

Київський національний університет технологій та дизайну

ТЕХНОЛОГІЇ ТА ІНЖИНІРИНГ

Науковий журнал

Том 26, № 2, 2025

Київ
2025

ISSN 2786-5371
e-ISSN 2786-538X

Засновник журналу:
Київський національний університет технологій та дизайну

Рік заснування: 1999 р.

Періодичність випуску: 6 разів на рік

*Рекомендовано до друку та поширення
через мережу Інтернет Вченою Радою
Київського національного університету технологій та дизайну
(протокол № 2 від 30 квітня 2025 р.)*

Державна реєстрація:
Ідентифікатор медіа: R30-05423
Рішення Національної ради України з питань телебачення і радіомовлення № 2520,
протокол № 22 від 8 серпня 2024 року

Журнал внесено до категорії «Б» Переліку наукових фахових видань України,
відповідно до якого можуть публікуватися результати дисертаційних робіт на здобуття наукових
ступенів доктора та кандидата наук зі спеціальностей:
технічні (наказ Міністерства освіти і науки України № 735 від 29 червня 2021 р.):
122 – Комп’ютерні науки, 131 – Прикладна механіка, 133 – Галузеве машинобудування,
151 – Автоматизація та комп’ютерно-інтегровані технології, 161 – Хімічні технології та інженерія,
132 – Матеріалознавство, 182 – Технології легкої промисловості, 141 – Електроенергетика,
електротехніка та електромеханіка, 144 – Теплоенергетика, 171 – Електроніка.

**Журнал представлено у міжнародних наукометричних базах даних,
репозитаріях та пошукових системах:** Національна бібліотека України імені В. І. Вернадського,
Crossref, Google Scholar, Litmaps, OUCI, Фахові видання України, EBSCO, Dimensions, UCSB Library,
University of Oslo Library, University of Hull Library, BASE, Ulrichsweb Global Serials Directory

Адреса редакції:
Київський національний університет технологій та дизайну
01011, вул. Мала Шияновська, 2, м. Київ, Україна
тел/факс: +38 (044) 256-84-27
E-mail: info@technologies-engineering.com.ua
<https://technologies-engineering.com.ua/uk>

Kyiv National University of Technologies and Design

TECHNOLOGIES AND ENGINEERING

Scientific Journal

Vol. 26, No. 2, 2025

Kyiv
2025

ISSN 2786-5371
e-ISSN 2786-538X

Founder:

Kyiv National University of Technologies and Design

Year of foundation: 1999

Frequency: 6 issues per year

*Recommended for printing and distribution
via the Internet by the Academic Council
of Kyiv National University of Technologies and Design
(Minutes No. 2 April 30, 2025)*

State registration:

Media identifier: R30-05423

Decision of the National Council of Television and Radio Broadcasting of Ukraine No. 2520,
Minutes No. 22 dated 08.08.2024

The journal is included in category “B” of the List of scientific professional publications of Ukraine, in which the results of dissertations for the degree of doctor and candidate of sciences can be published in the following specialties: technical (Order of the Ministry of Education and Science of Ukraine No. 735 dated 29.06.2021): 0612 – Database and network design and administration, 0715 – Mechanics and metal trades, 0714 – Electronics and automation, 0711 – Chemical engineering and processes, 0588 – Inter-disciplinary programmes and qualifications involving engineering, manufacturing and construction, 0723 – Textiles (clothes, footwear and leather), 0713 – Electricity and energy, 0714 – Electronics and automation

The journal is presented in the international scientometric databases, repositories and scientific systems: Vernadsky National Library of Ukraine, Crossref, Google Scholar, Litmaps, OUCI, Professional Publications of Ukraine, EBSCO, Dimensions, UCSB Library, University of Oslo Library, University of Hull Library, BASE, Ulrichsweb Global Serials Directory

Editors office address:

Kyiv National University of Technologies and Design
01011, 2 Mala Shyianovska Str., Kyiv, Ukraine
tel/fax: +38 (044) 256-84-27
E-mail: info@technologies-engineering.com.ua
<https://technologies-engineering.com.ua/en>

Редакційна колегія

Головний редактор

Ігор Панасюк	доктор технічних наук, професор, Київський національний університет технологій та дизайну, Україна
--------------	--

Заступник головного редактора

Борис Злотенко	доктор технічних наук, професор, Київський національний університет технологій та дизайну, Україна
----------------	--

Національні члени редколегії

Ірина Шведчикова	доктор технічних наук, професор, Київський національний університет технологій та дизайну, Україна
Володимир Стаценко	доктор технічних наук, професор, Київський національний університет технологій та дизайну, Україна
Інна Білоус	кандидат технічних наук, доцент, Національний технічний університет України «Київський політехнічний інститут імені Ігоря Сікорського», Україна
Сергій Денисюк	доктор технічних наук, професор, Національний технічний університет України «Київський політехнічний інститут імені Ігоря Сікорського», Україна
Валерій Жуйков	доктор технічних наук, професор, Національний технічний університет України «Київський політехнічний інститут імені Ігоря Сікорського», Україна
Володимир Мілих	доктор технічних наук, професор, Національний технічний університет «Харківський політехнічний інститут», Україна
Олег Ніконов	доктор технічних наук, професор, Київський національний університет технологій та дизайну, Україна
Володимир Осипенко	доктор технічних наук, професор, Національний університет біоресурсів і природокористування України, Україна
Дмитро Стаценко	кандидат технічних наук, доцент, Київський національний університет технологій та дизайну, Україна
Ірина Суходуб	кандидат технічних наук, доцент, Національний технічний університет України «Київський політехнічний інститут імені Ігоря Сікорського»
Олексій Чорний	доктор технічних наук, професор, Кременчуцький національний університет імені Михайла Остроградського, Україна
Олександр Шавьолкін	доктор технічних наук, професор, Київський національний університет технологій та дизайну, Україна
Людмила Галавська	доктор технічних наук, професор, Київський національний університет технологій та дизайну, Україна
Світлана Арабулі	кандидат технічних наук, доцент, Київський національний університет технологій та дизайну, Україна
Ольга Гараніна	доктор технічних наук, професор, Київський національний університет технологій та дизайну, Україна
Наталія Первая	доктор технічних наук, професор Київський національний університет технологій та дизайну, Україна
Яна Редько	доктор технічних наук, професор, Київський національний університет технологій та дизайну, Україна
Алла Славінська	доктор технічних наук, професор, Хмельницький національний університет, Україна
Анна Хімічева	доктор технічних наук, професор, Київський національний університет технологій та дизайну, Україна
Вікторія Плавач	доктор технічних наук, професор, Лауреат Національної премії України імені Бориса Патона, Київський національний університет технологій та дизайну, Україна
Олена Мокроусова	доктор технічних наук, професор, Лауреат Національної премії України імені Бориса Патона, Київський національний університет технологій та дизайну, Україна
Ольга Андреева	доктор технічних наук, професор, Київський національний університет технологій та дизайну, Україна
Володимир Бессарабов	доктор технічних наук, професор, Київський національний університет технологій та дизайну, Україна
Олена Іщенко	доктор технічних наук, доцент, Київський національний університет технологій та дизайну, Україна

Редакційна колегія

Євген Кузьмінський	доктор хімічних наук, професор, Національний технічний університет України «Київський політехнічний інститут імені Ігоря Сікорського», Україна
Володимир Левицький	доктор технічних наук, професор, Національний університет «Львівська політехніка», Україна
Богдан Савченко	доктор технічних наук, професор, Київський національний університет технологій та дизайну, Україна
Володимир Скорохода	доктор технічних наук, професор, Національний університет «Львівська політехніка», Україна
Владислав Страшний	доктор фармацевтичних наук, професор, Київський національний університет технологій та дизайну, Україна
Володимир Хоменко	доктор технічних наук, доцент, Лауреат Національної премії України імені Бориса Патона, Київський національний університет технологій та дизайну, Україна

Міжнародні члени редколегії

Анджей Борусевич	доктор наук, доцент, Міжнародний університет прикладних наук у Ломжі, Польща
Джасім Мохмед	доктор філософії, доцент, Технічний університет Аль-Фурат Аль-Авсат - Технічний коледж Аль-Муссаїб, Ірак
Юрай Герлічі	доктор технічних наук, професор, Жилінський університет, Словачка Республіка
Катерина Кравченко	доктор філософії, Жилінський університет, Словачка Республіка
Кей Берклінг	доктор філософії, професор, Кооперативний державний університет Баден-Вюртемберга (DHBW), Німеччина
Мірослав Скібневський	доктор технічних наук, професор, Школа інженерії ім. А. Джеймса Кларка Університету Меріленда, США
Олена Кизимчук	доктор технічних наук, професор, Дрезденський технічний університет, Німеччина
Лариса Очеретна	доктор філософії, доцент, Ліберецький технологічний університет, Чеська Республіка
Валейка Віргіліус	доктор філософії, професор, Каунаський технологічний університет, Литва

Editorial Board

Editor-in-Chief

Igor Panasiuk	Doctor of Technical Sciences, Professor, Kyiv National University of Technologies and Design, Ukraine
----------------------	---

Deputy Editor-in-Chief

Boris Zlotenko	Doctor of Technical Sciences, Professor, Kyiv National University of Technologies and Design, Ukraine
-----------------------	---

National Members of the Editorial Board

Iryna Shvedchykova	Doctor of Technical Sciences, Professor, Kyiv National University of Technologies and Design, Ukraine
Volodymyr Statsenko	Doctor of Technical Sciences, Professor, Kyiv National University of Technologies and Design, Ukraine
Inna Bilous	PhD in Technical Sciences, Associate Professor, National Technical University of Ukraine “Igor Sikorsky Kyiv Polytechnic Institute”, Ukraine
Sergiy Denisyuk	Doctor of Technical Sciences, Professor, National Technical University of Ukraine “Igor Sikorsky Kyiv Polytechnic Institute”, Ukraine
Valeriy Zhuikov	Doctor of Technical Sciences, Professor, National Technical University of Ukraine “Igor Sikorsky Kyiv Polytechnic Institute”, Ukraine
Volodymyr Milikh	Doctor of Technical Sciences, Professor, National Technical University “Kharkiv Polytechnic Institute”, Ukraine
Oleh Nikonov	Doctor of Technical Sciences, Professor, Kyiv National University of Technologies and Design, Ukraine
Volodymyr Osipenko	Doctor of Technical Sciences, Professor, National University of Bioresources and Nature Management of Ukraine, Ukraine
Dmytro Statsenko	PhD in Technical Sciences, Associate Professor, Kyiv National University of Technologies and Design, Ukraine
Iryna Sukhodub	PhD in Technical Sciences, Associate Professor, National Technical University of Ukraine “Igor Sikorsky Kyiv Polytechnic Institute”, Ukraine
Oleksiy Chorniy	Doctor of Technical Sciences, Professor, Kremenchutsk National University of Meny Mikhail Ostrogradskiy, Ukraine
Oleksandr Shavolkin	Doctor of Technical Sciences, Professor, Kyiv National University of Technologies and Design, Ukraine
Liudmyla Galavska	Doctor of Technical Sciences, Professor, Kyiv National University of Technologies and Design, Ukraine
Svitlana Arabuli	PhD in Technical Sciences, Associate Professor, Kyiv National University of Technologies and Design, Ukraine
Olga Garanina	Doctor of Technical Sciences, Professor, Kyiv National University of Technologies and Design, Ukraine
Natalia Pervaya	Doctor of Technical Sciences, Professor, Kyiv National University of Technologies and Design, Ukraine
Yana Redko	Doctor of Technical Sciences, Professor, Kyiv National University of Technologies and Design, Ukraine
Alla Slavinska	Doctor of Technical Sciences, Professor, Khmelnytsky National University, Ukraine
Anna Khimicheva	Doctor of Technical Sciences, Professor, Kyiv National University of Technologies and Design, Ukraine
Viktoriia Plavan	Doctor of Technical Sciences, Professor, Kyiv National University of Technologies and Design, Ukraine
Olena Mokrousova	Doctor of Technical Sciences, Professor, Kyiv National University of Technologies and Design, Ukraine
Olha Andreieva	Doctor of Technical Sciences, Professor, Kyiv National University of Technologies and Design, Ukraine
Volodymyr Bessarabov	Doctor of Technical Sciences, Associate Professor, Kyiv National University of Technologies and Design, Ukraine
Olena Ishchenko	Doctor of Technical Sciences, Associate Professor, Kyiv National University of Technologies and Design, Ukraine

