, спеціалізованої автоматизованої системи документообігу

## Assembly and reassembly processes of documents in the document system

**Tetiana I. Korobeinikova**

PhD, Associate Professor of Information Technology Security Department  
Lviv Polytechnic National University  
Stepana Bandery St, 12, Lviv, Ukraine, 79000  
ORCID 0000-0003-2487-8742

**Liudmyla A. Savytska**

PhD, Associate Professor of Computing Engineering Departmen  
Vinnytsya National Technical University  
Khmelnysk highway, 95, Vinnytsya, Ukraine, 21021  
ORCID 0000-0003-1130-2621

**Leonid V. Krupelnitskyi**

PhD, Associate Professor of Computing Engineering Departmen  
Vinnytsya National Technical University  
Khmelnysk highway, 95, Vinnytsya, Ukraine, 21021  
ORCID 0000-0001-7370-9772

**Abstract.** This work is dedicated to analyzing and improving document processing methods in the electronic document management environment, particularly methods of preserving the integrity and authenticity of documents and their automated generation. The focus is on document assembly and reassembly processes. The research is based on extensive experience with electronic document management systems and utilizes publicly available information on the latest methods and practices of processing, protecting, and generating documents for general use. During the literature review, an analysis of modern document management systems was conducted and a consideration of the manual document processing approach.

The review part of the work aimed to familiarize with existing implementations of electronic document management systems and to develop their comparative characteristics, highlighting their advantages and disadvantages. Specifically, such electronic document management systems as "DIA", "PandaDoc", and "GoogleDocs" were examined by the author. As a result of analyzing the current state of the issue in the field of automated and manual document processing, a technological chain of a specialized automated document management system was developed. Document assembly and reassembly mechanisms were designed and described, along with other processes accompanying this technological chain. The purpose of the technical part of this work is a detailed examination of the critical mechanisms of a specialized automated document management system and their overall interaction at the client-server level. In conclusion, the scientific novelty lies in improving the technological chain of a specialized automated document management system through software tools for document assembly and reassembly. During the research, an analytical description of software tools for document assembly and reassembly was proposed, considering the possibility of automated document generation.

**Keywords:** document circulation, electronic document management system, automated generation, document assembly and reassembly processes, specialized automated document management system.

**DOI:** <https://doi.org/10.31649/1999-9941-2024-60-2-51-65>

**Вступ.** У сучасному світі цифрові технології дедалі більше інтегруються в усі сфери діяльності, включаючи управління інформацією та документообіг. Електронні системи

документообігу (Ганжа та ін., 2022) стали основним засобом організації та автоматизації процесів обробки документів у державному (Копняк та Радзіховська, 2020) та приватному секторах (Койбічук та Гришко, 2021), і особливо, в галузі національної освіти України (Цибульник, Накорик і Півторак, 2024); Федотова та Хлань, 2023) та світовій медицині (Фон Геріх та ін., 2024); Кабельо-Колладо та ін., 2024), Фу та ін., 2023), також застосування автоматизованого підходу до формування документації все частіше звертаються фахівці у галузі інформаційних технологій та кібербезпеки (Атаїде та ін., 2023; Аладе, Семюел, 2023). Однією з основних проблем таких систем є забезпечення цілісності та автентичності документів, а також їх автоматизованої генерації, що є ключовими аспектами ефективного документообігу. У цьому контексті важливу роль відіграють процеси асемблювання та реасемблювання документів, що дозволяють зберігати логічну структуру документів та забезпечувати їх коректну обробку на всіх етапах життєвого циклу.

У цій статті основна увага приділяється процесам асемблювання та реасемблювання документів, що є невід'ємною частиною функціонування сучасних електронних систем документообігу. Проаналізовано сучасний стан розвитку цих технологій на прикладі таких систем, як «Дія», «PandaDoc» та «GoogleDocs», проведено порівняльний аналіз їх функціональних можливостей. Водночас, у статті розглядаються методи покращення процесів документообігу через вдосконалення технологічного ланцюжка автоматизованих систем з використанням новітніх програмних засобів для асемблювання та реасемблювання документів.

Загальний результат дослідження спрямований на розробку ефективної автоматизованої системи документообігу, що забезпечує підвищення продуктивності за рахунок вдосконалення ключових механізмів обробки документів, покращення їх захисту та збереження автентичності.

Щороку різноманітні освітні організації та вузи опрацьовують велику кількість запитів у формі документів. Кожен з цих документів повинен відповідати певним стандартам і створюватися вручну. Крім того, на основі цих документів секретарі складають протоколи засідань та витяги після їх розгляду та прийняття рішень на рівні кожної з організаційних інстанцій. Цей підхід призводить до значного навантаження та стає дуже вразливим до бюрократичних помилок, які можуть включати граматичні неточності або помилки під час надсилання документів для подальшої обробки. Усі вищезазначені аргументи стосуються процесу ручної обробки документів. Дослідження цієї теми є актуальною, оскільки відсутність адекватного науково-методичного апарату потребує розвитку напрямку вдосконалення та вивчення методів обробки документів, зокрема із застосуванням методів збереження цілісності та автентичності документів, а також їх автоматизованої генерації. Основна увага зосереджена на процесах асемблювання та реасемблювання документів.

Метою роботи є підвищення рівня відповідності запропонованої автоматизованої системи документообігу основним принципам електронних систем документообігу (ЕСД) за рахунок застосування вдосконалених процедур асемблювання та реасемблювання.

Для досягнення мети необхідно: виконати аналіз сучасних ЕСД; розробити технологічний ланцюжок спеціалізованої автоматизованої системи документообігу (САСД) з описом процесів асемблювання та реасемблювання; перевірити розробку на предмет відповідності основним принципам ЕСД.

**Огляд літератури.** Дослідження у цій галузі активно ведуться багатьма науковцями в Україні та за кордоном, адже питання електронного документообігу стає все більш актуальним, зокрема, процеси створення і генерації електронних файлів в цілому. Загалом, є стійка ідея у сучасному суспільстві, яку в роботі вдало описали Сімона Стернад, Сандра Джордан і Само Бобек, (2023), що цифровізація установ, зокрема, впровадження ЕСД, дозволяє впровадити безпаперовий бізнес, прискорить процеси, знизити бізнес-витрати та підтримувати заходи щодо сталого розвитку організацій. Одним із популярних напрямків

розвитку у сфері ЕСД є застосування хмарних технологій для створення системи керування файлами. У своїй роботі (Цзінь Хань та ін., 2020) пропонує архітектуру системи керування електронними файлами на основі хмарних обчислень для різного роду систем. На основі отриманих результатів у роботі (Авіаска Азміра та Луки Віджаянті, 2022) було виявлено, що можливість хмарних обчислень в електронному документообігу полягає в підвищенні ефективності управління електронними документами та процесами генерації таких документів, тобто, їх збирання (асемблювання) та перезбирання (реасемблювання). Про документообіг у цифровому вигляді, який постійно розвивається у всіх сферах життя і господарювання, пишуть у роботі (Койбічук та Гришко, 2021). У статті (Березного, 2024) розглядається роль хмарних технологій в організації моделей безперервного обліку документації для бухгалтерського обліку, враховуючи зміни традиційного менеджменту шляхом впровадження моделей безперервного обліку завдяки автоматизації таких рутинних завдань. А в галузі медицини електронний документообіг життєво необхідний для ефективного вирішення задач створення, зберігання і генерації документації, що доводять у своїх роботах (Фон Геріх Х. та ін., 2024), (Кабельо-Колладо та ін., 2024), (Фу та ін., 2023).

Українські науковці також роблять значний внесок у розвиток цієї сфери. Зокрема, для розкриття цієї роботи скористаємося дослідженнями науковців, що займаються питаннями електронного документообігу, цифровізації та ЕСД в Україні для закладів освіти. Застосування системи електронного документообігу та інших складових інтегрованої системи управління закладом вищої освіти розглядають у роботі (Копняк та Радзіховська, 2020) і наголошують на практичній цінності створення та функціонування інформаційно-освітнього середовища та його інтеграції з системою обліку навчального процесу та системою моніторингу результатів роботи викладача та студента. Варто зазначити, що розробка і впровадження автоматизованих систем генерації супровідної документації освітнього процесу активно розвиваються у вузах України, зокрема, на прикладі асемблювання силабусів це показують (Цибульник, Накорик і Півторак, 2024) у роботі, що дозволить забезпечити більш гнучку роботу навчальних закладів, створить сприятливе середовище для навчання (в тому числі під час карантинів або дистанційного навчання). В напрямку впровадження електронного документообігу активно працюють (Федотова та Хлань, 2023) і у своїй роботі розвивають ідею про створення єдиного інформаційного простору у межах вузу. Це може бути застосування електронних автоматизованих систем, проте зазначають, що процес переходу закладів вищої освіти на цифровий документообіг гальмується недосконалим семантичним аналізом документів відсутністю систем резервного копіювання, недосконалістю відомчих стандартів, вартістю міграції тощо.

Загалом, існує тенденція до розвитку досліджень, зосереджених на підвищення відповідності окремих автоматизованих систем документообігу основним принципам ЕСД; також актуальними є застосування процедур асемблювання та реасемблювання для генерації окремих документів, груп документів, звітності тощо.

**Матеріали та методи.** Для створення ефективних систем документообігу, зокрема для автоматизованої генерації документів, застосовуються різноманітні ресурси.

Програмні засоби. Для управління документами використовуються спеціалізоване ПЗ, таке як SM SharePoint, Alfresco, або OpenText, які забезпечують зберігання, обробку та управління електронними документами. Для забезпечення доступу до документів, їх зберігання та спільної обробки у режимі реального часу використовуються хмарні технології: Amazon Web Services (AWS), Google Cloud, або Microsoft Azure тощо. Для забезпечення автентичності і безпеки документів використовуються системи електронного підпису: DocuSign або Adobe Sign тощо. Для розпізнавання та обробки тексту використовуються технології оптичного розпізнавання символів, на кшталт ABBYY FineReader (дозволяє автоматично обробляти текст з відсканованих документів і інтегрувати їх у системи документообігу).

Методи автоматизації та автоматизованого контролю. Автоматизація процесів на основі алгоритмів (RPA – Robotic Process Automation) дозволяє використання RPA-систем для автоматизації рутинних завдань (генерація шаблонних документів, сортування документації тощо). Інструменти від компаній UiPath або Automation Anywhere дозволяють дуже ефективно автоматизувати ручні процеси в системах документообігу. Шаблонні системи (Template-based generation) автоматизовують процеси генерації документів на основі закладених раніше спеціалізованих шаблонів і дозволяють швидко генерувати документи на основі заздалегідь визначених структур, зменшуючи кількість ручної праці та помилок. Машинне навчання та штучний інтелект (AI/ML) є незамінними для аналізу великих обсягів документів, розпізнавання тексту, класифікації та автоматизованої генерації нових документів на основі контексту або попередніх шаблонів. Інструменти TensorFlow або OpenAI, допомагають автоматизувати створення складних документів. Технології блокчейну для збереження автентичності документів для забезпечення прозорого та незмінного запису історії змін у документах, що підвищує рівень безпеки. Інтеграція системам управління даними (ERP/CRM) з платформами управління підприємствами або клієнтами, такими як SAP, Salesforce, або Oracle, що дозволяє автоматично отримувати дані для створення документів.

З точки зору архітектурного підходу: клієнт-серверна архітектура – для централізованого управління даними та документами, де сервер обробляє запити та генерує документи, а клієнти мають доступ до необхідних функцій через інтерфейси; мікросервісна архітектура – окремі функції системи документообігу (генерація документів, їх зберігання, перевірка підписів тощо) є незалежними модулями, легко оновлюються та масштабуються.

Алгоритми та стандарти: XML/JSON – для структурування даних; цифрові сертифікати та криптографічні протоколи – для захисту документів та перевірки їх автентичності (SSL/TLS, цифрові сертифікати).

З точки зору кібербезпеки: шифрування даних (AES або RSA); контроль доступу та ролей (Role-Based Access Control, RBAC).

Ці ресурси можуть бути застосовані до створення ефективних систем документообігу, які забезпечують автоматизовану генерацію, безпеку та зручність роботи з документами в різних галузях діяльності.

**Аналіз сучасних систем документообігу.** Сьогодні підприємства мають зосереджувати увагу на операціях з даними. Витрати на бізнес-розвиток та документообіг шляхом мануальної обробки зросли до величезних розмірів. Традиційні методи обробки даних недостатньо ефективні через людський фактор і ускладнюються обробкою великих обсягів даних вручну (Prashant Vilas Kanade & Arun Kumar, 2021; Adanur Dedetürk B. & Bakir-Gungor, 2024).

Електронна система документообігу (ЕСД) – це таке програмне забезпечення, що керує процесом створення, управління та розповсюдженням електронних документів в організації (Hryshyn & Filirova, 2023; Kuts, 2023). Основні принципи такої системи такі:

1. Одноразова реєстрація документа, що забезпечує його унікальну ідентифікацію.
2. Можливість проведення декількох операцій одночасно для прискорення обігу документів.
3. Рух документа з можливістю визначити відповідальну особу на кожному етапі обробки.
4. Єдина база документів для уникнення їх дублювання.
5. Ефективна система пошуку документів, що дозволяє знаходити їх з мінімальною інформацією.
6. Система звітності для контролю руху документів та прийняття управлінських рішень.

Мануальна обробка документів призводить до деяких недоліків: обмежена швидкість введення даних; високі витрати на працю; більше виправлень та переробок.

В табл. 1 подано порівняльний аналіз мануальної та електронної обробки документів. Результати свідчать на користь електронної системи.

**Таблиця 1.** Порівняння мануальної і електронної обробки документів

|                    | Мануальна   | Електронна   |
|--------------------|---|--|
| Початкова вартість | Низька.   | Висока через вартість програмного забезпечення.  |
| Людське втручання  | Вимагає від працівників покрокового опрацювання   | Програми для автоматизованої обробки даних потребують мінімального втручання людини. Вони мають чітко визначені робочі процеси для обробки документів та введення даних. |
| Час                | Трудомісткий процес, і компанії мають знайти відповідних фахівців.                              | За допомогою попередньо навчених API займає кілька хвилин на обробку документів.   |
| Кількість помилок  | Ризик повторних записів, друкарські та граматичні помилки або введення неправильної інформації. | Автоматизоване введення даних може посылатися на досвід попередніх моделей даних і перевіряти введenu інформацію   |
| Обробка            | Аналіз даних мануально настільки ж громіздкий, як і їх обробка.                                 | Технологія автоматичного вилучення даних швидко читає, упорядковує та аналізує дані, що зберігаються.  |

**Джерело:** розроблено автором на основі (Korobeinikova, 2024).

Яскравим прикладом ЕСД є "Дія", і надає доступ до онлайн-послуг отримання електронного паспорту, водійського посвідчення, автоцивілки, студентського квитка тощо (Hryshyn & Filipova, 2023). Зараз «Дія» функціонує як цифрова сховище для документів користувача, які офіційно визнаються аналогами паперових або пластикових посвідчень. Сьогодні сервіс «Дія» пропонує понад 20 онлайн-послуг для громадян та підприємств. Для користування ними потрібно зареєструватися за допомогою кваліфікованого електронного підпису (Kuts, 2023). Більшість послуг, запропонованих у «Дії», реалізуються через сайти державних органів, що вимагає додаткової авторизації.

ЕСД «PandaDoc». PandaDoc – це інтернет-сервіс для створення та управління електронними документами, де можна створювати документи за допомогою вбудованого редактора, підписувати їх електронно та надсилати на затвердження (Nakoryk, 2024). Цей редактор спрямований на створення документів за допомогою шаблонів або з нуля і дозволяє створювати як офіційні документи, так і маркетингові матеріали. Містить включають функції створення, відстеження та виконання документів, а також можливості електронних підписів; інтегрується з платіжними системами, хмарними сховищами тощо. Дозволяє створювати пропозиції та угоди шаблонами з фірмовим стилем, темами та параметрами персоналізації. Дозволяє налаштовувати документи онлайн; інтегрується із службами Google, Dropbox, Vox та іншими, сумісна з Google Docs (Kuts, 2023).

ЕСД Google Docs – це онлайн-редактор для створення та форматування текстових документів, дозволяє, ділитися ними з колегами та спільно редагувати. Документи користувача зберігаються в хмарі, що забезпечує доступ до них з різних пристроїв. Google Docs має редактор тексту з повним набором функцій, який можна використовувати як у веб-браузері, так і на мобільних пристроях. Додатково, для розширення можливостей Google Docs можна поєднати з різними веб-додатками. Особливості Google Docs включають створення та редагування документів, спільне редагування в реальному часі, надсилання запрошень користувачам через посилання або електронну пошту, перегляд історії змін, роботу з форматом Word, а також можливість завантаження та переклад документів у різні мови (Vasylovskiy, 2023).

Основними перевагами системи електронного документообігу "Дія" є:

1. Зберігання офіційно визнаних цифрових копій документів в компактному форматі.
2. Можливість отримання державних послуг в режимі онлайн на різних рівнях.

Основними недоліками цієї системи є:

1. Складний процес авторизації, який передбачає ідентифікацію через технологію BankID, що може викликати незручності.

2. Неповне відображення всіх документів, що стає причиною скарг користувачів.

3. Неможливість підтвердження аутентичності документів у відсутності доступу до Інтернету та відсутності QR-кодів у офлайн-режимі.

Щодо системи електронного документообігу "GoogleDocs", її ключовими перевагами є:

1. Безкоштовність сервісу.

2. Зберігання всіх даних у хмарі, що забезпечує захист і доступність інформації.

3. Простий обмін файлами та можливість колективної роботи.

4. Зручна історія змін файлів та кросплатформеність.

Але є і недоліки:

1. Повільна робота системи, особливо з об'ємними текстами при поганому Інтернет-з'єднанні.

2. Залежність від облікового запису та наявності доступу до Інтернету.

Щодо системи "PandaDoc", її переваги полягають у такому:

1. Інтеграція з різними сервісами, включаючи Google, Dropbox і Box.

2. Можливість завантажувати та налаштовувати файли будь-якого типу.

3. Використання юридично обов'язкових електронних підписів та співпраця в режимі real-time.

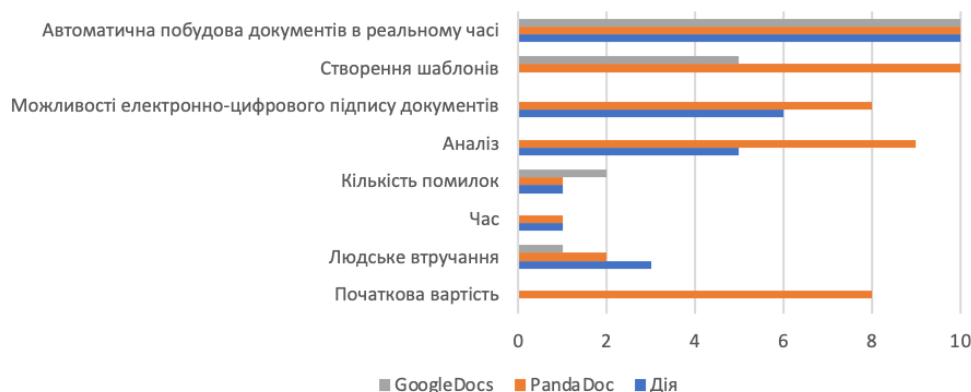
Але існують такі недоліки:

1. Відсутність системи контролю версій та погане форматування шаблонів.

2. Висока ціна сервісу.

За результатами наукового пошуку пропонується порівняння розглянутих ЕСД за такими критеріями: базова ціна, людський фактор, часові витрати, рівень похибок, оцінка, застосування ЕЦП, формування шаблонів, автоматичне формування документів. Оцінимо кожен із розглянутих систем за десятибальною шкалою і створимо візуалізацію отриманих даних (рис. 1).

Виходячи із огляду переваг та недоліків – лідером є «GoogleDocs», оскільки має простий і доступний інтерфейс, є безкоштовним, і підтримує версії документів.



**Рисунок 1.** Порівняння розглянутих ЕСД

Джерело: розроблено автором на основі відкритих джерел

**Технологічний ланцюжок спеціалізованої автоматизованої системи документообігу.** Дано визначення, що спеціалізованою автоматизованою системою документообігу (САСД) – це така система для ведення документообігу, базовою метою якої є автоматизація процесів оброблення документів та всіх процесів, що виконуються вручну (рис. 2). Підвищення ефективності обробки документів шляхом зменшення недоліків мануального підходу, використовуючи сучасні хмарні технології та найявні засоби парсингу

для автоматичного формування нових документів та швидкого отримання інформації, що програмно обробляється.

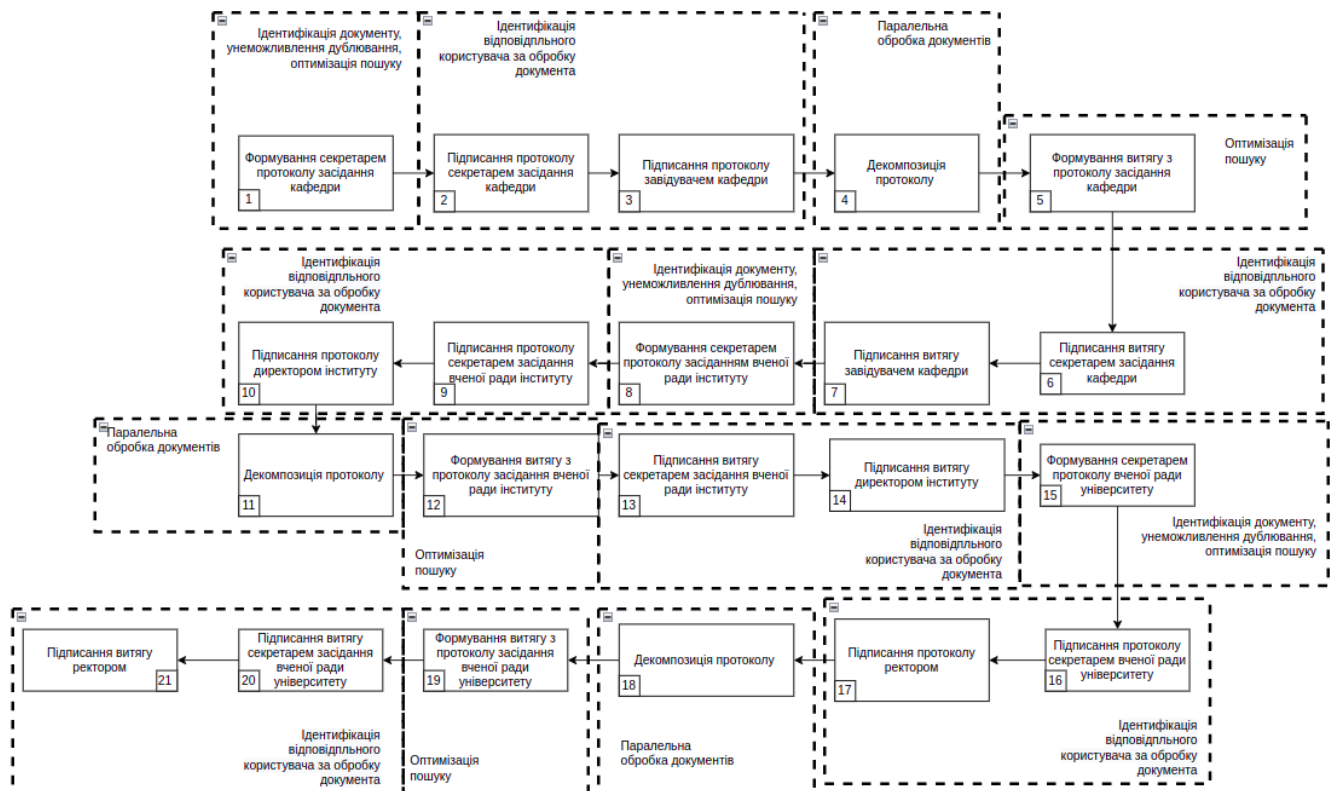


Рисунок 2. Технологічний ланцюжок САСД

Джерело: розроблено авторами

1) На першому етапі секретар формує протокол засідання, обираючи документи для розгляду. Після відбору дані автоматично обробляються (асемблюється готовий файл) протоколу засідання, зберігаються у хмарному середовищі та надсилаються для підпису. Доступ до протоколу мають авторизовані особи.

2) Після процесу формування протоколу секретар кафедри переглядає доступні документи та підписує необхідні. Його підпис зберігається в базі даних для подальшої верифікації.

3) Також завідувач кафедрою має доступ до документів та підписує потрібні. Його підпис також зберігається в базі даних для подальшої верифікації.

4) Під час декомпозиції протоколу кафедри запропонована система перевіряє наявність підписів секретаря та завідувача в базі даних для підтвердження забезпечення цілісності та автентичності документів. Після цього алгоритм виконує парсинг файлів з маркованими блоками (порядок денний, список присутніх та рішення) для формування витягів.

5) Система створює витяг та переносить дані з певного блоку протоколу. Новий файл знаходиться в хмарному середовищі та надсилається завідувачу та секретарю на підпис.

6) Також секретар переглядає надіслані витяги та підписує необхідні. Створений підпис зберігається в базі даних для можливості верифікації.

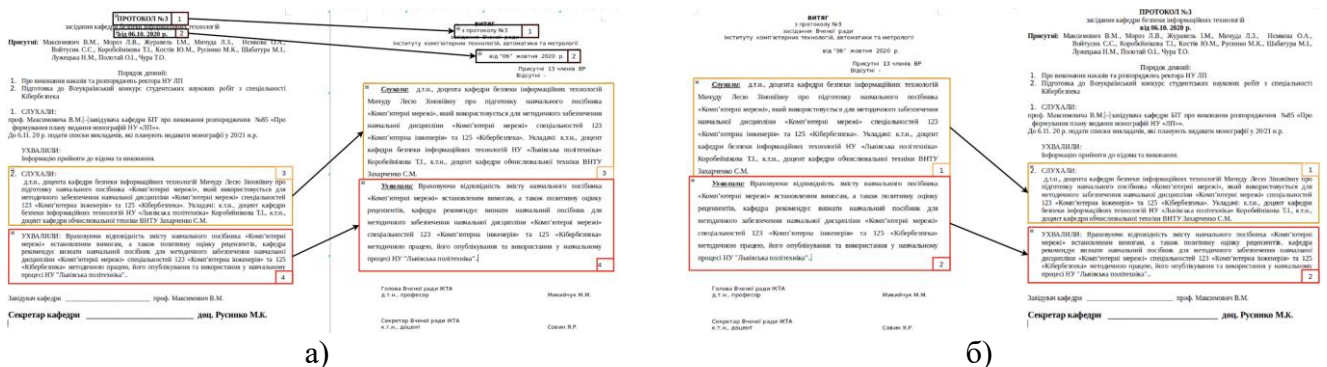
7) Завідувач кафедри переглядає витяги та підписує потрібні. Його підпис також зберігається в базі даних з метою верифікації.

8) На кроці формування протоколу для засідання вченої ради інституту секретар вибирає документи, які будуть розглядатись на засіданні та надсилає їх системі. Система автоматично формує протокол, завантажує його в хмару та надсилає секретареві та директору інституту для підпису.

9) Після створення протоколу секретар переглядає документи та підписує необхідні. Підпис зберігається в базі даних для подальшої верифікації.

- 10) Директор інституту після формування протоколу переглядає документи та підписує потрібні. Його підпис також зберігається в базі даних для можливості верифікації.
- 11) На етапі декомпозиції протоколу перевіряються підписи секретаря та директора для верифікації цілісності та автентичності. Коли перевірка виконана алгоритм парсить файл з маркованими блоками для формування витягів.
- 12) Під час формування витягу система створює витяг та заповнює дані з певного блоку протоколу як було зазначено на попередньому кроці. Файл зберігається в хмарі та надсилається секретареві та директору для підпису.
- 13) Після формування витягів з протоколу секретар переглядає витяги та підписує необхідні. Підпис зберігається в базі даних.
- 14) Директор інституту переглядає сформовані системою витяги та підписує потрібні. Його підпис зберігається в базі даних для верифікації.
- 15) При формуванні протоколів засідання Вченої ради університету секретар обирає документи які будуть розглядатись Вченою ради та надсилає їх системі. Система асемблює протокол засідання Вченої ради університету, завантажує його в хмару та надсилає уповноваженим особам для підпису. Розміщений в хмарному середовищі документ авторизовані особи отримують доступ до нього, а також можуть завантажити на комп'ютер для перегляду.
- 16) Після створення протоколу секретар переглядає та підписує необхідні документи. Підпис зберігається в базі даних для верифікації.
- 17) Також ректор університету після створення файлу протоколу переглядає необхідні документи та підписує їх. Його підпис також зберігається в базі даних для верифікації.
- 18) Під час декомпозиції протоколу системою перевіряються наявність підпису секретаря та ректора для верифікації автентичності та цілісності. Алгоритм парсить файл з маркованими блоками для формування витягів.
- 19) При формуванні протоколу система створює витяг та переносить дані з певного блоку протоколу. Файл зберігається в хмарі та надсилається секретареві та ректорові для підпису.
- 20) Після формування витягів протоколів секретар переглядає необхідні витяги та підписує їх. Підпис секретаря зберігається в базі даних подальшої верифікації документів..
- 21) Ректор університету після формування витягів переглядає та підписує потрібні. Його підпис зберігається в базі даних із метою подальшої верифікації підписаних документів.

**Процеси асемблювання та реасемблювання.** Процеси реасемблювання полягають у розподілі документа на окремі частини за допомогою спеціальних маркерів і регулярних виразів, що дозволяє подальше програмне опрацювання цих фрагментів. Реасемблювання передбачає перетворення документа на програмні об'єкти, які можуть оброблятися і змінюватися системою. Візуалізацію процесу реасемблювання представлено на рисунку 3, а.



**Рисунок 3.** Візуалізація процесів реасемблювання (а) та асемблювання (б)

Джерело: розроблено автором

Під час реасемблювання система автоматично розбиває протокол на складові частини для формування витягів. Зокрема, протокол ділиться на такі чотири частини:

- 1) Заголовок протоколу, що містить його номер. Ця інформація використовується для відображення номера у верхній частині витягу, що забезпечує подальшу ідентифікацію протоколу, з якого витяг сформовано.
- 2) Блок з датою проведення засідання, який пізніше вказується у верхній частині витягу як дата документу.
- 3) Опис питання, яке розглядалося на засіданні і є основою для формування витягу. Програма розміщує цей опис одразу під заголовком витягу, ближче до верхньої частини документа.
- 4) Прийняте рішення щодо розглянутого питання, розташованого в попередньому блоці. Це рішення автоматично розміщується у витягу під відповідним питанням, яке було обговорено.

Процес асемблювання, що є протилежним до реасемблювання, перетворює програмний об'єкт у фізичний документ для подальшої обробки. Для асемблювання зазвичай застосовується пустий шаблон, який програма заповнює потрібними даними. У результаті створюється документ, що об'єднує наявні дані та об'єкти програми. Візуалізацію процесу асемблювання наведено на рисунку 3,б.

На етапі асемблювання система вибирає витяги з питаннями, які слід включити до протоколу засідання, поступово додаючи їх до документа. Ось як програмно розкладаються витяги для перенесення до протоколу:

- 1) Перший блок у витягу – це опис питання, яке обговорювалося на засіданні. Він додається до протоколу відповідно до порядку денного.
- 2) Другий блок містить рішення, ухвалене щодо зазначеного питання. У протоколі це рішення розміщується одразу після відповідного питання

## Результати

**Аналітичний опис процедур асемблювання та реасемблювання.** Розглядається типова структура протоколу засідання кафедри (вважається, що це структурний підрозділ 1-го рівня СП I) як множини  $Pr^I$  складових цього протоколу (1):

$$Pr^I = \{N^I, SP^I, D^I, Pres^I, Agenda^I, Sign^I\}, \quad (1)$$

де  $N^I$  – ім'я документу, тут це «протокол засідання кафедри».

$SP^I$  – назва структурного підрозділу першого рівня СП<sup>I</sup>.

$D^I$  – дата засідання СП<sup>I</sup>.

$Pres^I$  – список присутніх СП<sup>I</sup>.

$Agenda^I$  – порядок денний СП<sup>I</sup>.

$Sign^I$  – підписанти СП<sup>I</sup>.

Список членів колективу структурного підрозділу рівня 1 СП<sup>I</sup> визначається множиною  $Pres^I$  (2), міститиме ПІБ працівників, де  $E_1^I, E_2^I$  – ПІБ головуючого та секретаря відповідно.

$$Pres^I = \{E_1^I, E_2^I, E_3^I, \dots, E_N^I\} \quad (2)$$

Для кожного працівника передбачено вказання його регалій  $R$ , як матрицю впорядкованих складових, що міститиме значення ПІБ  $E_N^I$ , посаду  $Pos_P^I$  та науковий ступінь  $Sc_S^I$  (3).

$$R^I = \begin{pmatrix} E_1^I & Pos_1^I & Sc_1^I \\ \dots & \dots & \dots \\ E_N^I & Pos_P^I & Sc_S^I \end{pmatrix} \quad (3)$$

Опишемо визначений порядок денний  $Agenda^I$  як матрицю впорядкованих елементів, що міститиме інформацію щодо заслуханих питань «слухали»  $Q_M^I$  та інформацію щодо відповідних ухвал «ухвалили»  $D_M^I$  (4).

$$Agenda^I = \begin{pmatrix} Q_1^I & D_1^I \\ \dots & \dots \\ Q_M^I & D_M^I \end{pmatrix} \quad (4)$$

Опишемо множину підписантів структурного підрозділу першого рівня  $СП^I$ , як  $Sign^I$ , що міститиме значення цифрових підписів (5).

$$Sign^I = \{DS_1^I, DS_2^I\} \quad (5)$$

За ідеєю власне декомпозиції відбувається процедура реасемблювання документу (тобто, протоколу засідання)  $СП^I$ , і потім згідно процедури асемблювання у витяг з протоколу  $СП^I$   $Pr_{EX}^I$  надійдуть елементи  $Pr^I$ , що допоможуть САСД сформувати витяг:  $SP^I, D^I, Q_m^I, D_m^I, DS_1^I, DS_2^I, E_1^I, Pos_1^I, Sc_1^I, E_2^I, Pos_2^I, Sc_2^I$ . Опишемо структуру витягу протоколу засідання  $СП^I$  як множину  $Ex^I$  складових (6):

$$Ex^I = \{Ex_{name}^I, SP^I, D^I, Q_m^I, D_m^I, \langle E_1^I, D_1^I \rangle, \langle E_2^I, D_2^I \rangle\} \quad (6)$$

де

$Ex_{name}^I$  – ім'я документу, тут це «витяг з протоколу засідання кафедри»;

$SP^I$  – назва структурного підрозділу першого рівня  $СП^I$ ;

$D^I$  – дата засідання;

$Q_m^I, D_m^I$  – інформація про заслухане питання, що винесене у витяг і його ухвала;

$\langle E_1^I, D_1^I \rangle$  та  $\langle E_2^I, D_2^I \rangle$  – кортежі, що містять цифрові підписи головууючого та секретаря  $СП^I$ , також їх ППБ, регалії тощо.

Опишемо загальну структуру протоколу засідання Вченої ради інституту або факультету (структурного підрозділу 2-го рівня  $СП^{II}$ ) як множину  $Pr^{II}$  складових цього документу (7):

$$Pr^{II} = \{N^{II}, SP^{II}, D^{II}, Pres^{II}, Agenda^{II}, Sign^{II}\} \quad (7)$$

де

$N^{II}$  – ім'я документу, тут це «протокол засідання ВРІ».

$SP^{II}$  – назва структурного підрозділу другого рівня  $СП^{II}$ .