Editorial Board

Evgen Kuzminsky	Doctor of Chemical Sciences, Professor, National Technical University of Ukraine “Igor Sikorsky Kyiv Polytechnic Institute”, Ukraine
Volodymyr Levytskiy	Doctor of Technical Sciences, Professor, National University “Lvivska Politechnika”, Ukraine
Bogdan Savchenko	Doctor of Technical Sciences, Professor, Kyiv National University of Technologies and Design, Ukraine
Volodymyr Skorokhoda	Doctor of Technical Sciences, Professor, National University “Lvivska Politechnika”, Ukraine
Vladyslav Strashnyi	Doctor of Pharmaceutical Sciences, Professor, Kyiv National University of Technologies and Design, Ukraine
Volodymyr Khomenko	Doctor of Technical Sciences, Associate Professor, Kyiv National University of Technologies and Design, Ukraine

International Members of the Editorial Board

Andrzej Borusiewicz	Doctor of Science, Associate Professor, International University of Applied Sciences in Lomza, Poland
Jasim Mohmed	PhD, Associate Professor, Al-Furat Al-Awsat Technical University – Al-Musssaib Technical college, Iraq
Juraj Gerlici	Doctor of Technical Sciences, Professor, University of Žilina, Slovak Republic
Kateryna Kravchenko	PhD, University of Žilina, Slovak Republic
Kay Berkling	Doctor of Philosophy, Professor, Cooperative State University of Baden-Württemberg (DHBW), Germany
Mirosław Skibniewski	Doctor of Technical Sciences, Professor, A. James Clark School of Engineering University of Maryland, USA
Olena Kizimchuk	Doctor of Technical Sciences, Professor, Technical University of Dresden, Germany
Larisa Ocheretna	Doctor of Philosophy, Associate Professor, Libertsy Technical University, Cheska Republic
Valeika Virgilijus	Doctor of Philosophy, Professor, Kaunas University of Technology, Lithuania

ЗМІСТ

В. Войтенко, Р. Поліщук

Децентралізована генерація та її роль у підвищенні живучості енергетичних островів та критичної інфраструктури: сучасні тенденції та перспективи 11

А. Каяфюк

Вплив обробки лог-файлів на швидкість навчання та точність класифікації дефектів 27

М. Киричек

Градієнтний бустінг як інструмент для вирішення задач класифікації в умовах обмежених даних 37

Д. Кучинська, О. Іщенко, Р. Качан, О. Сумська, О. Роїк

Антимікробні та сорбційні властивості полімерних плівок на основі модифікованого крохмалю з ксероформом для ветеринарного призначення 48

М. Овчарук, О. Топчій

Реологічні властивості муцину равлика *Achatina fulica* 58

М. Савка

Аналіз основних моделей, методів та засобів збору даних в Інтернеті речей 66

О. Салій, А. Бублик, М. Попова, Л. Фуклева

Дослідження впливу допоміжних речовин на фармако-технологічні параметри капсул з рибавірином методом випадкового балансу 79

В. Шука

Використання штучного інтелекту для створення інформаційних кампаній в умовах війни: візуальні інструменти протидії дезінформації 89

CONTENTS

V. Voitenko, R. Polishchuk

Decentralised generation and its role in enhancing the resilience of energy islands
and critical infrastructure: Current trends and prospects 11

A. Kaiafiuk

Impact of log file processing on learning speed and defect classification accuracy..... 27

M. Kyrychek

Gradient boosting as a tool for solving classification problems in data-constrained environments 37

D. Kuchynska, O. Ishchenko, R. Kachan, O. Sumska, O. Roik

Antimicrobial and sorption properties of polymer films based on modified starch
with xeroform for veterinary purposes 48

M. Ovcharuk, O. Topchii

Rheological properties of mucin from the snail *Achatina fulica* 58

M. Savka

Analysis of the key models, methods, and means of data collection in the Internet of Things 66

O. Saliy, A. Bublyk, M. Popova, L. Fukleva

Study of the effect of excipients
on the pharmacotechnological attributes of ribavirin capsules using the random balance method 79

V. Shchuka

The use of artificial intelligence for information campaigns in wartime:
Visual tools for countering disinformation 89



Decentralised generation and its role in enhancing the resilience of energy islands and critical infrastructure: Current trends and prospects

Vitalii Voitenko*

Postgraduate Student

National University of Life and Environmental Sciences of Ukraine

03041, 15 Heroiv Oborony Str., Kyiv, Ukraine

<https://orcid.org/0009-0005-0107-8357>

Roman Polishchuk

Postgraduate Student

National University of Life and Environmental Sciences of Ukraine

03041, 15 Heroiv Oborony Str., Kyiv, Ukraine

<https://orcid.org/0009-0009-0356-0226>

Abstract. Threats to critical infrastructure have heightened the relevance of autonomous energy supply and facilitated the transition to decentralised solutions. The study aimed to analyse the theoretical foundations of decentralised generation, its application in energy islands to enhance the resilience of critical infrastructure under emergency conditions. The methodological framework of the research included a theoretical synthesis of contemporary scientific approaches, real-world case studies of renewable technology development and autonomous energy supply models in the US, EU, and Ukraine, as well as a comparative analysis of autonomous energy supply models. The results demonstrated that solar and hybrid (solar-biogas) systems proved most effective. In Ukraine, over 5,000 residential solar installations were deployed in 2023, while in the village of Nyzhniy Bystryy (Zakarpattia), a hybrid energy island supplied electricity to 120 households. In the US, solar generation increased by 16% in 2023, with a 40% growth recorded in 2024. In the EU, the average levelised cost of solar electricity production decreased by 12% in 2023. In Tuscany (Italy), an energy island restored power supply within 12 hours in 2022 compared to 72 hours in the centralised system. In Ukraine, decentralised generation reduced recovery time to 3-6 hours in frontline regions (versus 12-18 hours traditionally), with electricity costs at 0.08 USD/kWh compared to 0.12 USD/kWh in the centralised system. The study revealed that decentralised generation enhanced grid flexibility and resilience by reducing transmission losses (up to -10%) and diversifying supply sources. Innovative technologies (digital control, artificial intelligence) were found to improve dispatching efficiency, enabling energy islands and microgrids to autonomously power critical facilities even during total grid collapse. The practical significance of the research lies in applying its findings to energy security strategies, microgrid development, and planning autonomous systems in vulnerable regions

Keywords: renewable energy sources; system autonomy; solar power; wind turbines; bioenergy; digital technologies

Introduction

Modern energy systems are undergoing transformation driven by climate change, decentralisation, digitalisation, and threats to critical infrastructure. In this context, decentralised generation (DG) is gaining prominence by offering localised energy production as an alternative to centralised models. This reduces transmission losses and

enhances flexibility and crisis resilience. Energy islands (autonomous or semi-autonomous microsystems within DG frameworks) ensure power system survivability, particularly during disruptions or overloads. Their value lies in rapid recovery, redundancy, and adaptability, guaranteeing basic energy supply in emergencies.

Suggested Citation:

Voitenko, V., & Polishchuk, R. (2025). Decentralised generation and its role in enhancing the resilience of energy islands and critical infrastructure: Current trends and prospects. *Technologies and Engineering*, 26(2), 11-26. doi: 10.30857/2786-5371.2025.2.1.

*Corresponding author



Copyright © The Author(s). This is an open access article distributed under the terms of the Creative Commons Attribution License 4.0 (<https://creativecommons.org/licenses/by/4.0/>)

The theoretical foundation for understanding the operation and benefits of such systems is actively developing in academic discourse. S.S. Yusuf & A.S. Abubakar (2023) conducted an in-depth analysis of grid-forming inverters, which are essential for stable microgrid operation. Their research highlighted the critical role of modern power electronics in enabling autonomous synchronisation of energy islands. M. Sadiq *et al.* (2025) focused on offshore microgrid management, demonstrating the efficacy of decentralised predictive control in complex environments such as ports or energy-intensive hubs. The study by M. Adnan *et al.* (2023) provided a comprehensive examination of transmission network planning methodologies in ultra-smart systems, which was pivotal for integrating large-scale decentralised generation into broader energy landscapes. P. Kadar & A. Vass (2023) demonstrated a practical implementation of an islanded urban microgrid operating under emergency conditions – a case study validating the feasibility and operational viability of energy islands in real-time scenarios.

Certain authors have focused on systemic approaches to reliability and energy system planning. Ye. Bardyk *et al.* (2024) investigated the operational reliability of power systems, particularly in the context of sustainable energy planning in Ukraine, and emphasised the importance of integrating risk assessment and adaptive mechanisms into the design of autonomous energy systems. Their research underscored the necessity of incorporating such considerations into decentralised energy system models. I. Blinov *et al.* (2023) substantiated the role of hydropower, specifically pumped storage capacities, in load balancing in Ukraine, which could be integrated into energy island models to enhance grid stability. This study is pivotal for understanding the potential of renewable energy sources (RES) in stabilising autonomous energy systems. E. Chomać-Pierzecka *et al.* (2022), in their review of the hydropower sector in Poland and the Baltic states, emphasised the potential of decentralised systems within sustainable development strategies. This study demonstrated how energy diversification policies could foster municipal-level energy autonomy.

Another key research direction involves assessing the potential of solar and wind energy for decentralised system development. I. Doronina *et al.* (2023) conducted a geospatial analysis to evaluate solar and wind energy potential in Ukraine, providing a foundation for decentralised energy system deployment. They also stressed the significance of geographical and climatic factors in energy island design. However, a research gap persists regarding the integration of these technologies in real-world conditions, particularly in Ukraine, where not only technical but also socio-economic challenges must be addressed. Existing studies often focus on individual system components, with insufficient attention paid to holistic approaches for assessing their effectiveness in highly unstable grid conditions.

The aim of this study was to examine the role of DG in enhancing the resilience of energy islands and critical infrastructure under emergency conditions. Special

attention was given to analysing the potential of such systems to ensure energy supply autonomy during crises, including military conflicts or natural disasters. The following key objectives were defined:

1. Evaluate theoretical approaches and practical developments in DG and energy islands, assessing their role in improving power system resilience during war or natural disasters;

2. Analyse global trends in resilient energy system deployment, examining case studies from various countries (the US, EU, and Ukraine) and assessing their effectiveness in disaster and conflict scenarios;

3. Investigate the impact of artificial intelligence (AI) and blockchain technologies on energy system and microgrid management, particularly their potential to enhance energy island stability during emergencies.

Materials and Methods

The study was conducted between September 2024 and January 2025 and focused on analysing aspects of distributed generation (DG) in the context of energy islands and their role in enhancing the resilience of critical infrastructure under emergency conditions. Primary attention was given to assessing the impact of various types of renewable energy sources (RES) – solar, wind, and bioenergy – on reducing CO₂ emissions, lowering energy consumption, and increasing the autonomy of energy systems. At the first stage, a comparative analysis of traditional centralised energy systems and modern decentralised energy systems was conducted to identify their advantages and disadvantages. The objective of the analysis was to evaluate how different energy supply models contribute to sustainable development amid growing energy demand, climate change, and increased vulnerability to external threats. The efficiency of the systems was analysed using content analysis to summarise their strengths and weaknesses, enabling an assessment of their potential for regional conditions and crisis scenarios.

The second stage involved evaluating the effectiveness of decentralised energy system integration in the United States (U.S. Energy Information Administration, 2023; 2024; International Renewable Energy Agency, 2024; Environmental Protection Agency, n.d.), the EU (Zakeri *et al.*, 2022; Frilingou *et al.*, 2023; International Energy Agency, 2023; 2024), and Ukraine (Shahini *et al.*, 2024; Hrytsiuk *et al.*, 2024; Diachenko, 2024). The selection of these countries was justified by their experience in implementing DG and RES, as well as the need to ensure energy autonomy due to vulnerability to external threats. The US, EU, and Ukraine have varying levels of energy system development and face different challenges, allowing for a comparison not only of the effectiveness of approaches to sustainable development and energy security under wartime or natural disaster conditions but also of socio-economic barriers.

To compare the efficiency of DG in the US, EU, and Ukraine, a comparative analysis method was applied using uniform criteria: recovery time, electricity costs, RES integration, and the scale of implemented solutions based

on statistical data collected from open reports. The analysis was based on a combination of quantitative (statistical indicators from official reports) and qualitative (content analysis of scientific publications and regulatory frameworks) approaches. Due to the publication policies of international organisations, which entail an annual delay in reporting, the analysis relied on the latest complete report from the International Renewable Energy Agency (2024). The 2024 report is expected no earlier than August-September 2025, while the 2025 report will be released in 2026.

The use of content analysis of publications and data analysis from international reports (International Renewable Energy Agency, International Energy Agency, Environmental Protection Agency) enabled the systematisation of approaches to DG implementation in the US, EU, and Ukraine and an assessment of DG efficiency under crisis conditions and its economic benefits. Inclusion criteria for sources required data on technological efficiency, environmental performance, and the relationship with DG in crisis scenarios. The regulatory framework was outlined through a review of key documents: Law of Ukraine No. 555-IV “On Alternative Energy Sources” (2003), Directive of the European Parliament and of the Council No. 2019/944 “On Common Rules for the Internal Market for Electricity and Amending Directive 2012/27/EU” (2019), Directive of the European Parliament and of the Council No. 2018/2001 “On the Promotion of the Use of Energy from Renewable Sources” (2018), Public Law No. 117-169 “Inflation Reduction Act of 2022” (2022), which regulate the development of DG (distributed generation) in the US, EU, and Ukraine, without conducting a comparative analysis of their legal approaches.

At the third stage, the practical application of modern digital technologies was described: Model Predictive Control, blockchain-based solutions for energy transactions, machine learning algorithms for load optimisation, as well as energy storage technologies (lithium-ion batteries, hydrogen storage systems, pumped hydro storage). The objective was to assess the potential for optimising resource allocation, reducing costs, and enhancing the overall resilience of energy systems. The reference manager Zotero was employed for source processing, facilitating systematic research organisation, accurate citation, and thematic

classification. The application of data search, analysis, and synthesis methods ensured a comprehensive study of DG’s impact on energy security, enabling the identification of key efficiency patterns and the formulation of evidence-based conclusions on improving system stability.

Results and Discussion

DG is defined as an energy production system that involves generating power directly at consumption sites, thereby reducing the load on centralised grids and improving energy efficiency. DG encompasses three primary types: solar, wind, and bioenergy, each with distinct theoretical foundations, advantages, and limitations that require critical evaluation to develop effective deployment strategies.

Solar energy, based on photovoltaic (PV) panels, is one of the most accessible DG types due to declining technology costs. This underscores its potential for integration into diverse infrastructure, from residential buildings to agricultural complexes. However, solar generation efficiency depends on insolation levels, limiting its applicability in regions with low solar irradiance, while also requiring significant land area for panel installation – a challenge in densely populated areas (Frilingou *et al.*, 2023). DG enables the creation of alternative energy sources that operate independently of centralised grids. Wind energy, which converts kinetic wind energy into electricity via turbines, holds significant potential in coastal regions with strong and consistent wind patterns. This indicates its capacity to deliver substantial energy output under optimal turbine placement. However, its implementation is hindered by high initial costs, noise pollution, and dependence on natural conditions, making it less versatile compared to other DG types. Bioenergy, derived from biomass (organic waste, agricultural residues, wood), is regarded as a key component of sustainable development, as it reduces greenhouse gas emissions and promotes waste recycling. Nevertheless, biomass utilisation faces logistical challenges in collection and transportation, as well as competition for resources with other agricultural needs, which may threaten food security, particularly in developing countries. Table 1 provides a concise comparative analysis of the main DG types – solar, wind, and bioenergy generation.

Table 1. DG types: Advantages, limitations, and optimal application conditions

DG types	Advantages	Limitations	Optimal application
Solar	Low technology costs, suitability for small-scale facilities.	Dependence on insolation, requirement for large installation areas.	Southern regions with high solar potential.
Wind	High potential in wind-rich zones, environmental sustainability.	Noise pollution, dependence on natural conditions, high initial costs.	Coastal zones and open plains.
Bioenergy	Waste recycling, CO ₂ emission reduction, resource stability.	Logistics of collection and transportation, competition with food production.	Rural areas with biomass availability.

Source: compiled by the authors based on M. Farghali *et al.* (2023), O. Diachenko (2024), E. Shahini *et al.* (2024)

Each type of DG has its optimal application scope, determined by both technical and geographical factors. Solar energy is viable in regions with high insolation levels;

however, it requires integration with battery storage systems to ensure stable power supply. Bioenergy, despite its resource stability and environmental benefits, necessitates

resolving logistical challenges, particularly in rural areas. Wind power demonstrates efficiency in coastal zones, but its economic viability is closely tied to investments in noise-reduction technologies and improving turbine efficiency.

The energy island concept involves the creation of an autonomous energy system capable of operating independently from the main grid and is categorised into fully autonomous (for local consumers) and integrated systems (with the capability to exchange energy with the centralised grid). Such systems are resilient to emergencies, such as natural disasters or military conflicts, which is particularly relevant for Ukraine under unstable conditions (Tröndle *et al.*, 2024). However, their implementation is complicated by economic constraints, including the high cost of establishing autonomous systems, as well as challenges in integrating with centralised grids under unstable demand conditions. To enhance the efficiency of energy islands, it is advisable to develop hybrid models that combine autonomy with integration capabilities, enabling adaptation to various usage scenarios.

The resilience of energy systems is defined as the ability to withstand disruptive influences, adapt to changes, and rapidly restore functionality, which is critically important in the context of decentralisation and the increasing share of RES, especially for Ukraine during wartime. There are three key characteristics of resilience: power redundancy, adaptability, and recovery speed. Backup capacities (diesel generators, batteries) ensure autonomy but require significant investment. Adaptability depends on the integration of RES with energy storage systems, enhancing grid flexibility; however, recovery speed is hindered in cases of large-scale damage (Rabocha *et al.*, 2023). Together, these characteristics ensure the energy system's ability to operate under disruption. A comparative analysis of approaches revealed that local autonomous systems ensure independence but are costly, whereas optimal dispatch control enhances flexibility but relies on forecasting.

The resilience of a power system is determined by three key characteristics: capacity redundancy, adaptability, and recovery speed. Backup capacities, such as diesel generators or batteries, ensure autonomy during disruptions but require significant investment. Adaptability, achieved through the integration of RES with energy storage systems, enhances grid flexibility, enabling the system to respond dynamically to failures. Recovery speed, although complicated in large-scale damage scenarios, is critically important for restoring normal operation. Together, these characteristics ensure the power system's ability to operate under disruption.

Critical infrastructure, encompassing healthcare, transport, telecommunications, and energy sectors, relies on the reliability of power supply. Theoretical analysis has shown that decentralised systems enhance its autonomy, minimising the risks of outages. However, the implementation of such systems in developing countries may be constrained by social factors, particularly the limited accessibility of technologies for small communities, which must be considered when designing energy development strategies.

The management of decentralised energy systems, particularly microgrids, is crucial for ensuring their flexibility and stability under variable loads and limited resources, especially in crisis situations, as seen in Ukraine during the war. There are two primary approaches: smart grids and model predictive control (MPC). Smart grids enable automatic regulation of power flows but require substantial investment and increase cyber risks, whereas MPC minimises losses and peak loads but depends on forecasting accuracy (Gonzalez-Reina *et al.*, 2024). An additional factor in stability is the involvement of prosumer households, which generate energy (via solar or biogas installations), enhancing autonomy and diversifying supply sources. However, integrating prosumers complicates management due to generation instability. Comparative analysis suggests that for crisis conditions, it is advisable to combine smart grids for baseline control and MPC for adaptation to changes, while simultaneously leveraging the potential of prosumers to increase autonomy.

The optimisation of DG integration is a key element for ensuring the stable operation of future energy systems, as it enables the effective combination of decentralised energy sources with centralised systems, reducing dependence on traditional grids and improving overall supply reliability. Successful DG integration depends on developing models that account for both technological and economic aspects, as well as the system's ability to adapt to changes in demand and supply. Load distribution models between centralised and decentralised sources involve strategies for the efficient use of solar panels, wind turbines, and other RES alongside conventional capacities. An important aspect is load management, which optimises energy use based on demand fluctuations and resource availability. However, such models have limitations: they often fail to account for RES intermittency, which can lead to system overloads, and require significant infrastructure modernisation investments to ensure synchronisation between centralised and decentralised sources (Table 2).