$D^{II}$  – дата засідання  $СП^{II}$ .

$Pres^{II}$  – список присутніх  $СП^{II}$ .

$Agenda^{II}$  – порядок денний  $СП^{II}$ .

$Sign^{II}$  – підписанти  $СП^{II}$ .

За аналогією з (2) та (3), список присутніх  $Pres^{II}$  засідання ВРІ та регалії учасників опишемо (8):

$$Pres^{II} = \{E_1^{II}, E_2^{II}, E_3^{II}, \dots, E_T^{II}\} \text{ та } R^{II} = \begin{pmatrix} E_1^{II} & Pos_1^{II} & Sc_1^{II} \\ \dots & \dots & \dots \\ E_U^{II} & Pos_V^{II} & Sc_W^{II} \end{pmatrix} \quad (8)$$

Порядок денний тут (за процедурою асемблювання) може формуватися із питань, розглянутих на засіданнях довільної кількості різних  $СП^I$  і доповнюватися питаннями, доданих секретарем засідання ВРІ в ручному режимі, тоді отримуємо (9):

$$Agenda^{II} = \begin{pmatrix} Q_1^I & D_1^I \\ \dots & \dots \\ Q_A^I & D_A^I \end{pmatrix} \times \begin{pmatrix} Q_1^{II} & D_1^{II} \\ \dots & \dots \\ Q_B^{II} & D_B^{II} \end{pmatrix} \quad (9)$$

Аналогічно з (5) визначаємо множину підписантів  $СП^{II}$ , як  $Sign^{II}$ , що оперує значеннями цифрових підписів секретаря та головуєчого ВРІ відповідно (10).

$$Sign^{II} = \{DS_1^{II}, DS_2^{II}\} \quad (10)$$

Далі, за основною ідеєю декомпозиції виконується процедура реасемблювання документу протоколу  $СП^{II}$ , і тоді відповідно процедурі асемблювання у підсумковий витяг  $СП^I Pr_{EX}^{II}$  надійдуть такі елементи  $Pr^{II}$ , що допоможуть запропонованій авторами САСД сформувавши витяг з протоколу засідання ВРІ:  $SP^{II}, D^{II}, Q_a^I, D_a^I, Q_b^I, D_b^I, DS_1^{II}, DS_2^{II}, E_1^{II}, Pos_1^{II}, Sc_1^{II}, E_2^{II}, Pos_2^{II}, Sc_2^{II}$ . Подано структуру витягу з протоколу засідання  $СП^{II}$  множиною  $Ex^{II}$  з такими складовими (11):

$$Ex^{II} = \{Ex_{name}^{II}, SP^{II}, D^{II}, Q_x^{II}, D_x^{II}, \langle E_1^{II}, D_1^{II} \rangle, \langle E_2^{II}, D_2^{II} \rangle\} \quad (11)$$

де

$Ex_{name}^{II}$  – назва документу, а саме «витяг з протоколу засідання ВРІ».

$SP^{II}$  – назва структурного підрозділу другого рівня  $СП^{II}$ .

$D^{II}$  – дата засідання.

$Q_x^{II}, D_x^{II}$  – значення питання, що винесене у витяг ВРІ і його відповідна ухвала.

$\langle E_1^{II}, D_1^{II} \rangle$  та  $\langle E_2^{II}, D_2^{II} \rangle$  – кортежі, що містять ППБ, регалії та цифрові підписи головуєчого та секретаря  $СП^{II}$  відповідно.

Тоді структуру протоколу засідання Вченої ради Університету (ВРУ, структурного підрозділу другого рівня  $СП^{III}$ ) можна описати як множину  $Pr^{III}$  складових цього документу (12):

$$Pr^{III} = \{N^{III}, SP^{III}, D^{III}, Pres^{III}, Agenda^{III}, Sign^{III}\}, \quad (12)$$

де

$N^{III}$  – ім'я документу, тут це «протокол засідання ВРУ».

$SP^{III}$  – назва структурного підрозділу другого рівня  $СП^{III}$ .

$D^{III}$  – дата засідання  $СП^{III}$ .

$Pres^{III}$  – список присутніх  $СП^{III}$ .

$Agenda^{III}$  – порядок денний  $СП^{III}$ .

$Sign^{III}$  – підписанти  $СП^{III}$ .

Тоді витяг з протоколу засідання ВРУ можна описати як (13)

$$Ex^{III} = \{Ex_{name}^{III}, SP^{III}, D^{III}, Q_y^{III}, D_y^{III}, \langle E_1^{III}, D_1^{III} \rangle, \langle E_2^{III}, D_2^{III} \rangle\} \quad (13)$$

де

$Ex_{name}^{III}$  – ім'я документу, тут це «витяг з протоколу засідання ВРУ»;

$SP^{III}$  – назва структурного підрозділу 3-го рівня (університету)  $СП^{III}$ ;

$D^{III}$  – дата засідання ВРУ;

$Q_y^{III}, D_y^{III}$  – значення заслуханого питання, що винесене у витяг ВРУ і його відповідна ухвала;

$\langle E_1^{III}, D_1^{III} \rangle$  та  $\langle E_2^{III}, D_2^{III} \rangle$  – кортежі, що містять ПБ, регалії та цифрові підписи ректора та секретаря ВРУ відповідно.

**Обговорення.** У запропонованій аналітичній частині цієї роботи основні принципи електронної системи документообігу (ЕСД) відображено у технологічному ланцюжку спеціалізованої автоматизованої системи документообігу (САСД) (рисунок 2). Принципи ЕСД реалізовані через виконання таких вимог:

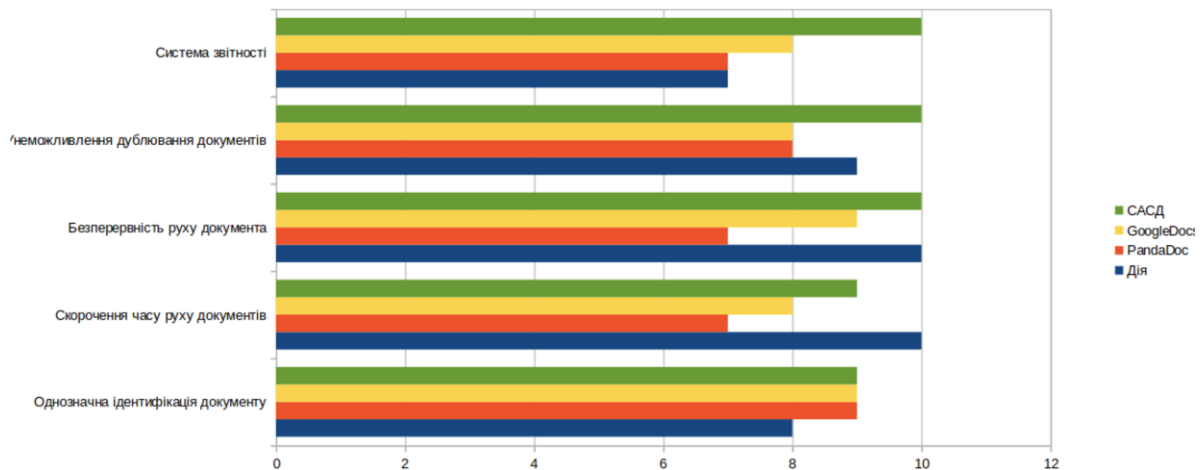
1) Система забезпечує унікальну ідентифікацію кожного документа за допомогою присвоєння унікального ID. Це значення є рядковим полем з 12 символів, що може містити цифри та літери латинського алфавіту.

2) Автоматизація процесів генерації протоколів та витягів з них дозволила зменшити час обробки документів та підвищити оперативність їх виконання.

3) Принцип безперервності документообігу реалізовано завдяки автоматизованому надсиланню документів до відповідних департаментів для обробки та отримання підписів.

4) Механізм запобігання дублюванню документів реалізовано через алгоритм хешування, що генерує унікальну строку для кожного документа. Цей набір символів є специфічним для кожного документа з конкретним вмістом, що гарантує відсутність дублікатів.

5) Функціонал звітності включає автоматичне накладання підписів секретарів та керівників засідань. Підпис генерується за допомогою алгоритму асиметричного шифрування RSA, а кожен користувач має приватний ключ, сформований на основі ID та секретної строки, що зберігається на сервері. Цей підпис підтверджує автентичність та цілісність документа, а також забезпечує можливість перевірки підпису авторами.



**Рисунок 4.** Візуалізація відповідності САСД основним принципам ЕСД

Джерело: розроблено авторами

Дані на рис. 4 свідчать, що запропонована САСД відповідає принципам ЕСД на 8% краще за найближчий аналог. Такого результату вдалося досягти завдяки впровадженню механізму запобігання дублюванню документів. Система унікальності реалізована для уникнення повторного розгляду однакових документів, що зберігає час на засіданнях. Перевірка унікальності документів здійснюється шляхом хешування їх вмісту алгоритмом SHA-256. Результат хешування, унікальна строка, зберігається в базі даних та прив'язується до документа. Загалом, результати дослідження підтверджують відповідність САСД основним принципам ЕСД.

**Висновки.** У цій роботі застосовано системний підхід для вирішення задач асемблювання та реасемблювання документів у спеціалізованій системі документообігу, зосередившись на структурних підрозділах закладу вищої освіти. Основна увага приділена аналізу та вдосконаленню методів обробки документів, зокрема методів забезпечення цілісності та автентичності документів, а також їх автоматичної генерації.

В результаті проведених досліджень вдалося підвищити рівень відповідності автоматизованої системи документообігу основним принципам електронних систем документообігу на 8%. Це стало можливим завдяки автоматизації процесів асемблювання та реасемблювання документів, а також поліпшенню підходів до збереження цілісності та автентичності документів.

У роботі виконано такі завдання: проведено огляд сучасних електронних систем документообігу, проаналізовано їх переваги та недоліки; розроблено технологічний ланцюжок спеціалізованої автоматизованої системи документообігу; побудовано аналітичний опис технологічного ланцюга САСД; перевірено відповідність розробки основним принципам ЕСД.

**Подяки** «Немає».

**Конфлікт інтересів** «Немає».

## References

- Adanur Dedetürk B. & Bakir-Güngör B. (2024). Aguhyper: a hyperledger-based electronic health record management framework. *PeerJ Computer Science*10:e2060.
- Alade, S. M. (2023). Design and implementation of a web-based document management system. *International Journal of Information Technology and Computer Science*, 15(2), 35–53.
- Ataide, M., Brahollı, O., & Siegele, D. (2023). Digital transformation of building permits: Current status, maturity, and future prospects. *Buildings*, 13, 2554.
- Azmir, A., & Wijayanti, L. (2022). Cloud computing opportunities and challenges in electronic document management. *Record and Library Journal*, 8, 248–258.
- Bereznyi, O. (2024). The role of cloud technologies in the organization of continuous accounting. *Economics, Management and Administration*, 2(108), 78–83. [https://doi.org/10.26642/ema-2024-2\(108\)-78-83](https://doi.org/10.26642/ema-2024-2(108)-78-83).
- Brad Hershowitz, Melinda Hodkiewicz, Tyler Bikaun, Michael Stewart & Wei Liu (2024). Causal knowledge extraction from long text maintenance documents. *Computers in Industry* (161).
- Cabello-Collado, C., Rodriguez-Juan, J., Ortiz-Perez, D., Garcia-Rodriguez, J., Tomás, D., & Vizcaya-Moreno, M. F. (2024). Automated generation of clinical reports using sensing technologies with deep learning techniques. *Sensors*, 24, 2751.
- Diyan V. R. (2024). *Methods of designing a university document management system*: master's thesis, specialty 123 Computer Engineering, KhNUof Radio Electronics, Kharkiv.
- Fedotova, O. O., & Khlan, A. S. (2023). Ways to optimize the clerical service in higher education institutions of Ukraine. In *Development of Education, Science and Business: Results*, Proceedings of the International Scientific and Practical Internet Conference, December 21-22, 2023. FOP Marenichenko V.V., Dnipro, Ukraine 222.

- Fu, Q., Lai, J., Zhong, T., et al. (2023). Design and implementation of clinical LIS360 laboratory management system based on AI technology. *International Journal of Computational Intelligence Systems*, 16, 33.
- Han, J., Wang, C., Miao, J., Lu, M., Wang, Y., & Jin, S. (2020). Research on electronic document management system based on cloud computing. *Computers, Materials & Continua*, 66, 2645–2654.
- Hanzha, A. S., Antonenko, S. V., & Izmailova, M. K. (2022). A review of existing automated management systems for educational institutions. *Actual problems of automation and information technology*, 26, . Oles Honchar Dnipropetrovsk National University. <https://doi.org/10.15421/432205>. [in Ukrainian]
- Hryshyn, V., & Filipova, L. (2023). Trends in the Development of Digitalisation of Document Management Systems and Services in Public Authorities. *Scientific journal “Library Science. Record Studies. Informology”*, Issue 3. <https://doi.org/10.32461/2409-9805.3.2023.290988>
- Jingqi Zhang & Shaohua Jiang. Systematic Literature Review on Knowledge Graphs in Construction Management from a Multi-Modal Perspective. *International Journal of Human–Computer Interaction* 0:0, 1-20.
- Koibichuk, V. V., & Grishko, R. A. (2021). *Automated enterprise management system*. Doctoral dissertation, Sheffield, Science and Education LTD.
- Kopnyak, K., & Radzikhovska, L. (2020). Integrated automated management system of a higher education institution. *Scientific of Academic Society of Michal Baludansky*, 8(3), 101–103.
- Korobeinikova, T. (2022). Analysis of modern document systems. In *Modern systems of science and education in the USA, EU and other countries*, 19–22.
- Korobeinikova, T. (2024). Modern solutions for assembling and reassembling documents in a specialized document management system. In *Science for modern man: Computer science, cybernetics and automation, security systems, physics and mathematics* (Monographic series «European Science», Book 26, Part 2, Chapter 4, pp. 84–114). European Science. ISBN 978-3-98924-032-2.
- Kuts, V. (2023). Technology of using digital signature in electronic document management systems.
- Lobbé, Q., Delanoë, A., & Chavalarias, D. (2022). Exploring, browsing and interacting with multi-level and multi-scale dynamics of knowledge. *Information Visualization*, 21(1), 17-37.
- Nakoryk, V. V. (2024). *Automated system for creating accompanying documents of the educational process* : Master's thesis : 151 Automation and computer-integrated technologies. Kyiv.
- Prashant Vilas Kanade & Arun Kumar (2021). Security Aspects of Electronic Health Records and Possible Solutions. *International Journal of Advanced Computer Science and Applications(IJACSA)*, 12(4).
- Sternad, S., Jordan, S., & Bobek, S. (2023). Managing document management systems' life cycle in relation to an organization's maturity for digital transformation. *Sustainability*, 15, 15212.
- Tsybulnyk, S., Nakoryk, V., & Pivtorak, D. (2024). Development of the prototype of the automated system for creating accompanying documents of the educational process. *Bulletin of the National Technical University «KhPI» Series: New Solutions in Modern Technologies*, 2(20), 30–37. <https://doi.org/10.20998/2413-4295.2024.02.05>
- Vasylovskiy I. K. (2023). *Information system of corporate document management*. Thesis: 126 Information systems and technologies; KhNU- Khmelnytskyi.
- Von Gerich, H., Lozada-Perezmitre, E., Pruinelli, L., & Peltonen, L.-M. (2024). Towards automated evaluation of patient-centered care: Assessing the potential of electronic health records. *Studies in Health Technology and Informatics*, 310, 344–348.

## КОМП'ЮТЕРНІ СИСТЕМИ ТА КОМПОНЕНТИ

UDC 004.85

### Model-based learning of coordinators of the decentralized multi-zone objects control systems

**Volodymyr M. Dubovoi**

Doctor of Science (Engineering), Professor,  
Professor of Computer Control Systems Department  
Vinnytsia National Technical University, Vinnytsia  
ORCID 0000-0003-0440-3643

**Abstract.** Decentralized control systems are gaining more and more expansion, which is due to the increase in the availability and power of microcontrollers. Decentralized control of multi-zone objects is associated with the need to coordinate the local control systems of zones state. Learning systems are preferred for implementation of the coordination methods, as they are able for flexibly adjust to the specifics of control of each zone. However, the training of coordinators is complicated task by the absence at the stage of a system creating of marked datasets for controlled multi-zonal objects. This article considers the creation of a dataset based on a simulation of a decentralized system and four scenarios for training neural coordinators. A model for simulation of a decentralized system was been created on the Scilab/Xcos platform using a pre-built library of blocks for simulating decentralized systems. The scenarios differ depending on the structure of the neural coordinators: a segmented network according to the structure of the coordinator simulation model or an integrated one, as well as on the training strategy: train all the coordinators of the decentralized system in parallel or only one coordinator and then clone the results. Experimental studies of the proposed method of training neural network coordinators, implemented on Python TensorFlow, were conducted. The study showed greater effectiveness of segmented coordinators parallel training. However, in the course of the study, the last step of the scenarios – fine tuning on a real physical object, was not performed. A preliminary evaluation suggests that after such additional training, the advantages of mono-neural coordinators will become more visible, since such additional training will correct the shortcomings of imitation.

**Key words:** machine learning, distributed control system, decentralized coordination, model-based learning.

**Модельно-орієнтоване навчання координаторів децентралізованої системи управління багатозональним об'єктом**

**Володимир Михайлович Дубовой**

доктор технічних наук, професор,  
професор кафедри комп'ютерних систем управління  
Вінницький національний технічний університет, Вінниця

**Анотація.** Децентралізовані системи керування набирають все більшого розповсюдження, що зумовлено збільшенням доступності і потужності мікроконтролерів. Децентралізоване керування багатозональними об'єктами пов'язане з необхідністю координації локальних систем керування станом зон. Для реалізації методів координації перевагу мають системи,

що навчаються, оскільки вони здатні гнучко налаштовуватися на особливості керування кожною зоною. Проте навчання координаторів ускладнюється відсутністю на стадії створення системи розмічених датасетів для керованих багатозональних об'єктів. У цій статті розглядається створення датасету на основі імітаційної моделі децентралізованої системи і чотири сценарії навчання нейронних координаторів. Імітаційна модель децентралізованої системи створена на платформі Scilab/Xcos з використанням попередньо створеної бібліотеки блоків для моделювання децентралізованих систем. Сценарії відрізняються залежно від структури нейронних координаторів: сегментована мережа відповідно до структури імітаційної моделі координатора або інтегрована мережа, а також від стратегії навчання: навчати паралельно усі координатори децентралізованої системи або тільки один і результати клонувати. Проведені експериментальні дослідження запропонованого методу навчання нейромережових координаторів, реалізованих на Python TensorFlow. Дослідження показало більшу ефективність паралельного навчання сегментованих координаторів. Проте в ході дослідження не виконувався останній етап сценаріїв – донавчання на реальному фізичному об'єкті. Попередня оцінка дозволяє припустити, що після такого донавчання переваги інтегрованих нейронних координаторів стануть помітнішими, оскільки таке донавчання дозволить виправити недоліки імітації.

**Ключові слова:** машинне навчання, розподілена система керування, децентралізована координація, модельно-орієнтоване навчання.

**DOI:** <https://doi.org/10.31649/1999-9941-2024-60-2-66-76>

**Introduction.** Decentralized control systems (DCS) are becoming more and more widespread, which is due to the increase in the availability and power of microcontrollers. Decentralized control of multi-zone objects (MZO) is associated with the need to coordinate local systems for control the state of zones. The peculiarity of multi-zone objects is the mutual influence of zones. As a result, their control becomes much more complicated, since the zone control algorithm must take into account the consequences not only for the controlled zone, but also for other zones of the multi-zone object. There are also other tasks within the framework of the problem of decentralized control of MZO: ensuring the stability of a multi-connected system that arises as a result of the interaction of MZO zones and local control systems (LCS); ensuring the reliability and safety of the DCS, which may decrease due to failures of elements, communication means and coordinators, etc.; ensuring the necessary speed of action, which is limited by the speed of the spread of impacts between MZO zones, which requires high-quality forecasting of processes, etc.

**State-of-the-Art.** Classical approaches to coordination are based on a centralized (for a small number of LCS and a small distance between them) or a hierarchical (for a large number of LCS or a large distance between them) architecture of the coordination system. These approaches were developed mainly for organizational management systems: organization management, military management, etc. For the coordination of local control systems, these principles began to be applied with the development of technological automation (Katrenko & Savka, 2008; Ladanyuk et al., 2012). However, such systems have a rigid structure of connections and are difficult to scale. This makes it difficult to use them in objects with frequent and rapid changes in requirements.

The decrease in the cost of automation tools, the development of the Internet of Things (IoT), and the exponential growth of the microcontrollers power have opened up a promising way to solve the problem: the use of decentralized coordination with smart coordinators in each local control system.

In order to investigate the phenomenon of coordination, a consortium comprising 12 European

research centers was established, tasked with conducting the Project Control for Coordination of Distributed Systems (CON4COORD or C4C) (Schuppen & Villa, 2015). An insightful review and analysis of distributed coordination systems is expounded in (Ge et al., 2020). These inquiries primarily focus on hierarchical systems, with particular emphasis on the allocation of functions across different levels.

Decentralized systems are also the subject of research. A feature of distributed decentralized systems is the significant uncertainty of subsystem interaction parameters, the incompleteness of the system, and the lack of complete information about the state of other subsystems that are in direct communication with a separate subsystem. The article (Bakule, 2008) reviews past and current results in the field of decentralized control of large-scale complex systems.

Depending on system typology, the responsibilities associated with decentralized system control encompass synchronization, decentralized stabilization, single-level coordination, peer control, among others (Gong & Aldeen, 1997). This spectrum of inquiry spans linear and nonlinear systems, continuous and discrete domains, and encompasses optimal and adaptive control paradigms, robust methodologies, and systems featuring elements of artificial intelligence (Shaikh et al., 2014).

The development of "small" energy, the appearance of a large number of solar wind and other power plants made it practically impossible to centralize control. As a result, a number of works dedicated to decentralized control in energy appeared (Zabet I. & Montazeri M. (2010); Aghdam et al., 2019).

The analysis of modern research in the direction of the creation and research of distributed systems of automatic control allows us to identify several urgent problems, in the solution of which great attention is paid to the application of decentralized systems:

- Scalability and efficiency. Distributed automatic control systems often work with a large amount of data and require high performance. The development of methods for optimizing the operation of such systems in order to increase speed and efficiency remains an urgent problem;
- Reliability and security. Ensuring the reliability of distributed systems is an important task. This covers both protection against accidental failures and ensuring resistance to failures of individual elements of the system;
- Adaptability to changes. Given the variability of conditions and inputs, it is important to develop methods that allow distributed automatic control systems to adapt to changes in real time and provide optimal control.

**Related works.** Decentralized control of distributed systems has certain advantages, but it raises a number of theoretical and practical problems. In many studies, attention is paid to the problem of stability and quality of control in decentralized systems. Works (Boyd et al., 1994; Šiljak & Stipanović, 2000). Use an approach based on the Lyapunov functions of the block-diagonal structure and the construction of systems of matrix linear inequalities based on them. Special tools for stability analysis are also being developed (Elmahdi et al., 2015).

Mirkin B. (1992) proposed the concept of adaptive decentralized control with model coordination is. At the same time, it is assumed that information about the state of reference models of all local subsystems is available to local controllers. In (Jianget et al., 2018). the problem of consensus for a class of heterogeneous linear multi-agent systems is investigated. The consensus problem is decomposed into a set of local tracking problems with local cost functions determined from the tracking errors. Based on game theory, the set of stable optimal policies of the entire network falls into a Nash equilibrium. In order to find the Nash solution, a distributed algorithm has been developed that calculates control strategies using an iterative process.

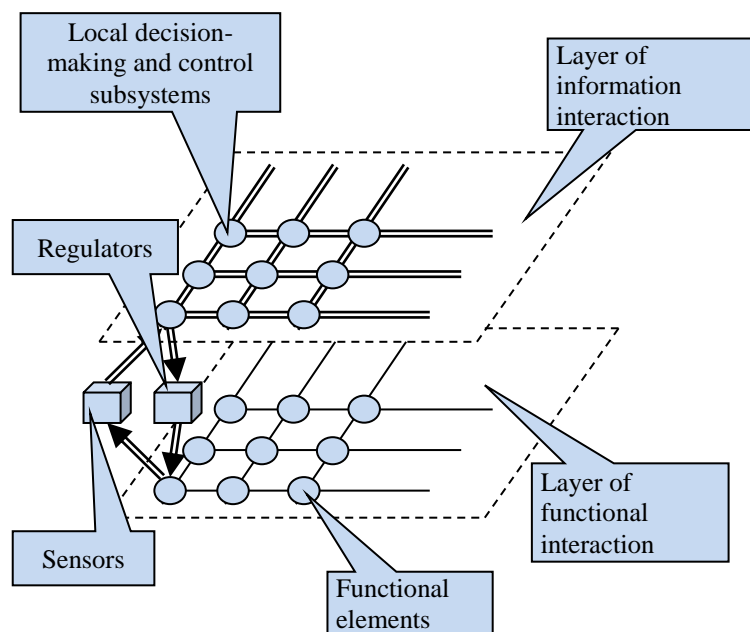
For the implementation of methods of coordination of control systems of multi-zone technological objects, learning systems are preferred, as they are able to flexibly adjust to the specifics of each zone control. However, a difficult problem for the application of such systems is the creation of training datasets. The article (Zhang et al., 2021). investigates model-based methods

in multi-agent reinforcement learning (MARL). The complexity of dynamic sampling and the complexity of component sampling in MARL are determined.

In modern works on the coordination of decentralized control in technical systems, the training of agents in multi-agent systems based on neural networks is mainly considered. At the same time, model-based multi-agent reinforcement learning is used. Wang et al. (2022) provides an overview of existing research on model-based MARL, including theoretical analysis, algorithms, and applications, and analyzes the advantages and potential of model-based MARL.

The article (Akramizadeh et al., 2010). developed model-based reinforcement learning for a group of agents with self-interests and sequential action selection based on traditional priority sorting. The learning process is considered as an extensive Markov game.

In previous works, the author of this article presented decentralized distributed control systems (DCS) for multi-zonal objects (MZO) as two interacting layers (Fig. 1) (Dubovoi & Yukhymchuk, 2022): a layer of physical interaction of MZO zones and a layer of information interaction of local control systems (LCS). The communication between MZO zones and the corresponding LCS is carried out through regulators (executive devices) and sensors (feedback). The LCS implements both the implementation of a specific control law (relay or linear) and coordination with other LCSs in order to optimize the state of the MZO according to a given global criterion. In a decentralized system, optimization is not carried out simultaneously by all LCSs, but in a sliding mode based on a combined local-global criterion (Yukhymchuk & Dubovoi & Kovtun, 2022).



**Figure 1.** Generalized image of a decentralized control system

In preceding research (Dubovoi & Yukhymchuk, 2022) specific facets of the architecture pertaining to the decentralized coordination system governing multi-zone thermal entities were examined. The constituents comprising the state control system of the DCS utilizing neural networks are delineated as follows:

- DCS itself, which is divided into control and control zones;
- a set of agents implementing the sliding decentralized coordination algorithm;
- datasets characterizing the given and actual states of the DCS;
- training procedure for coordinators;
- user interface - system operator;
- stream processors for control of data exchange between system components.

The coordinator agent in (Dubovoi & Yukhymchuk, 2022) uses the following principles,

## methods and algorithms

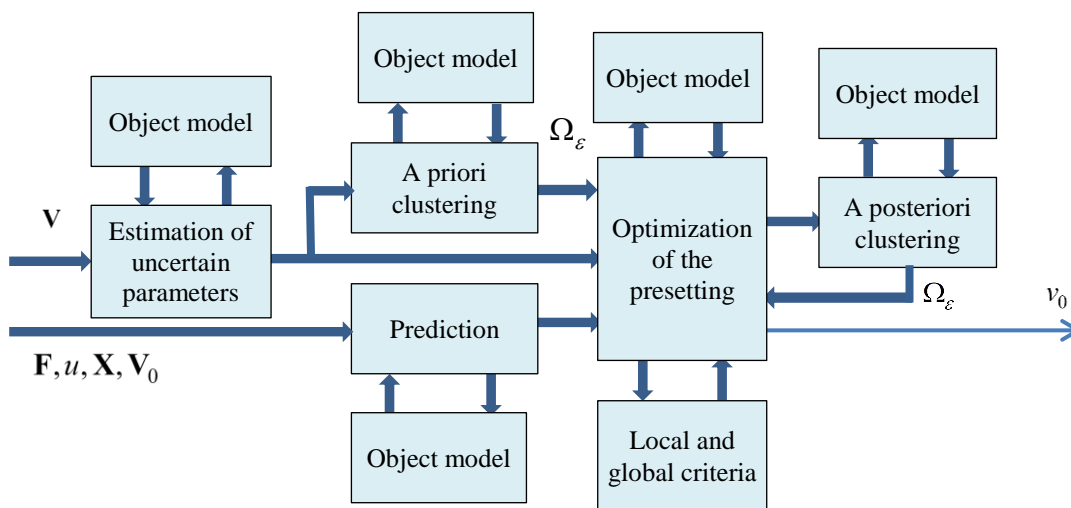
- A model of a distributed cyber-physical system with decentralized control, in which each element of a multi-zone object is controlled by a local control system, which in turn receives optimal control settings from a local coordinator;
- A sliding algorithm for control of the coordination sequence to ensure the stability of the system;
- The principle of close action, which defines a set of coordinators (cluster), which provide information to the sliding coordination center;
- Methods of estimating uncertain parameters;
- Methods of forecasting the effects and states of the system and the use of forecasting during coordination control.

All the specified elements of the coordinator are combined in a synergistic interaction as modules shown in fig. 2. As a result, the coordination function is implemented by a software complex, but another version of the implementation of such an architecture using neural networks is also possible.

The initial data for optimal coordination for MZO with  $n$  zones are:

- Vector  $\mathbf{F}[n]$  of given zone states;
- Vector  $\mathbf{V}_0[n]$  of optimal states of zones (vector of LCS presettings is the initial data and objective of coordination recursively);
- Vector  $\mathbf{V}[n]$  of actual zone states (resource accumulated in the zone);
- Vector  $\mathbf{X}[n]$  of the amount of input raw materials to each MZO zone, for the processing of which the accumulated resource is spent;
- State  $u[1]$  of the surrounding environment.

At the output of the coordinator, only one value  $v_0[1]$  is been obtained - the presetting for the corresponding LCS.



**Figure 2.** Scheme of interaction of coordinator modules

**Objectives and problems.** The conducted analysis showed that the existing approaches to solving the problem of creating effective decentralized control systems for multi-zonal objects have certain shortcomings. These shortcomings are mainly caused by too much idealization of the MZO model. And since the MZO model, according to the scheme of connections of the coordinator modules in Fig. 2, is used to perform all procedures: assessment, forecasting, optimization, clustering, their accuracy is insufficient.

The problem can potentially find resolution through the development of coordinators based on

neural networks (NN). However, another problem arises here: the training of such coordinators. After all, training a neural network of coordinator requires a substantially extensive marked dataset due to the complexity of the coordination task. Unfortunately, in most cases such datasets do not exist for MZO.

The purpose of this work is to develop approaches to the creation of training datasets and training procedures for DCS neuron coordinators (NC).

### **Proposed Method.**

**Statement of research.** The volume and structure of the training dataset depends on the type and size of the NN, which in turn depend on the characteristics of the problem for which they are intended to be solved. Therefore, the first task of the work is to formulate an approach to choosing the type and size of NN. The training method also affects the formation of the training dataset. Given that multi-zone objects can differ significantly from each other in terms of spatial structure, static and dynamic characteristics, final pre-operational training of the system is impossible. Therefore, in this work, we will plan a two-stage training: preliminary training based on the DCS simulation (model-based training) and further training in the process of operation. Accordingly, it is necessary to formulate an approach to the creation of datasets and training procedures at both stages.

**Basic approach.** Model-based training of the neural coordinator can be performed according to various scenarios, depending on the structure of the neural coordinator and the features of the structure of the multi-zone object.

According to the architectural depiction of the coordinator illustrated in Fig. 2, neural networks must implement the following functions:

- DCS modeling;
- Clustering;
- Optimal estimation of parameters;
- DCS state predicting;
- Optimization of the presetting according to the local-global criterion.

These functions can be implemented by separate neural networks (Separated Neural Coordinator, SNC) and the NN of the architecture most suitable for this task should be chosen for the performance of each function or with the help of one NN (Mono-Neural Coordinator, MNC) which has  $4n+1$  inputs, 1 output and architecture united all the mentioned tasks in accordance with Fig. 2.

GRU (Gated Recurrent Unit) is the most suitable type of neural network for solving the problem of vector time series forecasting.

To solve the problem of optimal estimation of parameters, where one of the vectors is calculated by solving a system of differential equations, it is advisable to apply the modification of recurrent neural networks GRU (Gated Recurrent Unit)

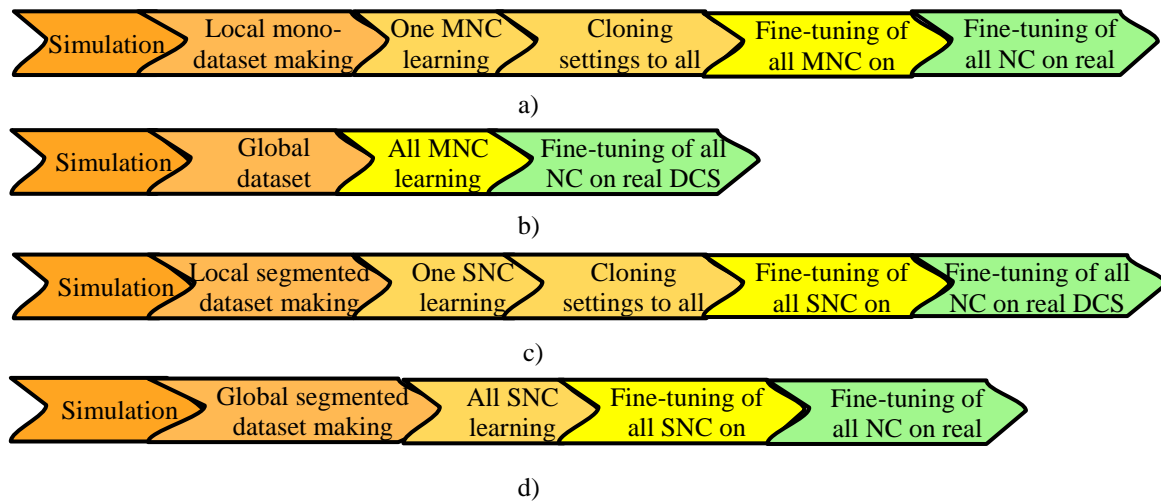
Convolutional regression neural network is the most suitable type of neural network for solving the problem of presetting optimization.

For the task of clustering a set of zones given by a weighted graph, the most suitable type of neural network is GNN (Graph Neural Network). Among the modifications of GNN, GATs (Graph Attention Networks) are the most acceptable - these networks pay attention to the proximity of vertices in the graph and can take edge weights into account.

The most difficult task is the creation of a neural model of DCS, since it consists of interconnected models of zones, resource flows and LCS, and each of them are described by differential equations of different orders. As a result, the neural model should contain several recurrent layers, several (at least two) layers with lateral connections, as well as a Convolutional Neural Network with Vector Output.

The MNC coordinator solves the problem comprehensively, therefore, taking into account the order of the DCS dynamics model, the need to predict states, as well as the mutual influence of zones and coordinators, the MNC architecture should be built on the base of a GRU network with an additional recurrent layer and convolutional layer. Note that the MNC coordinator must receive all  $4n+1$  input data and have an appropriate number of items in each layer, while NNs for individual tasks have a much smaller dimension.

Thus, there are 4 basic scenarios of NC learning, shown in Fig. 3, and their combinations in the case of partial merging of NNs.



**Figure 3.** Basic scenarios for training neural coordinators

According to the scenario in Fig. 3a, one MNC coordinator is trained. The dataset is created for this purpose with the help of simulation of one NC of the most characteristic LCS, i.e., the one that is affected by the largest number of external influences. Based on the training results of this NC, other DCS coordinators are cloned. In the next step, individual adjustment of all MNCs is carried out on the simulation model and further training on the physical object. The last stage of additional training is already in the process of implementing the system.

According to the scenario in Fig. 3b, MNC coordinators of all LCS are simultaneously trained. For this, a global dataset of inputs and outputs of all NCs is created using simulation. This method does not require additional training on a simulation model, but requires a much larger dataset. In the last step, additional training is carried out at the physical object.

According to the scenario in Fig. 3c, one SNC coordinator is trained. Since the NC is segmented, that is, each function of the coordination task is performed by a separate NN, and a corresponding segmented dataset for the most characteristic LCS is created with the help of simulation for this. Based on the results of the training of all segments of the NN of this NK, other DCS coordinators are cloned. In the next step, individual fine-tuning of all relevant SNC segments is carried out on a simulation model and then fine-tuning on a physical object.

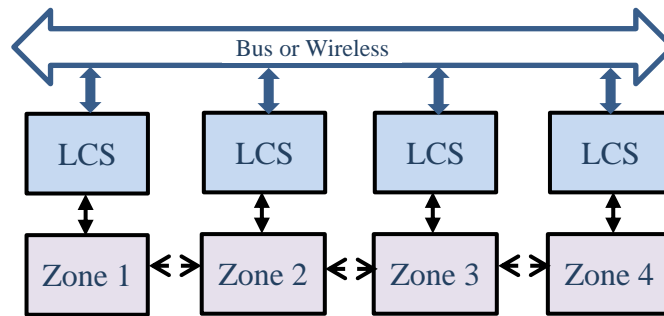
According to the scenario in Fig. 3d, SNC coordinators of all LCSs are simultaneously trained. For this, a global segmented dataset is created for each module of all NCs using simulation. This method requires fine-tuning the joint action of all modules of each SNC on a simulation model. In the last step, additional training is carried out at the physical object.

### Numerical Experiments, Results and Discussion

The study of the effectiveness of these scenarios was carried out on the basis of a decentralized MZO control system with 4 zones located in series (Fig. 4).

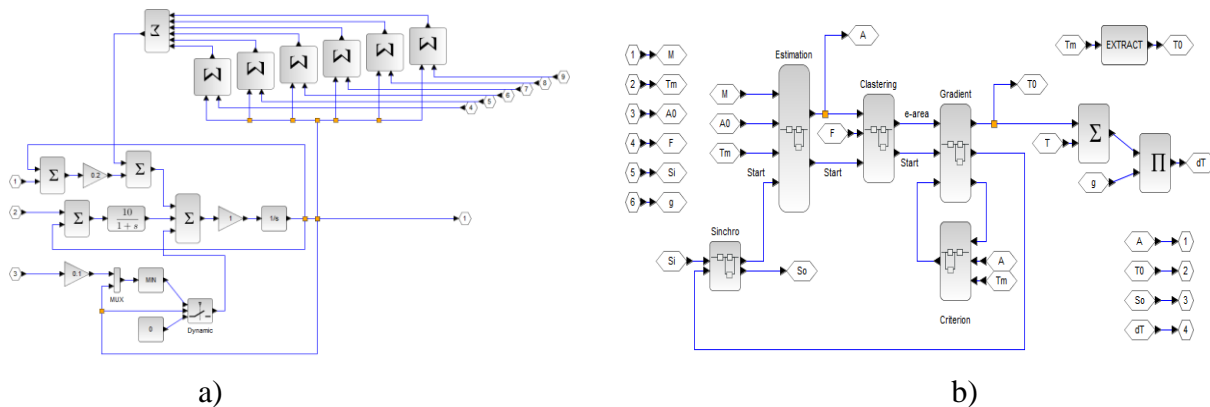
Zones 1...4 affect each other by exchanging resource flows (heat, raw materials, etc.). Local zone control systems exchange information necessary for coordination through a data transmission

system.



**Figure 4.** Basic structure for simulation

The simulation model of the system was created on the Scilab/Xcos platform using the block library for modeling distributed control systems (Dubovoi & Yukhymchuk, 2022; Yukhymchuk, 2022). Fig. 5a shows the model of one zone with LCS, Fig. 5b shows the model of the coordinator.



**Figure 5.** Simulation models of the controlled zone and the coordinator

Neural coordinators MNC were created using ChatGPT with the Tensor-flow library. An example of the network generation function is shown in Fig. 6.

```
def create_model():
    model = Sequential([
        Input(shape=(5, 17)), # Input layer for historical data
        GRU(64, return_sequences=True), # Recurrent layer GRU
        GRU(32), # The second recurrent layer GRU
        Dense(1) # Output layer with one neuron
    ])
    return model
```

**Figure 6.** Function for MNC generation from ChatGPT

Let's calculate the number of configurable parameters for the proposed model:

- 1) First layer GRU:
  - Parameters for the input vector:  $17 \times 64 + 64 = 1088$
  - Parameters for recurrent matrix:  $64 \times 64 + 64 = 4160$
  - Total:  $N_1 = 1088 + 4160 = 5248$
- 2) Second layer GRU:

- Parameters for the input vector:  $64 \times 32 + 32 = 208064 \times 32 + 32 = 2080$ ;
- Parameters for recurrent matrix:  $32 \times 32 + 32 = 105632 \times 32 + 32 = 1056$ ;
- Total:  $N_2 = 2080 + 1056 = 31362080 + 1056 = 3136$ .

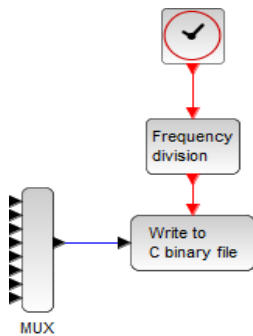
3) Output fully connected layer:

- Parameters for scales:  $N_3 = 32 \times 1 + 1 = 3332 \times 1 + 1 = 33$

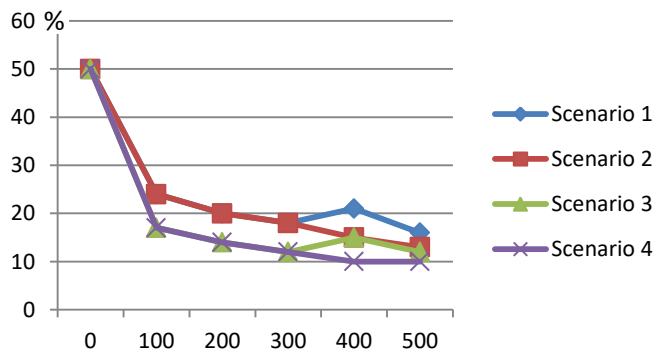
Thus, the total number of configurable parameters for this model is:

$$N_p = 5248 + 3136 + 33 = 84175248 + 3136 + 33 = 8417$$

To create a dataset from the input, output and internal parameters of the coordinators, the module for outputting simulation data to a file shown in Fig. 7 was used. Pre-normalization of the simulation data was not required because the stability of the system had been verified in previous work and the parameter ranges were specified in the simulation presetting. The resulting data array was divided into a training (80%) and testing (20%) subset.



**Figure 7.** Module for outputting simulation data to a file



**Figure 8.** Comparison of the effectiveness of the scenarios

Dependencies of the RMSE of the output vector of presettings of the neural network relative to the same vector of presetting of the simulation on the testing sample and the number of training epochs were studied. The results of the study are shown in Fig. 8. The dataset was generated by a simulation with random input influences  $F[n]$ ,  $X[n]$ ,  $u$  with a uniform distribution and an exponential correlation function. The probability distribution of the components of the vector  $F$  of a given state of the zones is normal, with mathematical expectation  $m_f = 37$ , variation  $\frac{\sigma_f}{m_f} = 20\%$ , correlation interval  $\tau_f = 5$  simulation cycles. The probability distribution of the components of the raw material flow vector  $X$  is uniform, with mathematical expectation  $m_x = 10$ , variation  $\frac{\sigma_x}{m_x} = 10\%$ , correlation interval  $\tau_x = 5$  simulation cycles. The probability distribution of the state of the environment  $u$  is uniform, with a mathematical expectation  $m_f = 15$ , range is  $[0; 30]$ , correlation interval  $\tau_u = 15$  simulation cycles.

Let's estimate the required number of dataset blocks. To prevent overfitting, we will use the "10x rule". This means that the number of training data should be 80,000, and the total number of simulation cycles and, accordingly, data blocks (including test ones) should be  $N_d = 10^5$ .

Research results have shown that with a sufficiently large number of epochs, all scenarios lead to approximately the same result. However, segmented NCs approach the final learning level faster. In the cases of Scenarios 1 and 3, where one coordinator was first trained for 300 epochs and then the results were cloned to other coordinators, the overall RMSE increased immediately after cloning, but quickly decreased to close to Scenarios 2 and 4 as a result of fine-tuning.

At the same time, the training of segmented coordinators requires the creation of similarly segmented datasets: separately for each task which complicates the training process.

**Conclusions.** The studies allow us to conclude about the possibility of transition from the algorithmic solution of the problem of coordination of local control systems of multi-zonal objects to the solution using neural networks. Two variants of neural coordinators are considered: segmented and non-segmented. The advantage of the segmented version is a slightly higher learning speed. However, the non-segmented version is more versatile because it is not tied to a defined coordinator structure.

In the course of the study, the last stage of the scenarios - supplementary training on a real physical object - was not performed. However, it can be assumed that after such fine-tuning, the advantages of MNC will become even more noticeable, since such fine-tuning will correct the shortcomings of imitation.

Further research is expected to be directed to the task of choosing the optimal architecture of a neural network for decentralized coordination of MZO.

**Acknowledgements.** The author gratefully acknowledge the invaluable support and expertise provided by the staff of the Department of Computer Control Systems at Vinnytsia National Technical University. Their help and resources were essential to the successful completion of this research. Special thanks are due to colleagues who offered constructive insights that significantly enhanced the work.

**Conflicts of Interest.** The author declares no conflict of interest.

## References

- Aghdam A.G., Miller D.E., & Davison E.J. (2019). *Decentralized Control of Large-Scale Systems*, Springer, ISBN/EAN: 9781441960146.
- Akramizadeh, A., Afshar, A., B Menhaj, M., & Jafari, S. (2010). Model-based Reinforcement Learning in Multiagent Systems in Extensive Forms with Perfect Information. *IFAC Proceedings Volumes*, Vol. 43, Issue 8, pp. 487–494. Elsevier BV. <https://doi.org/10.3182/20100712-3-fr-2020.00080>
- Bakule, L. (2008). Decentralized control: An overview. *Annu. Rev. Control.*, 32, pp. 87-98. [Online]. Available: <https://api.semanticscholar.org/CorpusID:205659095>.
- Boyd, S., El Ghaoui, L., Feron, E., & Balakrishnan, V. (1994). *Linear Matrix Inequalities in System and Control Theory*. Society for Industrial and Applied Mathematics. SIAM, Philadelphia, 1994. <https://doi.org/10.1137/1.9781611970777>
- Coordination Control of Distributed Systems*. (2015). In J. H. van Schuppen & T. Villa (Eds.), *Lecture Notes in Control and Information Sciences*. Springer International Publishing. <https://doi.org/10.1007/978-3-319-10407-2>
- Dubovoi V.M. & Yukhymchuk M.S. (2022). *Decentralized coordination control of distributed cyber-physical systems with continuous objects*, VNTU, Vinnytsia. [in Ukrainian]
- Elmahdi, A., Taha, A. F., Sun, D., & Panchal, J. H. (2015). Decentralized Control Framework and Stability Analysis for Networked Control Systems. *Journal of Dynamic Systems Measurement and Control*, vol. 137. doi: 10.1115/1.4028789.
- Ge, X., Han, Q.-L., Ding, L., Wang, Y.-L. & Zhang, X.-M. (2020). Dynamic Event-Triggered Distributed Coordination Control and its Applications: A Survey of Trends and

- Techniques. *IEEE Transactions on Systems, Man, and Cybernetics: Systems*. Vol. 50, Number 9, pp. 3112–3125. <https://doi.org/10.1109/tsmc.2020.3010825>
- Gong Z. & Aldeen M. (1997). Stabilization of Decentralized Control Systems, *Journal of Mathematical Systems, Estimation, and Control*, Vol. 7, No. 1, pp. 1–16,
- Jiang, H., Liu, X.-K., He, H., Yuan, C., & Prokhorov, D. (2018). Neural Network Based Distributed Consensus Control for Heterogeneous Multi-agent Systems. In *2018 Annual American Control Conference (ACC). 2018 Annual American Control Conference (ACC)*. IEEE. pp. 5175–5180. <https://doi.org/10.23919/acc.2018.8431744>
- Katrenko, A.V., & Savka I.V., (2008). Mechanisms of coordination in complex hierarchical systems. *Bulletin of Lviv Polytechnic National University*. Vol. 631, pp. 156–166. [Online]. Available: [http://vlp.com.ua/files/16\\_1.pdf](http://vlp.com.ua/files/16_1.pdf). [in Ukrainian]
- Ladanyuk, A., Shumygai, D. & Boiko, R. (2012). System task of co-ordination in the technological complexes of continuous type. *Collected Works of Kirovohrad National Technical University Machinery in agricultural production, industry machine building, automation*. Vol. 25-1. pp. 288-294. [Online]. Available: [http://dspace.nuft.edu.ua/jspui/bitstream/123456789/4444/1/Sh\\_3.pdf](http://dspace.nuft.edu.ua/jspui/bitstream/123456789/4444/1/Sh_3.pdf). [in Ukrainian]
- Mirkin B.M. (1992). A new decentralized model reference adaptive control scheme for large scale systems. in *Prepr. 4th IFAC International Symposium on Adaptive systems in control and signal processing*, Grenoble, France, pp. 645–650.
- Shaikh P.H., Nor.B.M., Nallagownden P., Elamvazuthi I., & Ibrahim T. (2014). A Review on Optimized Control Systems for Building Energy and Comfort Control of Smart Sustainable Buildings. *Renewable and Sustainable Energy Reviews*, Vol. 34, pp. 409–429. <https://doi.org/10.1016/j.rser.2014.03.027>.
- Šiljak, D.D., & Stipanović, D.M. (2000). Robust stabilization of nonlinear systems: The LMI approach. *Mathematical Problems in Engineering*, vol. 6, pp. 461–493.
- Wang, X., Zhang, Z., & Zhang, W. (2022). Model-based Multi-agent Reinforcement Learning: Recent Progress and Prospects (Version 1). *arXiv*. <https://doi.org/10.48550/ARXIV.2203.10603>
- Yukhymchuk, M., Dubovoi, V., & Kovtun, V. (2022). Decentralized Coordination of Temperature Control in Multiarea Premises. *Complexity*, vol. 2022, Article ID 2588364. <https://doi.org/10.1155/2022/2588364>.
- Zabet I. & Montazeri M. (2010). Decentralized control and control systems for power industry via multiagent systems technology, in *PEOCO 2010 - 4th International Power Engineering and Optimization Conference, Program and Abstracts*, pp. 549–556, IEEE. doi: 10.1109/PEOCO.2010.5559254.
- Zhang, W., Wang, X., Shen, J., & Zhou, M. (2021). Model-based Multi-agent Policy Optimization with Adaptive Opponent-wise Rollouts (Version 3). *arXiv*. <https://doi.org/10.48550/ARXIV.2105.03363>.

УДК 519.2:004.94

## Evaluating Fast Charging of Electric Vehicles Along Motorways Using Finite Multi-Server Queueing System Simulation

**Maria S. Forkaliuk**

Bachelor, student,

Vinnitsia National Technical University

Khmelnyske shose 95, 21021, Vinnitsia, Ukraine

ORCID: 0009-0004-4533-5282

**Gerald Franzl**

PhD, senior researcher,

University for Continuing Education Krems, Department for Integrated Sensor Systems

Viktor Kaplan Str. 2/E, 2700 Wiener Neustadt, Austria

ORCID: 0000-0003-0260-9912

**Oleg V. Bisikalo**

PhD, full professor,

head of department of Automation & Intelligent Information Technologies,

Vinnitsia National Technical University,

Khmelnyske shose 95, 2102, Vinnitsia, Ukraine

ORCID: 0000-0002-7607-1943

**Abstract** Fast DC charging sites are required along motorways to abrogate the car drivers' anxiety of long-distance travels when driving electric vehicles (EVs) with batteries optimised for efficient average reach. This is important to facilitate the mobility transition to EVs. In this study, a queueing model-based approach to simulate and evaluate fast charging sites equipped with many DC charging points is presented. Charging sites are modelled as multi-server queueing systems with finite waiting space, where the servers represent the charging points and the waiting space the parking area available for EVs waiting for service. To evaluate also arrival and service time distributions that are non-Markovian, the queueing system is evaluated using event based simulation. Exemplary results and a comparison with analogous simulation tools complete the presentation of the simulation approach.

On one hand, the simulation reveals the mean potential waiting time per EV before charging can start due to the temporary occupation of all charging points. On the other hand, the tool analyses the aggregated power demand of all charging points. Based on latter, the smart charging mechanism reduces dynamically the individually available charging power if needed to stay below the power grid access limit. This smart charging mechanism causes a small decline in the charging performance at high EV traffic loads when all charging points are maximally occupied. In combination with the state-of-charge depending power demand, the tool provides the user critical insights into realistically expectable waiting times and decreased charging volumes when many EVs charge in parallel. Experimenting with different number of charging points and grid power limitations helps the tool-user, the systems designer, to dimension charging sites along motorways that can efficiently handle future traffic loads.

**Keywords:** Fast Charging, Smart Charging, Charging Sites, Event-based Simulation, Power Histograms.

**DOI:** <https://doi.org/10.31649/1999-9941-2024-60-2-77-90>

## Оцінка швидкої зарядки електромобілів вздовж автомагістралей за допомогою моделювання скінченної багатосерверної системи масового обслуговування

**Марія Сергіївна Форкалюк**

бакалавр, студент

Вінницький національний технічний університет

Хмельницьке шосе 95, 21021, Вінниця, Україна

ORCID: 0009-0004-4533-5282

**Геральд Францл**

PhD, старший науковий співробітник

Університет безперервної освіти Кремс, відділ інтегрованих сенсорних систем,

вул. Віктора Каплана 2/Е, 2700 Вінер-Нойштадт, Австрія

ORCID: 0000-0003-0260-9912

**Олег Володимирович Бісікало**

доктор техн. наук, професор, завідувач кафедри АІТ,

Вінницький національний технічний університет

Хмельницьке шосе 95, 21021, Вінниця, Україна

ORCID: 0000-0002-7607-1943

**Анотація.** Швидкі зарядні станції постійного струму необхідні вздовж автомагістралей, щоб позбавити занепокоєння водіїв щодо поїздок на великі відстані на електромобілях з батареями, оптимізованими для ефективного середнього запасу ходу. Це важливо для подальшого сприяння переходу на електромобілі. У цьому дослідженні представлено підхід на основі моделі черг для моделювання та оцінки майданчиків швидкої зарядки, обладнаних багатьма зарядними пунктами постійного струму. Зарядні станції моделюються як багатосерверні системи масового обслуговування з обмеженим простором очікування, де сервери представляють пункти зарядки, а простір очікування-паркувальну зону, доступну для електромобілів, які очікують на обслуговування. Щоб оцінити також розподіл часу прибуття та обслуговування, який не є марковським, система масового обслуговування оцінюється за допомогою імітаційного моделювання на основі подій. Прикладні результати та порівняння з аналогічними інструментами моделювання завершують презентацію підходу до моделювання. З одного боку, симуляція демонструє середній потенційний час очікування електромобіля до початку заряджання через тимчасову зайнятість усіх зарядних точок. З іншого боку, інструмент аналізує сукупний попит на електроенергію усіх зарядних станцій. На основі останнього, механізм розумної зарядки динамічно зменшує індивідуально доступну потужність зарядки, щоб вона не перевищувала ліміт доступу до електромережі. Цей інтелектуальний механізм призводить до зниження продуктивності зарядки при високому навантаженні трафіку електромобілів, коли всі зарядні точки зайняті. У поєднанні з потребою в енергії, що залежить від стану заряду, інструмент надає користувачеві критично важливу інформацію про реалістично очікуваний час очікування та зменшення обсягів заряджання, коли багато електромобілів заряджаються паралельно. Експерименти з різною кількістю точок зарядки та обмеженнями потужності мережі допомагають користувачеві інструменту та проектувальнику системи визначити розміри зарядних майданчиків уздовж автомагістралей, які зможуть ефективно впоратися з майбутнім транспортним навантаженням.

**Ключові слова:** швидка зарядка, розумна зарядка, зарядні станції, моделювання на основі подій, гістограми потужності.

**Introduction.** To significantly reduce greenhouse gas emissions in the mobility sector and thereby mitigate climate change (Erbach 2024), the transition to electric vehicles (EVs) appears imperative (Tang 2023). EV sales rise steeply and reached 20% or more in many markets, often supported by governmental incentives. Overcoming the range limitations of current EV technology with larger batteries is probably neither efficient nor sustainable because for the majority of distances travelled (Eurostat 2021), the extra weight and cost are not required. Still, for the occasional longer-distance travel, recharging needs to be integrated into the travel planning if recharging opportunities are scarce and recharging occurs comparably slow. To be reliable, a robust, freely accessible, and trusted fast charging infrastructure is required especially along motorways (EUR-Lex 2023) to ensure that travel remains convenient and time-efficient, and to alleviate the common range anxiety among non-EV users.

Recent literature covers mostly the deployment of a distributed AC charging infrastructure in residential areas (Huang 2023, Jansson 2022). Research results on the integration of fast direct current (DC) charging along motorways, where fast charging is most needed, are hardly found and often very complex, as in (Witt 2023). Yet, some companies plan recharging sites with megawatt grid access needs. Are the thereby triggered grid expansions really necessary, or can smart charging considerably reduce the required grid access needs? The answer to this research question depends heavily on the actual EV traffic load expected and the dynamics of fast EV charging (Witt 2023). In addition to the technical and economic challenges of installing and operating such charging sites, also the impact on the electric grid shall be encompassed (Lee 2019).

This research is grounded in the imperative necessity to address climate change through the transition to a sustainable transport sector, in which the strategic shift towards EVs is a critical element (Erbach 2024). As the global automotive market gradually shifts away from fossil-fuelled combustion engines, the establishment of a comprehensive charging infrastructure becomes paramount (EUR-Lex 2023). For an economic deployment, the designers of these infrastructures, the systems engineers, need to anticipate future demands, and cannot cater to the immediate needs of today's EV users only.

This research ventures beyond the common discourse on distributed AC charging solutions at home and at work, which are needed because these cover the vast majority of the recharging events (IEA 2024). The EV adoption is not solely an economic issue, the anxiety concerning the occasional long-distance travel needs to be solved equally for broad acceptance (Sirapa 2022). Delving into the realm of high-capacity, fast-charging sites capable of supporting long-distance EV travel raises other issues, in particular regional grid access limitations (Csanyi 2018). It is therefore critical to assess the potential of smart charging solutions to mitigate the impact on the electrical grid, thereby avoiding or at least postponing extensive and costly grid upgrades (Rahila 2024). The focal point of the presented simulation tool and exemplary study resides in the elaboration of efficient and effective DC charging sites along motorways, pivotal for overcoming the notable barriers to EV adoption, such as range anxiety and long recharging times (Sirapa 2022).

Exploring the interrelation of fast and smart charging technologies, the study adopts a methodology from the telecommunications sector employing event-based simulation and queueing theory (Yang 2018) to efficiently model the dynamics of smart and fast charging sites. This approach offers fresh insights into optimizing charging site performance, managing the actual power demand, and an effective step-by-step expansion path for high-capacity charging sites. In essence, this investigation not only highlights the technical and logistical challenges inherent in the deployment of a high-capacity DC charging infrastructure but also situates these challenges within the larger context of the electricity distribution grid, the complex energy system as a whole, urban planning, and environmental policymaking. By providing a detailed analysis of the requirements and impacts of DC charging sites, the study seeks to inform a wide range of stakeholders, from policymakers to industry leaders, thereby supporting the collective move towards a more sustainable, electrified future.

The purpose that triggered the development of the simulation tool is the need to accurately

assess the performance and grid impact of modern DC charging and to use this information for the design of efficient DC charging infrastructures (charging sites) along motorways. The R&D aim is to balance the EV serving capacity and the power demand challenge, ensuring that future EV charging sites, yet planned and soon deployed, deliver efficient, reliable service without overburdening the electrical grid. Through simulation, this study seeks to support the development of smart charging strategies that achieve scalable, sustainable solutions for the mobility transition to a low-carbon mobility future, which requires the rapid integration of a steeply rising number of EVs into the transportation network every day.

The primary research objective is to simulate charging sites with any number of DC charging points considering:

- the dynamic, state-of-charge dependent power demand of DC charging,
- the dynamic charging power limits introduced by smart charging technology,
- visualising the statistical results in publishable quality as scalable vector graphics for increasing traffic loads,
- saving complete scenarios including set parameters in txt and XML format together with all statistical results in CSV and graphical format in a single zip-file for research documentation and open data support.

**State-of-the-art and analogous simulation tools.** The *Electric Vehicle Queueing Simulation* designed by Ken Lau in 2017 (Lau 2017), is a visualization tool created to model and analyse the traffic intensity at electric vehicle charging stations. It calculates waiting times based on car arrival rates and charging speeds, assuming a Poisson process for the randomised arrivals and negative exponential distributed charging times. The tool users can adjust parameters like mean arrival rate, charging rate, and the number of charging points, to study their effects on key performance indicators such as queue lengths and waiting times. The simulation of this M/M/n queueing system serves as a valuable tool for understanding and optimizing charging station throughput and customer waiting times. However, the electric load and the charging performance are not evaluated, and other than Markovian arrival and service distributions cannot be chosen to better approximate a more realistic charging time distribution.

The C++ simulation of a multi-server queueing system, made public available on GitHub by Zedrex in 2021 (Kabeer 2021), leverages an event-driven approach for efficiently managing customer flows. By simulating customer arrivals, service processes, and departures across multiple servers, it adeptly mimics real-world queuing scenarios. It uses an exponential distribution for generating random inter-arrival and service times, reflecting real-world unpredictability by assuming a memory-less distribution process. Customers are queued and served based on server availability, with the system tracking each customer's journey from arrival to departure. Key methods include event scheduling, handling arrivals/departures, and statistical logging for analysis. This setup allows for detailed insights into queuing dynamics and server efficiency, although it simplifies real-world complexities and relies on an exponential distribution of both, the arrival and the service process.

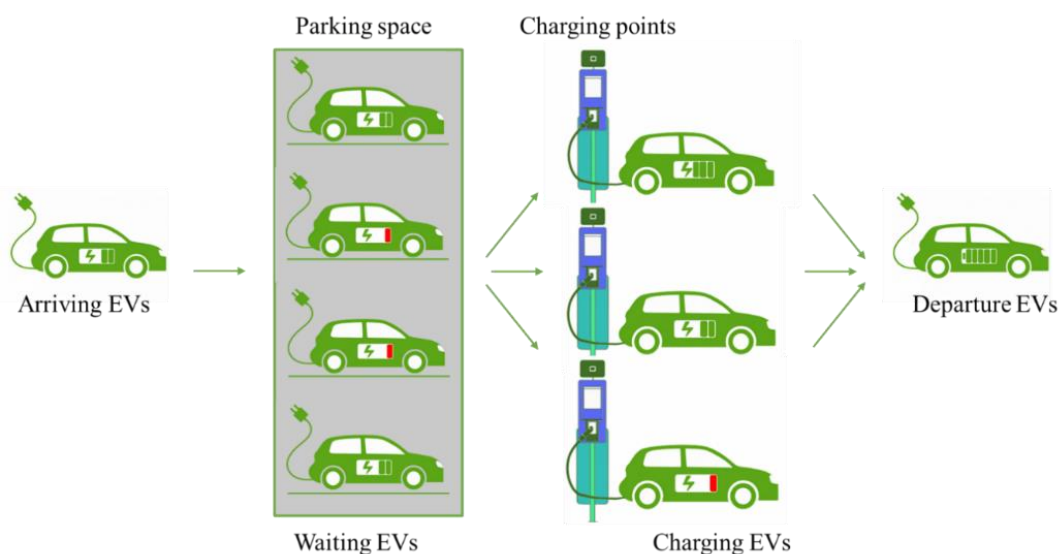
The *Queueing Simulation* tool implemented in Java and published in open source by Adnan Ansari 2019 (Ansari 2019), aims to model the process of car servicing with two service types: one assumed for Sedans and another for SUVs. It aims to model the process of servicing different EVs (Sedans and SUVs) using a queuing system, where the inter-arrival time and the two service time distributions can be specified by any number of discrete probability densities, which allows it to mimic measured distributions. The main objective is to analyse the performance of the EV service centre (charging site), i.e., average waiting probability and time as well as server utilisation, to assess the efficiency of the configured service resources. The two servers are assumed to have their own queues and are predestined to serve the assigned clients, Sedans or SUVs. However, if the according server is busy and the other server is idle, the over server also serves EVs of the other type. Still, the individual queues are filled with the assigned EV type only. This simulation tool allows to determine key performance indicators of the service and helps to improve service

processes, reduce waiting time for customers and optimise the utilisation of resources. On the downside, to evaluate the performance over increasing load (arrival rate over service rate) the distribution definitions need to be accordingly adjusted in the source code. Without prior calculation, the simulated load is difficult to predict and similarly to compose curves over increasing load.

A different approach is commonly taken when analysing the impact of EV charging on electricity distribution and supply. Most studies rely on real-world data and model all EV charging throughout a whole region, based on modelled, assumed, or empiric driving patterns for rural and urban residents. These studies need to include charging at the owners' premises and at long-term parking spaces, where most charging events occur, and where AC charging suffices because speed is at these locations of minor relevance. These studies, e.g. (Witt 2023), (Grigorev 2021), and (Yang 2021), focus on the impact of electric vehicles (EVs) on the regional electricity supply, with considerations ranging from traffic congestion to the dynamics of the EV charging demand. Grigorev and colleagues explore how EVs could influence traffic congestion and energy consumption through an integrated modelling approach, revealing potential shifts in urban mobility patterns (Grigorev 2021). Witt's research utilizes real traffic data and discrete event-based simulation to determine the required number of charging stations along a German motorway, providing essential insights for infrastructure planning (Witt 2023). Meanwhile, Yang and co-authors offer a dynamic model for the real-time management of a system of EV fast-charging stations, addressing the challenges of meeting EV charging demand without overburdening the electrical grid (Yang 2021). Each study assumes varying degrees of smart charging control, underscoring the critical role of intelligent charging strategies in harmonizing EV integration with existing electricity networks.

The simulation tool and study closest in its aim to the customised tool developed for the R&D project *eAlloc* (eAlloc 2021), is the tool presented in (Witt 2023). However, the reported results are only mean values, lacking confidence intervals and variance to be convincing. To learn about the system behaviour at different loads the simulation needs to be repeatedly executed and separately visualised with the help of some extra software tool.

**Finite multi-server queueing system simulation.** The simulation developed in Java models charging sites as finite multi-server queueing systems (Kleinrock 1975) as depicted in Figure 1. A dedicated number of charging points and limited waiting (parking) space determine the site configuration.



**Figure 1.** Finite multi-server queueing system modelling a charging site with three charging points and waiting space for four EVs

Markov processes, renowned for their memoryless property, offer a solid foundation for the

analysis of queueing systems. This distinctive feature ensures that the future state of a system is determined solely by its current state, rather than the sequence of events leading up to it. When applied to EV charging stations, this principle states that the arrival of vehicles and their respective charging duration are both independent memoryless processes. This simplification aids in the analysis yet effectively captures the core dynamics of people's arrival and queueing behaviour.

Utilizing the theory of Markov processes to model the queueing system for EV charging involves defining the system's states based on the number of EVs either charging or awaiting charge. Transitions between these states are governed by probabilities, i.e., the intensity of EV arrivals and the times spent for charging. Such a methodical approach enables the examination of the queueing system's dynamics, thereby facilitating the identification of an optimal design. This modelling technique simplifies stochastic dynamics into manageable mean values and variances, making it easier to analyse and optimize the operation of for example EV charging sites.

For memoryless (Markovian) arrival processes and independent (also Markovian) charging durations, the state probabilities  $P_i$  of such a queueing system can be calculated analytically. These systems are composed of  $n$  charging points and  $(N-n)$  places for EVs to wait, as shown in Figure 2.

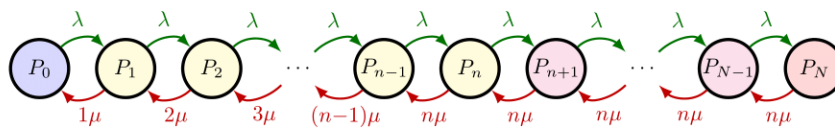


Figure 2. State-transition diagram

In Kendall’s notation, this is an M/M/n/N system: the first “M” denotes that the arrivals result from a Poisson process with negative exponentially distributed inter-arrival times (memoryless), the second “M” indicates that the service times are also negative exponentially distributed (also memoryless),  $n$  represents the number of servers (charging points in this context), and  $N$  is the total system capacity, including both charging points and waiting space. Thus, an M/M/5/15 system can accommodate up to 15 EVs at once, with 5 being charged and 10 waiting. If all waiting places are occupied, arriving EVs are deflected (blocked), i.e., are assumed to continue to an alternative charging site.

The theory of queueing systems has become popular with the advent of information and communication technology (ICT) to predict system occupation and client waiting times. However, the theory intensely studied by L. Kleinrock in (Kleinrock 1979, Liu 2021), can be applied to many similar systems where clients arrive, occupy a resource, and after some time leave the system. Based on this theory, the mean number of waiting clients equals the sum of the probabilities that the system is in a waiting state multiplied by the number of clients waiting in each state. *Little's law*  $E\{N\} = \lambda \cdot E\{T\}$  yields that the average waiting time  $T_{wait}$  equals the mean number of waiting clients divided by the arrival rate  $\lambda$ , as shown in equation (1). Summing over all waiting state probabilities  $P_{i>n}$  yields the waiting probability  $P_{wait}$ , as shown in equation (2).

$$T_{wait}^{M/M/n/N} = \frac{1}{\lambda} \sum_{i=n+1}^N (i - n) P_i^{M/M/n/N} \tag{1}$$

$$P_{wait}^{M/M/n/N} = \sum_{i=n+1}^N P_i^{M/M/n/N} \tag{2}$$

The individual state probabilities  $P_i^{M/M/n/N}$  can be calculated by solving the following equation system:

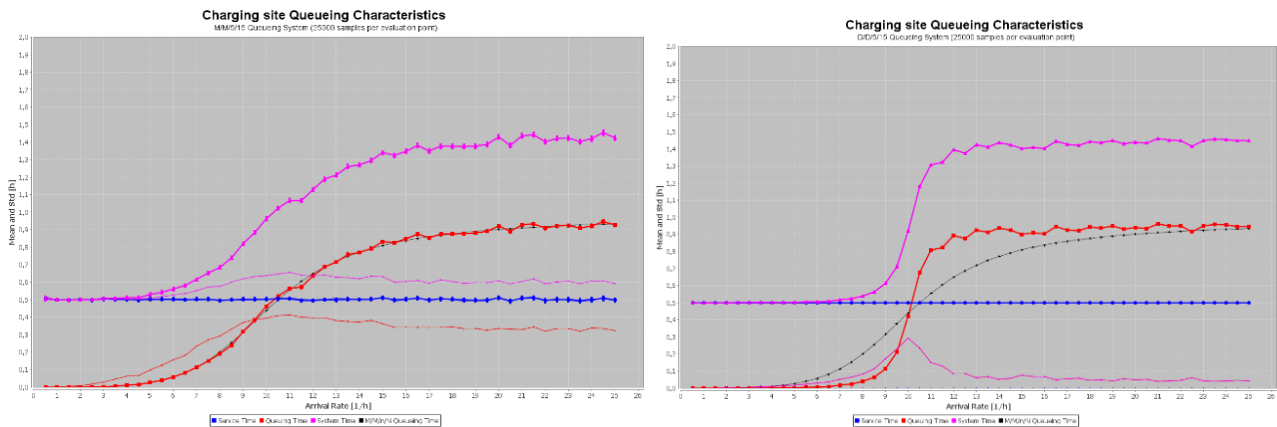
$$P_i|_{i=1}^5 = \frac{1}{i!} \frac{\lambda^i}{\mu^i} P_0 = \frac{\rho^i}{i!} P_0 \tag{3}$$

$$P_i|_{i=6}^{15} = \frac{\rho^i}{n! \cdot n^{i-n}} P_0 \tag{4}$$

$$P_0 = 1 - \sum_{i=1}^{15} P_i \tag{5}$$

where the load factor  $\rho = \frac{\lambda}{n\mu}$  is introduced, and equation (5) states that the system needs to be in one of the possible states at all times, i.e.,  $\sum P_i = 1$ .