Table 2. Comparative analysis of approaches to the management and organisation of decentralised energy systems

Category/Approach	Advantages	Limitations/Challenges	Optimal application context
Autonomous energy islands	Complete grid independence, crisis resilience	High implementation costs, isolation	Crisis situations, isolated communities
Integrated energy islands	Flexibility, ability to exchange energy with the grid	Need for complex grid synchronisation	Regions with unstable demand, partial grid access
Smart grids	Automatic load balancing, scalability	High investment costs, cyber risks	High investment costs, cyber risks

Table 2. Continued

Category/Approach	Advantages	Limitations/Challenges	Optimal application context
Model predictive control	Load optimisation, precise forecasting	Dependence on forecast accuracy, complexity	Systems with variable demand, wartime or instability
Hybrid model (smart grids+model predictive control)	Balance between stability and adaptability	High implementation complexity, infrastructure requirements	Crisis conditions, strategic facilities
Local autonomous systems	Energy independence, backup capability	High cost, logistical challenges	Critical infrastructure facilities
Dispatch control	Flexibility, adaptability to changes	Dependence on forecasting systems	Urban energy systems
Prosumer integration	Source diversification, increased autonomy	Integration complexity, generation instability	Small communities, households

Source: created by the authors based on Y.A. Veremiichuk et al. (2024), C. Rehtanz et al. (2024), A.E. Gonzalez-Reina et al. (2024), I. Hrytsiuk et al. (2024)

A comparative analysis of theoretical approaches to DG integration reveals specific advantages and limitations, which largely depend on the context of their application. Combining smart grids for baseline management with MPC for adaptive response to fluctuations enables a balance between stability and flexibility. Such a hybrid approach is particularly relevant in unstable conditions, such as during wartime in Ukraine, due to significant resource availability fluctuations, infrastructure destruction, and unpredictable changes in energy flows. Consequently, the practical effectiveness of integration models

increases under the condition of their adaptability to real-world operational environments.

Global trends in renewable energy and DG indicate a steady increase in the role of new energy sources in power systems. Progress in reducing technology costs, the growing share of RES in the energy balance, and the optimisation of grid losses create prerequisites for the rapid deployment of DG across various regions worldwide. Table 3 presented key global indicators illustrating the development dynamics of this sector and confirms the relevance of transitioning towards more flexible and sustainable energy supply models.

Table 3. Key global statistical indicators in renewable and DG-based electricity (2023)

Indicator	Value
Share of RES in global electricity generation (2023)	In 2023, RES accounted for 30% of global electricity production.
Share of solar and wind generation (2023)	The combined share of solar and wind generation in 2023 reached approximately 13% of total electricity production.
Projected share of RES by 2030	The share of RES in the global energy mix is expected to exceed 50% by 2030.
Reduction in solar PV generation costs (2023)	In 2023, the levelised cost of solar PV generation decreased by 12% compared to 2022.
Reduction in onshore wind energy costs (2023)	In 2023, the levelised cost of onshore wind generation declined by 3% compared to 2022.
Reduction in solar panel costs (2010-2023)	Between 2010 and 2023, solar panel prices fell by 89%.
Reduction in grid losses due to DG	Through the implementation of DG, grid losses have been reduced to 10%.

Source: compiled by the authors based on International Energy Agency (2023; 2024), International Renewable Energy Agency (2024), Environmental Protection Agency (n.d.)

The presented data demonstrated not only the current level of RES adoption but also their rapid global expansion. The substantial reduction in generation costs and equipment expenditures (solar panels and wind turbines) indicates growing economic viability of decentralised solutions. When combined with reduced grid losses, these trends create favourable conditions for transitioning towards more sustainable, localised energy systems, particularly in countries with uneven access

to centralised power supply. Environmental and social aspects of DG play a pivotal role in shaping sustainable energy policies globally. The integration of RES not only contributes to greenhouse gas emission reductions but also generates additional opportunities for socio-economic development, particularly through improved electricity access and job creation. Table 4 summarises current statistical indicators illustrating DG's potential environmental and societal impact.

Table 4. Statistical indicators of DG's environmental and social impact

Indicator	Value
Global CO ₂ emissions from energy (2022)	In 2022, energy-related CO ₂ emissions reached 37 gigatonnes.
Projected CO ₂ emission reduction by 2050	International climate strategy envisages reducing energy-related CO ₂ emissions to net zero by 2050.
Population without electricity access (2023)	As of 2023, approximately 675 million people worldwide lacked access to electricity.
Projected new RES jobs by 2030	Over 30 million new jobs will be created in the RES sector by 2030.

Source: compiled by the authors based on International Energy Agency (2023; 2024)

The presented indicators demonstrated that DG holds not only technical but also strategic significance in the context of achieving global sustainable development goals. The reduction in CO₂ emissions, expanded access to electricity, and the creation of millions of new jobs indicate that DG serves as a crucial instrument for ecological transformation and enhancing energy equity. In the context of the war in Ukraine, where missile strikes can disable centralised generation facilities, DG gains particular importance due to its decentralised nature, which ensures the resilience of the power system. This is especially relevant for countries with limited access to centralised resources or those recovering from crises.

The implementation of DG demonstrates not only technical but also substantial social and environmental benefits, particularly in regions with limited access to centralised grids. In Ukraine, where the war leads to frequent power outages, DG becomes a key tool for meeting the

basic needs of the population in conflict zones, enabling thousands of people to maintain daily life through autonomous RES-based systems. In Global South countries, such as Sub-Saharan Africa and South Asia, more than half of the population still lacking electricity could be reached precisely through microgrids or autonomous RES-based solutions. Government programmes and donor initiatives in these regions prove the effectiveness of DG as a tool for rapid electrification with minimal environmental impact.

In developed countries, DG takes on a different role – it enhances the resilience of critical infrastructure during emergencies, such as hurricanes, armed conflicts, or disruptions in centralised grids. Thus, the potential of DG should be considered not merely as a technological solution but as a comprehensive socio-climatic development tool adaptable to diverse conditions, ranging from global instability to regional challenges (Table 5).

Table 5. Implemented case studies of DG deployment: environmental and social impact

Country/Region (project)	Implemented solution	Social/Environmental impact
India (Saubhagya Programme)	Solar microgrids for electrifying over 25 million rural households.	Large-scale improvement in electricity access for remote regions.
USA (California, Fire Station, 2021)	Local microgrid for fire service with autonomous energy supply during outages.	Enhanced resilience of critical infrastructure during crises.
Nigeria (Deutsche Gesellschaft für Internationale Zusammenarbeit (GIZ) + Government Initiatives)	Off-grid power systems for educational and medical facilities in remote areas.	Reduction in CO ₂ emissions and fossil fuel dependence.
South Africa (Rural Schools and Clinics)	Solar PV with energy storage in schools and clinics under municipal programmes.	Increased energy independence and improved learning/healthcare conditions.
Ukraine (DG Implementation in Communities, 2022-2023)	Solar systems and biogas plants for rural communities and temporary medical facilities in conflict zones.	Improved living conditions in affected regions via energy access for communities and medical points; reduced fossil fuel reliance through RES.

Source: compiled by the authors based on Environmental Protection Agency (n.d.), U.S. Energy Information Administration (2023), International Energy Agency (2024), E. Shahini et al. (2024), Y.A. Veremiichuk et al. (2024), O. Diachenko (2024)

The EU's experience in the development of DG serves as an important model for countries seeking to integrate RES into their energy systems. One of the key strategic documents in this regard is the European Green Deal, which envisages achieving climate neutrality by 2050 (Directive of the European Parliament..., 2019). This policy emphasised the development of smart grids and local energy sources. A crucial role in the legal framework for RES development is played by Directive (EU) 2018/2001 of the European Parliament and of the Council "On the Promotion of the Use of Energy from Renewable Sources" (2018), which sets binding targets for member

states regarding the share of renewable energy – at least 32% by 2030. To achieve this goal, countries are required to establish favourable conditions for DG development, including through "feed-in tariff" mechanisms for private investors. This approach stimulates the deployment of small and medium-sized power plants, reduces reliance on centralised generation, and enhances system resilience.

These efforts have already yielded tangible results, as evidenced by statistical indicators. The declining cost of technologies, the increasing share of RES, the implementation of energy islands, and the reduced power supply

restoration time all demonstrate the effectiveness of EU policies (the lack of publicly available DG statistics exhibits regional variations. In the EU, this is attributed to data fragmentation caused by the diversity of small-scale installations and the lack of unified reporting standards. In the US, DG statistics are limited due to a predominant focus on

centralised generation and insufficient integration of distributed system data. In Ukraine, DG data is unavailable due to the prioritisation of urgent energy system restoration during wartime and a lack of resources for detailed monitoring). Table 6 presented statistical indicators illustrating key trends in DG development in the EU.

Table 6. Statistical data on DG implementation in the EU

Category	Description	Data
DG implementation in the EU	Reduction in solar PV costs in 2023, fostering DG growth.	In 2023, the average levelised cost of electricity (LCOE) for solar PV generation decreased by 12% compared to 2022.
Energy island efficiency	Power restoration time in Tuscany (Italy) during the 2022 energy crisis.	In the city of Pontedera (Tuscany), power supply was restored within 12 hours using an energy island, whereas centralised restoration took 72 hours.
Regional trends in the EU	Share of solar PV and wind power in EU electricity generation in 2024.	In 2024, the combined share of solar and wind generation in the EU's electricity mix exceeded the total share of coal and natural gas-based generation.
DG deployment example	Increased adoption of small hydropower plants in Poland and the Baltic states.	The estimated installed capacity potential of small hydropower plants is 1.2 GW in Poland and 0.8 GW in the Baltic states.
Economic viability	Reduction in onshore wind costs in 2023, enhancing DG economic benefits.	In 2023, the average LCOE for onshore wind generation decreased by 3% compared to 2022.
System resilience impact	Doubling of renewable capacity growth rates in 2024-2030, necessitating grid upgrades.	According to projections, the growth rate of new renewable capacity in the EU will double between 2024 and 2030 compared to the previous decade, creating a need for enhanced distributed grid development.

Source: compiled by the authors based on E. Chomać-Pierzecka et al. (2022), B. Zakeri et al. (2022), International Energy Agency (2024), International Renewable Energy Agency (2024)

The United States is actively developing decentralised energy systems, particularly through support for solar and wind generation via federal subsidies and tax incentives. This contributes to reduced energy transmission costs, enhances energy security, and lowers emissions, notably through the implementation of microgrids and energy islands. A key regulatory framework is Public Law No. 117-169 "Inflation Reduction Act of 2022" (2022), which provides substantial tax incentives for the development of renewable energy sources (RES), including distributed generation

systems and microgrids. Additionally, according to data from the US Environmental Protection Agency, the document Environmental Protection Agency (n.d.) highlights the advantages of decentralised solutions in reducing emissions and minimising energy losses during transmission. This approach helps decrease energy transmission costs, improves energy security, and reduces environmental impact, particularly through the deployment of microgrids and energy islands. Table 7 presented key statistical indicators reflecting the scale of DG deployment in the United States.