We use equation (1) later on with the presented results to calculate the waiting time of an equivalent M/M/n/N queueing system depicted as narrow black curve. Comparing the simulated waiting time with that calculated for the Markovian system reveals the impact of the actually used distributions compared to the memoryless system that can be analytically solved. Figure 3 proves that the calculated waiting time equals the result of the simulation if we simulate the Markovian system where both, the inter-arrival and the service times are negative exponential distributed.



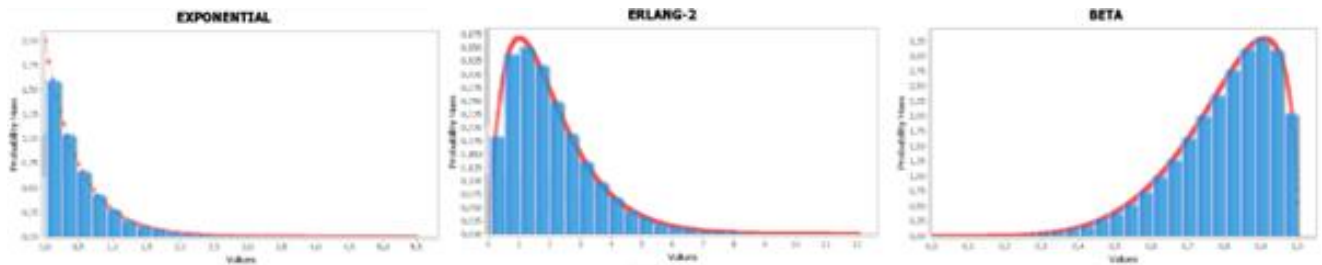
**Figure 3.** Waiting time over EV arrival rate for M/M/5/15 and U/D/5/15 queueing system

The results for the M/M/5/15 queueing system shown on the left in Figure 3, prove that the statistically derived waiting times (red curve) achieved from simulating the queueing system outlined in the next section equal the waiting times calculated analytically (narrow black curve) using equation (1). The results on the right side in Figure 3 show that for uniformly distributed arrivals (equally probable inter-arrival times between zero and twice the mean) and deterministic service times (all equal), i.e., a U/D/5/15 queueing system, the performance up to the critical load is better, but much worse if the traffic load exceeds the total service capacity ( $n\mu$ ). This example shows that the analytic approach is not always feasible and the simulation of the queueing system is more versatile in the case that the inter-arrival and service time distributions are known and do not fulfil the memoryless property required to be Markovian.

**Charging site analysis based on simulation.** EVs are assumed to arrive with negative-exponentially distributed inter-arrival times, which models the independent arrivals of vehicles to a charging site. Considering the typical similarity of charging times (service time), which results from the commonly faster charging possibility of larger batteries, we assume the service time to be Erlang-2 distributed. The charging demand is assumed to be Beta distributed between 0% and 100% of the battery capacity, assuming a mean state of charge (SoC) when EVs arrive at the charging site between 10% and 30% SoC. Figure 4 depicts the analytically calculated characteristic probability density function (pdf) of these distribution functions as red lines and the histograms of the generated random samples via blue bars.

The EV charging time is in general not negative exponential distributed but peaks at some rather short duration between 15 minutes and one hour, as shown in the presentation in (Lau 2017). This can be approximated with an embedded Markov chain and solved analytically, but easier and applicable for any distribution is the simulation of the queueing system using the event-based simulation technique. This enables the analysis of general G/G/n/N queueing systems (for any inter-arrival and service time distribution that can be generated) and in addition, to directly evaluate also the dynamic interactions among decreasing charging power demand (the SoC dependence) and smart charging (dynamic power limit adjustments) at the charging site. This involves considering

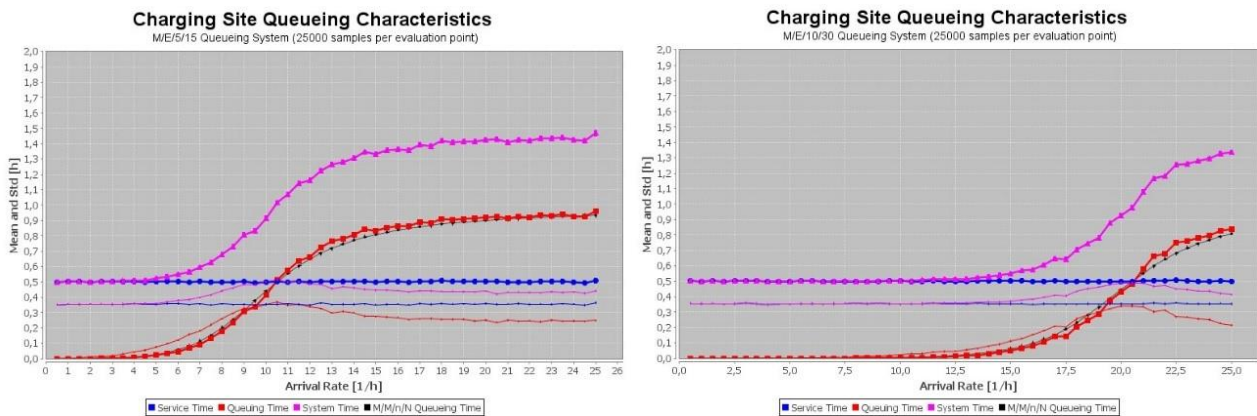
the fast-charging curve of EVs, which is not constant, in particular not when the SoC rises above 80% (Witt 2023). Key metrics of the individual charging and the aggregated electricity grid load are recorded, analysed and visualised to reveal the charging site's performance as exemplarily shown in the results presented next.



**Figure 4.** Random distributions characterising negative exponential distributed EV arrivals (left), Erlang-2 distributed charging times (middle) and Beta distributed charging demands (right)

**Exemplary simulation results.** Using the above outlined simulation of a finite multi-server queueing system, we can analyse EV charging sites as they soon may become deployed along motorways. According to the expected increase of EVs on the road, also the traffic load served by these charging sites will accordingly rise.

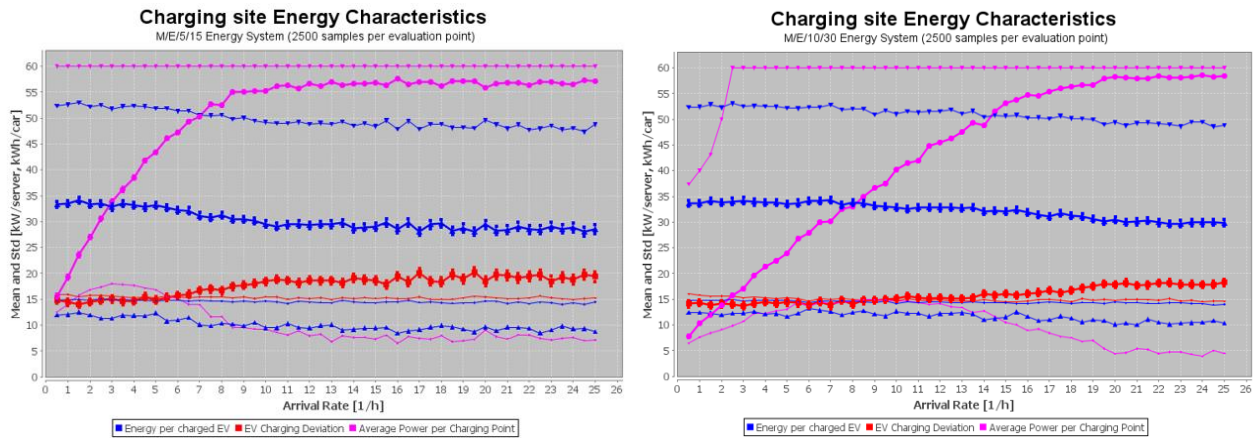
The recorded performance samples are statistically analysed and visualized to determine the efficiency and effectiveness of different charging site configurations. For the EV drivers, the waiting times and the amount charged in the intended charging time are of prime relevance. For the charging site operator and the adjacent distribution grid operator, the aggregated load and the efficiency of the peak power limitation provided by the dynamic load limiting implemented by smart charging are of prime concern. In Figure 5 the waiting time performance for two differently configured charging sites is compared.



**Figure 5.** Waiting time over increasing EV arrival rate for two exemplary charging site configurations

The charging site shown on the left side provides five charging points and ten waiting places, the site on the right has ten charging points and twenty waiting spaces, i.e., twice the size. In consequence, the site shown on the left becomes overloaded at an arrival rate of ten EVs per hour, whereas the site on the right remains uncritical for up to twenty EV arrivals per hour. Practically, waiting times up to few minutes are commonly accepted by the EV drivers, thus five charging points yield in average acceptable waiting times up to six EV/h and ten charging points for up to 15 EV/h, which shows the efficiency of size effect that here twice the number of charging points can more than twice the traffic load with the same performance.

Figure 6 shows the change of the site power demand for increasing EV arrival rates.

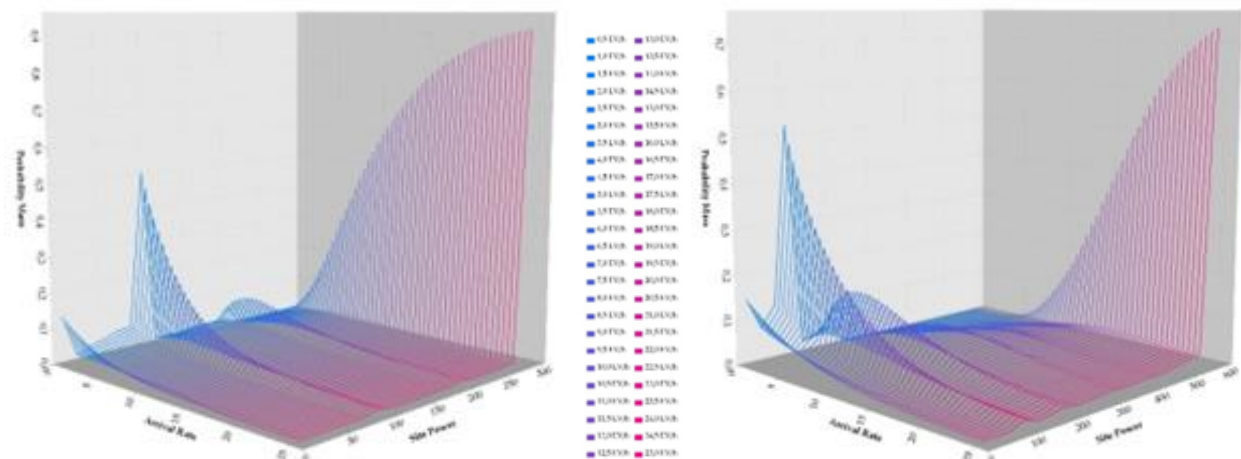


**Figure 6.** Waiting time over increasing EV arrival rate for two exemplary charging site configurations

The bold blue line in Figure 6 (Energy per charged EV) shows the average amount of energy charged. This amount drops slightly with increasing traffic load (EV/h) due to the dynamic site power limitation. The narrow blue lines with triangle markers above and below show the 90% and 10% percentiles of the samples. The bold red line shows the average battery capacity not charged when the EV leaves the charging site after completing the randomly determined charging time. This curve shows the opposite behaviour compare to the blue line because the sum of the two needs to equal the mean battery capacity of the charged EVs. We see that in average the battery is loaded to nearly 80%, which indicates that the simulated mean charging time is in average sufficient.

The bold magenta line in Figure 6 (Average power per charging point) shows the average power consumed by each charging point. It is calculated as  $\bar{P} = \frac{\sum P_i}{n}$  and increases as the frequency of vehicle arrivals increases causing parallel charging events. Because the total site power is limited, which is shown by the narrow magenta line on the top, the dynamic power limiting (smart charging) causes that it levels off slightly below the maximum possible. For the smaller dimensioned charging site on the left side in Figure 6 this line rises rapidly as arrival rates (the frequency of cars arriving for charging) increase. Doubling the capacity shifts the rise to higher arrival rates, which is shown on the right side.

The narrow lines without markers show the standard deviation in respect to the equally coloured results. A higher deviation indicates higher variability and accordingly less stable charging performance. To better visualise the variance of the site power, Figure 7 shows the histograms of the site power for increasing arrival rates.



**Figure 7 –** Site power demand histograms over arrival rate for the two exemplary charging site configurations

Notably, we see ripples caused by the maximum power the charging points offer, here 100 kW each, which are in particular visible at low arrival rates. The EVs per se, could charge with higher power. Still, the maximum site power, which for five charging points is limited to 300 kW (left side) and to 600 kW for ten charging points (right side), is in both cases dominant at high arrival rates indicating a reduced charging performance per EV.

If we configure a charging site with five charging points that provide 150 kW each and 500 kW maximum site power, and consider a more diverse mix of EV types including 30% hybrid EVs that cannot use fast charging, i.e., are limited to 22 kW maximum charging power, we see in Figure 8 that the site power limit is not so often reached even though five times 150 kW would be beyond the site power limit. This example represents a good charging site configuration, where performance and effort are better balanced.

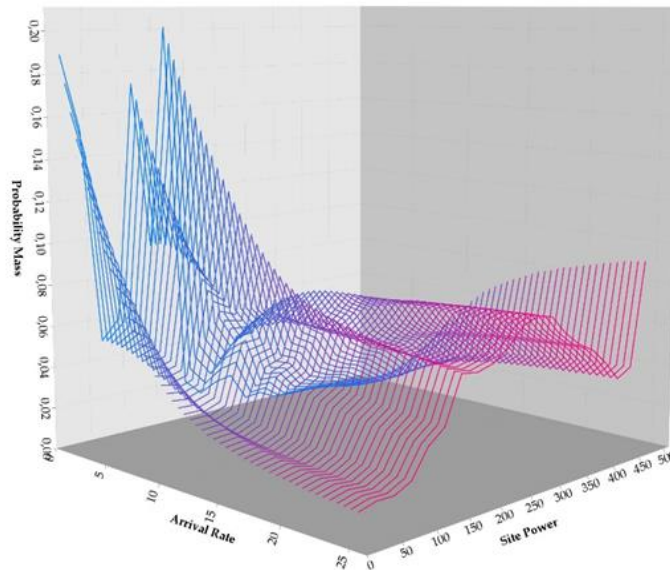


Figure 8 – Site power histogram over increasing EV arrival rate for a well-configured charging site

By analysing the different performances of different charging site configurations, potential bottlenecks and areas for improvement in the EV charging infrastructure can be identified. Visualization of key statistical measures such as mean, standard deviation, and confidence intervals are utilized to provide a comprehensive view of the system's performance for increasing traffic loads.

**Comparison with analogous simulation options.** The customised tool fulfils the criteria required to directly generate publishable R&D results, as requested. Tables 1 to 3 show that analogous simulation tools provide different features but not the precise set required by the client, the *eAlloc* project team, for whom the software tool is developed exclusively. The tools compared serve quite different purposes, including a PhD thesis on the topic (Jansson 2022), simple visualisation and on-line calculators for multi-server queueing systems (Lau 2017, Kabeer 2021, Ansari 2019) and complex real-world simulation tools (Witt 2022, Lee 2019, Liu 2021) interlinked with other simulation tools.

Table 1 reveals that no other tool supports selecting a distribution function to model the arrival and charging time distribution. KTH and EVCsim import pdfs and patterns respectively, which is very versatile. Witt and ACN rely exclusively on gathered field data, which focuses the studies on real scenarios of today. Smart Charging in respect to the grid access limitation  $P_{max}^{site}$  was nowhere found. Most studies considering the energy demand focus primarily on energy costs but consider no electricity grid access power limitation or peak power costs to be paid by the site owner.

**Table 1.** Features of exemplary EV charging simulation tools

| features<br>---<br>tool | traffic load and charging time modelling |     |      |         |          | charging technology |    |                  |
|-------------------------|--|-----|------|---------|----------|---------------------|----|------------------|
|                         | mean                                     | pdf | data | pattern | function | AC                  | DC | $P_{site}^{max}$ |
| eAlloc                  | +  | -   | -    | -       | +        | +                   | +  | +                |
| KTH                     | -  | +   | -    | -       | -        | +                   | -  | -                |
| Witt                    | -  | -   | +    | -       | -        | +                   | +  | -                |
| ACN                     | -  | -   | +    | -       | -        | +                   | -  | -                |
| EVQS                    | +  | -   | -    | -       | -        | +                   | -  | -                |
| MSQS                    | +  | -   | -    | -       | -        | +                   | -  | -                |
| Qsim                    | +  | -   | -    | -       | -        | +                   | -  | -                |
| EVCsim                  | -  | -   | -    | +       | -        | +                   | -  | -                |

Table 2 reveals that the waiting time  $T_{wait}$  is evaluated by most tools, only MSQS focuses on the charging time  $T_{ch}$ . Some tools also provide the waiting probability  $p_{wait}$  and EVQS the mean queue length  $L_{queue}$ . Naturally, the plain queueing system tools EVQS, MSQS, and Qsim do not yield information on the charging performance and the power demand. Those that model also the charging itself, focus on the charged energy  $E_{ch}$ , the SoC or  $\Delta E_{ch}$  at departure. The total power demand  $\sum P_{ch}$  is also a topic in most of the studies, even though no grid access limit is commonly considered.

**Table 2.** Simulation aims of exemplary EV charging simulation tools

| aims<br>---<br>tool | queueing performance |            |             |          |              | charging performance |     |               |
|---------------------|----------------------|------------|-------------|----------|--------------|----------------------|-----|---------------|
|                     | $T_{wait}$           | $P_{wait}$ | $L_{queue}$ | $T_{ch}$ | $T_{in-out}$ | $E_{ch}$             | SoC | $\sum P_{ch}$ |
| eAlloc              | +                    | -          | -           | +        | +            | +                    | +   | +             |
| KTH                 | +                    | +          | -           | +        | -            | +                    | +   | -             |
| Witt                | +                    | -          | -           | -        | -            | +                    | -   | +             |
| ACN                 | +                    | -          | -           | -        | -            | -                    | +   | +             |
| EVQS                | +                    | -          | +           | -        | -            | -                    | -   | -             |
| MSQS                | -                    | -          | -           | +        | +            | -                    | -   | -             |
| Qsim                | +                    | +          | -           | +        | -            | -                    | -   | -             |
| EVCsim              | +                    | +          | -           | +        | -            | +                    | -   | -             |

Table 3 shows that all except one lack the scientific rigour to include information on the stochastic quality of the derived statistical results: Non includes confidence intervals (ci) or standard deviation (std), only EVCsim (Liu 2021) states at least mean and maximum, as well as the median. Lacking access to the source code of some tools, some properties could not be determined and needed to be derived from the provided results. ACN-Sim does not bother to perform a statistical evaluation and delivers simply the gathered traces, delegating the data processing to the user.

The plain queueing system tools without charging performance and power demand considerations assume negative exponential inter-arrival and service time distributions, i.e., a Markovian system, for which the mean values can be analytically calculated (Kleinrock 1975). K. Lau mentions in the presentation in the EVQS (Lau 2017) documentation on GitHub that in practice the service time appears not to be negative exponential distributed and suggests to use a Gamma distribution instead. Actually, the empirically achieved histogram presented shows similarity with an Erlang-7 distribution, which is more smooth than the Erlang-2 distribution assumed for the above presented results.

**Table 3.** Outputs of exemplary EV charging simulation tools

| outputs<br>---<br>tool | data quality |    |     |        |       | means and formats |     |     |
|------------------------|--------------|----|-----|--------|-------|-------------------|-----|-----|
|                        | mean         | ci | std | traces | peaks | GUI               | CSV | SVG |
| eAlloc                 | +            | +  | +   | -      | +     | +                 | +   | +   |
| KTH                    | +            | -  | -   | -      | -     | +                 | -   | -   |
| Witt                   | +            | -  | -   | -      | -     | -                 | -   | -   |
| ACN                    | -            | -  | -   | +      | -     | +                 | -   | -   |
| EVQS                   | +            | -  | -   | -      | -     | +                 | -   | -   |
| MSQS                   | +            | -  | -   | -      | -     | +                 | +   | -   |
| Qsim                   | +            | -  | -   | -      | -     | +                 | -   | -   |
| EVCsim                 | +            | -  | -   | -      | +     | +                 | +   | -   |

**Conclusions.** The customised charging site simulation tool developed, attempts to fill the existing research gap by enabling the user to predict the demand for, and performance of, fast charging along motorways. It focuses on key parameters such as the aggregated power demand and expected charging performance, and the expected waiting time and need for parallel recharging facilities (charging points). Calculated expectable waiting times are for example used for the multi-objective EV assignment optimizer, developed as part of the R&D project *eAlloc* (eAlloc 2021), to consider the EV driver's common wish not to wait for the recharging. The presented study seeks to enhance the current understanding of fast EV charging infrastructure requirements and to facilitate more efficient and user-friendly charging solutions and related services that support widespread EV adoption. The project objectives are:

- to consider the dynamic power demand of fast charging,
- to simulate charging sites offering many charging points in parallel,
- to conveniently visualize the statistical results,

such that both, the characteristics of different charging site configurations and the electricity grid access demands, become understandable and apparent. The results show the demand for the number of charging points at charging sites along motorways and the expected power demand for any future EV traffic load. Thereby, the strategic planning for electric vehicle infrastructure development is supported and stakeholders can make informed decisions regarding the deployment of charging infrastructure extensions to meet future needs.

The developed tool enables future research that provides valuable insights into the dynamics and requirements of smart DC charging as it will probably be required along motorways in the near future when mobility is primarily based on electric vehicles. The presented approach shows a comprehensive tool to better understand the properties and performance issues of EV charging sites along motorways. The findings highlight the importance of adequate planning and investment in charging infrastructures to conveniently accommodate the anticipated electrification of the mobility sector. High-quality visualization of the analysed site's performance using vector graphics provides the quality necessary for high-quality publications in scientific media.

The event-based simulation of the finite multi-server queueing system that models a charging site enables thorough statistical analysis, including not only mean values but also confidence intervals (ci) and higher moments, e.g., the standard deviation. Approximating possible future traffic loads by generated random traffic using different distribution functions that can be chosen to mimic the statistical properties (mean and histogram) of recorded traffic data enables upscaling by simply increasing the mean value. Each implemented distribution can be checked by generating a histogram from a generated sample and comparing that with the analytically calculated pdf as shown in Figure 4. The event-based simulation itself can be checked by simulating a Markovian queueing system and comparing the statistically gained waiting time curve with the analytically calculated as shown in Figure 3.

**Acknowledgements.** The work presented has been conducted in the course of the national R&D project *eAlloc* funded by the Austrian Climate and Energy Fund (KLIEN) in the 3<sup>rd</sup> call of the *Zero Emission Mobility* program, administered by the Austrian Research Promotion Agency (FFG) under project No 885026.

**Conflict of interest.** All authors are with academia and there exists no conflict of interest.

## References

- Ansari A. (2019). Queueing Simulation – A Java program that simulates the working of a multi-server queueing model. Retrieved from: <https://github.com/psyclone20/Queueing-Simulation>.
- Csanyi E. (2028). The essentials of electrical distribution systems every engineer should know. *Electrical Engineering Portal*. Retrieved from <https://electrical-engineering-portal.com/electrical-distribution-systems>
- eAlloc (2021). *Dynamically Optimizing the Allocation of e-cars to Charging Sites*. R&D project, 2021–2025, Austrian Climate and Energy Fund (KLIEN), 3rd call Zero Emission Mobility. Retrieved from <https://projekte.ffg.at/projekt/4031006>.
- Erbach, G., & Jensen, L. (2024). Fit for 55 package. In EPRS Briefings (PE 733.513), European Parliamentary Research Service. Retrieved from [https://www.europarl.europa.eu/thinktank/en/document/EPRS\\_BRI\(2022\)733513](https://www.europarl.europa.eu/thinktank/en/document/EPRS_BRI(2022)733513).
- EUR-Lex (2023). Regulation (EU) 2023/1804 of the European Parliament and of the Council of 13 September 2023 on the deployment of alternative fuels infrastructure, and repealing Directive 2014/94/EU (AFIR). *Official Journal of the European Union*, article number 32023R1804. Retrieved from <http://data.europa.eu/eli/reg/2023/1804/oj>.
- Eurostat Statistics Explained (2021). *Passenger mobility statistics* in ISSN 2443-8219. Retrieved from [https://ec.europa.eu/eurostat/statistics-explained/index.php?title=Passenger\\_mobility\\_statistics](https://ec.europa.eu/eurostat/statistics-explained/index.php?title=Passenger_mobility_statistics).
- Grigorev A., Mao T., Berry A., Tan J., Purushothaman L., & Mihaita A.-S. (2021), How will electric vehicles affect traffic congestion and energy consumption: an integrated modelling approach. In *2021 IEEE International Intelligent Transportation Systems Conference – ITSC* (pp. 1635-1642), Indianapolis, IN, USA, doi: [10.1109/ITSC48978.2021.9564561](https://doi.org/10.1109/ITSC48978.2021.9564561).
- Huang Q., Yang L., Hou C., Zeng Z., & Qi Y. (2023). Event-Based EV Charging Scheduling in a Microgrid of Buildings. *IEEE Transactions on Transportation Electrification*, 9(1), 1784–1796. doi: [10.1109/TTE.2022.3201084](https://doi.org/10.1109/TTE.2022.3201084).
- International Energy Agency (2024). Trends in electric vehicle charging. *Global EV Outlook 2024*, Retrieved from <https://www.iea.org/reports/global-ev-outlook-2024>.
- Jansson D., & Niklasson N. (2022). Dimensioning of Charging Infrastructure Using Model-Based Systems Engineering. (Doctoral dissertation, KTH, School of Engineering Sciences (SCI), Mathematics (Dept.), Mathematics (Div.), Stockholm, Sweden). Retrieved from <https://urn.kb.se/resolve?urn=urn:nbn:se:kth:diva-325912>.
- Kabeer A.A. (2021). Multi-Server Queueing System (C++ Simulation). Retrieved from: <https://github.com/zedrex/msqs>.
- Kleinrock L. (1975). *Queueing Systems – Volume 1: Theory*. Wiley-Interscience, ISBN: 978-0-471-49110-1.
- Lau K. (2017). Electric Vehicle Queuing Simulation – A visualization app to model electric vehicle traffic intensity. Retrieved from <https://kenlau177.github.io/Electric-Vehicle-App/>.
- Lee Z.J., Johansson D., & Low S.H. (2019). ACN-Sim: An Open-Source Simulator for Data-Driven Electric Vehicle Charging Research. In *2019 IEEE International Conference on*

- Communications, Control, and Computing Technologies for Smart Grids – SmartGridComm (pp. 1–6). Beijing, China: IEEE Communications Society, arXiv: 2012.02809v2.
- Liu W., Shi X., Zhao J., Zhang X.-P., & Xue Y. (2021). Electric Vehicle Charging Simulation Framework Considering Traffic, User, and Power Grid. *Journal of Modern Power Systems and Clean Energy*, 9(3), 602–611, doi: [10.35833/MPCE.2020.000208](https://doi.org/10.35833/MPCE.2020.000208).
- Rahila J., Soundra Devi G., Radhika A., & Singh G. (2024). Electric vehicle smart charging with network expansion planning using hybrid COA-CCG-DLNN approach. *Optim Control Appl Meth.*, 45(4), 1524-1545. doi: [10.1002/oca.3108](https://doi.org/10.1002/oca.3108).
- Sirapa S., Bivek B., Malesh S., Sailesh C., Bim P.S. (2022). Measures to resolve range anxiety in electric vehicle users. *International Journal of Low-Carbon Technologies*, 17, 1186–1206, doi: [10.1093/ijlct/ctac100](https://doi.org/10.1093/ijlct/ctac100).
- Tang C., Tukker A., Sprecher B., & Mogoll'on J.M. (2023). Assessing the European electric-mobility transition: Emissions from electric vehicle manufacturing and use in relation to the EU greenhouse gas emission targets. *Environmental Science & Technology*, 57(1), 44–52, pMID: 36574507. doi: [10.1021/acs.est.2c06304](https://doi.org/10.1021/acs.est.2c06304).
- Witt A. (2023). Determination of the number of required charging stations on a German motorway based on real traffic data and discrete event-based simulation. *LOGI – Scientific Journal on Transport and Logistics*, 14(1), 1–11, doi: [10.2478/logi-2023-0001](https://doi.org/10.2478/logi-2023-0001).
- Yang D., Sarma N.J., Hyland M.F., & Jayakrishnan R. (2021). Dynamic modeling and real-time management of a system of EV fast-charging stations. In *Transportation Research Part C: Emerging Technologies*, vol. 128, paper number 103186. Retrieved from <https://www.sciencedirect.com/science/article/pii/S0968090X21002023>.
- Yang X.-S. (2018). Queueing Theory and Simulation. In *Optimization Techniques and Applications with Examples* (pp. 227–247), Wiley, doi: [10.1002/9781119490616.ch10](https://doi.org/10.1002/9781119490616.ch10).

УДК 004.7

## Багатовимірна матриця класифікації інформації для оцінки ризиків інформаційної безпеки

**Тетяна Іванівна Коробейнікова**

к.т.н., доцент кафедри безпеки інформаційних технологій  
Національний університет «Львівська Політехніка»  
79013, м. Львів. вул. Степана Бандери, 12  
ORCID 0000-0003-2487-8742

**Андрій Богданович Ямнич**

аспірант кафедри безпеки інформаційних технологій  
Національний університет «Львівська Політехніка»  
79013, м. Львів. вул. Степана Бандери, 12  
ORCID 0009-0005-7226-1896

**Анотація.** У даній роботі розглядається одна з ключових задач щодо комплексної системи оцінки ризиків інформаційної безпеки для персоналу під час розмежування доступу до інформаційних ресурсів компанії. Актуальність дослідження підтверджується численними випадками витоків інформації, які демонструють недостатню ефективність традиційних методів класифікації та контролю доступу. Дослідження передбачає аналіз існуючих стратегій класифікації інформаційних ресурсів компанії та розробка додаткового методу, що базується на постійному аналізі доступу і динамічному коригуванні класифікації таких ресурсів. Для досягнення цієї мети було використано методи аналізу поточних стратегій класифікації інформації, комбінування різних методів класифікації та впровадження графічного методу поєднання традиційної класифікації ресурсів із динамічною складовою за допомогою багатовимірної матриці. Основні результати дослідження передбачають розробку вдосконаленого методу, що дозволяє постійно аналізувати доступ персоналу до інформаційних ресурсів компанії і в динамічному режимі коригувати класифікацію інформаційних ресурсів залежно від правил розмежування цього доступу. Запропонований підхід дозволяє включати довільну кількість показників на графіку у вигляді набору векторів та в подальшому розраховувати загальні оцінки ризиків на основі суми або різниці цих векторів. Практична цінність роботи полягає в тому, що запропонований підхід дозволяє в повній мірі використовувати сучасні технології контролю доступу і може бути основою для подальших досліджень, наприклад, таких, які підтримують автоматизовану класифікацію інформації за допомогою тренування нейронних мереж. Крім того, у рамках даного дослідження було проведено детальний огляд існуючих методів оцінки ризиків інформаційних ресурсів компанії та виявлено ключові недоліки, що притаманні традиційним підходам. Зокрема, були проаналізовані методи, що базуються на фіксованих рівнях доступу та використанні статичних правил для контролю доступу. Виявилось, що такі методи не здатні адекватно реагувати на динамічні зміни в поведінці користувачів та зміни у важливості інформаційних ресурсів. Таким чином, запропонований підхід дозволяє забезпечити більш гнучкий та адаптивний контроль доступу до інформаційних ресурсів. Це досягається шляхом постійного моніторингу доступу та автоматичного коригування рівнів доступу на основі аналізу поведінкових даних користувачів та зміни контексту використання інформаційних ресурсів.

**Ключові слова:** динамічна класифікація інформації, візуалізація стану критичних ресурсів, багатовимірна матриця класифікації, класифікаційний стек, інтегральні оцінки ризиків.

## Multidimensional classification matrix for information security risk assessment

**Tetiana I. Korobeinikova**

PhD, Associate Professor of Information Technology Security Department  
Lviv Polytechnic National University  
Stepana Bandery St, 12, Lviv, Ukraine, 79000  
ORCID 0000-0003-2487-8742

**Andrii B. Yamnych**

Postgraduate Student of Information Technology Security Department  
Lviv Polytechnic National University  
Stepana Bandery St, 12, Lviv, Ukraine, 79000  
ORCID 0009-0005-7226-1896

**Abstract.** In this study, we address one of the key challenges related to a comprehensive risk assessment system for information security concerning personnel during access delineation to company information resources. The relevance of this research is confirmed by numerous instances of information leaks, which highlight the insufficient effectiveness of traditional classification and access control methods. The research aims to analyze existing classification strategies for company information resources and develop an additional method based on continuous access analysis and dynamic adjustment of resource classification. To achieve this goal, we employed methods such as analyzing current information classification strategies, combining various classification techniques, and implementing a graphical method that combines traditional resource classification with a dynamic component using a multidimensional matrix. The main results of the study involve the development of an enhanced method that allows continuous analysis of personnel access to company information resources and dynamic adjustments to resource classification based on access delineation rules. The proposed approach allows for the inclusion of any number of indicators in a graph as a set of vectors, subsequently calculating overall risk assessments based on the sum or difference of these vectors. The practical value of this work lies in its ability to fully utilize modern access control technologies and serve as a foundation for further research, such as automated information classification using neural network training. Additionally, within this study, we conducted a detailed review of existing risk assessment methods for company information resources, identifying key limitations inherent in traditional approaches. Specifically, we analyzed methods based on fixed access levels and the use of static rules for access control. It became evident that such methods are inadequate in responding to dynamic changes in user behavior and the evolving importance of information resources. Thus, the proposed approach allows for more flexible and adaptive access control to information resources, achieved through continuous access monitoring and automatic adjustments based on behavioral user data and contextual changes in resource utilization.

**Keywords:** dynamic information classification, visualization of critical resources, multidimensional classification matrix, classification stack, integral risk assessments.

**DOI:** <https://doi.org/10.31649/1999-9941-2024-60-2-91-106>

**Вступ.** Центральним ризиком інформаційної безпеки завжди є несанкціонований доступ до критичної інформації та/або її витік за межі організації (Velmurugan et al., 2024; Pitafi et al., 2023; Al Qahtani et al., 2024). Залежно від того, наскільки критичною була інформація (Mikuletič et al., 2024), збитки можуть варіюватися від звичайних до ланцюжка наслідків, які потягнуть шкоду не лише для конкретної організації, а й для її клієнтів, контрагентів, а іноді навіть для громадян, які виявилися супутніми жертвами витоку

(Wiedemann et al., 2024; Alotibi, 2024) (особливо якщо йдеться про персональні дані). Традиційним способом запобігання або мінімізації таких випадків є класифікація інформації та обмеження доступу до неї лише певним колом довірених осіб (Ramamurthy et al., 2022; Wang & Gu, 2024).

Процес класифікації може бути ускладнений додатковими факторами, як-от рівнем доступу персоналу до інформаційних ресурсів підприємства. Приватні організації часто економлять на персоналі, і кожен працівник в них може відповідати одночасно за кілька напрямків. Це створює значні ризики для інформаційної безпеки (Shmatko et al., 2020).

Таким чином, існує потреба додати у комплексну систему оцінки ризиків інформаційної безпеки (для персоналу під час розмежування доступу до інформаційних ресурсів компанії) динамічну складову класифікації інформації. Це дозволить визначити, які саме ресурси є найбільш затребуваними, які саме ресурси найчастіше делегуються тимчасовим працівникам і, нарешті, які з них необхідно цифровізувати в першу чергу.

Метою роботи є аналіз та вдосконалення способів візуалізації стану критичних інформаційних ресурсів на підприємстві за допомогою динамічного підходу із застосуванням багатовимірної матриці класифікування інформації для оцінки ризиків інформаційної безпеки.

Для досягнення поставленої у даній роботі мети необхідно: 1. виконати аналіз засад класифікації інформації та обґрунтувати доцільність динамічного підходу; 2. запропонувати динамічну класифікацію на базі аналізу доступу до ресурсів; 3. запропонувати багатовимірну матрицю класифікації ресурсів.

Наукова новизна роботи: пропонується багатовимірна матриця класифікування інформації для оцінки ризиків інформаційної безпеки, яка, на відміну від аналогів, дозволяє візуалізувати кожен ресурс у формі вектора, пропорційному максимальній зміні динамічного показника, спрямованого в бік цієї зміни. Це дозволить візуалізувати миттєві піки для малозатребуваних ресурсів.

**Огляд літератури.** В роботі (Velmurugan et al., 2024) зазначається, що класифікація інформації є важливим етапом для запобігання витоків даних. У праці (Pitafi et al., 2023) подано дослідження про технічне убезпечення витоків даних, а у роботі (Al Qahtani et al., 2024) основна увага приділена організаційним та соціально-інженерним причинам витоків даних. Також, під час роботи із персональними даними часто особисті переконання та етика стають чинниками безпеки, про що йдеться у статті (Mikuletič et al., 2024), яка досліджує методи захисту конфіденційних медичних даних. Часто в результаті втрати чутливих даних жертвами є треті особи, що досліджують (Wiedemann et al., 2024) та (Alotibi, 2024). Таким чином, провідні учені (Ramamurthy A. et al., 2022), (Wang & Gu, 2024) пропонують обмежити доступ до чутливої конфіденційної інформації, водночас поділивши її на певні групи, тобто, пропонують класифікувати інформацію та обмежити до неї доступ.

Однак у державних підприємствах та військових установах такий підхід підкріплено додатковими заходами безпеки та загрозою кримінальної відповідальності, а на рівні приватних підприємств він базується лише на персональній відповідальності працівника. Як правило, підписується лише спеціальне доповнення до трудового договору, яке обумовлює відповідальність за порушення режиму доступу до даних або розголошення комерційної таємниці. Якщо підприємство має доступ до державних даних у рамках державно-приватного партнерства (зазвичай, це стосується персональних даних), то різниця між рівнями забезпечення режиму секретності на державному та організаційному рівнях може створити значну дірку в системі безпеки даних (Arslan et al., 2024; Emmanuel et al., 2024; Venn et al., 2024; Song et al., 2024).

Тому приватні підприємства значно більше залежать від точності класифікації інформації, а також більш щільного контролю за доступом до конкретних ресурсів. Чим більшою і різноманітнішою інформаційною базою володіє організація, тим складнішою буде

задача класифікації ресурсів, що до неї входять (Barnawi et al., 2024, Gambarelli et al., 2021; Mazzola et al., 2021).

Загалом, існує тенденція до розвитку досліджень, зосереджених на підвищенні точності класифікації інформаційних ресурсів та адаптації підходів до сучасних викликів інформаційної безпеки.

**Матеріали та методи.** Основні етапи роботи включали аналіз сучасних методів класифікації інформаційних ресурсів, обґрунтування доцільності використання динамічної класифікації, розробку багатовимірної матриці класифікації ресурсів. Методологія включала використання статистичних методів для аналізу даних про доступ до ресурсів, а також розробку графічних інструментів для візуалізації динаміки доступу. Аналіз сучасних методів класифікації інформаційних ресурсів починався з вивчення існуючих підходів до класифікації інформації в різних організаціях, зокрема державних підприємствах та приватних компаніях. Було досліджено стандарти ISO 27001 та QGISCF, що застосовуються для класифікації інформаційних ресурсів (QGISCF, 2023; ISO 27001, 2022). Вивчалися методи класифікації за критеріями чутливості, цілісності та доступності інформації.

Обґрунтування доцільності використання динамічної класифікації базувалося на вивченні динаміки доступу до інформаційних ресурсів на основі даних журналювання доступу в сучасних файлових системах. Було проведено порівняння статичних методів класифікації, які використовують фіксовані точки перегляду, з динамічними методами. Виявлено ключові показники, які впливають на критичність інформаційних ресурсів, такі як частота доступу, протокол доступу та тип користувачів.

Розробка багатовимірної матриці класифікації ресурсів включала створення багатовимірної матриці, що поєднує різні показники критичності ресурсів, включаючи частоту доступу, рівень конфіденційності та ступінь цифровізації. Використовувалися методи візуалізації для відображення змін частоти доступу та інших динамічних показників у вигляді графіків і діаграм. Було розроблено методологію для адаптації класифікації в режимі реального часу на основі даних про доступ до ресурсів.

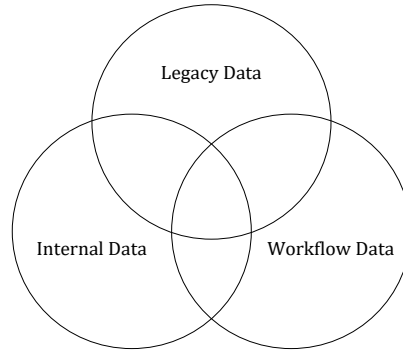
Опис стратегії та критеріїв формування вибірки включав вибірку інформаційних ресурсів на основі їх значущості для діяльності організації. Визначалися ключові показники для аналізу, такі як кількість користувачів, типи доступу та частота запитів до ресурсів. Використовувалися статистичні методи для аналізу даних вибірки та оцінки ризиків інформаційної безпеки.

Експериментальна база дослідження включала проведення експериментів з використанням журналювання доступу до інформаційних ресурсів у реальних умовах організації. Аналізувалися отримані дані для виявлення патернів доступу та аномалій, що можуть свідчити про потенційні загрози інформаційній безпеці. Отримані результати використовувалися для коригування методів класифікації та підвищення ефективності заходів безпеки.

Для забезпечення повної картини ходу дослідження використовувалися додаткові методи. Метод статистичного аналізу застосовувався для обробки великих обсягів даних про доступ до інформаційних ресурсів, а також для виявлення статистично значущих відхилень у патернах доступу. Графічні методи візуалізації використовувалися для представлення результатів дослідження у вигляді діаграм та графіків, що дозволяло наочно відобразити динаміку доступу до ресурсів та виявлені аномалії.

## **Результати**

**Аналіз засад класифікації інформації та обґрунтування динамічного підходу.** Перш ніж розглядати класифікацію інформації з безпекового ракурсу, необхідно зрозуміти, як виглядає структура даних в межах організації. Простіше за все розглядати їх з точки зору щоденного робочого процесу. Тоді всі дані підприємства можуть бути поділені на три групи (Malchiodi et al., 2024; Zhang et al., 2018; Robinson, 2024) (рис. 1).



**Рисунок 1.** Класифікація даних за робочим процесом

**Джерело:** розроблено авторами на основі досліджень (Malchiodi et al., 2024; Zhang et al., 2018; Robinson, 2024)

*Legacy Data (архівні дані).* Архівні дані та документи минулих періодів, які можуть зберігатись в повністю нецифровій формі або в форматах, які втратили актуальність і не є сумісними із сучасними інформаційними системами. Залежно від значущості цих даних для поточної роботи, ці дані можуть бути як повноцінними, структурованими документами (що дозволить зберегти історичність даних, відстежувати тенденції в роботі установи чи комерційної фірми, застосовувати Big Data для аналізу тощо), або вноситись в системи обліку як певний підсумковий залишок. Як правило, це найбільш архаїчна частина інформаційної бібліотеки, із найменшим відсотком цифровізації;

*Internal Data (внутрішні дані).* Внутрішня документація, яка використовується у повсякденній роботі, але не виходить за межі підприємства. Зазвичай, це накази, положення, циркуляри та інша нормативно-адміністративна документація. Тут найчастіше зустрічається цифровізація, оскільки вони можуть існувати в формі обміну повідомленнями через електронну пошту чи корпоративні системи спільної роботи;

*Workflow Data (потокові дані).* Робоча документація, яка використовується у виробничій діяльності і є офіційною документацією, яка може бути використана контролюючими, судовими та іншими установами як правочинна чи підставова документація. Ці дані діляться на первинні (створюються або вносяться працівниками як результат їх професійної діяльності) та вторинні (формуються з даних, взятих з первинних документів). Вони можуть бути цифровізовані як медіа-документи (зазвичай, сканлейти або фотокопії) та/або замінені стандартними електронними документами в форматах текстових процесорів (таких як Microsoft Office, OpenOffice та ін.). Але вони також можуть бути на папері, у формі записок, ордерів та інших щоденних розхідних документів.

Набір ресурсів, які знаходяться на перетині всіх трьох груп, як правило, є найбільш критичними для функціонування підприємства (Robinson, 2024). Сюди може входити як аудиторська інформація (в ній "сходяться" всі найбільш важливі показники діяльності), так і інформація, необхідна для забезпечення виробництва або просто щоденної роботи інформації (наприклад, персональні дані клієнтів, замовлення тощо).

В силу різної специфіки роботи різних установ межі класифікації документів на вищезгадані групи не є чіткими, тому обсяги перетинів можуть бути різними, а це впливає як на критичність інформації, так і на безпекову стратегію та стратегію цифровізації. Тому як правило, її доповнюють додатковими класифікаторами, які дозволяють більш чітко окреслити межі і більш чітко оцінити критичність.

Підходів до створення конкретизуючих класифікаторів три:

1. на основі змісту інформації. Відповідно, така класифікація має передбачати методику оцінки чутливості чи критичності її для функціонування організації. Наприклад, в стандарті Queensland Government Information Security Classification Framework (QGISCF) (Queensland Government, 2024), який застосовується в Австралії, для подібних класифікаторів рекомендується інтегрований параметр Business Impact Level, який

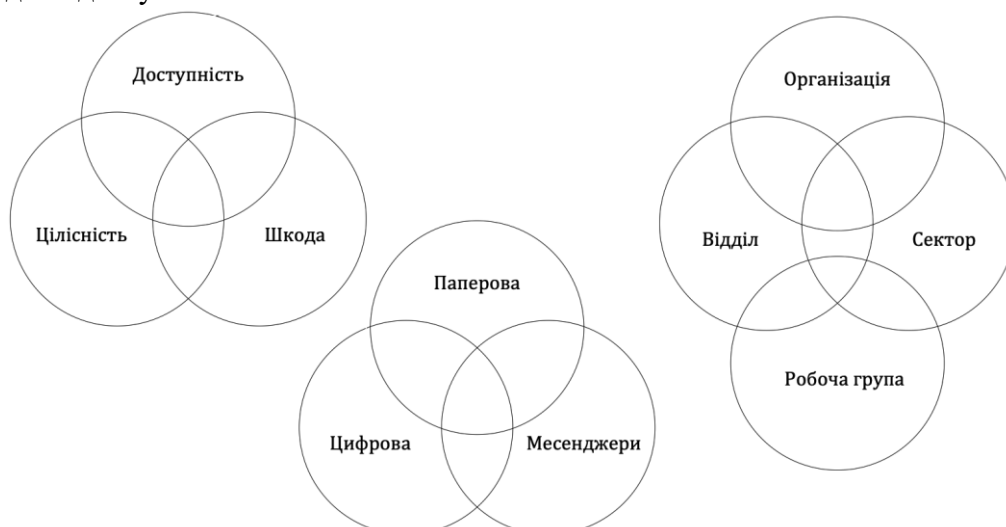
- складається із трьох параметрів – потенційна шкода, цілісність/частковість та доступність інформації;
- на основі застосування інформації. Відповідно, так класифікація повинна відповідати технологічним, організаційним, або навіть кадровим ланцюжкам всередині підприємства. Як правило, йдеться про класифікацію внутрішньої інформації за організаційними відділами, а робочої документації за напрямками діяльності;
  - на основі користувача інформації. Така класифікація має базуватись на чіткому визначенні, яка саме інформація необхідна для тієї чи іншої роботи і кому саме. Знову ж таки, в рамках QGISCF виокремлюються власники (створювачі, розпорядники) інформації, користувачі, менеджмент, перевіряючі, офіцери безпеки даних (у випадку державно-приватного партнерства) та ін.

Можна згадати ще про підхід, який уже частково згадувався: на основі форми представлення інформації. Стандарт ISO 27001 (Irwin, 2022) містить спрощений варіант такої класифікації із чотирьох класів: паперові документи, цифрові документи на локальних/знімних носіях даних, цифрові документи у віддалених сховищах та електронні повідомлення в системах миттєвого обміну інформацією.

Всі ці класи мають нечіткі межі, і так само, як і класифікація ресурсів за робочим процесом, описується круговими діаграмами із перетинами. Таким чином, незалежно від типу і набору класифікаторів ресурсів, які використовуються в організації, вони всі мають нечіткі межі і перекриваються між собою. Інтегральна оцінка критичності ресурсу з точки зору інформаційної безпеки, відповідно, буде набором "перетинів перетинів", і критична інформація знаходитиметься в одному із них. Для інтегральної оцінки, як правило, використовують табличний метод. Рядки в ньому відповідають різним класифікаторам інформації, а стовпчики – одному із рівнів доступу. Параметри оцінюються за певною чисельною шкалою.

Можна скласти таблицю на основі вищенаведених типів класифікації, взявши систему рівнів доступу із стандарту ISO 27001, однак заповнення таблиці відповідати емпіричним показникам – за шкалою "низький-середній-високий-найвищий" (New Zealand Protective Security Requirements, 2024).

Найкраще табличний метод поєднує систему рівнів доступу із BIL (Business Impact Levels) – там лінійна залежність, що дозволяє застосувати систему балів – наприклад таку, яку використовують в Новій Зеландії (Aroga et al, 2023). Однак адекватно, на нашу думку, такий підхід дозволяє оцінити лише такі показники як шкода (чутливість інформації) та безпосередньо доступність.



**Рисунок 2.** Кругові діаграми додаткових класифікаторів інформації

**Джерело:** розроблено авторами на основі розробки Бюро безпеки урядового зв'язку Н. Зеландії (2024), досліджень (Aroga et al, 2023).

Найбільша доступність буде у публічної інформації, найменша – у конфіденційної. Найбільша шкода від розголошення конфіденційної інформації, а найменша – публічної. Решта показників сильно залежить від кількості документів, зокрема, цілісність.

Для демонстрації загального підходу достатньо буде неконкретизованого прикладу. Наприклад, в дослідженні пошукових систем, які використовують публічні та приватні дані, у якості прикладу інформаційно достатньої вибірки для висновку на основі аналізу приватних та публічних даних наведено електронні листи та новинарна стрічка Wikipedia – 47 тис. електронних листів та 5,2 млн. новин. Що дає нам співвідношення між публічними та приватними даними в 0,0011.

Емпірично можна припустити, що подібне співвідношення має існувати між будь-якою парою ступенів приватності. Тобто, на 10000 листів публічних документів має бути 11 листів приватних, із яких, в свою чергу, максимум по одному – обмеженого та конфіденційного характеру. Таке ж співвідношення можна використати як емпіричну оцінку цілісності – як умовну кількість листів, необхідних для видачі адекватного висновка. Поєднання системи рівнів доступу із застосуванням інформації таке: найбільш застосовною завжди є інформація організаційного рівня – і вона, як правило, завжди є публічною, максимум внутрішнього рівня. А обмежена і конфіденційна інформація не виходить за межі відділів та навіть окремих робочих груп.

Класифікацією в залежності від користувачів інформації складніша. Наприклад, зрозуміло, що автор чи власник інформації буде мати до неї доступ незалежно від присвоєного рівня. Користувачу доступ може бути наданий на підставі посади, але може фактично він може доступитись до критичних через суміщення посад чи пряму передачу цих даних від автора чи власника. Аудитори можуть мати постійний доступ, а можуть – тимчасовий. Нарешті, офіцер безпеки даних може відповідати не за всі дані свого рівня доступу, а тільки за певну їх підмножину – яка має відношення до державно-приватного партнерства (перетин класифікатора застосування інформації та користувача). Пітер Вуд в своїй статті для Computer Weekly (Wood, 2013) рекомендує комплексне оцінювання, засноване на оцінці мотивації та можливостей користувачів (акторів), які оцінюються по п'ятибальній шкалі плюс нуль як відсутність загрози. Це дає змогу побудувати таку матрицю оцінок загроз з боку акторів (табл. 1).

**Таблиця 1.** Матриця загроз з боку акторів

|           | Можливість   |           |      |          |        |       |
|-----------|--------------|-----------|------|----------|--------|-------|
|           |              | Дуже мала | Мала | Обмежена | Значна | Повна |
| Мотивація | Відсутня     | 0         | 0    | 1        | 1      | 2     |
|           | Зацікавлення | 0         | 0    | 1        | 2      | 3     |
|           | Інтерес      | 0         | 1    | 2        | 3      | 4     |
|           | Робота       | 1         | 1    | 2        | 4      | 4     |
|           | Фокус        | 1         | 3    | 3        | 4      | 5     |

**Джерело:** розроблено авторами на основі досліджень (Wood, 2013).

За допомогою цієї матриці ми можемо емпірично визначити, що загроза від автора завжди буде максимальна, тому що він має повний доступ до інформації і він сфокусований на темі. В той же час він практично не зацікавлений в розголосі, тому що це поставить під удар його інтереси. Таким чином, його загроза як актора буде варіюватись в залежності від того, наскільки сильним буде цей удар. Точно так же, в залежності від особистої шкоди, буде залежати загроза розголошення на рівні користувача.

Найбільш невизначеними будуть оцінки високих рівнів безпеки – тому що там вступають в дію фактори, які складно оцінити. Наприклад, критичність певного ресурсу в залежності від користувача в такому випадку емпірично можна оцінити так:

– публічна інформація критичною не є, так як користувач має завжди до неї доступ і у нього

- відсутня зацікавленість в розголосі (0);
- внутрішня інформація має 1-бальну критичність, тому що користувач має замалі можливості ознайомитись із нею (випадкові);
  - обмежена і конфіденційна інформація має 4-бальну критичність, перш за все через малі можливості користувача її отримати (максимум в форматі "необхідно знати"), однак як нам відомо з оцінки цілісності, достатньо всього 1 листа приватної інформації для отримання відповіді (нанесення шкоди).

Що, в свою чергу, ставить необхідність долучення до вищенаведеної матриці третій вимір – як мінімум шкоду для актора, яка має коригуватись відносно шкоди для підприємства.

Точно таким же чином можна оцінити і ризики від застосування інформації в залежності від організаційного рівня, для якого призначені документи.

Всі документи організаційного рівня мають найвищу критичність (оскільки шкода від них може вплинути на функціонування підприємства), а найменшу – документи рівня робочої групи, тому що вони демонструють лише матеріали проєктів, або навіть лише певну їх частину – якщо проєкт сильно диверсифікований.

Найбільші проблеми – якщо до оцінки додається форма представлення. Найбільш захищеними, як не дивно, є паперова і цифрова локальна форми, тому що доступ до них можна контролювати організаційними заходами – і відповідно, можна оцінити ризики інформаційної безпеки. А до віддалених сховищ доступ може мати невстановлена кількість осіб (наприклад, обслуговуючий персонал фірми, яка забезпечує хостинг даних), а електронна пошта та месенджери є безпеково слабкими сховищами (вони можуть бути підсилені шифруванням і політикою підприємства до рівня цифрових локальних сховищ, але це виходить за рамки тематики даної статті).

Нарешті, частина документів може дублюватись в різних формах і форматах – або з метою більшої збереженості, або просто для спрощення роботи (що, до речі, додатково ускладнює безпекове питання – тому що за копіями складніше услідкувати).

Матриця Пітера Вуда також дозволяє зробити таку оцінку – вона навіть наведена в статті (Wood, 2013). Зведемо всі вищенаведені оцінки в спільну таблицю 2.

**Таблиця 2.** Інтегральна оцінка критичності ресурсів

|                                       | Public | Internal | Restricted | Confidential |
|---------------------------------------|--------|----------|------------|--------------|
| <b>Business Impact Levels(BIL)</b>    |        |          |            |              |
| Доступність                           | 1      | 2        | 3          | 4            |
| Шкода                                 | 1      | 2        | 3          | 4            |
| Цілісність                            | 10000  | 11       | 1          | 1            |
| <b>Застосування інформації</b>        |        |          |            |              |
| Організаційний                        | 0      | 3        | 5          | 5            |
| Секторальний                          | 0      | 3        | 4          | 5            |
| Відділ                                | 0      | 3        | 3          | 5            |
| Робоча група                          | 0      | 3        | 2          | 5            |
| <b>Користувач інформації</b>          |        |          |            |              |
| Автор/Власник                         | 0      | 5        | 5          | 5            |
| Користувач                            | 0      | 1        | 4          | 4            |
| Перевіряючий                          | 0      | 3        | 5          | 5            |
| Офіцер БД                             | 0      | 2        | 5          | 5            |
| <b>Форма представлення інформації</b> |        |          |            |              |
| Паперова                              | 0      | 2        | 3          | 5            |
| Цифрова (локальна)                    | 0      | 2        | 4          | 5            |
| Цифрова (віддалена)                   | 0      | 3        | 5          | 5            |
| Електронна пошта                      | 0      | 4        | 5          | 5            |

**Джерело:** розроблено авторами на основі досліджень (Wood, 2013).

Ця таблиця добре показує, наскільки різноманітними є показники, які

використовуються для оцінки критичності інформаційних ресурсів. Також, вона показує, що таблична методика дозволяє ефективно поєднувати не більше двох, максимум трьох класифікаторів. Цього достатньо для забезпечення мінімальної інформаційної безпеки, але не достатньо для забезпечення її в належних обсягах.

Крім того, наскільки б не була продумана структура таблиці і наскільки деталізованими не були б класифікатори, вони всі представляють собою статичну оцінку "на певну фіксовану точку" – наприклад, на момент аудита, або на момент чергової посадової атестації.

В проміжку між цими фіксованими точками ані сама класифікація ресурсів, ані безпосередні переліки класифікованих ресурсів не переглядаються. Між тим, швидкість зміни ринкової кон'юнктури, законодавчого поля, навіть політичної інформації, дедалі зростає і уже зараз вона швидша, ніж будь-який період між аудитами чи атестаціями. Не кажучи про те, що між фіксованими точками може змінитись кількість людей, яким надано доступ до тієї чи іншої інформації. Таким чином, певний формат динамічної класифікації стає життєво необхідним.

**Динамічна класифікація на базі аналізу доступу до ресурсів.** Одним із перспективних варіантів динамічної класифікації ресурсів може стати профілювання доступу до них. Найкраще цей метод спрацює для цифрових документів, оскільки сучасні операційні системи дозволяють обмежувати доступ не лише в розрізі дисків та окремих каталожних елементів (папок, архівів, типів файлів тощо), а й для конкретно взятих ресурсів. Розглянемо лише два додаткові параметри: *частота і протокол доступу до ресурсу*.

Обидва ці параметри є універсальними і багатозначними, при належному використанні здатні висвітлити цілий набір корисних показників як для оцінки ризиків, так і для покращення безпеки організації. Наприклад, якщо проаналізувати частоту доступу для обмежених і конфіденційних ресурсів, легко побачити, які із них є більш критичними для роботи підприємства (до них буде доступ майже постійним), а які – ні (до них доступ буде спорадичним, або й взагалі не буде за період скринінга).

Якщо частота доступу вимірюється періодично, то можна скласти так званий "стандартний паттерн доступу" – тривимірний графік доступу до критичних ресурсів. За його допомогою можна відстежити аномальне зростання доступу до конфіденційних ресурсів. В залежності від ситуації це може бути пов'язано із окремим проектом, реорганізацією підприємства, аудитом... або прихованим делегуванням доступу іншим працівникам. Це допоможе більш чітко визначити права доступу і навіть перемістити деякі ресурси з менш захищеної категорії в більш захищену.

Наприклад, якщо за основу взяти організаційну структуру, то можна емпірично показати, що частота звертання до документів відповідного рівня будуть приблизно відповідати кількості персоналу, який зайнятий у підрозділах, які мають до нього доступ.

Тобто, якщо в даному підприємстві існує три робочі групи по 10 працівників, 3 менеджери і 1 директор, і по одному документу кожного типу, відповідно, доступи розподіляться так (табл. 3).

**Таблиця 3.** Розподіл доступу між персоналом

|                | Public | Internal | Restricted | Confidential |
|----------------|--------|----------|------------|--------------|
| Організаційний | 34     | 4        | 1          | 1            |
| Секторальний   | 34     | 4        | 15         | 1            |
| Робоча група   | 34     | 4        | 12         | 1            |

**Джерело:** розроблено авторами на основі досліджень Вуда П. (2024).

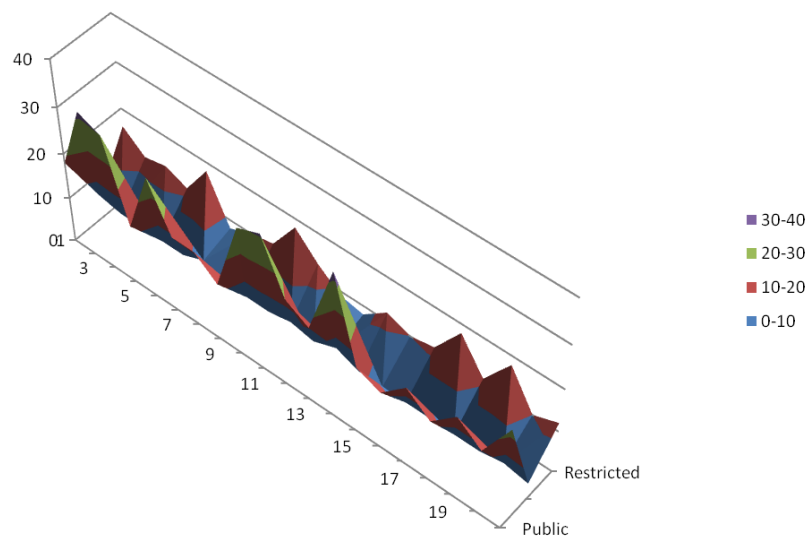
Зрозуміло, що до публічного документів матимуть доступ всі. До внутрішніх – весь керівний склад (тобто, 3 менеджери та директор). До документів обмеженого доступу на рівні робочої групи матимуть доступ працівники, менеджер і директор, на рівні сектора –

працівники робочої групи, всі менеджери і директор. А на організаційному рівні до всіх документів обмеженого доступу матиме доступ лише директор, як і до конфіденційного документа.

Серед сучасних файлових систем виділяється окрема група так званих "журнальованих файлових систем", які дозволяють вести реєстр транзакцій на рівні окремих файлів. Транзакція – це будь-яка елементарна дія, наприклад, читання, запис, копіювання та переміщення. В поєднанні із системою авторизації журналювання може чітко показати, хто і наскільки часто звертався до того чи іншого ресурсу.

Це дозволяє побудувати аналогічні стандартні патерн для доступу за дозволом чи за частотою копіювання. Якщо періодично вимірювати частоту доступу, а потім сформувати часовий графік, вийде наступна поверхня (тривимірний графік). Піки покаже аномально високу частоту, яку можна трактувати або як делегування повноважень, або несанкціонований доступ (рис. 3).

Фактично на одному такому графіку можна розміщувати лише співмірні за показниками групи ресурсів. Наприклад, на діаграмі 3 насправді розміщено не дві категорії, а три – Public, Internal та Restricted. Однак друга категорія має настільки незначний доступ, що просто не видна на поверхні і відстежити можна лише зовсім аномальний доступ – на рівні масивного витоку даних. Що, в свою чергу робить невідстежувані дрібні або одиничні втрусння в безпеку на рівні малозатребуваних ресурсів, або інформації, яка не повинна виходити за межі дуже обмеженого кола – а це не лише внутрішня документація, а й конфіденційна.



**Рисунок 3.** Стандартний патерн доступу

Джерело: розроблено авторами

Знову ж таки, аномальне збільшення звертання до раніше малозатребуваного ресурсу може означати або приховане делегування доступу, або передачу файла іншим працівникам. Однак його необхідно якось відстежити, а на загальному патерні це помітити складно.

Що стосується пікової частоти для дійсно часто затребуваного ресурсу – то у разі постійних перевищень патернів необхідно або переглядати класифікацію як таку (розглянути, наприклад, пониження ступеню захисту), або розділити контент на менш захищену (із переводом в менш захищену категорію) та більш захищену.

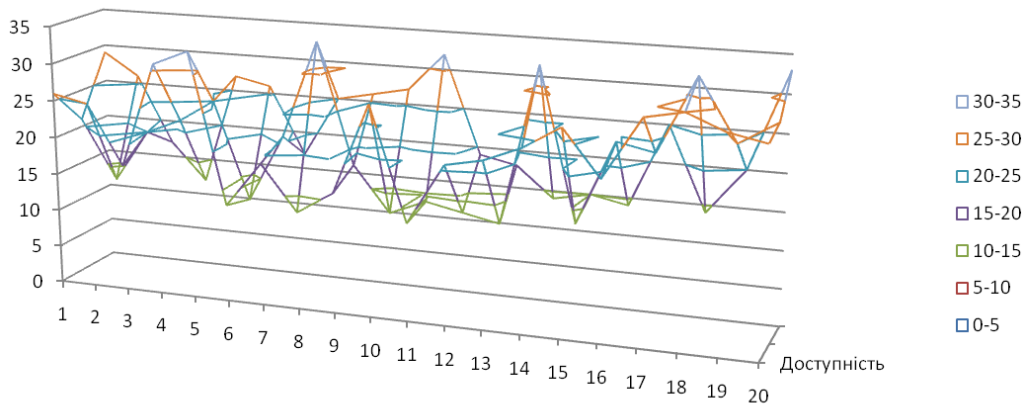
Для нецифрових критичних документів висока частота доступу може бути достатньою підставою пришвидшити їх цифровізацію, оскільки так простіше запобігти їх витоку і запровадити протокол (моніторинг) доступу до них.

Це робить метод стандартних патернів достатньо простим, щоб їх можна було

вимірювати з достатньо високою періодичністю і зробити базою для коригування інших класифікаторів ресурсів. Однак у нього є суттєвий недолік. Він добре підходить для автоматизованих систем і для часто використовуваних документів, але вкрай погано підходить, якщо на підприємстві існує розгалужена і різноманітна класифікація документів за рівнями доступу – для малозатребуваних даних цей метод покаже лише масові витоки (Ху, et al., 2024).

Це породжує задачу такої візуалізації кількох показників ризиків, які б полегшували аналіз насамперед малозатребуваних ресурсів.

**Багатовимірна матриця класифікації ресурсів.** З метою спрощення використовуються два класифікатори – основний, в рамках якого визначаються рівні доступу, і доповняльний, який дозволяє чіткіше визначити межі класів і конкретизує класифіковану інформацію. Тому найбільш поширеним є табличне класифікування. Для варіанта одночасного застосування трьох класифікаторів використовується "класифікаційна сітка". Це більш гнучка система представлення класифікації, оскільки замість лінійних узагальнених блоків ресурсів дозволяє розглядати їх кластери. Наприклад, побудуємо таку сітку, відклавши по вісі Z частоту доступу до деякого набору документів ресурсу за певний перевірючий період (рис. 4).



**Рисунок 4.** Класифікаційна сітка

**Джерело:** розроблено авторами

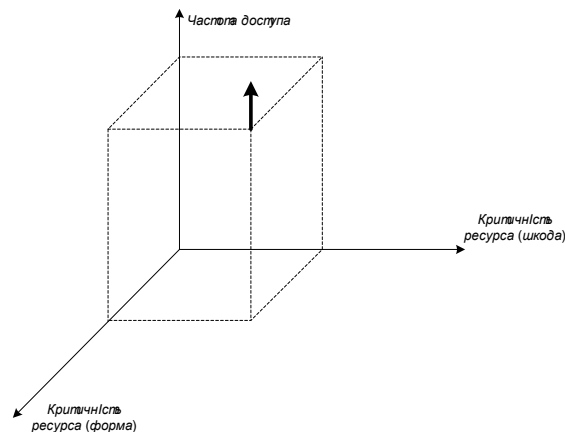
На рис. 4 видно, які ресурси використовуються найчастіше, а які – найрідше. Водночас на будь-якому рівні можна взяти переріз і отримати все ту ж традиційну табличну форму.

Класифікаційна сітка використовується і для класифікації, і також для аналізу її коректності та коригування. Варіюючи сітку по вертикальній вісі, можна будувати класифікаційні кластери. А найбільшого перегляду (як і найбільшої уваги) заслуговують ресурси, до яких найбільша і найменша частота доступу. Щоб адекватно їх проаналізувати, необхідно додати на графік динамічну складову.

Класифікаційна сітка є складною, а чим більше показників, які поєднуються під час аналізу, тим більше вимірів матиме поверхня і тим складніше буде її аналізувати. На нашу думку, для спрощення аналізу і більшої наглядності найкраще використати форму "класифікаційного вектора", що дозволить проаналізувати динамічну складову, і при цьому не обтяжити графік додатковими даними.

Ідея така: кожен ресурс необхідно показувати не в формі статичної точки на графіку, а в формі вектора, пропорційному максимальній зміні динамічного показника, спрямованого в бік цієї зміни. Це дозволить, наприклад, візуалізувати миттєві піки для малозатребуваних ресурсів, які в інших варіантах візуалізацій складно розрізнити. Крім того, перевагою такого підходу є можливість показувати практично будь-яку кількість показників та їх комбінацій і

при цьому не втрачати динамічної складової. Його можна називати багатовимірною матрицею класифікації ресурсів. Спрощений приклад такої матриці – класифікаційний стек – показаний на рисунку 5.



**Рисунок 5.** Спрощений варіант багатовимірної матриці класифікації ("класифікаційний стек")

**Джерело:** розроблено автором

На рис. 5 наведено ресурс, який має середнє значення ризику потенційної шкоди і середнє значення форми (цифрове локальне представлення), які показані на пласкій частині діаграми. Вона відповідає табличній формі класифікатора ресурсів. Очікувана частота доступу до ресурсу відкладена по вісі аплікат, а максимальна зміна цієї частоти за перевірючий період показана у вигляді стрілочки-вектора, пропорційної цій зміні.