Table 7. Key indicators of DG implementation in the United States

Category	Description	Data
DG implementation in the US	Advantages of DG in reducing electricity transmission losses.	DG (e.g., rooftop solar panels) reduces losses, which are significant in centralised systems.
Energy island efficiency	Potential of DG to support the delivery of clean and reliable energy.	DG can facilitate clean energy delivery and reduce transmission losses.
Regional trends in the US	Growth in solar generation in 2023-2024.	In 2023, solar generation in the US increased by 16%, with a further 40% rise projected in 2024 compared to the previous year.
DG implementation example	Addition of new solar capacity in 2024, with a significant share from small-scale installations.	In 2024, the US added 36 GW of new solar capacity, a substantial portion of which came from distributed installations.
Economic efficiency	Decline in the Levelised Cost of Electricity (LCOE) for solar PV in 2023, enhancing the economic benefits of DG.	In 2023, the average cost of solar energy decreased by 12% compared to 2022.
Impact on system resilience	Share of wind and solar generation in total electricity production.	In 2023, solar and wind generation accounted for 16% of total electricity in the US, with a projected increase to 18% in 2024.

Source: compiled by the authors based on International U.S. Energy Information Administration (2023), Renewable Energy Agency (2024), International Renewable Energy Agency (2024), U.S. Energy Information Administration (2024), Environmental Protection Agency (n.d.)

Global trends in the deployment of sustainable energy systems indicate a rapid increase in the share of RES. According to the International Energy Agency (2023; 2024), the share of RES in global electricity generation reached 30% in 2023, with projections suggesting an increase to 50% by 2030, driven by the development of DG and smart grids. In Ukraine, these trends are reflected in the growth of RES to 10% in 2023, aligning with the global shift towards

energy system decarbonisation. The current scale of DG utilisation demonstrates significant growth – in 2023, DG accounted for 12% of global electricity generation, 60% of which was derived from solar and wind energy. This expansion is supported by regional initiatives, particularly in the EU, the US, and Ukraine, where DG is being actively integrated into local energy systems. Table 8 presents consolidated data on the effectiveness of DG implementation in Ukraine.

Table 8. Implementation and efficacy of DG in Ukraine

Category	Description	Data
DG implementation in Ukraine	Number of installed residential solar systems for autonomous energy supply in 2023.	In 2023, 5,000 residential solar systems for autonomous energy supply were installed in Ukraine.
Energy island efficacy	Time taken to restore energy supply to critical infrastructure in Volyn Oblast during the 2023 blackout.	During the 2023 blackout in Volyn Oblast, DG restored power to critical facilities within 4-6 hours, whereas centralised recovery took 24 hours.
Regional trends in Ukraine	Share of RES in electricity generation in 2023 and number of biogas plants in rural communities.	In 2023, RES accounted for 10% of electricity generation, while 2,000 biogas plants in rural communities met up to 15% of their energy demand.
Wartime application	Energy island in Irpin (2022) for a field hospital.	In 2022, an energy island in Irpin provided uninterrupted power to a field hospital for 72 hours, saving 300 lives.
DG deployment case study	Energy island in Nyzhniy Bystryi village (Zakarpattia) in 2023.	In 2023, Nyzhniy Bystryi village (Zakarpattia) implemented an energy island comprising solar panels (50 kW) and a biogas plant (30 kW), supplying 120 households.
Economic efficiency	Comparison of LCOE for DG and centralised supply in 2023.	In 2023, Ukraine's DG LCOE stood at USD 0.08/kWh, below centralised supply costs (USD 0.12/kWh).
System resilience impact	Reduction in energy supply restoration time in frontline regions (2022).	In 2022, frontline regions saw a 60% reduction in power restoration time (from 12-18 to 3-5 hours) due to DG.

Source: compiled by the authors based on H. Pivniak et al. (2022), E. Shahini et al. (2024), I. Hrytsiuk et al. (2024)

The presented data on DG deployment in the EU, US, and Ukraine enable a comparative analysis, revealing shared trends and regional specificities within energy security, sustainable development, and technological integration contexts. In the EU, DG is integrated through advanced regulatory frameworks that enhance economic viability, system resilience, and faster power restoration, though grid modernisation is required to accommodate growing RES contributions. The US prioritises distributed systems supported by state incentives, reducing transmission losses and enabling rapid capacity scaling (particularly

via small-scale installations), albeit technological complexity hinders integration in less developed regions.

In Ukraine, DG plays a strategic crisis-response role, ensuring critical infrastructure and rural community autonomy through energy islands and biogas plants, yet scaling remains constrained by infrastructural and funding limitations. Comparative analysis demonstrates that DG success hinges on balancing technological integration, economic feasibility, and local context adaptation, underscoring the need for regionally tailored strategies to maximise its environmental and social benefits (Table 9).

Table 9. Comparative analysis of DG implementation in the EU, the US, and Ukraine

Criterion	EU	US	Ukraine
Level of DG support	High, due to regulatory frameworks and the objectives of the European Green Deal	High, through tax incentives and stimulus programmes	Moderate, constrained by resources but increasing due to wartime demands
Development focus	Integration of RES, grid modernisation	Small-scale distributed installations, microgrids	Community self-sufficiency, critical infrastructure
Grid resilience	Increasing, though dependent on grid upgrades	High, owing to the autonomy of distributed systems	Key objective: ensuring supply in crisis-affected regions
Primary DG sources	Solar, wind, small hydropower	Solar, wind	Solar, biogas, small hydropower

Table 9. Continued

Criterion	EU	US	Ukraine
Economic efficiency	High: reduced operational costs, shorter recovery times	High: reduced losses, increased DG share	Notable in regions where the centralised system is inefficient
Implementation challenges	Need for grid modernisation	Integration in less developed regions, cybersecurity	Limited funding, infrastructure destruction

Source: compiled by the authors based on M. Farghali et al. (2023), International Energy Agency (2023; 2024), U.S. Energy Information Administration (2023; 2024), C. Rehtanz et al. (2024), International Renewable Energy Agency (2024), Environmental Protection Agency (n.d.)

The implementation of DG in each region reflects distinct national priorities and challenges. In the EU, it is driven by technological modernisation under the green transition; in the US, by the pursuit of flexibility through distributed systems; and in Ukraine, by the strategic necessity of ensuring autonomy amid wartime instability. A common trend across all three contexts is the growing role of RES and the need for adaptive development models. This underscores the importance of a holistic approach to DG,

combining economic efficiency with social resilience and technological integration.

The power supply recovery time is a key indicator of energy system viability during crises. A comparative analysis of case studies from Italy, the US, and Ukraine demonstrates that decentralised systems enable significantly faster restoration of critical infrastructure. This confirms the feasibility of deploying microgrids and energy islands as components of resilient energy infrastructure (Fig. 1).

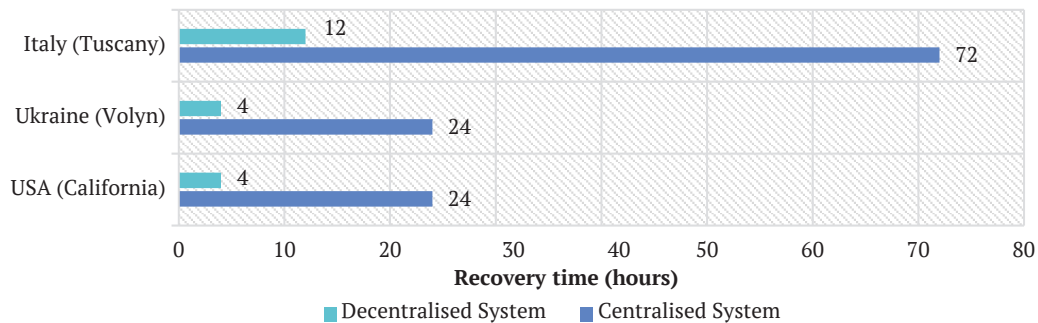


Figure 1. Power supply restoration time in centralised and decentralised systems (case studies of Tuscany, California, and Volyn)

Source: compiled by the authors based on Environmental Protection Agency (n.d.), U.S. Energy Information Administration (2023), International Renewable Energy Agency (2024)

As illustrated in Figure 1, the power supply restoration time when employing decentralised systems was 4–6 hours (USA, Ukraine) and 12 hours (Italy), which is significantly shorter compared to centralised models, where this indicator reached 24–72 hours. From an economic perspective, this reduces losses associated with disruptions to critical infrastructure operations, limits emergency response costs, and avoids the deployment of large-scale reserve capacities. Decentralised solutions also mitigate the risks of cascading grid failures in the event of damage to centralised nodes. However, the implementation of such systems entails higher initial costs, the necessity of integration with energy storage systems, and increased demands on load management. Despite this, under conditions of heightened infrastructure vulnerability, the advantages of decentralisation may offset the associated costs by enhancing energy resilience.

In the context of the war in Ukraine, the survivability of energy systems relies on mobile generators and decentralised energy sources, which ensure the autonomy of critical infrastructure (hospitals, water supply, communications) and support the population in the event of the destruction

of main grids due to military actions or man-made disasters. Mobile generators enable rapid power restoration, but their efficiency is constrained by fuel costs and logistical challenges. DG systems (solar, wind, bioenergy installations) enhance energy security by forming autonomous energy islands. Bioenergy contributes to stability in rural regions; however, generation intermittency and infrastructure requirements complicate their deployment. In wartime conditions, it is advisable to combine mobile generators for rapid response with DG systems for long-term autonomy, ensuring both prompt recovery and system resilience.

Damage to central energy infrastructure, particularly during armed conflicts or natural disasters, poses significant challenges to ensuring uninterrupted energy supply. However, the development of DG represents a crucial step towards achieving energy independence and resilience. In the context of centralised infrastructure destruction, especially in Ukraine during wartime, DG becomes critically important. Local solar, wind, and bioenergy installations provide power to vital facilities (hospitals, water supply systems, communication networks) even in the event of a complete grid

blackout. The development of DG at the community level enables the creation of energy-independent hubs, enhancing the resilience of the energy system – that is, its ability to maintain or rapidly restore critical energy supply following destruction or attacks. The integration of microgrids with energy islands minimises dependence on the centralised grid and ensures energy supply during emergencies.

In Ukraine, DG development is supported through legislative initiatives aimed at integrating RES into the national energy grid. The Law of Ukraine No. 555-IV “On Alternative Energy Sources” (2003) establishes key legal frameworks for the development of energy systems incorporating solar and wind installations. A significant milestone was the introduction of the “Green Tariff” for investors, ensuring financial support and stability for RES-related projects. As of 2025, Ukraine is in the process of drafting regulatory acts governing the integration of mini- and microgrids, particularly to support energy islands operating in autonomous mode, especially in crisis situations. However, the primary challenges for DG development in Ukraine include complex licensing and permitting procedures, which are further complicated under martial law and crises. Frequent legislative changes create legal uncertainty, reducing the investment appeal of such projects. Additionally, there is no unified approach to regulating local energy systems, particularly in emergency operation modes, which requires clear standards and regulatory adjustments for the effective implementation of RES.