Якщо такий вектор буде надзвичайно великим (тобто, в 1.5 і більше разів), його використання слід промоніторити і підтвердити один із трьох можливих висновків:

- потреба в документі дійсно виросла і необхідно переглянути його класифікацію;
- має місце делегування доступу особам, які зазвичай його не мають – це може не становити суттєвого ризику (у випадку матеріалів обмеженого доступу), а може становити суттєвий ризик;
- відбувся несанкціонований доступ.

Таким чином, в класифікаційному стекові "спливають на поверхню" всі ресурси із проблемною класифікацією чи проблемами із захистом, і "потонуть" всі ресурси, чия безпека не викликає суттєвих питань. Варіюючи масштаб матриці, а також варіюючи пари пласкої частини діаграми, можна порівняно легко визначити основні ризики інформаційної безпеки і мінімізувати їх за рахунок більш точної класифікації та контролю доступу.

Крім того, в такий тип діаграми можна додавати будь-яку кількість додаткових показників у вигляді різноспрямованих векторів – що полегшить обрахунок інтегральних показників, подібних тому, який було наведено вище для загроз з боку акторів. Показники-складові можна показати у вигляді векторів, а сума цих векторів буде шуканим інтегральним показником.

При належному використанні такий підхід дозволяє вирішити задачі, які ставились в цій роботі, і одночасно є зручним способом візуалізації стану критичних інформаційних ресурсів на підприємстві.

**Обговорення.** Основний матеріал дослідження включає результати застосування динамічної класифікації інформаційних ресурсів та їх візуалізацію за допомогою багатовимірної матриці. Аналіз проводився на основі даних, отриманих від реальних компаній, що використовують сучасні методи контролю доступу до інформаційних ресурсів.

Рисунок 3 демонструє стандартний патерн доступу до критичних інформаційних ресурсів. Цей тривимірний графік показує частоту звернень до ресурсів залежно від рівня

конфіденційності. На графіку видно, що найбільша частота доступу спостерігається до внутрішніх даних (Internal Data), тоді як до архівних даних (Legacy Data) доступ здійснюється значно рідше.

Встановлено, що частота доступу до обмежених і конфіденційних ресурсів дозволяє визначити їхню критичність для роботи підприємства. Наприклад, ресурси з постійним високим рівнем доступу є найбільш критичними, тоді як ресурси з низькою частотою доступу мають меншу значущість. Таблиця 4 подає результати статистичної обробки даних про доступ до ресурсів.

**Таблиця 4.** Частота доступу до інформаційних ресурсів за рівнем конфіденційності

| Рівень конфіденційності  | Частота доступу (середня кількість запитів на місяць) |
|--------------------------|---|
| Публічна інформація      | 10,000 ± 500  |
| Внутрішня інформація     | 5,000 ± 300   |
| Обмежена інформація      | 1,000 ± 50  |
| Конфіденційна інформація | 100 ± 10  |

**Джерело:** розроблено автором на основі досліджень (Oseghale, 2023; Lipps & Schotten, 2022)

За результатами аналізу було встановлено, що використання динамічної класифікації дозволяє виявити аномалії в доступі до ресурсів. Такі аномалії в доступі можуть бути пов'язані з виконанням специфічних проектів або реорганізацією компанії, однак вони також можуть свідчити про потенційні загрози інформаційній безпеці. Використання багатовимірної матриці класифікації дозволяє виявляти та моніторити такі зміни в режимі реального часу, що підвищує ефективність захисних заходів.

Рисунок 5 демонструє спрощений варіант багатовимірної матриці класифікації, так званий, "класифікаційний стек"), який показує відносне розташування ресурсів за їх критичністю та частотою доступу. Ресурси з високою критичністю та частотою доступу розміщені у верхній частині матриці, тоді як менш критичні ресурси – у нижній частині. Ця візуалізація дозволяє ідентифікувати ресурси, які потребують підвищеного рівня захисту, та ресурси, які можуть бути цифровізовані для покращення їх захисту. Використання багатовимірної матриці також дозволяє автоматизувати процес класифікації, що зменшує ризики, пов'язані з людським фактором.

Обговорення результатів включає порівняння отриманих даних з результатами інших досліджень. Наприклад, дослідження (Wiedemann et al., 2024), (Ramamurthy et al., 2022) підтверджують, що динамічна класифікація забезпечує більш високий рівень захисту інформаційних ресурсів порівняно зі статичними методами. Робота (Mazzola L. et al., 2021) показує доцільність створення інструменту, який доповнює механізм безпеки, підтримуючи експертів у виявленні незвичайних патернів та подій, пов'язаних із безпекою, які слід відстежувати та перевіряти механізмом класифікації подій.

Також виявлено, що запропонований підхід із динамічною складовою дозволяє більш ефективно реагувати на зміни у використанні інформаційних ресурсів, що підтверджується результатами експериментів, проведених у реальних умовах. Використання динамічної класифікації сприяє підвищенню точності оцінки ризиків та забезпечує більш ефективний захист критичних даних.

Порівняння отриманих результатів з даними інших досліджень свідчить про те, що запропонований підхід має значний потенціал для впровадження у системах інформаційної безпеки. Використання багатовимірної матриці класифікації дозволяє підвищити точність і швидкість реагування на загрози, що є важливим фактором у сучасних умовах швидкого розвитку інформаційних технологій.

Загалом, результати дослідження підтверджують ефективність запропонованого підходу та його практичну цінність для підвищення рівня інформаційної безпеки.

**Висновки.** В даній роботі обґрунтовано та запропоновано покращений підхід до

візуалізації стану критичних інформаційних ресурсів на підприємстві за допомогою динамічного підходу із застосуванням багатовимірної матриці класифікування інформації для оцінки ризиків інформаційної безпеки.

Зокрема, в роботі було сформульовано і проаналізовано задачу класифікації інформаційних ресурсів як ключову задачу під час оцінки ризиків інформаційної безпеки. Основну увагу приділено питанням класифікації, оскільки невірна класифікація може призвести не лише до витоків даних, а й до несанкціонованого доступу до широкого спектру чутливої інформації.

Було проаналізовано табличні та графічні методи класифікації інформаційної безпеки та її перегляду в процесі діяльності організації чи підприємства. Зроблено висновок, що переважна більшість методів класифікації надзвичайно слабо враховує динамічну складову – насамперед частоту доступу до тих чи інших ресурсів, яка може слугувати основним показником реальної класифікаційної належності інформаційного ресурсу – особливо враховуючи сучасні технічні можливості, які дозволяють контролювати і журналювати доступ навіть на рівні окремих документів.

Для періодичного перегляду класифікації та відстеження динаміки доступу запропоновано використовувати багатовимірну класифікаційну матрицю, яка у спрощеній формі "класифікаційного стеку" дозволяє відстежити ресурси, що становлять потенційний ризик через некоректну чи застарілу класифікацію.

Запропонований підхід містить значний потенціал і може стати базою для подальших досліджень, оскільки за його допомогою можна розмістити на графіку будь-яку кількість показників у вигляді набору векторів та формувати інтегральні оцінки ризиків на основі суми чи різниці векторів. Також запропонований підхід потенційно може слугувати базою для досліджень із тренування нейромережевих засобів автоматизованої класифікації.

**Подяки** «Немає».

**Конфлікт інтересів** «Немає».

## References

- Al Qahtani, E., Story, P., & Shehab, M. (2024). The impact of risk appeal approaches on users' sharing confidential information. In *Proceedings of the CHI Conference on Human Factors in Computing Systems (CHI '24)* (Article 579, pp. 1–21). Association for Computing Machinery. <https://doi.org/10.1145/3613904.3642524>
- Alotibi, G. (2024). A cybersecurity awareness model for the protection of Saudi students from social media attacks. *Engineering, Technology & Applied Science Research*, 14(2), 13787-13795.
- Arora, S., Lewis, P., Fan, A., Kahn, J., & Ré, C. (2023). Reasoning over public and private data in retrieval-based systems. *Transactions of the Association for Computational Linguistics*. Retrieved from <https://transacl.org/index.php/tacl/article/view/4705>
- Arslan, M., & Cruz, C. (2024). Business text classification with imbalanced data and moderately large label spaces for digital transformation. *Applied Network Science*, 9, 11. <https://doi.org/10.1007/s41109-024-00623-5>
- Barnawi, A., Kumar, K., Kumar, N., & Alzahrani, B., & Almansour, A. (2024). A deep learning approach for landmines detection based on airborne magnetometry imaging and edge computing. *Computer Modeling in Engineering & Sciences*, 139(2), 2117-2137. <https://doi.org/10.32604/cmescs.2023.044184>

- Emmanuel, I., Sun, Y., & Wang, Z. (2024). A machine learning-based credit risk prediction engine system using a stacked classifier and a filter-based feature selection method. *Journal of Big Data*, 11, 23. <https://doi.org/10.1186/s40537-024-00882-0>
- Gambarelli, G., Gangemi, A., & Tripodi, R. (2023). Is your model sensitive? SPEDAC: A new resource for the automatic classification of sensitive personal data. *IEEE Access*, 11, 10864-10880. <https://doi.org/10.1109/ACCESS.2023.3240089>
- Irwin, L. (2022, August 30). What is ISO 27001 information classification? *IT Governance*. Retrieved from <https://www.itgovernance.co.uk/blog/what-is-information-classification-and-how-is-it-relevant-to-iso-27001>
- Lipps, C., & Schotten, H. D. (2022). Physical layer security: About humans, machines and the transmission channel. In *Proceedings of the 21st European Conference on Cyber Warfare and Security* (Vol. 21, No. 1, pp. 161-169). Academic Conferences International Limited. <https://doi.org/10.34190/eccws.21.1.403>. Retrieved from <https://papers.academic-conferences.org/index.php/eccws/article/view/403/357>
- Malchiodi, D., Raimondi, D., Fumagalli, G., et al. (2024). The role of classifiers and data complexity in learned Bloom filters: Insights and recommendations. *Journal of Big Data*, 11, 45. <https://doi.org/10.1186/s40537-024-00906-9>
- Mazzola, L., et al. (2021). Security rules identification and validation: The role of explainable clustering and information visualisation. In Stephanidis, C., Antona, M., & Ntoa, S. (Eds.), *HCI International 2021 - Posters*. HCII 2021. Communications in Computer and Information Science, vol 1420. Springer. [https://doi.org/10.1007/978-3-030-78642-7\\_58](https://doi.org/10.1007/978-3-030-78642-7_58)
- Mikuletič, S., Vrhovec, S., Skela-Savič, B., & Žvanut, B. (2024). Security and privacy oriented information security culture (ISC): Explaining unauthorized access to healthcare data by nursing employees. *Computers & Security*, 136, 103489. <https://doi.org/10.1016/j.cose.2023.103489>
- New Zealand Protective Security Requirements. (2024, May 10). Applying business impact levels. Retrieved from <https://nzism.gcsb.govt.nz/assets/Previous-versions/v3-2/NZISM-Part-One-v3.2-December-2018.pdf>
- Oseghale, O. (2023). Digital information literacy skills and use of electronic resources by humanities graduate students at Kenneth Dike Library, University of Ibadan, Nigeria. *Digital Library Perspectives*, 39(2), 181-204. <https://doi.org/10.1108/DLP-09-2022-0071>
- Pitafi, S., Anwar, T., Widia, I. D. M., & Yimwadsana, B. (2023). Revolutionizing perimeter intrusion detection: A machine learning-driven approach with curated dataset generation for enhanced security. *IEEE Access*, 11, 106954-106966. <https://doi.org/10.1109/ACCESS.2023.3318600>
- Queensland Government. (2024, May 10). Information security classification framework (QGISC) – Queensland Government guidelines. Retrieved from <https://www.forgov.qld.gov.au/information-and-communication-technology/qgea-policies-standards-and-guidelines/information-security-classification-framework-qgisfc>
- Ramamurthy, A., Sathya, V., Rochman, M. I., & Ghosh, M. (2022). ML-based classification of device environment using Wi-Fi and cellular signal measurements. *IEEE Access*, 10, 29461-29472. <https://doi.org/10.1109/ACCESS.2022.3158056>
- Robinson, P. (2024). Data classification? Definition, levels & examples – Lepide data security. *Lepide*. Retrieved from <https://www.lepide.com/blog/what-is-data-classification-and-how-to-do-it/>
- Shmatko, O., Balakireva, S., Vlasov, A., Zagorodna, N., Korol, O., Milov, O., Petrov, O., Pohasii, S., Rzayev, K., & Khvostenko, V. (2020). Development of methodological foundations for designing a classifier of threats to cyberphysical systems. *Eastern-European Journal of Enterprise Technologies*, 3(9), 6-19. <https://doi.org/10.15587/1729-4061.2020.205702>
- Song, X., Liu, Z., & Jiang, B. (2024). Adaptive boosting with fairness-aware reweighting technique

- for fair classification. *Expert Systems with Applications*, 250, 123916. <https://doi.org/10.1016/j.eswa.2024.123916>
- Velmurugan, S., Prakash, M., Neelakandan, S., et al. (2024). Provably secure data selective sharing scheme with cloud-based decentralized trust management systems. *Journal of Cloud Computing*, 13, 86. <https://doi.org/10.1186/s13677-024-00634-8>
- Venn, B., Leifeld, T., Zhang, P., et al. (2024). Temporal classification of short time series data. *BMC Bioinformatics*, 25, 30. <https://doi.org/10.1186/s12859-024-05636-6>
- Wang, G., & Gu, Y. (2024). Multi-task scenario encrypted traffic classification and parameter analysis. *Sensors*, 24(10), 3078. <https://doi.org/10.3390/s24103078>
- Wiedemann, N., Janowicz, K., Raubal, M., et al. (2024). Where you go is who you are: A study on machine learning based semantic privacy attacks. *Journal of Big Data*, 11, 39. <https://doi.org/10.1186/s40537-024-00888-8>
- Wood, P. (2013, January). Business priorities: What to protect, monitor and test. *Computer Weekly*. Retrieved from <https://www.computerweekly.com/feature/Business-priorities-what-to-protect-monitor-and-test>
- Xu, A., Gao, J., Sui, X., Wang, C., & Shi, Z. (2024). LiDAR dynamic target detection based on multidimensional features. *Sensors*, 24(5), 1369. <https://doi.org/10.3390/s24051369>
- Zhang, Y., Deng, Q., Liang, W., & Zou, X. (2018). An efficient feature selection strategy based on multiple support vector machine technology with gene expression data. *BioMed Research International*, 2018. <https://doi.org/10.1155/2018/1234567>

## БІОЛОГІЧНІ ТА МЕДИЧНІ ПРИЛАДИ І СИСТЕМИ

УДК 681.5:613

### **Construction Guidelines for Optical-Electronic Expert Systems in Blood Rheology**

**Li Jinqiong,**

Postgraduated Student

Biomedical Engineering and Optic-Electronic Systems Department,

Vinnitsia National Technical University,

95, Khmelnytske shosse, 21021, Vinnitsia, Ukraine

ORCID: 0000-0002-0505-621X

**Sergii Pavlov,**

D.Sc., Professor

Biomedical Engineering and Optic-Electronic Systems Department,

Vinnitsia National Technical University,

95, Khmelnytske shosse, 21021, Vinnitsia, Ukraine

ORCID: 0000-0002-0051-5560

**Oleksandr Poplavskyi,**

Ph.D., Associate Professor of the Department of Information Technologies,

Kyiv National University of Construction and Architecture.

03680, Kyiv, Povitrflotskyi Avenue, 31

ORCID: 0000-0003-0465-6843

**Oleksandr Kaduk**

Ph.D., Associate Professor of the Department of Information Technologies,

Vinnitsia National Technical University,

95, Khmelnytske shosse, 21021, Vinnitsia, Ukraine

ORCID: 0009-0001-2388-9813

**Abstract.** Building specifically designed optical-electronic information processing expert systems for blood rheology bioimage analysis requires a painstaking, subtle approach. Such systems provide essential support for diagnostic operations and require an understanding of experimental properties such as the rheology of blood and bioimage analysis. To properly build these systems, guidelines are needed for improving imaging methods, image processing routines, and application of expert knowledge so the blood's rheological properties can be analyzed precisely. Information features (information parameters) for the analysis of the biomedical images, in particular, for the assessment of the rheologic properties of the blood, are formed. Algorithm and optical-electronic expert system for the analysis of the rheological properties of the blood are suggested, they are used for the increase of the diagnostic validity which is a determining factor in the biomedical diagnostics. The main focus of modern clinical hemorheology is the search diagnostic and prognostic criteria for various diseases and rheological correction methods violations. Changes in the rheological parameters of blood are one of the significant mechanisms of the formation of insufficient blood supply in the early stages the development of the disease. Main pathological effects violations of rheological properties in the blood can lead to micro-flow failure circulation, the extreme manifestation of which may lead to a decrease in trophism and the development of ischemic

syndrome, a violation of micro-rheology and an increase in the viscosity of blood, which causes an increase in total peripheral resistance and the development of arterial hypertension syndrome, to atherosclerotic changes in blood vessels, to a violation of hemorheology, which contributes to increased thrombosis.

**Keywords:** biomedical information, optical – electronic expert system, membership functions, complex hierarchical structure, hemodynamic parameters, analysis of blood rheology

## **Принципи побудови оптико-електронних експертних систем для дослідження реології крові**

**Лі Цзіньюн,**

аспірант кафедри біомедичної інженерії та оптико-електронних систем,  
Вінницький національний технічний університет,  
м. Вінниця, Хмельницьке шосе, 95, Україна  
ORCID: 0000-0002-0505-621X

**Сергій Володимирович Павлов,**

д.т.н., професор,  
професор кафедри біомедичної інженерії та оптико-електронних систем,  
Вінницький національний технічний університет,  
м. Вінниця, Хмельницьке шосе, 95, Україна  
ORCID: 0000-0002-0051-5560

**Олександр Анатолійович Поплавський,**

к.т.н., доцент,  
доцент кафедри інформаційних технологій,  
Київський національний університет будівництва і архітектури,  
Київ, Україна,  
ORCID: 0000-0003-0465-6843

**Олександр Володимирович Кадук,**

к.т.н., доцент кафедри обчислювальної техніки,  
Вінницький національний технічний університет,  
м. Вінниця, Хмельницьке шосе, 95, Україна  
ORCID: 0009-0001-2388-9813

**Анотація.** Створення спеціально розроблених оптико-електронних експертних систем обробки інформації для реологічного аналізу біозображень крові вимагає ретельного аналізу. Такі системи забезпечують необхідну підтримку для діагностичних операцій і вимагають розуміння експериментальних властивостей, таких як реологія крові та аналіз біозображень. Щоб належним чином побудувати ці системи, потрібні рекомендації щодо вдосконалення методів візуалізації, процедур обробки зображень і застосування експертних знань, щоб можна було точно аналізувати реологічні властивості крові. Формуються інформаційні ознаки (інформаційні параметри) для аналізу біомедичних зображень, зокрема, для оцінки реологічних властивостей крові. Запропоновано алгоритм та оптико-електронну експертну систему для аналізу реологічних властивостей крові, які використовуються для підвищення діагностичної валідності, що є визначальним фактором у біомедичній діагностиці. Основним напрямком сучасної клінічної гемореології є пошук діагностичних і прогностичних критеріїв

різних захворювань і порушень методів реологічної корекції. Зміна реологічних показників крові є одним із суттєвих механізмів формування недостатності кровопостачання на ранніх стадіях розвитку захворювання. Основні патологічні наслідки порушення реологічних властивостей крові можуть призвести до порушення мікрореології кровообігу, крайній прояв якого може призвести до зниження трофіки та розвитку ішемічного синдрому, порушення мікрореології та підвищення в'язкості крові. крові, що зумовлює підвищення загального периферичного опору та розвиток синдрому артеріальної гіпертензії, до атеросклеротичних змін у судинах, до порушення гемореології, що сприяє посиленню тромбоутворення.

**Ключові слова:** біомедична інформація, оптико-електронна експертна система, складна ієрархічна структура, гемодинамічні параметри, аналіз реології крові.

**DOI:** <https://doi.org/10.31649/1999-9941-2024-60-2-107-121>

**Introduction.** Building specifically designed optical-electronic information processing expert systems for blood rheology bioimage analysis requires a painstaking, subtle approach. Such systems provide essential support for diagnostic operations and require an understanding of experimental properties such as the rheology of blood and bioimage analysis (Begum et al, 2020). To properly build these systems, guidelines are needed for improving imaging methods, image processing routines, and application of expert knowledge so the blood's rheological properties can be analyzed precisely.

Mastering the imaging process is thus a critical part of building these systems. Careful control of resolution, contrast, and image acquisition protocols is essential if relevant rheological information held by bioimages is to be reliably obtained (Begum et al, 2020). Recommendations also stress the use of suitable imaging modalities and methods suited to fully capturing blood sample rheological characteristics. The second key principle is to develop strong image-processing algorithms (Abdeldaim et al., 2018). Further, image processing algorithms are used when processing bioimages for analysis of blood rheology properties. Such algorithms must take into account variations in the quality of images, noise removal, separation of blood components, and extraction of rheological parameters (Singha et al., 2021). Powerful analyses improve rheological property analysis, which leads to more precise diagnostic capabilities.

Furthermore, these systems must absorb expert knowledge to allow blood rheology analysis. Domain-specific know-how and biomedical principles are integrated in such a way that the system can accurately interpret bioimages, taking into account blood component interactions and characteristics (Singha et al., 2021). Algorithms based on expert knowledge enable computers to interpret the rheological differences and analyze them with clinical significance (Crow, 2012). Expert optical-electronic systems require a full grasp of both imaging and biomedical principles before their implementation can be considered. This interdisciplinary approach brings together experts in imaging technology with biomedical engineers and clinicians (Begum et al, 2020). It helps to guarantee that the systems are by clinical demands, meet biomedical requirements, and combine imaging technology of international standards.

Rapid development of the information needs puts forward new requirements to the means and methods for the analysis of biomedical information. To meet these needs there appears the necessity to apply new physical principles and technologies. Principally new solution of the problem is the development of the expert optical-electronic information system, intended for the processing of the biomedical data, based on new computation technologies. Objects in the medical images are complex and of multifactorial nature and these factors stipulate high requirements concerning the reliability, accuracy and validity of their research results. Usage of the computational facilities and mathematical methods in this sphere enables not only to accelerate the processing of the biomedical images but improve the accuracy of the result obtained. Greater part of the software products does

not give sufficiently complete and accurate information regarding the quantitative and qualitative character of the medical images. That is why, there exists the problem of the development of highly efficient information system, which is able to meet the modern requirements of the problem put forward, the system must be constructed on the modern optical-electronic element base. Much attention is paid to the development of the efficient methods and means of the work with the images in medical diagnostic systems (Abdeldaim et al., 2018; Ahmed et al., 2023; Ansari et al., 2023). The concept of the expert optical-electronic information system provides the creation of the basically new structural organization both of the channels for the collection of the primary information and means, intended for the processing, proceeding from the assumption that they must simulate biomedical processes. This means that, for instance, parallel optical channels must interact with the neurolike networks (Riva, & Feke, 1981; Riva, et al., 1992; Lerche et al., 1989).

Such system as the basic model of the complex noninvasive diagnosis assumes that this is engineering systems which acquires the information, sent in the form of the visual medium of the arbitrary shape, allocates certain features of the bioobject, processes them and makes the decision automatically or with the participation of the operator (Harris et al., 1994; Kozhemiako et al., 2010). In particular, systems for the analysis of the blood rheology are of great interest for the specialists (Arber et al., 2017). Such optical electronic information system carries out the transfer of the biomedical image to the computer where, applying the corresponding software, this image is processed. Having obtained the biomedical image of the blood content, it is necessary to know how to process correctly this information to obtain necessary, from the point of view of medicine, data regarding the state of the studied object (Stoltz et al., 1987; Pryzwan et al., 2024; Michelson et al., 1996).

In the last decade, significant advances have been made in the development of fundamentally new approaches to the study of the rheological properties of blood as fundamental issues in the theory of haemoreology, and problems in the methodology for diagnosing and correcting haemoreological disorders in clinical practice with the aim of preventing vascular occlusion and ischemia of vital organs (Riva & Feke, 1981; Pryzwan, 2024; Kozhemiako et al., 2010).

Disorders of the rheological properties of blood are an important pathogenetic factor in the development of many diseases. Under conditions of pathology, reduced blood turnover can become the primary cause of disorders in the microcirculatory and venous channels. The severity of these violations often determines the severity of the patient's condition and sometimes also the prognosis of the outcome of the disease. Each category of patients has its own characterized haemoreological profile, due to the original pathology. It is therefore urgent to make early diagnosis of the rheological properties of blood in the early diagnosis of vascular conditions and to improve the results of treatment, especially in cardiac patients, which will make it possible to correct blood flow in a timely and adequate manner according to the vessels, preventing dysfunctional disorders in vital organs (Lerche et al., 1989).

Of course, rheological dysfunction of the blood is also an important pathogenetic factor in the development of many diseases. In conditions of pathology, the reduction of blood turnover can become the primary cause of the impairment of internal organ functions, which in many ways can determine the severity of the patient's condition and further prognosis. This underlines the diagnostic value of a laboratory hemoreological evaluation that enhances the pathogenetic orientation (Lerche et al., 1989; Harris et al., 1994).

**Biomedical image processing methods for expert systems.** The application of biomedical image processing techniques by expert systems in extracting critical diagnostic information is crucial. These methodologies are essential to turning biomedical images into information, augmenting the decision-support capabilities of medicine (Singha et al., 2021). Various types of image processing are used to extract meaning from these images, such as image segmentation and feature extraction; pattern recognition and machine learning also play an important part. Segmentation of images is an essential step in biomedical image processing, which involves the

separation and identification of a certain area or structure within the picture. Such a process enables relevant areas of interest, for example, corrupt tissues or cells to be extracted from the surrounding environment and analyzed specifically by expert systems (Abdeldaim et al., 2018). Another important step is feature extraction, which isolates and selects the appropriate attributes or features of segmented biomedical images. Extracting such discriminative features is vital for diagnoses of actual clinical relevance to be possible using expert systems. The methods used to find patterns in biomedical images and interpret them are known as pattern recognition techniques (Begum et al, 2020). These techniques allow patterns, aberrations, or groupings of features associated with acute leukemia diagnosis to be identified.

The addition of machine learning algorithms to expert systems makes them even more diagnostic. Neural networks, support vector machines, and deep learning methods allow the systems to learn patterns from biomedical images (Singha et al., 2021). Many of these algorithms can detect complex relationships within images that would not be evident to the naked eye, increasing precision and diagnosis. Applying these approaches to image processing within expert systems, for instance, enables the appropriate interpretation and implementation of visual biomedical data in acute leukemia diagnosis (Begum et al, 2020). Through the use of these advanced methods, such systems can derive more sophisticated information from biomedical images and assist clinicians in diagnosis, prognosis, and treatment planning.

**Evaluating the efficacy of expert optical-electronic systems in blood rheology.** For example, research into methods to compare the accuracy of diagnostic tests using expert optical-electronic systems in analyzing biomedical information related to blood rheology is crucial. It also requires careful evaluation processes to determine how effective these systems are in assessing the blood's rheological properties, which can aid in the diagnosis of acute leukemia (Singha et al., 2021). At the core of these studies is an experiment to determine system performance, diagnostic accuracy, and reliability in detecting leukemia by analyzing bioimages related to blood rheology (Begum et al, 2020). Such tests serve to determine the accuracy, sensitivity, and specificity as well as the overall performance of such expert systems in interpreting bioimages related to blood rheology.

Researchers hope that by subjecting expert optical-electronic systems to rigorous analysis, the effectiveness and clinical applicability of these methods for diagnosing acute leukemia will be confirmed. Such studies include comprehensive assessments of how closely these systems match those typically employed by doctors in distinguishing blood properties between healthy and leukemic individuals (Singha et al., 2021). The assessment focuses on the sensitivity of their systems to detecting various minor differences, abnormalities, or vascular rheological changes related to leukemia. Comparing the system's output to diagnostic standards and expert evaluations is a comprehensive analysis. These kinds of comparative analyses can determine the extent to which these expert systems are reliable and accurate in identifying such rheological anomalies as may indicate acute leukemia (Begum et al, 2020). It also requires statistical studies to establish figures and prove their diagnostic capabilities.

In addition, this kind of research is typically validated through the use of samples from patients with different stages and subtypes of acute leukemia. The rigor and broad applicability of the systems' diagnostic capabilities have been proven with these extensive tests, strengthening their clinical utility (Rehman et al., 2018). The results of these research projects are valuable in helping the expert optical-electronic system used to diagnose acute leukemia gain validity and recognition. These positive findings on high accuracy, reliability, and correlation with established clinical criteria serve to further affirm the appropriateness of these systems in supporting clinicians (Begum et al, 2020). This kind of verification can be regarded as an important foundation for recognizing these systems as aids to clinical decision-making in acute leukemia diagnosis.

**Reliability metrics analysis for decision support.** Reliable decision support is of paramount importance for biomedical image analysis, especially when it comes to diagnosing acute leukemia.

These include expert judgment, algorithms, and diagnostic results (Rehman et al., 2018). They are taken together to make a final assessment of the reliability of decision support systems that depend on them for accuracy and credibility. Reliability indexes must be analyzed. Precision, sensitivity, specificity, and inter-rater agreement are important factors used to verify how accurate biomedical image diagnosis can be relied upon regarding issues of consistency or credibility. Precision sets the bar for how often test results are positive. Sensitivity assesses accuracy in detecting acute leukemia, while specificity measures the accurate diagnosis of patients without it (Singha et al., 2021). Inter-rater assessment measures the degree of consistency among various system elements or reviewers, reflecting on the stability and accuracy of diagnostic reporting.

It is these indicators that can best direct the clinician in interpreting system output, and thus help with accurate diagnosis of acute leukemias. High precision, sensitivity, and specificity will help the expert system to produce more accurate diagnostic information (Rehman et al., 2018). This helps healthcare professionals to have confidence in relying on the system's results when considering clinical action. Analysis of reliability metrics also provides a vehicle for continued development and fine-tuning of the decision support system (Abdeldaim et al., 2018). Scientists should examine these indicators to assess the condition of the system and look for areas that need adjustment or improvement (Singha et al., 2021). This way clinicians and developers alike can make improvements based on strengths and weaknesses discovered so that reliability improves with each rendition of the system.

Reliability metrics can help set standards for decision support systems in biomedical image analysis. These benchmarks also help establish standards for system validation, so that clinical implementation will only occur when the accuracy and reliability of a given system has been proven before its use. With such confidence in reliable indicators, decision support systems are accepted and used by clinical practitioners (Rehman et al., 2018). With these kinds of reliability measures showing that things are consistent, credible, and in line with what the experts set out for people, clinicians will very much be able to accept recommendations from such a system in their decision-making.

**Modern methods in information technology creation.** Taking advantage of modern techniques is crucial in designing information systems suitable for acute leukemia diagnosis. These technologies include cutting-edge equipment and techniques for analyzing data, improving diagnostic capabilities, and assisting decision-making (Singha et al., 2021). Incorporating modern methods makes it more adaptive to the continuous change and multifarious nature of medical data, maximizing its value for diagnoses such as various types of acute leukemia (Arber et al., 2017). Harnessing the power of artificial intelligence (AI), machine learning and data analysis is at the core of modern approaches to developing information technologies for use in diagnosing acute leukemia. It can quickly process enormous amounts of data, with AI-driven algorithms that find useful patterns and spot markers to differentiate the various leukemia subtypes (Rehman et al., 2018). What is more, machine learning models are trained to learn from the information they receive and grow progressively smarter when applied for future diagnostic purposes.

By using cloud computing and taking advantage of big data technologies in acute leukemia diagnosis, scientists can say that they are fundamentally changing the way many people approach working with large amounts of biomedical information (Rehman et al., 2018). Seamless storage, management, and analysis of large-scale biomedical data is based on cloud computing. Through its scalability and accessibility, it enables efficient data processing, providing decision support to real-time operations. Such accessibility allows healthcare professionals to get the information they need about a patient from anywhere when making quick critical decisions (Singha et al., 2021; Davis et al., 2014). Cloud-based big data These cloud sorts of tools also provide incredible capabilities for thoroughly analyzing large biomedical databases. It can unearth complex relationships and patterns in the data that traditional methods would overlook. This advanced analysis reveals obscure relationships among various factors, illuminating the complex nature of acute leukemia (Rehman et al., 2018). Big data analysis produces its breakthroughs by examining huge amounts of information

to reveal minute indicators or links that can guide correct diagnosis and treatment.

In addition, the incorporation of contemporary imaging technologies makes these systems even more powerful diagnostic tools. Modern imaging techniques and optical-electronic systems have a big impact on the detection of acute leukemia by increasing the resolution and refining the precision of images taken (Singha et al., 2021). Both technologies provide high-resolution image analyses, allowing a detailed examination of cellular shape and detecting characteristic abnormalities found in different types of acute leukemia. The clear rendering of the microscopic structure assists in identifying tiny differences or abnormalities imperative for diagnosis and treatment (Rehman et al., 2018). All these modern imaging technologies are of tremendous benefit for investigating the pathogenesis of leukemia.

**Algorithmic Software Development.** Designing computer programs specifically for acute leukemia that allow analysis of this type is a necessary component in improving diagnostic accuracy and speed. The software contains advanced computer programs that can analyze and assess complicated biomedical data, helping to make timely diagnoses of various forms of acute leukemia (Singha et al., 2021). But the most important part is developing algorithms that can handle various types of data, such as genetic information and blood profiles along with imaging scans. These kinds of algorithms are designed to learn how to identify patterns, abnormalities, and certain markers representative of different subtypes of leukemia (Rehman et al., 2018). Improved pattern recognition techniques allow one to detect slight variations or anomalies in the biomedical data which may indicate certain special features of leukemia.

Through the employment of machine learning, there has been a continuous development in the type of algorithm that responds to changing data sets. What makes these algorithms especially impressive is that they can learn as new information becomes available, improving their diagnostic powers and increasing their accuracy (Rehman et al., 2018). Using neural networks or deep learning enables the system to extract useful insights from tricky biomedical data, revealing subtle relationships that provide important diagnostic clues for acute leukemia.

One of the key factors behind improvements in diagnostic accuracy with time is machine learning's adaptive nature when confronted by new data. These algorithms pick up new knowledge from the updated data and as a result can better recognize minute markers, along with distinguishing different subtypes of leukemia (Singha et al., 2021). This adaptive learning mechanism increases the precision of the system's prognosis, improving assessment. The use of neural networks or deep learning multiplies the strength and effectiveness of analysis by allowing a greater understanding of biomedical data. These models are good at finding patterns, correlations, and associations that may not be apparent otherwise (Rehman et al., 2018). Through the use of these models, the system can draw out some diagnostic insights from bone marrow and build a better overall picture of acute leukemia's complex etiology.

Furthermore, decision support algorithms also strengthen the diagnostic process and provide valuable assistance to clinicians. After detailed analysis of the data, these algorithms can recommend tests and treatments or make differential diagnoses (Singha et al., 2021). Closing the gap between software and decision support, in this way, the system reduces the workload for searching through different types of acute leukemia. Algorithmic software communicating with decision support algorithms works together to maximize the efficiency and effectiveness of diagnostic procedures by providing clinicians with capable information (Rehman et al., 2018). The fusion of these cutting-edge algorithms and decision support systems allows healthcare providers to gain a comprehensive portrait of the diagnostic information, making diagnosis more efficient, precise, and informed about acute leukemia subtypes.

**Hardware support in leukemia diagnosis.** Information technology abounds, and the most important point in using acute leukemia diagnostics is to develop robust hardware support so that system performance can be optimized as well as data processing capacity or imaging. Advances in hardware make it easier to analyze and interpret data, providing aid for more accurate diagnosis of

all forms of acute leukemia (Singha et al., 2021). When these high-performance computing systems are combined, they allow rapid and detailed examination of large amounts of biomedical information. Such systems use high-speed processors and large memory capacity to crunch through the relevant calculations, quickly providing information that is useful in diagnoses (Rehman et al., 2018). Moreover, parallel computing architectures also speed up the processing of data and facilitate real-time analysis and decision support.

Progress in hardware will be crucial to developing the implementation of imaging technologies specifically designed for treating acute leukemia. The high-resolution microscopes and optical-electronic systems used as imaging hardware are state-of-the-art, greatly enhancing sampling precision (Singha et al., 2021). These new developments provide unprecedented opportunities for precise and detailed observation, making it easier to see cellular morphology or irregularities that may be related to specific disorders of acute leukemia. The application of high-resolution microscopes and optical electronic systems, moreover enables clearer images to be obtained that allow for better visualization of the structures related to acute leukemia. They also allow for the careful analysis of cellular structure, which can help determine abnormalities and irregularities unique to each type of leukemia (Rehman et al., 2018). This precision in imaging is crucial to the diagnosis and characterization of various forms of acute leukemia.

In addition, hardware accelerators like GPUs add a great boost to the efficiency of acute leukemia analysis. It's just that GPUs are specially designed to speed up the kinds of computation involved in biomedical imaging or with large data sets (Arber et al., 2017). What these accelerators do is increase the speed and efficiency of the system's mathematical processing, allowing data to be processed even more quickly. Consequently, the system performs better in the high-speed evaluation of vast amounts of biomedical data, and diagnostic accuracy is improved for the analysis of acute leukemia (Williams et al., 2019). In the system architecture, taking advantage of hardware accelerators (GPUs) results in significant improvements to computational efficiency. In biomedical applications, their parallel processing capabilities make them quicker at performing complex algorithms needed to analyze images and speed up the diagnostic process. With such accelerated computing power, data can be interpreted and analyzed much faster than before (Rehman et al., 2018). This not only ensures increased accuracy in the analysis of acute leukemia samples but also results in greater efficiency as well.

**Experimental studies & expert system framework.** Experimental studies and the development of an expert system in clinics are also major elements in constructing resilient acute leukemia technology. These studies include in-depth validations and assessments, while the structure of an expert system gives it a sound foundation for reliable diagnosis (Rehman et al., 2018). In addition, there must be numerous experimental studies of the validation and comparison of algorithms against standards for analyzing data and assessing diagnostics. The purpose of these studies is to verify the validity, precision, and reliability of the newly developed technologies in testing for different types of acute leukemia (Singha et al., 2021). Vigorous experimentation guarantees that the system's performance will meet clinical needs and helps to give it credibility.

Creating the skeleton of an expert system involves several different aspects, including how its structure will be organized and how decisions are made. Such an end-to-end system includes numerous components that have been proven individually through algorithms, data processing techniques, and decision support systems (Rehman et al., 2018). The shell way inside the expert system encompasses several vital components, such as data storage cartridges, and operational commands in evidence-based procedures for use by doctors to judge acute leukemia cases; all these are carefully engineered to work together (Shah et al., 2021). Expert system architecture is precisely tailored to combine the most rigorously tested and verified algorithms and methods. These parts are all brought together to work in concert, which requires precise and coordinated operations (Williams et al., 2019). The framework also provides data storage mechanisms to allow the structured entry and retrieval of extensive sets of information, which is essential for making sound biomedical decisions.

An important part of the expert system's architecture is that it requires user-friendly interfaces, specifically for doctors. These interfaces are intuitive, allowing the user to interact with them naturally (Singha et al., 2021). They allow clinicians to enter appropriate patient data, read the diagnostic output of machines, and deduce recommendations or findings generated by computers. These interfaces, which focus on user-friendliness to support healthcare professionals using them effectively and quickly, are highly innovative. In addition, decision support modules incorporated into the framework enhance diagnostic accuracy by supplying doctors with actionable information and guidance based on smart data analysis (Davis et al., 2014). These modules help practitioners understand the system's output, providing differential diagnoses and treatment recommendations as well as clinical data based on information analyzed.

The expert system skeleton consists of a mechanically organized structure in which verified congruous algorithms are tightly integrated with data input and processing procedures as well as decision support mechanisms and highly user-responsive interfaces. The fact that components are so well integrated within the system framework itself produces coherence and stability in support of accurate and actionable diagnoses of acute leukemia (Rehman et al., 2018). This user-centric design makes it easy to interact with, allowing healthcare professionals to enter and read key diagnostic information freely.

**Methodology.** By the blood rheology (hemorheology) we mean the study of the biophysical features of the blood as the viscous fluid. The blood viscosity is characterized by the conventional viscosity factor, which is referred to the certain conditions of the fluid flow (for instance, pressure, velocity).

Macrology treats blood as an integer, devoid of structure, although it is a dispersed system that is viscous, with a number of characteristics: viscosity of whole blood, viscosity of plasma, concentration of hemoglobin). Microrheology examines the rheological behavior of blood according to the properties of its components in particular erythrocytes, which is also characterized by a number of indicators: erythrocyte aggregation, erythrocyte deformation. In the area of the Newtonian current in the arterial channel, where the speeds are high ( $250\div 270\text{ s}^{-1}$ ), the red blood cells are completely disaggregated - it is in this zone that their deformation begins giving them the shape of ellipsoids. The cardiac function and the mechanical properties of the vascular walls are decisive for blood flow. In arterioles and pre-capillaries ( $70\text{ s}^{-1}$ ), large aggregates are first broken down into separate chains - coin bars that navigate along the flow, reduce their length in accordance with the increase in speed, not too strong, aggregates rapidly disintegrate, The largest aggregates are delayed (Riva & Feke, 1981; Kozhemiako et al., 2010).

In capillaries, the red blood cells move one after another. In the region of the Kesson current, in the venous channel ( $2,5\div 10\text{ s}^{-1}$ ) the most intensive aggregate formation takes place, after which partial aggregates are destroyed, in the heart at the rate of shift «0» practically all the erythrocytes are assembled in aggregates (Riva & Feke, 1981).

Dependence of the viscosity force on the velocity gradient becomes non-linear (Kozhemiako et al., 2010):

$$F_{visc} = f\left(\left(\eta \frac{dV}{dZ}\right)^2\right), \quad (1)$$

where  $dV/dZ[1/s]$  – is the velocity gradient that shows how the velocity  $V$  changes at the change by the unit of the distance in the direction of the transition from the layer to layer, otherwise the shift velocity;  $\eta$  [Pa·s] is the coefficient of the dynamic viscosity of the fluid, that characterizes the resistance of the fluid to the displacement of the layers.

The properties of such environment first of all depend of the fluid. The system in general will have other, greater viscosity that depends on the form and concentration of the particles. For the case of small concentrations of  $C$  particles, the formula (Riva & Feke, 1981) is valid:

$$\eta = \eta (1 + KC), \quad (2)$$

For the ellipsoids  $K$  increases and is determined by the values of its semi axes and their relations. If the structure of the particles changes (for instance, when the conditions of the process change), the coefficient  $K$  in (3) and the viscosity of such suspension changes.

To study the viscosity of blood and plasma, the method of rotational or capillary viscometry is most often used (Arber et al., 2017; Riva & Feke, 1981) with the determination of viscosity in the range of shear rates from 1 to 200  $s^{-1}$ . The choice of working shear rates is determined by the fact that in this range all the effects of nonlinear blood behavior are observed. The data obtained are indicated as apparent viscosity  $\eta$ .

**Realization of expert optical-electronic information system for the processing of the biomedical information.** The main problem, solved by the expert optical-electronic information system for the processing of the biomedical information is the automatic allocation of the arbitrary objects, their counting and calculation of the morphologic parameters. The system performs the processing of the biomedical images, in particular, counting of the number of the red blood cells (RBC) on the biomedical image (Fig.1). The search of the objects is performed completely in the automatic mode. The automatic mode operates with a certain error (practical error on the images of rather large area is approximately 10%).



Figure 1. Biomedical image of the blood rheology

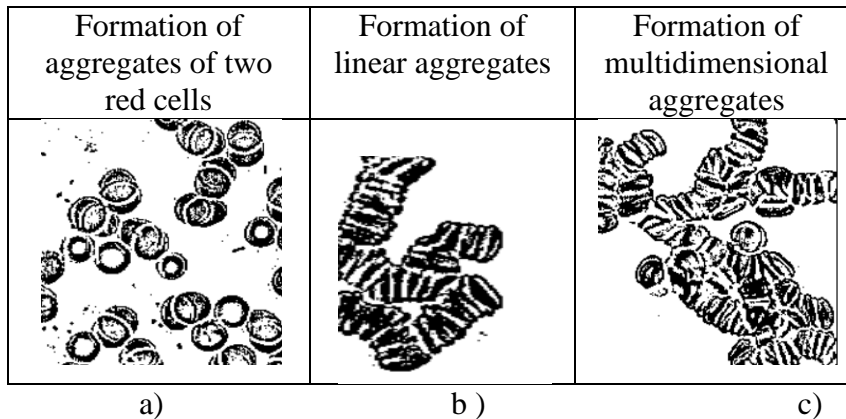
Hemodynamic indices of the blood flow are defined by the biophysical parameters of the cardiovascular system activity (for instance, stroke volume), structural characteristics of the vessels (their radius and flexibility) and directly by the properties of the blood itself (viscosity) (Riva & Feke, 1981; Kozhemiako et al., 2010).

By the blood rheology (hemorheology) we mean the study of the biophysical features of the blood as the viscous fluid. The blood viscosity is characterized by the conventional viscosity factor, which is referred to the certain conditions of the fluid flow (for instance, pressure, velocity).

Macrology treats blood as an integer, devoid of structure, although it is a dispersed system that is viscous, with a number of characteristics: viscosity of whole blood, viscosity of plasma, concentration of hemoglobin). Microrheology examines the rheological behavior of blood according to the properties of its components in particular erythrocytes, which is also characterized by a number of indicators: erythrocyte aggregation, erythrocyte deformation. In the area of the Newtonian current in the arterial channel, where the speeds are high ( $250 \div 270 s^{-1}$ ), the red blood cells are completely disaggregated - it is in this zone that their deformation begins giving them the shape of ellipsoids. The cardiac function and the mechanical properties of the vascular walls are decisive for blood flow. In arterioles and pre-capillaries ( $70 s^{-1}$ ), large aggregates are first broken down into separate chains - coin bars that navigate along the flow, reduce their length in accordance with the increase in speed, not too strong, aggregates rapidly disintegrate, The largest aggregates are delayed (Riva & Feke, 1981; Kozhemiako et al., 2010).

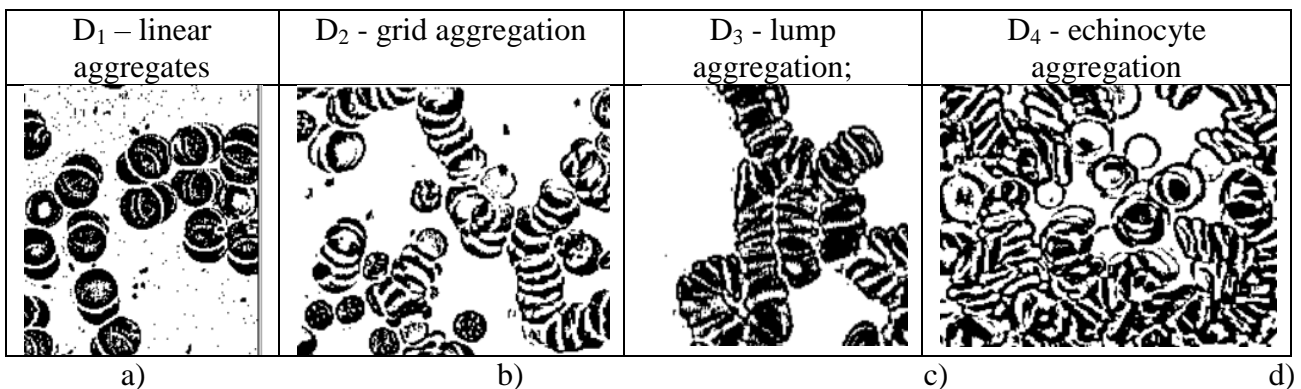
One of the main characteristics of the rheological properties of blood is the aggregation of red blood cells, which is determined the nature of superficial intererythrocyte interaction. States of erythrocytes spontaneous aggregation are classified (Riva & Feke, 1981): the formation of

aggregates of two red cells; the formation of linear aggregates; the formation of multidimensional aggregates (fig. 2).



**Figure 2.** States of erythrocytes spontaneous aggregation

Erythrocyte aggregates is also divided of types (Riva & Feke, 1981; Lerche et al., 2010):  $D_1$  – linear aggregates;  $D_2$  - grid aggregation;  $D_3$  - lump aggregation;  $D_4$  - echinocyte aggregation (fig.3).



**Figure 3.** Types of erythrocyte aggregates

Normal physiological aggregation has the character of linear chains in the form of coin columns consisting of 4-5 cells of Fig. 3.a.

On average, human blood samples are disaggregated at shift speeds of  $50^{-1}$  seconds, allowing for complete hydrodynamic disaggregation of erythrocytes in the vascular bed. At very low shear speeds, even the normal red blood cells are almost entirely grouped into 3.b coin bars.

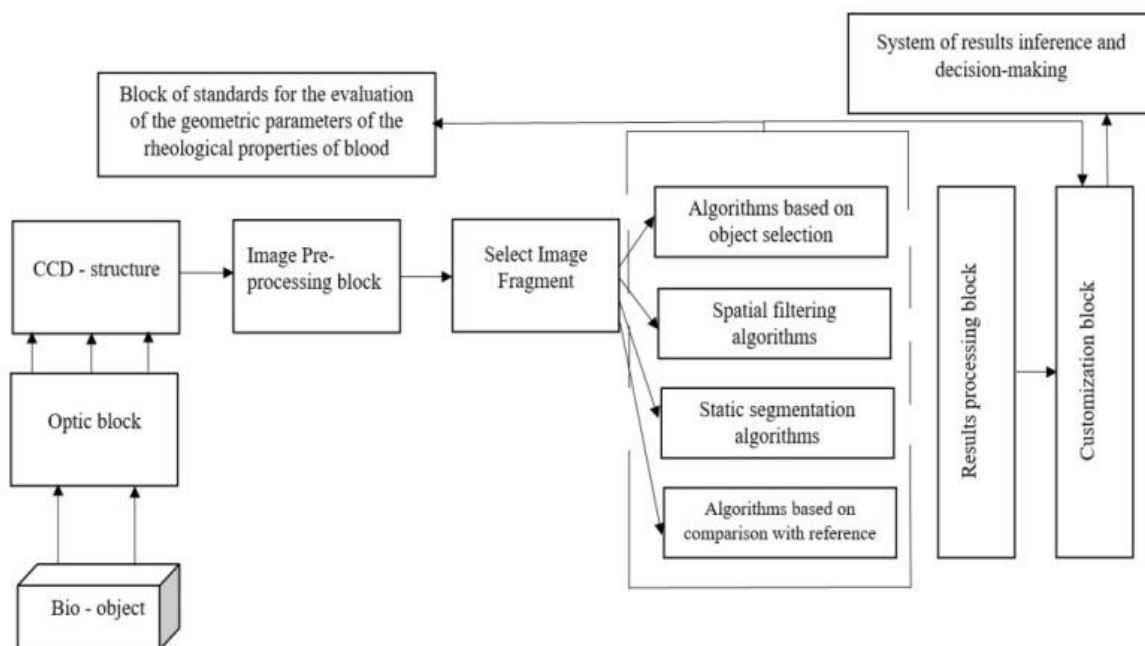
At a certain rate of shift, the coin bars are completely destroyed and a suspension flow from individual cells is observed (Riva & Feke, 1981; Lerche et al., 2010). The main feature of a pathological aggregation is a lump aggregation with increased adhesion strength between erythrocytes maintain aggregation even at ( $\gamma = 250^{-1}$  seconds). Lump assemblies are preserved even at high shear speeds, which makes normal blood flow impossible in micro-vessels (fig. 3.c). The resulting stable assemblies are dropped through a system of shunts past the capillary channel into the venous system, thus ensuring continuity of blood flow. This leads to a discrepancy between the volumetric blood flow through the main vessels and the level of microcirculation and the formation of plasma capillaries free of erythrocytes. A direct consequence of the pathological aggregation of red blood cells is the centralization of blood flow and the insufficiency of tissue perfusion (Riva & Feke, 1981; Stoltz et al., 2010).

Changing the shape from discoid to spherical leads to the impossibility of free packing of red blood cells, i.e., to an increase in the area of contact (aggregation). Echinocytic transformation

significantly increases the strength of aggregates (Fig. 3, d).

**Realization of the optical-electronic system for the analysis of blood rheology.** In general case, the input variables, used in the expert system, can be presented either in qualitative or quantitative form. By means of optical-electronic expert system (OEES) the introduction of the input variables, conversion of the quantitative variables into qualitative is performed. The given expert system on the base of the fuzzy logic introduction also realizes the functions of the collection, storage, correlation analysis and usage of knowledge, obtained by the experts in order to analyze the rheologic properties of the blood (Posudin, 2014; Kostruba et al., 1996; Saldan et al., 2017).

Structure scheme of the fuzzy-expert optical-electronic system for the analysis of blood rheology is shown in Fig. 4. OEES consists of the optical block, photosensitive matrix for the registration of the biomedical object of the research, unit for the image preprocessing of biomedical information, block for select of image fragment, blocks for performance of algorithms based on object selection, spatial filtering algorithms, static segmentation algorithms and algorithms based on comparison with references, results proceeding block, customization block, block of standards for evaluation of geometric parameters of the rheological properties of blood, system of results inference and decision-making. Unit of the fuzzy processing and output, used in the expert system allows to process complex hierarchical structures of the input variables, which can be presented in the form of the tree.



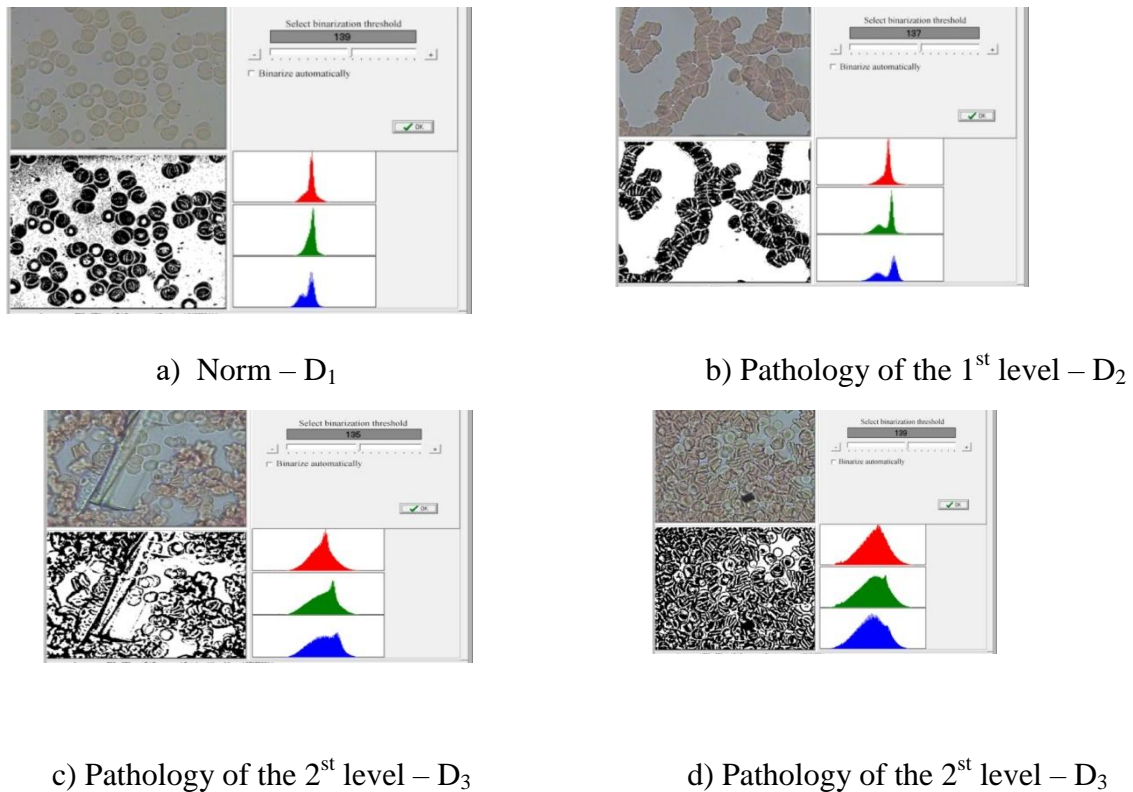
**Figure 4.** Structure scheme of the fuzzy-expert optical-electronic system for the analysis of blood rheology

**Results and Discussion.** For instance, after the examination by means of optical-electronic expert system, of the patients with the possible pathological changes, the results, shown in Fig. 5 are fixed.

To a great extent it is connected with the fact that the blood has the internal structure, that is the suspension of the formed elements in the solution – plasma. As 93% of the formed elements are the red blood cells, in case of the simplified model examination the blood is the suspension of the red blood cells in the physiological salt solution.

Changes in the rheological parameters of blood are one of the significant mechanisms of the formation of insufficient blood supply in the early stages the development of the disease. Main pathological effects violations of rheological properties in the blood can lead to micro-flow failure circulation, the extreme manifestation of which may lead to a decrease in trophism and the

development of ischemic syndrome, a violation of micro-rheology and an increase in the viscosity of blood, which causes an increase in total peripheral resistance and the development of arterial hypertension syndrome, to atherosclerotic changes in blood vessels, to a violation of hemorheology, which contributes to increased thrombosis (Riva & Feke, 1981).



**Figure 5.** Results of the biomedical images processing

**Conclusion.** It enabled to perform complex efficient diagnostics and increase by 10% the validity and speed of the diagnostics. Information features (information parameters) of the analysis of the biomedical images, in particular, for the assessment of the rheologic properties of the blood, are formed. Algorithm and optical-electronic fuzzy expert system for the analysis of the rheological properties of the blood are suggested, they are used for the increase of the diagnostic validity which is a determining factor in the biomedical diagnostics. Changes in the rheological parameters of blood are one of the significant mechanisms of the formation of insufficient blood supply in the early stages the development of the disease. In this way, the main focus of modern clinical hemorheology is the search diagnostic and prognostic criteria for various diseases and rheological correction methods violations.

**Acknowledgements.** Підготовлено та видано за грантової підтримки Національного фонду досліджень України в рамках проєкту 2022.01/0135 “Розробка лазерно-фотонного лікувально-діагностичного комплексу медичної реабілітації пацієнтів з політравмами різного ступеня важкості”

**Conflicts of Interest.** The authors declare no conflict of interest.

## References

- Abdeldaim, A. M., Sahlol, A. T., Elhoseny, M., & Hassanien, A. E. (2018), «Computer-aided acute lymphoblastic leukemia diagnosis system based on image analysis. *Advances in Soft Computing and Machine Learning in Image Processing*, 131-147.
- Ahmed, I. A., Senan, E. M., Shatnawi, H. S. A., Alkhraisha, Z. M., & Al-Azzam, M. M. A. (2023). Hybrid techniques for the diagnosis of acute lymphoblastic leukemia based on fusion of CNN features. *Diagnostics*, 13(6), 1026.
- Ansari, S., Navin, A. H., Sangar, A. B., Gharamaleki, J. V., & Danishvar, S. (2023). A customized efficient deep learning model for the diagnosis of acute leukemia cells based on lymphocyte and monocyte images. *Electronics*, 12(2), 322.
- Arber, D. A., Borowitz, M. J., Cessna, M., Etzell, J., Foucar, K., Hasserjian, R. P., ... & Vardiman, J. W. (2017). Initial diagnostic workup of acute leukemia: guideline from the College of American Pathologists and the American Society of Hematology. *Archives of pathology & laboratory medicine*, 141(10), 1342-1393.
- Blondal, R., Sturludottir, M.K., Hardarson, S.H., Halldorsson, & Stefansson G.H.,E. (2011). Reliability of vessel diameter measurements with a retinal oximeter. *Graefes. Arch. Clin. Exp. Ophthalmol*, 249(9), 1311-7.
- Cree, M.J., Gamble, E., & Cornforth D. (2005). Color normalisation to reduce inter-patient and intra-patient variability in microaneurysm detection in color retinal images. *WDIC2005 ARPS workshop on digital image computing*, Brisbane, Australia, 163-168.
- Gilmore, E.D., Hudson, C., Preiss, D., & Fisher J. (2005). Retinal arteriolar diameter, blood velocity, and blood flow response to an isocapnic hyperoxic provocation. *Am. J. Physiol. Heart. Circ. Physiol*, 288(6), H 2912-7.
- Harris A., Sergolt R.C., Spaeth G.L. et al. (1994). Color Doppler analysis of ocular vessel blood velocity in normal tension glaucoma. *Am. J. Ophthalmol*, 118, 642-649.
- Kozhemiako V. P., Pavlov S.V., Burdenyuk I. I., & Hamdi R.R. (2010). *Optoelectronic technology analysis of biomedical image: monograph*, Vinnitsya: VNTU.
- Lerche D., Baumler H., Kucera W. et al. (1989). Die Fließeigenschaften von Blut und ihre Charakterisierung mittels hamorheologischer Methoden. *Folia Haematol*, 116(5), 631—652.
- M. Kostruba, & Vlokh O.G.(1997). Device for measurement of micro-object linear dimension . in *International Conference on “Optical Storage, Imaging and Transmission of Information”*, Kiev Ukraine: SPIE the International Society for Optical Engineering, (pp. 73 – 77).
- Michelson, G., Schmauss G., & Langhans M.J. (1996). Principle, validity and reliability of scanning laser Doppler flowmetry. *J. Glaucoma*, 5, 99-105.
- Pavlov, S.V., Vassilenko, V.B., Saldan, I.R., Vovkotrub D.V., & Poplavskaya, A.A. (2016). Methods of processing biomedical image of retinal macular region of the eye, *Proc. SPIE 9961, Reflection, Scattering, and Diffraction from Surfaces V*, 99610X.
- Posudin Y. I. (2014). *Physics with Fundamentals of Biophysics*; ed. by S. J. Kays. 2nd Ed. Kiev : Printline.
- Pryzwan, Tomasz & Dolibog, Patrycja & Kierszniok, Klaudia & Pietrzyk, Barbara. (2024). Blood rheological properties and methods of their measurement. *Annales Academiae Medicae Silesiensis*, 78, 1-10.
- Riva, C.E. & Feke, G.T. (1981). Principles of retinal laser Doppler velocimetry. The characteristics of the scattering electric field incident to the photodetector surface. *The biomedical lasers: Technology and clinical applications*. Ed. by L. Goldmann. New York, Berlin: Heidelberg, Springer-Verlag.

- Riva, C.E., Harino, S., Petrig, B.L. & Shonat R.D. (1992). Laser Doppler flowmetry in the optic nerve. *Exp. Eye Res*, 55, 499-506.
- Rotshtein A. (1998). Design and Tuning of Fussy IF – THEN Vuly for Medical Didicol Diagnosis in Fuzzy and Neuro-Fuzzy Systems in Medicine, USA. *CRC-Press*. pp. 235–295.
- Rovira, R., Bayas, M.M., Pavlov, S.V., Kozlovskaya, T.I., Kisała, P., et al. (2015). Application of a modified evolutionary algorithm for the optimization of data acquisition to improve the accuracy of a video-polarimetric system, *Proc. SPIE 9816, Optical Fibers and Their Applications*, 981619 (December 18, 2015).
- Saldan Y. R., Pavlov S. V., Vovkotrub D.V., & Wójcik W. et al. (2017). Efficiency of optical-electronic systems: methods application for the analysis of structural changes in the process of eye grounds diagnosis, *Proc. SPIE 10445, Photonics Applications in Astronomy, Communications, Industry, and High Energy Physics Experiments*, 104450S.
- Stoltz J.F., Streiff F., & Larcan A. et al. (1987). Interet de la mesure de la deformabilite et de l'agregation des hematies en biologie clinique. *Ann. med. Nancy et Est.* 26(6), 431—436.
- Stoltz, J.F., Streiff, F., Larcan A. et al. Interet de la mesure de la deformabilite et de l'agregation des hematies en biologie clinique. *Ann. med. Nancy et Est.* 26(6), 431—436.
- Timchenko, L.I., Pavlov, S.V., Wójcik, W., et al. (2017). Bio-inspired approach to multistage image processing, *Proc. SPIE 10445, Photonics Applications in Astronomy, Communications, Industry, and High Energy*, 104453M.
- Wong, T.Y., Klein, R., Sharrett A.R., et al.(2014). Retinal arteriolar diameter and risk for hypertension. *Ann. Intern. Med*, 140(4), 248-255.

## МАТЕМАТИЧНЕ МОДЕЛЮВАННЯ ТА ОБЧИСЛЮВАЛЬНІ МЕТОДИ

УДК 004.02, 004.62

### Застосування методів статистичного аналізу для вибору виконавця задачі у розподіленій обчислювальній системі

**Роман Віталійович Слободян**

аспірант,

кафедра автоматизації та інтелектуальних інформаційних технологій

Вінницький національний технічний університет

21021, м. Вінниця, вул. Хмельницьке шосе, 95

ORCID 0009-0005-9834-9802

**Ілона Віталіївна Богач**

к.т.н., доцент,

доцент кафедри автоматизації та інтелектуальних інформаційних технологій

Вінницький національний технічний університет

21021, м. Вінниця, вул. Хмельницьке шосе, 95

ORCID 0000-0001-9398-8529

**Марія Володимирівна Барабан**

к.т.н., доцент,

доцент кафедри автоматизації та інтелектуальних інформаційних технологій

Вінницький національний технічний університет

21021, м. Вінниця, вул. Хмельницьке шосе, 95

ORCID 0000-0002-6745-351X

**Анотація.** Робота присвячена оптимізації процесу розподілу завдань у розподілених обчислювальних системах. За допомогою застосування методів статистичного аналізу розроблено підхід до автоматизації вибору виконавців, що дозволяє підвищити ефективність розподілу завдань, покращити щоденну продуктивність та задоволеність працівників. Дослідження показує, що використання оптимізованого підходу дозволило зменшити середню тривалість обробки звернень користувачів обраного типу з 34 до 31 хвилини, що на 7% ефективніше порівняно з випадковим розподілом задач, тим самим покращуючи якість обслуговування та продуктивність.

Запропонована уніфікована модель для оптимізованого розподілу завдань враховує такі ключові аспекти, як профілі внутрішніх користувачів, рівень їх навантаження, пріоритетність задач, взаємодію між виконавцями та інші доступні ресурси системи. Ця модель забезпечує баланс між компетентністю працівників та швидкістю обробки завдань, що суттєво підвищує продуктивність всієї системи.

Особливу увагу приділено авторській методології, побудованій на основі інструментів Salesforce CRM, яка дозволяє ефективно використовувати історичні дані щодо продуктивності працівників для визначення найбільш підходящих виконавців. У поєднанні зі статистичними методами аналізу великих обсягів даних цей підхід сприяє не лише оптимізації розподілу завдань, але й прогнозуванню часу їх виконання, виявленню аномалій у процесах та розробці гнучких стратегій розподілу. Врахування компетенцій та продуктивності працівників сприяє високій якості виконання завдань, скороченню часу обробки та зниженню навантаження, що критично важливо для ефективної роботи розподілених систем.

Загалом, запропоноване дослідження підтверджує, що застосування статистичного аналізу та інструментів CRM сприяє підвищенню ефективності роботи розподілених обчислювальних систем. Це відкриває перспективи для впровадження оптимізованих стратегій розподілу задач у різних галузях, враховуючи постійний ріст обсягу даних та складність бізнес-процесів.

**Ключові слова:** розподіл завдань, оптимізація, обчислювальні системи, статистичний аналіз, профілі користувачів, пріоритизація.

## **Statistical analysis methods application for a task distributor selection in a distributed computing system**

**Roman V. Slobodian**

Postgraduate student of automation and intelligent information technologies department  
Vinnytsia National Technical University  
95 Khmelnytske shose, Vinnytsia, Ukraine, 21021  
ORCID 0009-0005-9834-9802

**Ilova V. Bogach**

PhD, associate professor of automation and intelligent information technologies department  
Vinnytsia National Technical University  
95 Khmelnytske shose, Vinnytsia, Ukraine, 21021  
ORCID 0000-0001-9398-8529

**Maria V. Baraban**

PhD, associate professor of automation and intelligent information technologies department  
Vinnytsia National Technical University  
95 Khmelnytske shose, Vinnytsia, Ukraine, 21021  
ORCID 0000-0002-6745-351X

**Abstract.** This paper focuses on optimizing the task distribution process in distributed computing systems. By applying statistical analysis methods, a strategy has been developed to automate the selection of task performers, improving the efficiency of task distribution, daily productivity, and employee satisfaction. The research shows that the optimized approach reduced the average processing time for specific user requests from 34 to 31 minutes, which is 7% more effective compared to random task allocation, thereby enhancing service quality and overall productivity.

The proposed unified model for optimized task distribution considers key factors such as internal user profiles, their workload levels, task priority, interaction among performers, and other available system resources. This model balances employee competencies with the speed of task processing, significantly improving the system's overall performance.

Particular attention is given to the methodology based on Salesforce CRM tools, which allows for the effective use of historical data on employee performance to identify the most suitable task performers. Combined with statistical data analysis methods, this approach not only optimizes task distribution but also enables accurate time prediction for task completion, identification of process anomalies, and the development of flexible distribution strategies. Considering both competencies and productivity ensures high-quality task execution, reduces processing time, and minimizes workload, which is critical for the efficient operation of distributed systems.

In overall, the proposed study confirms that the use of statistical analysis and CRM tools enhances the efficiency of distributed computing systems. This opens opportunities for the implementation of

optimized task distribution strategies across various sectors, especially in the context of the growing volume of data and the complexity of business processes.

**Key words:** task distribution, optimization, computing systems, statistical analysis, user profiles, prioritization.

**DOI:** <https://doi.org/10.31649/1999-9941-2024-60-2-122-133>

**Вступ.** З кожним днем рівень складності повсякденних завдань, що постають перед працівниками у сучасних компаніях продовжує зростати. Це зумовлено збільшенням обсягів даних, еволюцією технологій та ускладненням бізнес-процесів. За таких умов вирішення завдань вимагає не лише розуміння домену проблеми чи певних технічних знань, але й стратегічного підходу.

Перш ніж приступати до виконання завдання, критично важливо впевнитись, що працівник має всі необхідні вхідні дані. Це включає чіткий опис задачі, поточний стан системи, очікуваний результат, а також розуміння можливого впливу на інші процеси. Недостатність, нечіткість або недостовірність цієї інформації може стати причиною затримок, помилок та додаткових витрат ресурсів.

Через зростання складності завдань, усе частіше трапляються ситуації, коли їх важко або ж взагалі неможливо вирішити самостійно. У таких випадках наступним кроком є визначення необхідних ресурсів для успішного виконання завдання. Це можуть бути як індивідуальні фахівці, так і групи працівників, окремі відділи компанії або навіть сторонні вендори, якщо внутрішніх знань та навичок недостатньо.