The integration of modern digital technologies, artificial intelligence (AI), energy storage systems (ESS), and blockchain solutions plays a pivotal role in enhancing the efficiency, flexibility, and reliability of DG and energy islands. Energy storage technologies are a critical component for the stable operation of decentralised systems and energy islands, particularly in autonomous conditions. The most widely used are lithium-ion batteries, which offer fast charging, durability, and high efficiency. For long-term storage, hydrogen batteries are a promising solution, while pumped hydro storage (PHS) and compressed air energy storage (CAES) are suitable for large-scale systems. These technologies enable the accumulation of surplus energy from RES and its utilisation during peak hours or periods of unstable generation. In Ukraine, projects integrating battery storage with solar and wind installations are already being implemented, increasing the autonomy and resilience of local energy systems.

AI is actively employed in microgrid management due to its ability to process vast amounts of real-time data, forecast demand, and optimise energy flow distribution. Intelligent systems enable automatic load balancing between different energy sources (RES, batteries, backup generators), taking into account weather forecasts, consumption dynamics, and grid operating conditions. The implementation of AI in microgrid management ensures automation with minimal manual intervention, reduces energy losses, improves demand and generation forecasting, and enhances system stability, decreasing reliance on backup sources.

AI gains particular relevance in unpredictable crisis situations where traditional dispatching approaches become ineffective. However, AI has limitations – high implementation costs and cybersecurity risks may hinder its application in unstable regions. AI should be integrated with other control technologies to ensure a balance between automation and security in crisis conditions.

Blockchain technologies provide a decentralised infrastructure for managing energy exchanges between consumers, producers, and energy suppliers. Such a system eliminates the need for a centralised intermediary, reducing costs while ensuring transparency and security in energy transactions. Blockchain solutions in the energy sector facilitate the creation of “smart contracts” for the automated execution of operations under predefined conditions, transparent accounting of energy transactions, and the minimisation of fraud risks and centralised errors. This underscores their potential to enhance the efficiency of decentralised systems; however, high implementation costs and the need for technical infrastructure limit their widespread adoption. Blockchain is best utilised in combination with other technologies to ensure transparency and security in energy operations.

In the context of distributed generation (DG) development in Ukraine, blockchain technologies are regarded as a promising tool for establishing transparent and autonomous local energy markets. Although Ukraine has not yet implemented large-scale pilot projects integrating blockchain into local energy markets, the existing regulatory framework for RES and decentralised generation creates favourable conditions for the future deployment of such solutions. Blockchain could become a key element in ensuring financial transparency between prosumers and suppliers, as well as in enhancing trust in energy operations. Given Ukraine’s commitment to energy decentralisation, blockchain technologies are viewed as a potentially effective instrument for improving the transparency and autonomy of micro-energy systems (Oum, 2024).

The scaling of DG is becoming increasingly relevant in developing countries due to growing energy demands and climate challenges. DG serves as a key instrument for ensuring energy security, reducing dependence on fossil fuels, and enhancing the resilience of energy systems in crisis situations. The primary driver of this process is the declining cost of technologies, particularly photovoltaic panels and wind turbines. The deployment of RES is outpacing projections, and there is growing interest in local energy systems that enable independence from centralised grids. Microgrids, in particular, have proven effective in supplying energy to hospitals, schools, and agricultural cooperatives. However, the scaling of DG faces several systemic barriers. A key challenge in Ukraine is the lack of flexible dispatchability of decentralised energy sources, which is virtually non-existent: under current conditions, DG is not integrated into grid operational management, preventing its full utilisation for balancing generation and consumption. Thus, the absence of flexible DG dispatchability in

Ukraine constitutes one of the critical systemic barriers to its large-scale integration into the energy system. This issue impedes effective generation-consumption balancing, reduces the adaptability of energy infrastructure to variable conditions, and hinders the formation of fully functional energy islands. Given the growing role of DG in decarbonised and decentralised energy systems, the development of mechanisms for flexible dispatchability and automation of DG management processes should become a key focus of future research. Establishing the necessary digital, regulatory, and technological frameworks is essential for improving the resilience and operational efficiency of future energy systems. In most developing countries, outdated grid infrastructure remains incapable of effectively responding to dynamic changes in demand and generation.

Integration of DG into the overall energy infrastructure, particularly through the use of smart grids and microgrids, is accompanied by a significant increase in cyber risks. In the context of the growing digitalisation of the energy sector, ensuring cybersecurity has become one of the key challenges for the stable operation of power systems, especially in crisis or wartime conditions. Modern energy systems are actively implementing digital tools for monitoring, managing, and forecasting loads. Technologies such as smart grids, blockchain solutions, the Internet of Things (IoT), and AI enhance efficiency but simultaneously open new vectors for potential attacks. Cyber threats can not only disrupt the operation of individual microgrids but also lead to large-scale systemic failures, particularly in cases where critical facilities – hospitals, communication hubs, or water supply systems – lose control.

Therefore, the implementation of DG is accompanied by substantial investment barriers. Deploying infrastructure for generation, storage, and energy management requires significant capital expenditures, which may pose a challenge for countries with limited access to financial resources. One of the key conditions for sustainable development in the field of DG is the establishment of “green” financing instruments, such as energy bonds, international climate funds, and public-private partnerships. The deployment of DG serves as a multidisciplinary solution, simultaneously encompassing environmental, social, and economic aspects. Emissions reduction, increased energy autonomy, the emergence of new labour markets, and financial challenges – all these require an integrated approach to policy planning in decentralised energy. This study has systematised existing approaches to assessing the environmental efficiency of DG, particularly in crisis situations. A comparative analysis of different countries demonstrated how national characteristics, economic conditions, and technological solutions influence the adoption of DG, energy islands, and the integration of AI and blockchain technologies. An important objective of the research was to determine how these solutions can enhance the resilience of energy systems, reduce costs and CO₂ emissions, and improve energy security in wartime conditions.

The study revealed that DG significantly enhances the resilience of energy systems in crisis conditions, reducing power restoration time by 60% in frontline regions of Ukraine (3–5 hours compared to 12–18 hours for centralised systems). This finding aligns with the work of V. Kaplun *et al.* (2022), who emphasised that renewable energy islands enable rapid power restoration through local generation. However, unlike the authors’ study, which focused on price forecasting, the present research highlights the practical impact of DG in real wartime conditions, making these conclusions more contextually relevant.

The obtained results partially coincide with the study by R. Khalid *et al.* (2024), who analysed mini- and micro-hydropower plants in Pakistan. They noted that local systems reduce outage response times, which supports the present findings. However, their study did not account for the specifics of wartime conditions, where infrastructure destruction is systematic, making these conclusions more relevant for similar contexts. The research established that DG is more economically advantageous compared to centralised supply. This result corresponds with the conclusions of M. Li *et al.* (2024), who examined prosumer households in Ukraine and noted that solar and biogas installations reduce costs through local production. However, unlike their study, which focused on economic benefits for households, the present research underscores the broader impact of DG on energy security in crisis conditions, making the conclusions more universal.

The current study also confirmed the economic potential of DG. O. Borodina *et al.* (2022) proposed a conceptual model of decentralised energy efficiency, integrating economic and technological approaches. The present results complement their model, demonstrating that the economic efficiency of DG can be achieved even in crisis conditions if focused on local solutions such as energy islands. The results highlighted the environmental and social benefits of DG, including CO₂ emissions reduction and improved energy access in rural regions. These conclusions correlate with the work of A. Zaporozhets *et al.* (2025), who investigated the optimisation of renewable energy systems. They described how DG creates new jobs, supporting the observed data on social benefits. However, the present study additionally notes that in crisis conditions, these benefits may be limited due to a lack of investment.

Current findings partially aligned with the conclusions of M. Iurchenko *et al.* (2024), who investigated the use of renewable energy sources (RES) for sustainable economic growth. The authors emphasised that decentralised systems contribute to environmental sustainability; however, this study focused on their role in ensuring energy security in crisis situations, which is particularly relevant for Ukraine. The work also highlighted the importance of energy islands in crisis conditions for ensuring energy supply autonomy. At the same time, these results partially contradict the work of Z.V. Derii *et al.* (2024), who modelled the potential of RES in Ukraine. They argued that solar and wind generation could fully replace centralised systems in

the long term; however, the current study demonstrated that in the short term (under wartime conditions), a complete replacement is unfeasible due to generation instability and the need for backup capacity.

AI and blockchain technologies enhance the efficiency of microgrid management but have limitations. AI enables automation, reduces losses, and adapts to demand fluctuations, yet its implementation is complicated by cyber risks and costs. These conclusions correlated with the study by M. Massaoudi *et al.* (2025), who analysed the use of machine learning in energy grid management. They noted that AI can optimise energy distribution but requires protection against cyberattacks, which confirmed the observational data. The current results further underscored the relevance of AI in crisis conditions, where traditional dispatching methods become inefficient, making the present findings more contextually oriented. Blockchain technologies ensure transparent accounting and reduce transaction costs, but their implementation is constrained by technical complexity. This partially aligned with the work of S.D. Rodrigues & V.J. Garcia (2023), who investigated transactional energy systems in microgrids. The authors indicated that blockchain could optimise energy flows; however, this study additionally noted that under wartime conditions in Ukraine, priority should be given to simpler solutions, such as local control systems, due to resource constraints.

The scaling of DG in developing countries opens significant prospects for ensuring energy security and sustainable development. Increased investment in RES (solar and wind technologies) will help reduce dependence on fossil fuels and improve energy access in rural and remote regions. The development of microgrids and energy islands will enable the creation of autonomous systems that ensure stability in crisis conditions while simultaneously contributing to CO₂ emission reductions and job creation. This also opens opportunities for the formation of new market models. Local energy markets based on blockchain technologies have the potential to ensure transparent accounting, optimise energy flows, and reduce transaction costs. Such technologies can be highly effective in contexts where traditional solutions are financially inaccessible.

To counter cyber threats, the implementation of a multi-layered cybersecurity approach is necessary, including cryptographic protection of data exchange, user and device authentication, AI and machine learning-based behavioural analytics systems, and continuous infrastructure monitoring to detect anomalies. In the context of DG, it is also essential to develop “edge-of-network” protection systems that enable local data processing and rapid response to breaches. Countries with limited resources face challenges in establishing effective cybersecurity. In this regard, strengthening international technical and analytical cooperation, as well as participation in cyber resilience programmes at the level of the EU, the International Energy Agency (IEA), the IAEA, and other institutions, is crucial. The joint development of standards, security protocols,

and risk assessment methodologies will enhance the resilience of decentralised energy systems on a global scale.

The expansion of DG is not only a technical and energy challenge but also a key element of environmental and socio-economic development strategies. The integration of RES in the form of microgrids and energy islands has the potential to significantly alter both the structure of the energy balance and the social dynamics of individual regions, particularly in developing countries and post-crisis recovery contexts. One of the primary environmental benefits of decentralised systems is the reduction of greenhouse gas emissions, particularly CO₂, through the replacement of fossil fuels with RES. The deployment of solar, wind, and bioenergy generation not only reduces environmental impact but also decreases the energy sector's dependence on hydrocarbon imports, which is critical during periods of geopolitical instability.