Зважаючи на вище згадані виклики, зростає потреба у автоматизованих рішеннях для розподілу завдань. Тут на допомогу приходять методи статистичного аналізу, які дозволяють оптимізувати цей процес.

Використання методів статистичного аналізу спрощує вибір виконавців, адже дозволяє автоматично оцінювати продуктивність, завантаженість і швидкодію працівників, ефективно розподіляючи завдання для оптимального використання ресурсів. Крім того, такі підходи допомагають прогнозувати час виконання, виявляти можливі аномалії та формувати стратегії розподілу, що враховують різноманітні фактори, таким чином підвищуючи швидкість і якість виконання, знижуючи витрати, та покращуючи конкурентоспроможність компанії в умовах сучасного ринку.

**Огляд літератури.** Проблема постійного росту рівня складності повсякденних задач турбує дослідників з декількох причин.

По-перше, у міру того, як завдання стають складнішими, працівники компаній все частіше стикаються з необхідністю освоювати нові навички, адаптуватися до нових інструментів та методик. Зазвичай це призводить до росту витрат часу необхідного для виконання завдань, а іноді і до зниження якості роботи та навіть підвищення стресу. Дослідження (Trautmann et al., 2011) доводить те, що невідповідність між компетенціями працівників і вимогами завдань призводить до низької продуктивності та задоволення роботою, що є серйозним викликом для компаній.

По-друге, зі збільшенням складності завдань також збільшується і об'єм роботи, що необхідно виконати, а разом з ним і ймовірність помилок, що так чи інакше необхідно виправити. Така ситуація не тільки підвищує навантаження на працівників, а й за часту призводить до негативних ітерацій у циклі вирішення задач – коли компанії витрачають більше ресурсів на вирішення проблем чи їх наслідків, ніж на їх попередження (Trautmann et al., 2011; Zhang, 2023).

По-третє, із розвитком обчислювальних систем з'являється необхідність постійного навчання працівників та оптимізації робочих процесів (Dery & Sebastian, 2017). Також,

необхідно зауважити, що деякі зміни можуть негативно вплинути на мотивацію працівників та їх здатність ефективно виконувати призначені завдання, тому вносити їх потрібно зважено, а негативні наслідки пом'якшувати шляхом впровадження процесу підтримки внутрішніх користувачів.

І нарешті, зі зростанням складності завдань, працівники компаній можуть відчувати все більший стрес та напруження, що безпосередньо впливає на їх рівень задоволення поточною посадою та на їх ефективність протягом тривалого часу (Dery & Sebastian, 2017). Цей аспект є важливим адже без ефективних та зручних у використанні рішень зростає ймовірність вигорання та втрати кваліфікованих працівників.

Перелічені вище фактори змушують як роботодавців так і дослідників шукати нові підходи для організації та оптимізації робочих процесів, інвестувати в навчання персоналу, розробляти та впроваджувати автоматизовані рішення для заощадження часу працівників, підвищення якості згенерованих рішень, тощо.

Проблема параметризації задач перед подальшим пошуком їх рішень привертає увагу науковців з кількох ключових причин.

По-перше, вона сприяє ефективному розв'язанню складних завдань, дозволяючи структурувати їх у вигляді моделей та оптимізувати процес виконання. Це особливо важливо в системах з великою кількістю обчислень та великим обсягом даних.

По-друге, параметризація є ключовим елементом у розробці систем автоматизації, які розподіляють завдання між виконавцями з метою оптимізації використання ресурсів та підвищення продуктивності. Це дозволяє підвищити точність та надійність результатів, а також досліджувати нові методи й моделі для вирішення різноманітних завдань.

Наприклад, у роботі (Huang et al., 2023) стверджується, що параметризація задач є важливою у контексті генерації траєкторій для роботів у робото-технічній галузі. Основна мета полягає в досягненні стабільності, надійності та продуктивності у виконанні завдань маніпуляції. Проблеми, які часто виникають у плануванні часово-оптимальних траєкторій, включають неоднорідність швидкості та прискорення, похибки у з'єднаннях, вібрації роботів та пошкодження двигунів. Вирішення цих проблем включає увагу до такого параметру, як розрив, що є похідною за часом прискорення. Однак нелінійні та неконвексні особливості обмежень на розрив ускладнюють оптимізацію траєкторій, що вимагає значних обчислювальних витрат.

У роботі (Xie et al., 2023) стверджується, що параметризація повсякденної для водіїв задачі щодо визначення полос руху на дорозі має вирішальне значення у випадку, якщо транспортний засіб керується за допомогою штучного інтелекту. У мережах інтелектуальних транспортних систем це дозволяє виявляти та контролювати рух транспортних засобів для забезпечення безпеки на дорозі. Параметризація є ключовим елементом алгоритмів виявлення полос руху. Такі алгоритми можуть базуватися на традиційних методах обробки зображень або на глибокому навчанні.

У дослідженні (Shi et al., 2023) йде мова про необхідність параметризації у задачах сегментації екземплярів, семантичної сегментації, оцінки монокулярної глибини, оцінки пози людини та відповідності форми.

У статті (Li et al., 2023) стверджується, що параметризація задач необхідна для адаптації роботів до змін у сцені в умовах неструктурованого середовища. У статті розглядається метод навчання за демонстрацією, який поєднує динамічні моделі руху та параметризоване завдання з метою покращення загальної продуктивності та розв'язання завдань маніпулювання об'єктами. Введення параметрів завдання дозволяє роботу навчитися більше локальних геометричних особливостей та узагальнити навчальну модель для невідомих ситуацій, забезпечуючи кращу адаптацію до змін в навколишньому середовищі.

У роботі (Mahmood et al., 2023) підкреслюється твердження про те, що параметризація задач є необхідною для вивчення складності обчислень в контексті вирішення проблеми в

штучному інтелекті. Розглядаються різні підходи та проблеми рішення, але основна увага приділяється параметризованій складності, яка визначається через параметризацію проблеми, що є важливою для практичного використання та має велику важливість для здійснення рішень у контексті ефективності за умови постійного або повільного зростання параметрів.

У доповіді (Perez-Villeda et al., 2023) йде мова про те, що параметризація задач є необхідною для досягнення підтримки правильної траєкторії руху та її адаптації до різних умов виконання завдань. Також наводиться приклад того, як використання параметризації дозволяє роботу адаптуватися до різних умов, що може бути важливо для успішної роботи в різних сценаріях. Наприклад, розмір посудини, куди робот наливає рідину, або її об'єм є важливими досягнення певного рівня наповнення. Такий підхід допомагає уникнути проблем з виконанням завдань поза областю даних для навчання, допомагаючи роботу досягти більш широкої генералізації та адаптивності при виконанні завдань.

У праці (Fu et al., 2023) йде мова про те, що параметризація задач є необхідною для досягнення більш гнучкої, універсальної поведінки виконавця задачі в різних сценаріях. У тексті зазначено, що off-policy meta-RL методи успішно узагальнюються в середовищах, де відмінності між завданнями можуть бути описані одним вектором (наприклад, зміни у швидкості цілі). Однак для завдань із більш різноманітними варіаціями (наприклад, тягнути кружку, натиснути кнопку, відкрити двері тощо) вище згадані методи показують гіршу універсальність порівняно з on-policy методами та методами на основі файн-тюнінгу. Параметризовані навички, які розширюють дискретні навички за допомогою неперервних параметрів, стають більш універсальними і придатними до використання для вирішення схожих проблем, оскільки їх можна легко адаптувати до різних ситуацій.

У статті (Gao et al., 2023) стверджується, що параметризація задач є важливою для поліпшення здатності невеликих попередньо навчених моделей до універсального застосування, а використання матричних методів декомпозиції дозволяє збільшити кількість параметрів під час подальшої підгонки. Це створює можливість розширення потенційно важливих параметрів моделі, поліпшуючи її здатність узагальнення.

У доповіді (Celber, 2023) розглядається необхідність використання параметризації задач для автоматичного створення організаційних структур систем шляхом оцінки поставлених цілей та доступності агентів-виконавців. Це дозволяє генерувати організаційні позиції та структури, що відображають вимоги до ресурсів та можливості їх заповнення агентами. Використання параметрів, таких як цілі та наявність агентів, допомагає автоматизувати процес створення організацій, роблячи його більш ефективним та гнучким.

Виходячи зі сказаного вище стає зрозуміло, що проблема параметризації задач для вибору виконавця існує у різних сферах науки та техніки, від робототехнічних галузей до штучного інтелекту, а її важливість важко переоцінити, оскільки правильно параметризована задача значно спрощує процеси пошуку можливих рішень.

**Матеріали та методи.** Мета дослідження полягає в розробці алгоритму розподілу завдань між виконавцями у розподілених обчислювальних системах. Автори докладно аналізують методи та моделі для вибору виконавців, використовуючи статистичний аналіз та методологію, засновану на Salesforce CRM. Основним завданням є зменшення часу обробки завдань, підвищення ефективності, якості та ефективного використання ресурсів у системах масового обслуговування. Зазначається важливість розробки уніфікованої моделі, що враховує ключові фактори, що можуть впливати на розподіл задач у тих чи інших системах масового обслуговування, а саме: варіативність профілів клієнтів; різні навички та завантаженість агентів; наявність інструментів, необхідних для швидкого підлаштування процесу під поточні задачі задля забезпечення їх оптимального та ефективного розподілу у реальному часі.

Задачі дослідження:

1. Розробка уніфікованої моделі розподілу завдань шляхом врахування характеристик задач та виконавців. Завданням цього етапу є створення комплексної моделі, яка інтегруватиме аналіз характеристик завдань (складність, терміновість, пріоритетність) та компетенції виконавців (досвід, навички, поточне завантаження). Така модель повинна враховувати не лише технічні параметри завдання, а й можливість його успішного виконання конкретними фахівцями або командами. Мета полягає в тому, щоб забезпечити баланс між навантаженням виконавців та якістю виконання завдань.

2. Застосування методів оптимізації для покращення ефективності та зменшення часу обробки завдань. На цьому етапі передбачається впровадження методів оптимізації, таких як лінійне програмування, евристичні алгоритми та методи штучного інтелекту. Оптимізація полягатиме в тому, щоб знизити затримки у виконанні завдань, використовуючи передові математичні та алгоритмічні підходи для розподілу задач між виконавцями. У результаті це зменшить витрати на виконання завдань, збільшуючи продуктивність та економічну ефективність компанії.

3. Використання статистичного аналізу для вибору виконавців. Цей етап є результатом виконання попереднього та пропонує використання статистичного аналізу для автоматизовано вибору виконавців задачі шляхом аналізу їх продуктивності на основі результатів виконання схожих завдань у минулому. Аналіз включає оцінку швидкості, точності, досвіду, а також завантаженості працівників. Використання цих даних дозволить забезпечити найбільш ефективний розподіл завдань, уникаючи перевантаження виконавців, таким чином запобігаючи можливому вигоранню, а також підвищує якість отриманих результатів.

4. Оцінка ефективності розподільчої системи. Цей етап полягає у оцінці ефективності системи шляхом аналізу ключових показників продуктивності, таких як час виконання завдань, кількість помилок та задоволеність клієнтів. Збір і аналіз цих даних дозволять виявити проблемні аспекти системи та підвищити її продуктивність. Також, оцінка допоможе визначити, чи реалізована система відповідає очікуванням і чи є необхідність у її подальшому вдосконаленні.

5. Вивчення динаміки змін у системі. Оскільки системи можуть змінюватися з плином часу або ж під впливом нових технологій або ж через збільшення обсягу завдань, важливо дослідити, як ці зміни впливатимуть на загальну продуктивність працівників компанії. Динаміка змін дозволить виявити, наскільки система адаптивна до нових викликів, таких як зміни у навантаженні, поведінці або ж зміни навантаження через збільшення об'єму роботи, котрий необхідно виконати кожному окремо взятому агенту.

6. Аналіз взаємозв'язку факторів, таких як навички, обмеження та інші, на продуктивність працівників у розподільчій системі. Вивчення впливу різних факторів на продуктивність виконавців є важливим етапом, оскільки на ньому відбувається аналіз таких аспектів, як рівень професійних навичок, доступність ресурсів, обмеження за часом та робочий графік. Виконання цього дослідження дозволить визначити, які чинники найбільш впливають на продуктивність і як вони можуть бути враховані в розподільчій системі для досягнення найбільш оптимальних результатів.

7. Розробка стратегій оптимального використання людських та технічних ресурсів для максимізації продуктивності та забезпечення оптимального розподілу завдань. На думку авторів, стратегії повинні бути спрямовані на максимізацію продуктивності компанії за рахунок ефективного використання як людських, так і технічних ресурсів. Це включає збалансований розподіл завдань, автоматизацію рутинних процесів та використання передових технологій для забезпечення гнучкості та швидкості виконання. Таким чином, ці стратегії допоможуть мінімізувати втрати часу та ресурсів, зменшити рівень стресу працівників, та підвищивши загальну ефективність роботи компанії у довготривалій перспективі.

**Результати та Обговорення.** Застосування методів статистичного аналізу дозволяють вирішити проблему вибору найбільш підходящого виконавця задачі у розподілених обчислювальних системах адже, враховуючи невпинно зростаючу складність обчислень, пошук кандидата для виконання завдання вимагає не лише технічних, а й стратегічних рішень.

У статті розкривається аналіз результатів експериментів та пропонується методологія вибору, покликана оптимізувати процес пошуку виконавця задачі у системі масового обслуговування (Amaral et al., 2023). на базі Salesforce CRM базуючись на кореляції даних про власне задачі, що були вирішені, та даних, що відображають продуктивність та доступність працівників.

Ретроперспективний статистичний аналіз інформації про виконані завдання є корисним інструментом для:

1. Оцінка ефективності, виявлення трендів і аналіз варіативностей. Статистичний аналіз допомагає виявити зміни в продуктивності з часом, що дозволяє приймати більш обґрунтовані управлінські рішення. За його допомогою можна аналізувати історичні дані для відстеження прогресу або проблем у виконанні завдань, що допомагає виявити ключові тренди, такі як стабільність процесів або повторювані помилки (Pham & Yang, 2023; Wang & Lin, 2023; Lee, & Kim, 2023). Це також надає можливість оцінити, які методи управління чи підходи найкраще працюють у певних умовах, що підтверджено в дослідженнях процесуального контролю та управління якістю.

2. Визначення підпроцесів, що тривають найдовше для їх подальшого покращення. Аналізуючи зібрані дані можна виявити етапи роботи, які споживають найбільше часу, що є основою для подальшого їх удосконалення. Це дає можливість виявити «вузькі місця» в процесі і визначити, які саме підпроцеси слід оптимізувати. Дослідження (Gupta & Bharti, 2020; Michaud; Ivanisenko & Radivilova, 2015) показали, що визначення тривалих підпроцесів є ключовим для зниження затримок та підвищення продуктивності в багатьох галузях.

3. Оптимізації розподілу завдань, надання додаткової підтримки найменш ефективним виконавцям або вдосконалення процесів роботи. Використання ретроспективного аналізу дає змогу не лише оцінити ефективність виконавців, але й допомогти тим, хто має труднощі з виконанням завдань. Це дозволяє керівникам вжити заходів щодо покращення навичок цих виконавців, що в свою чергу підвищує загальну ефективність. Такий підхід сприяє більш точному і ефективному розподілу ресурсів, а також підтримці працівників, які мають труднощі (Pham & Yang, 2023; Ivanisenko & Radivilova, 2015; Hu, Wu & Dong, 2023).

4. Розуміння потреб у ресурсах та часі, що мають бути вирішені для розв'язання різних видів завдань, а ретроспективний аналіз надає точні дані щодо того, скільки ресурсів і часу вимагають різні типи повторюваних завдань (Michaud F.). Наявність такої інформації значно полегшує планування та розподіл ресурсів компанії як у поточних, так і у майбутніх проектах. Ба більше, ретроперспективний аналіз дозволяє передбачити можливі ризики, затримки та оптимізувати використання людських і технічних ресурсів, що підтверджується дослідженнями у сфері управління проектами та аналізу затримок (Gupta & Verma, 2022; Jones, 2021).

З тверджень, наведених вище, стає зрозуміло, що ретроспективний статистичний аналіз допомагає оптимізувати робочі процеси, підвищуючи ефективність і знижуючи витрати через більш точне планування і розподіл ресурсів.

Для дослідження у даній роботі було виконано аналіз задач користувачів розподіленої обчислювальної системи на базі Salesforce CRM. Серед усього списку задач було зроблено вибірку тих, що стосувались проблем доступу кінцевих клієнтів до персонального кабінету на порталі. Усі обрані задачі вимагали схожих рішень, як от відправки системного сповіщення з інструкціями для зміни пароллю, скидання налаштувань двох-факторної

автентифікації, тощо. Приблизні приклади звернень наведено у таблиці 1.

**Таблиця 1.** Приблизні приклади звернень користувачів

| Номер, п.п. | Заголовок                        | Текст звернення   |
|-------------|----------------------------------|---|
| 1           | Unable to Access Account         | Hi, I'm having trouble accessing my account. I've tried entering my username and password multiple times, but it keeps saying my credentials are incorrect. Can you please assist me in resolving this issue?   |
| 2           | Réinitialisation de Mot de Passe | Bonjour, j'ai oublié mon mot de passe pour me connecter à mon compte. J'ai essayé le lien de réinitialisation, mais je ne reçois aucune instruction de réinitialisation par e-mail. Pouvez-vous m'aider à regagner l'accès à mon compte?                              |
| 3           | Probleme mit der Anmeldung       | Guten Tag, ich habe Probleme, auf mein Konto zuzugreifen. Ich habe meinen Benutzernamen und mein Passwort mehrmals eingegeben, aber es wird immer angezeigt, dass meine Anmeldeinformationen falsch sind. Können Sie mir bitte bei der Lösung dieses Problems helfen? |

**Примітки:** У прикладі №1 користувач намагається кілька разів ввести логін і пароль, але система постійно виводить повідомлення про невірні облікові дані. Запит на допомогу у відновленні доступу до акаунту. У прикладі №2 користувач забув свій пароль і спробував скористатися функцією скидання пароля, але не отримав електронного листа з інструкціями. Він просить допомоги у відновленні доступу до свого облікового запису. У прикладі №3 Користувач кілька разів вводив правильний логін і пароль, але система постійно показує помилку. Він запитує допомогу у вирішенні проблеми з доступом до свого акаунту.

**Джерело:** Таблицю розроблено авторами на основі дослідження популярних запитів та часто запитуваних питань до служби підтримки користувачів різних обчислювальних систем.

На першому етапі дослідження, розподіл звернень здійснювався вручну, без урахування будь-яких характеристик задач, як от мови звернення, чи навичок працівників, як от володіння мовами. При аналізі виконання задач працівниками спостерігались як позитивні, так і негативні ознаки. Наприклад, кожен працівник мав задачі для обробки та так чи інакше справлявся з їх вирішенням, але іноді це вимагало додаткових затрат у часі. Спостерігались суттєві витрати зусиль на комунікацію працівників компанії з клієнтами у тих випадках, коли працівник на високому рівні не володів мовою клієнта і використовував онлайн сервіс для перекладу повідомлень (див. таблицю 2).

**Таблиця 2.** Порівняння середнього часу витраченого працівниками на обробку звернень на різних мовах

| Мова звернення →<br>Мова працівника ↓ | Англійська | Французька | Німецька |
|---------------------------------------|------------|------------|----------|
| Англійська                            | 25хв       | 45хв       | 37хв     |
| Французька                            | 32хв       | 23хв       | 45хв     |
| Німецька                              | 34хв       | 36хв       | 24хв     |

**Примітки:** Таблиця показує середній час, витрачений працівниками на обробку звернень, залежно від мови звернення та мови, якою володіє працівник. Це дозволяє проаналізувати, наскільки швидко обробляються звернення, коли мови звернення і працівника збігаються або відрізняються.

**Джерело:** розроблено авторами на основі аналізу власних експериментальних даних та досліджень (Churkina et al., 2023; Canestrino et al., 2022).

Проаналізувавши результати, наведені у Таблиці 2, можна дійти до висновку, що

середня тривалість обробки клієнтського звернення працівником компанії – 34хв. Але, порівнюючи отримане значення з середнім часом витраченим на аналогічні операції в оптимальних умовах (коли працівник вільно володіє мовою користувача, 24хв), стає зрозуміло, що різниця – суттєва. Працівник компанії витрачає на ~29% більше часу, що в свою чергу є індикатором неефективного розподілу роботи, котрий необхідно усунути.

Врахування продуктивності працівників при розподілі задач необхідне з кількох причин, а саме:

1. Оптимізація ефективності шляхом урахування здібностей, досвіду, талантів, навичок, тощо.
2. Підвищення якості результатів. Наприклад, вирішення задач, що вимагають уваги до деталей, зазвичай можуть більш бути більш якісно вирішені кроріткими робітниками.
3. Зменшення часу, необхідного для виконання завдань. Потенційно, більш досвідчені працівники можуть вирішити проблему за коротший час, ніж працівники-початківці.
4. Підвищення мотивації та задоволення працівників, адже оператори можуть відчувати позитивні емоції працюючи над задачами, котрі вони виконують краще за інших з тих чи інших причин.

На другому етапі дослідження було проведено визначення рівня володіння іноземними мовами серед працівників. Це було необхідно для подальшої зміни процесу розподілу звернень кінцевих користувачів виходячи з рівнів володіння мовами працівниками компанії таким чином, щоб звернення надходили в першу чергу тим, хто вільно володіє мовою звернення, а в випадку їх недоступності (наприклад, коли вони працюють над іншими задачами) задачі були делеговані тим, хто володіє мовою звернення краще за інших.

Шляхом оптимізованого розподілу було зменшено середню тривалість обробки звернення користувача до 31хв, що на 7% ефективніше випадкового розподілу.

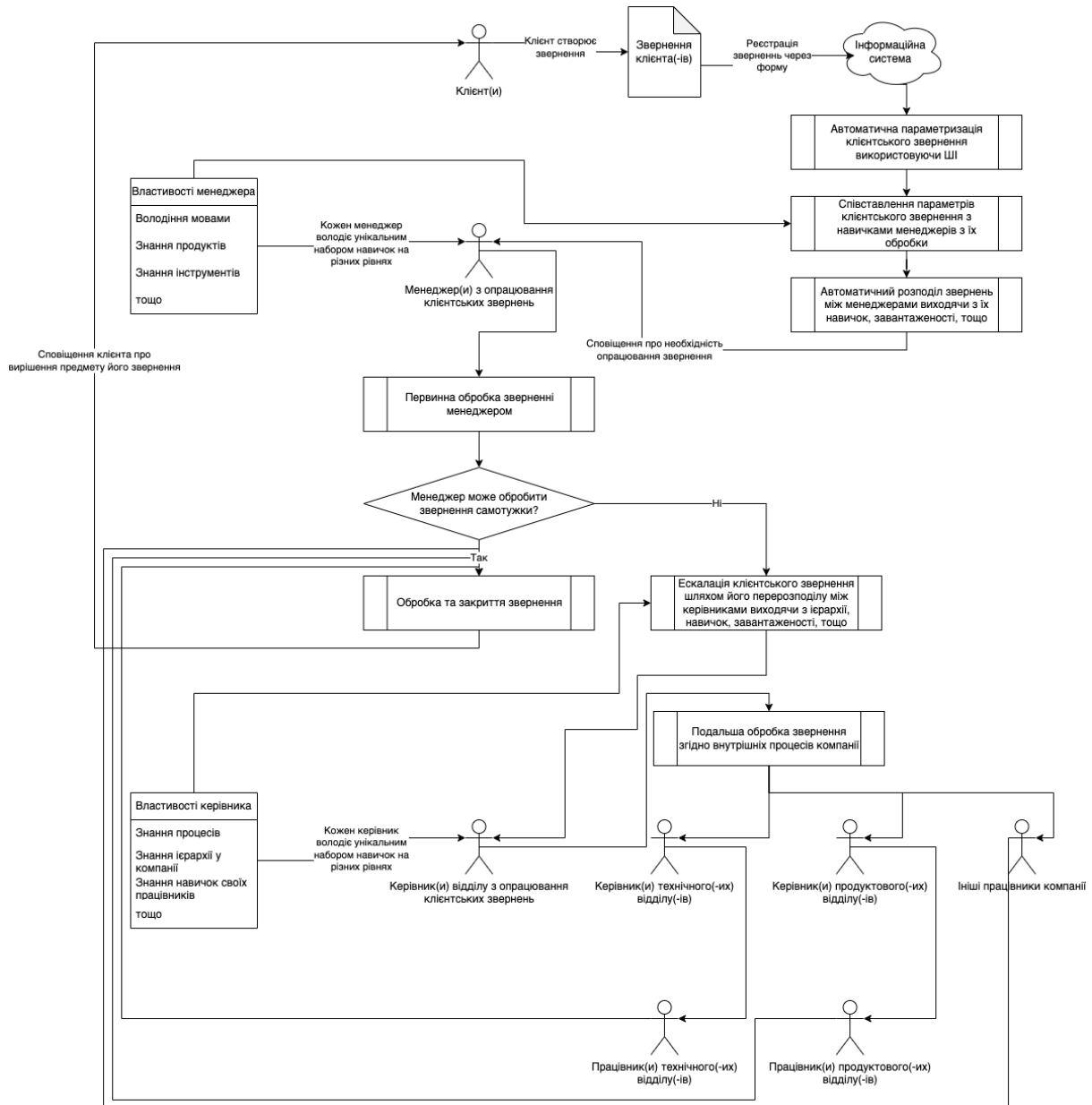
Однак, системи масового обслуговування не завжди є сталими, адже їх функціонал та характеристики можуть змінюватись з часом. Тому, виникає необхідність розробці уніфікованої моделі оптимізованого розподілу задач.

При розробці уніфікованої моделі оптимізованого розподілу задач важливо звернути увагу на кілька ключових аспектів:

1. Профіль користувача та компетентність: Розуміння навичок, вмінь та обмежень кожного виконавця. Це може включати експертність у певних областях, швидкість виконання різних видів завдань тощо.
2. Навантаження та обсяг завдань: Аналіз обсягу та типів завдань, що виникають у системі. Це важливо для розподілу завдань з урахуванням їх складності та терміновості.
3. Пріоритетність завдань: Визначення важливості та терміновості кожного завдання для оптимального розподілу в залежності від поточних обставин.
4. Взаємодія та комунікація: Забезпечення ефективної взаємодії між виконавцями та замовниками завдань. Це включає чіткість спілкування, обмін інформацією та спільне розуміння очікувань.
5. Адаптивність моделі: Здатність моделі пристосовуватися до змін у навантаженні, нових виконавців або змін у пріоритетності завдань. Гнучкість та можливість швидкого адаптування є ключовими для успішної роботи системи розподілу задач.
6. Ефективність ресурсів: Оптимізація використання ресурсів, включаючи людські та технічні, для максимізації продуктивності та забезпечення оптимального розподілу завдань.

Ці аспекти допомагають побудувати модель, що враховує специфіку системи та сприяє ефективному та оптимізованому розподілу завдань між виконавцями.

Приклад уніфікованої моделі вибору виконавця задачі у СМО що базується на результатах, отриманих вище, зображена на рисунку 1.



**Рисунок 1.** Приклад уніфікованої моделі вибору виконавця задачі у СМО

**Примітки:** Клієнт створює звернення (або задачу), котре реєструється через форму або іншу інформаційну систему. Далі автоматизована система виконує параметризацію поставленої задачі, використовуючи попередньо налаштовані критерії, як от тип звернення, терміновість, ключові параметри. Після реєстрації завдання відбувається порівняння параметрів звернення з навичками менеджерів. Система автоматично визначає, хто з менеджерів має відповідні навички для обробки звернення. Якщо менеджер може обробити звернення самостійно, відбувається його обробка та подальше закриття. Якщо ж потрібна додаткова допомога, відбувається ескаляція звернення на вищій рівень. Якщо звернення вимагає спеціальної уваги або залучення більш досвідчених працівників, воно передається на наступний рівень – керівникам або іншим спеціалістам. Система враховує не лише ієрархічну структуру, але й навички, завантаженість та інші важливі чинники. Після передачі задачі керівнику або спеціалістам, звернення проходить подальшу обробку згідно з внутрішніми процесами компанії. В результаті завершується обробка звернення та відправляється повідомлення клієнту про результат.

**Джерело:** авторська розробка.

**Висновки.** Застосування статистичного аналізу у виборі виконавців завдань у розподілених обчислювальних системах виявляється критичним через постійно зростаючу складність обчислень. У даному дослідженні, аналізуючи результати експериментів,

використано методологію вибору на основі Salesforce CRM. Ця методика ґрунтується на кореляції даних щодо виконаних завдань та продуктивності працівників. Підходячи до ретроспективного статистичного аналізу, вдалося виявити тенденції та зміни в продуктивності з часом, що стало основою для управлінських рішень у процесі оптимізації роботи.

Аналізуючи процес розподілу завдань вручну на початковому етапі, було виявлено значні недоліки. Не враховувалися основні характеристики завдань та навички працівників, що призводило до витрат часу на комунікацію або використання онлайн-сервісів для перекладу. Це призвело до істотних різниць у часі обробки, особливо при недостатній володінні мовою клієнта.

Врахування продуктивності працівників у розподілі завдань має кілька вагомих переваг. Це не лише оптимізує ефективність, але й покращує якість результатів, скорочує час виконання завдань та сприяє мотивації працівників. Оптимізована модель розподілу завдань, яка враховує профілі користувачів, навантаження, пріоритети, комунікацію та адаптивність, може стати ключовим інструментом для підвищення ефективності робочих процесів у системах масового обслуговування.

**Подяки.** Немає.

**Конфлікт інтересів.** Немає.

## References

- Amaral, C.J., Hübner, J.F. & Cranefield, S. (2023). Generating and choosing organisations for multi-agent systems. *Auton Agent Multi-Agent Syst*, 37, 41. <https://doi.org/10.1007/s10458-023-09623-8>.
- Canestrino, R., Magliocca, P., & Li, Y. (2022). The Impact of Language Diversity on Knowledge Sharing Within International University Research Teams: Evidence From TED Project. *Frontiers in Psychology*, Vol. 13. <https://doi.org/10.3389/fpsyg.2022.879154>
- Celber, D. (2023). Generating and Choosing Organizations for Multi-Agent Systems. *Autonomous Agents and Multi-Agent Systems*. Retrieved from <https://link.springer.com/article/10.1007/s10458-023-09623-8>.
- Churkina, O., Nazareno, L., & Zullo, M. (2023). The labor market outcomes of bilinguals in the United States: Accumulation and returns effects. *PLoS ONE* 18(6): e0287711. <https://doi.org/10.1371/journal.pone.0287711>
- Dery, K., & Sebastian, I. (2017). Employee Experience: Culture, Engagement and Leadership in the Digital Age. *MIT Center for Information Systems Research*. Retrieved from [https://c isr.mit.edu/publication/2017\\_0601\\_EmployeeExperience\\_DerySebastian](https://c isr.mit.edu/publication/2017_0601_EmployeeExperience_DerySebastian).
- Fu, H., Yu, S., Tiwari, S., Littman, M., & Konidaris, G. (2022). Meta-learning parameterized skills. *arXiv preprint*, arXiv:2206.03597.
- Gao, Z.-F., Zhou, K., Liu, P., Zhao, W. X., & Wen, J.-R. (2023). Small Pre-trained Language Models Can be Fine-tuned as Large Models via Over-Parameterization. In *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics*, Vol.1, (pp. 3819–3834). <https://doi.org/10.18653/v1/2023.acl-long.212>
- Gupta, S. K., & Verma, M. (2022). A Comprehensive Review of Statistical Approaches in Task Scheduling for Distributed Systems. *Advances in Statistical Computing*.
- Gupta, S., & Bharti, D. (2020). *Application of Statistical Techniques in Project Monitoring and Control*. In: Kapur, P.K., Singh, O., Khatri, S.K., Verma, A.K. (eds) Strategic System Assurance and Business Analytics. Asset Analytics. Springer, Singapore.

- [https://doi.org/10.1007/978-981-15-3647-2\\_7](https://doi.org/10.1007/978-981-15-3647-2_7).
- Hu, Q., Wu, X. & Dong, S. (2023). A Two-Stage Multi-Objective Task Scheduling Framework Based on Invasive Tumor Growth Optimization Algorithm for Cloud Computing. *J Grid Computing* 21, 31 <https://doi.org/10.1007/s10723-023-09665-y>
- Huang, H., Liu, H., Xia, C., Mei, H., Gao, X., & Liang, B. (2023). Sampling-based time-optimal path parameterization with jerk constraints for robotic manipulation. *Robotics and Autonomous Systems*, Vol. 170, p. 104530). Elsevier BV. <https://doi.org/10.1016/j.robot.2023.104530>
- Ivanisenko, I. N., & Radivilova, T. A. (2015). Survey of major load balancing algorithms in distributed system. In *2015 Information Technologies in Innovation Business Conference (ITIB) IEEE*. (pp. 89–92). <https://doi.org/10.1109/itib.2015.7355061>
- Jones, M. (2021). Exploring the Role of Statistical Analysis in Distributed System Optimization. *International Journal of Distributed Systems*.
- Lee, J. H., & Kim, T. Y. (2023). Optimization Strategies for Distributed Task Scheduling Using Statistical Methods. *Applied Computing*.
- Li, J., Cong, M., Liu, D. and Du, Y. (2023). Enhanced task parameterized dynamic movement primitives by GMM to solve manipulation tasks. *Robotic Intelligence and Automation*, Vol. 43, No. 2, pp. 85-95. <https://doi.org/10.1108/RIA-07-2022-0199>.
- Mahmood, Y., Meier, A., & Schmidt, J. (2023). Parameterized Complexity of Logic-based Argumentation in Schaefer’s Framework. *ACM Transactions on Computational Logic*, Vol. 24, Issue 3, pp. 1–25. ACM. <https://doi.org/10.1145/3582499>
- Michaud F. A statistical review of delay analysis methods used over the last decade. Retrieved from: <https://www.hka.com/a-statistical-review-of-delay-analysis-techniques-used-over-the-last-decade/>.
- Perez-Villeda, H., Piater, J., & Saveriano, M. (2023). Learning and extrapolation of robotic skills using task-parameterized equation learner networks. *Robotics and Autonomous Systems*, Vol. 160, p. 104309. Elsevier BV. <https://doi.org/10.1016/j.robot.2022.104309>.
- Pham, X. L., & Yang, Z. (2023). Task Scheduling in Distributed Computing Environments with Heuristic Algorithms. *Journal of Parallel and Distributed Computing*.
- Shi, H., Jiang, L., Zheng, J., & Zeng, J. (2023). Self-Parameterization Based Multi-Resolution Mesh Convolution Networks. In *Computer-Aided Design* (Vol. 162, p. 103550). Elsevier BV. <https://doi.org/10.1016/j.cad.2023.103550>
- Trautmann, M., Voelcker-Rehage, C. & Godde, B. (2011). Fit between workers’ competencies and job demands as predictor for job performance over the work career. *ZAF* 44, 339–347 <https://doi.org/10.1007/s12651-011-0078-2>
- Wang, H., & Lin, Y. (2023). Dynamic Task Allocation for Cloud Computing Systems Using Hybrid Algorithms. *Cloud Computing Journal*.
- Xie, T., Yin, M., Zhu, X., Sun, J., Meng, C., & Bei, S. (2023). A Fast and Robust Lane Detection via Online Re-Parameterization and Hybrid Attention. *Sensors*, Vol. 23, Issue 19, p. 8285). MDPI AG. <https://doi.org/10.3390/s23198285>.
- Zhang, L. (2023). The Changing Role of Managers. *American Journal of Sociology*, Vol. 129, Issue 2, pp. 439–484. University of Chicago Press. <https://doi.org/10.1086/727145>