At the same time, alongside the advantages, it is crucial to consider the potential limitations of DG implementation. Power generation based on RES, such as solar and wind energy, is weather-dependent and subject to natural variability, necessitating the use of energy storage systems or backup capacity. The high initial capital costs of deploying micro- and mini-grids may pose a barrier to their rollout, particularly in post-crisis economic conditions. Furthermore, the regulatory and technical integration of such systems into existing power grids requires comprehensive solutions and legislative amendments. In regions affected by conflicts or with limited infrastructure, decentralised facilities may be vulnerable to damage or require additional protective measures. Thus, the effective utilisation of DG potential demands not only technical implementation but also appropriate institutional, financial, and security support.

The development of microgrids incorporating RES, combined with local energy storage, enhances energy access in rural areas where centralised grids are technically or economically unviable. Moreover, the implementation of such projects stimulates job creation in the fields of design, installation, and operation of local energy systems, which is vital for regional development. Another significant social benefit is the reduction in energy costs for consumers in the medium to long term. Although the upfront costs of DG deployment are high, the overall lifecycle of the system demonstrates competitiveness compared to traditional energy models.

The integration of digital technologies, such as AI and blockchain, optimises energy flow management and enhances transparency in local energy markets, though this necessitates strengthened cybersecurity due to increased cyber risks. To overcome infrastructure barriers, it is advisable to engage international cooperation and “green” financing, which would facilitate the modernisation of outdated grids. Priority should be given to adaptive strategies that account for local conditions, ensuring a balance between technological innovation and economic feasibility for sustainable energy system development. This study holds significant importance for the advancement

of energy systems in crisis conditions, as it demonstrates that DG and energy islands can provide not only economic efficiency but also system resilience in real wartime scenarios. Unlike many previous studies focused on theoretical modelling, this research provides practical examples confirming the effectiveness of DG in real-world applications. This makes the current findings valuable for countries facing similar challenges, such as infrastructure instability or armed conflicts.

Conclusions

The conducted study confirmed the practical effectiveness of solar and hybrid systems in the context of decentralised energy supply. Specifically, in 2023, over 5,000 residential solar installations were deployed in Ukraine, demonstrating the widespread adoption of this generation type at the household level. Hybrid solutions, such as the energy island in the village of Nyzhniy Bystryi, ensured full autonomy for 120 households by combining 50 kW of solar and 30 kW of biogas generation. Similar trends were observed in the US, where solar generation increased by 16% in 2023 and reached 40% by the end of 2024. In the EU, a 12% reduction in the cost of solar energy generation during 2023 created additional economic incentives for transitioning to decentralised generation (DG). Furthermore, a case study from Tuscany demonstrated that energy supply restoration time based on an energy island was reduced to 12 hours compared to 72 hours for centralised systems. In frontline regions of Ukraine, the implementation of DG reduced restoration time from 12-18 to 3-6 hours, which is critically important for ensuring the operation of critical infrastructure.

The study confirmed the significant role of DG in enhancing the reliability of energy systems, particularly in crisis situations such as war or natural disasters. DG substantially reduced energy supply restoration time, which is crucial for the uninterrupted operation of critical facilities. This is especially vital for frontline zones, where centralised power grids often cannot function due to damage. The application of DG in such regions facilitated energy supply restoration within a few hours, significantly faster than traditional methods, where recovery could take days. The economic benefits of DG were also confirmed, as the LCOE generated by renewable energy sources (RES) was lower compared to centralised systems (0.08 USD/kWh for DG versus 0.12 USD/kWh for centralised systems). This

improved cost efficiency in energy supply, a key consideration for rural and remote regions where centralised grids are either absent or economically unviable. The study also revealed that energy islands, which provided autonomous energy supply in the absence of or damage to centralised grids, significantly enhanced the resilience of energy systems. They ensured energy supply even in crisis situations where traditional energy systems could not operate. However, large-scale deployment of such systems required addressing challenges related to infrastructure investments and advancements in energy storage technologies.

Despite positive results, the study identified a number of limitations, such as insufficient infrastructure development for the integration of DG into national power systems. In particular, in Ukraine, the complete absence of flexible DG dispatchability makes their effective integration into power system operational management impossible, reducing the ability of these sources to stabilise energy supply under variable demand or crisis conditions. The high costs of implementing cutting-edge technologies (batteries and energy flow management systems) remain a significant barrier to their large-scale deployment. Further research should focus on improving the mechanisms for DG integration into national power systems, advancing energy storage technologies, and developing flexible solutions for demand-side management. This will ensure the stability and efficiency of future power systems while enhancing resilience to extreme situations, particularly in wartime or natural disasters. The adoption of DG in Ukraine presents opportunities for energy independence; however, its potential is constrained by the lack of flexible dispatchability, which limits its full-fledged role in the national energy sector. This complicates stable energy supply during wartime and fluctuating demand. The development of automated control mechanisms for these sources is a critically important task for the future of the country's energy sector.

Acknowledgements

None.

Funding

None.

Conflict of Interest

None.

References

- [1] Adnan, M., Ghadi, Y., Ahmed, I., & Ali, M. (2023). Transmission network planning in super smart grids: A survey. *IEEE Access*, 11, 77163-77227, [doi: 10.1109/ACCESS.2023.3296152](https://doi.org/10.1109/ACCESS.2023.3296152).
- [2] Bardyk, Ye., Bondarenko, O., & Bolotnyi, M. (2024). Operational reliability analysis for sustainable energy system planning development. *Renewable Energy*, 4(79), 46-58. [doi: 10.36296/1819-8058.2024.4\(79\)46-58](https://doi.org/10.36296/1819-8058.2024.4(79)46-58).
- [3] Blinov, I., Olefir, D., Parus, E., & Kyrylenko, O. (2023). Improving the efficiency of HPP and PSHPP participation in the electricity market of Ukraine. In O. Kyrylenko, S. Denysiuk, D. Derevianko, I. Blinov, Ie. Zaitsev & A. Zaporozhets (Eds.), *Power systems research and operation* (pp. 51-74). Cham: Springer. [doi: 10.1007/978-3-031-17554-1_3](https://doi.org/10.1007/978-3-031-17554-1_3).
- [4] Borodina, O., Kryshchal, H., Hakova, M., Neboha, T., Olczak, P., & Koval, V. (2022). A conceptual analytical model for the decentralized energy-efficiency management of the national economy. *Energy Policy Journal*, 25(1), 5-22. [doi: 10.33223/epj/147017](https://doi.org/10.33223/epj/147017).

- [5] Chomać-Pierzecka, E., Kokieli, A., Rogozińska-Mitruć, J., Sobczak, A., Soboń, D., & Stasiak, J. (2022). Hydropower in the energy market in Poland and the Baltic States in the light of the challenges of sustainable development – an overview of the current state and development potential. *Energies*, 15(19), article number 7427. doi: [10.3390/en15197427](https://doi.org/10.3390/en15197427).
- [6] Derii, Z.V., Tkalenko, S.I., Liubachivska, R.Z., Hrytsku-Andriesh, Y.P., & Timish, R.Y. (2024). Modelling and forecasting the production potential of renewable energy sources in the context of sustainable development. *IOP Conference Series: Earth and Environmental Science*, 1415, article number 012116. doi: [10.1088/1755-1315/1415/1/012116](https://doi.org/10.1088/1755-1315/1415/1/012116).
- [7] Diachenko, O. (2024). Investments in solar energy in a decentralized energy system as a factor in strengthening the energy security of Ukraine during wartime. *Renewable Energy*, 2(77), 73-78. doi: [10.36296/1819-8058.2024.2\(77\)73-78](https://doi.org/10.36296/1819-8058.2024.2(77)73-78).
- [8] Directive of the European Parliament and of the Council No. 2018/2001 “On the Promotion of the Use of Energy from Renewable Sources”. (2018, December). Retrieved from <https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A32018L2001>.
- [9] Directive of the European Parliament and of the Council No. 2019/944 “On Common Rules for the Internal Market for Electricity and Amending Directive 2012/27/EU”. (2019, June). Retrieved from <https://eur-lex.europa.eu/eli/dir/2019/944/oj/eng>.
- [10] Doronina, I., Nehrey, M., & Putrenko, V. (2023). Reengineering of the Ukrainian energy system: Geospatial analysis of solar and wind potential. In Z. Hu, Q. Zhang & M. He (Eds.), *Advances in artificial systems for logistics engineering III* (pp. 404-415). Cham: Springer. doi: [10.1007/978-3-031-36115-9_37](https://doi.org/10.1007/978-3-031-36115-9_37).
- [11] Environmental Protection Agency. (n.d.). *Distributed generation of electricity and its environmental impacts*. Retrieved from <https://www.epa.gov/energy/distributed-generation-electricity-and-its-environmental-impacts>.
- [12] Environmental Protection Agency. (n.d.). *Interconnection and net metering standards*. Retrieved from https://www.epa.gov/sites/default/files/2015-08/documents/gta_chapter_7.3_508.pdf?utm_source=.
- [13] Farghali, M., Osman, A.I., Mohamed, I.M., Chen, Z., Chen, L., Ihara, I., Yap, P.-S., & Rooney, D.W. (2023). Strategies to save energy in the context of the energy crisis: A review. *Environmental Chemistry Letters*, 21, 2003-2039. doi: [10.1007/s10311-023-01591-5](https://doi.org/10.1007/s10311-023-01591-5).
- [14] Frilingou, N., et al. (2023). Navigating through an energy crisis: Challenges and progress towards electricity decarbonisation, reliability, and affordability in Italy. *Energy Research & Social Science*, 96, article number 102934. doi: [10.1016/j.erss.2022.102934](https://doi.org/10.1016/j.erss.2022.102934).
- [15] Gonzalez-Reina, A.E., Garcia-Torres, F., Girona-Garcia, V., Sanchez-Sanchez-de-Puerta, A., Jimenez-Romero, F.J., & Jimenez-Hornero, J.E. (2024). Cooperative model predictive control for avoiding critical instants of energy resilience in networked microgrids. *Applied Energy*, 369, article number 123564. doi: [10.1016/j.apenergy.2024.123564](https://doi.org/10.1016/j.apenergy.2024.123564).
- [16] Hrytsiuk, I., Volynets, V., Komenda, N., Hrytsiuk, Y., & Hadai, A. (2024). Modelling the optimal switching scheme of the Ukrainian power grid during blackout (Volyn region). *Machinery & Energetics*, 15(2), 95-105. doi: [10.31548/machinery/2.2024.95](https://doi.org/10.31548/machinery/2.2024.95).
- [17] Public Law No. 117-169 “Inflation Reduction Act of 2022”. (2022). Retrieved from https://www.congress.gov/bill/117th-congress/house-bill/5376/text?utm_source=.
- [18] International Energy Agency. (2023). *Net zero roadmap: A global pathway to keep the 1.5°C goal in reach*. Paris: IEA publications.
- [19] International Energy Agency. (2024). *Renewables 2024*. Retrieved from <https://www.iea.org/reports/renewables-2024>.
- [20] International Renewable Energy Agency. (2024). *Renewable power generation costs in 2023*. Abu Dhabi: IRENA.
- [21] Iurchenko, M., Nyzhnychenko, Y., Rudyk, N., Zolotarova, O., & Stakhurska, S. (2024). Harnessing renewable energy for sustainable economic growth and environmental resilience. *Grassroots Journal of Natural Resources*, 7(3), s52-s69. doi: [10.33002/nr2581.6853.0703ukr03](https://doi.org/10.33002/nr2581.6853.0703ukr03).
- [22] Kadar, P., & Vass, A. (2023). An isolated microgrid supply block in an urban network for emergency situations. In *Proceedings of the 21st world symposium on applied machine intelligence and informatics* (pp. 245-250). Herlany: IEEE. doi: [10.1109/SAMI58000.2023.10044538](https://doi.org/10.1109/SAMI58000.2023.10044538).
- [23] Kaplun, V., Osypenko, V., & Makarevych, S. (2022). Forecasting the electricity pricing of energy islands with renewable sources. *Machinery & Energetics*, 13(4), 38-47. doi: [10.31548/machenergy.13\(4\).2022.38-47](https://doi.org/10.31548/machenergy.13(4).2022.38-47).
- [24] Khalid, R., Basit, A., Sohail, M., Ahmad, T., & Muhammad, N. (2024). Community energy and socio-technical infrastructure resilience: Analysis of mini/micro hydro power projects in Khyber Pakhtunkhwa, Pakistan. *Environmental Research: Infrastructure and Sustainability*, 4(3), article number 035015. doi: [10.1088/2634-4505/ad7886](https://doi.org/10.1088/2634-4505/ad7886).
- [25] Law of Ukraine No. 555-IV “On Alternative Energy Sources”. (2003, February). Retrieved from <https://zakon.rada.gov.ua/laws/show/555-15#Text>.
- [26] Li, M., Pysmenna, U., Petrovets, S., Sotnyk, I., & Kurbatova, T. (2024). Managing the development of decentralized energy systems with photovoltaic and biogas household prosumers. *Energy Reports*, 12, 4466-4474. doi: [10.1016/j.egyr.2024.10.011](https://doi.org/10.1016/j.egyr.2024.10.011).

- [27] Massaoudi, M., Ez Eddin, M., Ghayeb, A., Abu-Rub, H., & Refaat, S.S. (2025). Advancing coherent power grid partitioning: A review embracing machine and deep learning. *IEEE Open Access Journal of Power and Energy*, 12, 59-75. doi: [10.1109/OAJPE.2025.3535709](https://doi.org/10.1109/OAJPE.2025.3535709).
- [28] Oum, K.R. (2024). Empowering energy transition through blockchain technology. In A.R. Jamader, M. Selvam & B.R. Acharya (Eds.), *Blockchain and cryptocurrency: Management systems and technology challenges* (pp. 105-116). Boca Raton: CRC Press. doi: [10.1201/9781003453109](https://doi.org/10.1201/9781003453109).
- [29] Pivniak, H., Aziukovskyi, O., Papaika, Y., Lutsenko, I., & Neuberger, N. (2022). Problems of development of innovative power supply systems of Ukraine in the context of European integration. *Scientific Bulletin of the National Mining University*, 5, 89-103. doi: [10.33271/nvngu/2022-5/089](https://doi.org/10.33271/nvngu/2022-5/089).
- [30] Rabocha, T., Maslii, O., Robochyi, V., Frolov, O., & Pizintsali, L. (2023). Ukraine's energy supply in the defense sector: The first lessons of war. *Sustainable Engineering and Innovation*, 5(2), 219-246. doi: [10.37868/sei.vi.id236](https://doi.org/10.37868/sei.vi.id236).
- [31] Rehtanz, C., Ulbig, A., Palaniappan, R., Faulwasser, T., Saidi, S., Schmeink, A., & Wietfeld, C. (2024). Towards holonic power and energy systems – a novel ICT architecture as enabler for resilience. *International Journal of Electrical Power & Energy Systems*, 162, article number 110283. doi: [10.1016/j.ijepes.2024.110283](https://doi.org/10.1016/j.ijepes.2024.110283).
- [32] Rodrigues, S.D., & Garcia, V.J. (2023). Transactive energy in microgrid communities: A systematic review. *Renewable and Sustainable Energy Reviews*, 171, article number 112999. doi: [10.1016/j.rser.2022.112999](https://doi.org/10.1016/j.rser.2022.112999).
- [33] Sadiq, M., Su, C.-L., Ali, Z., Rouhani, S.H., Straka, M., Buzna, L., Micallef, A., & Parise, G. (2025). Decentralized model predictive control for offshore wind-powered seaport DC microgrids with electric vehicle stations. *IEEE Transactions on Industry Applications*, 61(2), 2258-2270. doi: [10.1109/TIA.2025.3532585](https://doi.org/10.1109/TIA.2025.3532585).
- [34] Shahini, E., Fedorchuk, M., Hruban, V., Fedorchuk, V., & Sadovoy, O. (2024). Renewable energy opportunities in Ukraine in the context of blackouts. *International Journal of Environmental Studies*, 81(1), 125-133. doi: [10.1080/00207233.2024.2320021](https://doi.org/10.1080/00207233.2024.2320021).
- [35] Tröndle, T., Melnyk, O., Tutova, O., Porieva, V., Neumann, F., Staffell, I., & Patt, A. (2024). Rebuilding Ukraine's energy supply in a secure, economic, and decarbonised way. *Environmental Research: Infrastructure and Sustainability*, 4(3), article number 031002. doi: [10.1088/2634-4505/ad6738](https://doi.org/10.1088/2634-4505/ad6738).
- [36] U.S. Energy Information Administration. (2023). *International energy outlook 2023*. Washington: CSIS.
- [37] U.S. Energy Information Administration. (2024). *Short-term energy outlook*. Retrieved from <https://www.eia.gov/outlooks/steo/archives/jan24.pdf>.
- [38] Veremiichuk, Y.A., Opryshko, V.P., Prytyskach, I.V., & Yarmoliuk, O.S. (2024). Prospects for autonomous low-power renewable energy communities. *IOP Conference Series: Earth and Environmental Science*, 1415, article number 012120. doi: [10.1088/1755-1315/1415/1/012120](https://doi.org/10.1088/1755-1315/1415/1/012120).
- [39] Yusuf, S.S., & Abubakar, A.S. (2023). A comprehensive review on grid-forming inverter: Potential and future trends. *Majlesi Journal of Electrical Engineering*, 17(1), 1-27. doi: [10.30486/mjee.2023.1970278.0](https://doi.org/10.30486/mjee.2023.1970278.0).
- [40] Zakeri, B., et al. (2022). Pandemic, war, and global energy transitions. *Energies*, 15(17), article number 6114. doi: [10.3390/en15176114](https://doi.org/10.3390/en15176114).
- [41] Zaporozhets, A., Kulyk, M., Babak, V., & Denysov, V. (2025). *Structure optimization of power systems with renewable energy sources*. Cham: Springer. doi: [10.1007/978-3-031-83697-8](https://doi.org/10.1007/978-3-031-83697-8).

Децентралізована генерація та її роль у підвищенні живучості енергетичних островів та критичної інфраструктури: сучасні тенденції та перспективи

Віталій Войтенко

Аспірант

Національний університет біоресурсів і природокористування України

03041, вул. Героїв Оборони, 15, м. Київ, Україна

<https://orcid.org/0009-0005-0107-8357>

Роман Поліщук

Аспірант

Національний університет біоресурсів і природокористування України

03041, вул. Героїв Оборони, 15, м. Київ, Україна

<https://orcid.org/0009-0009-0356-0226>

Анотація. Загрози для критичної інфраструктури посилили актуальність автономного енергозабезпечення та сприяли переходу до децентралізованих рішень. Метою дослідження був аналіз теоретичних основ децентралізованої генерації, її застосування в енергетичних островах для підвищення живучості критичної інфраструктури в умовах надзвичайних ситуацій. Методологічна база дослідження включала теоретичне узагальнення сучасних наукових підходів, реальні кейси розвитку відновлювальних технологій та моделей автономного енергопостачання у США, ЄС та Україні, а також порівняння моделей автономного енергозабезпечення. У результаті встановлено, що найбільш ефективними виявилися сонячні та гібридні (сонячно-біогазові) системи. В Україні у 2023 році було встановлено понад 5 000 домашніх сонячних установок, а в селі Нижній Бистрий (Закарпаття) гібридний енергетичний острів забезпечив енергопостачання 120 домогосподарств. У США сонячна генерація зросла на 16 % у 2023 році, а у 2024 році було зафіксовано зростання на 40 %. У ЄС у 2023 році середня собівартість виробництва електроенергії з сонячної енергії знизилася на 12 %. У Тоскані (Італія) відновлення енергопостачання енергетичним островом у 2022 році тривало 12 годин замість 72 у централізованій системі. В Україні децентралізована генерація дозволила скоротити час відновлення до 3-6 годин у прифронтових регіонах (проти 12-18 годин традиційно), а собівартість електроенергії становила 0,08 доларів США/кВт·год порівняно з 0,12 доларів США/кВт·год у централізованій системі. Дослідження показало, що децентралізована генерація дозволяла підвищити гнучкість і стійкість енергосистем за рахунок скорочення втрат у мережі (до -10 %) і диверсифікації джерел постачання. Було виявлено, що інноваційні технології (цифрове управління, штучний інтелект, підвищували ефективність диспетчеризації, що дозволяло енергетичним островам і мікромережам автономно забезпечувати критичні об'єкти навіть у разі повного руйнування мережі. Практична цінність дослідження полягала у використанні його висновків для стратегій енергетичної безпеки, розвитку мікромереж і планування автономних систем у вразливих регіонах

Ключові слова: відновлювані джерела; автономність систем; сонячна енергія; вітрові турбіни; біоенергетика; цифрові технології



Impact of log file processing on learning speed and defect classification accuracy

Anton Kaiafiuk*

Postgraduate Student

Kyiv National University of Technologies and Design

01011, 2 Mala Shyianovska Str., Kyiv, Ukraine

<https://orcid.org/0009-0003-9917-0834>

Abstract. The purpose of the study was to investigate the effect of automatic testing log file preprocessing on the speed of vectorisation and training of machine learning models. The HDFS_v3_TraceBench set was used, which contains more than 370 thousand traces collected in the Hadoop Distributed File System Environment. Processing included noise removal, lemmatisation, and duplication reduction. The data was vectorised using the Term frequency – inverse document frequency method, and then the RandomForestClassifier model was trained. The experimental results showed that optimising the input data reduced the total processing time by almost five times. The time required for text vectorisation and model training has been reduced, which helped to speed up work with large volumes of logs. However, the classification accuracy was not only preserved, but also showed a slight improvement: the F1-score and Matthews correlation coefficient indicators remained consistently high. There was also a decrease in the Log Loss value, which indicated an increase in the model's confidence in its own forecasts. This is especially important in the context of unbalanced classes that are characteristic of defect classification problems. A detailed analysis showed that a significant part of the service and repetitive information in the logs is not critical for training the model, and its removal, on the contrary, improves the quality of data preparation. In the course of the study, it was also confirmed that the resulting target labels for logs correspond to typical error classes. Implemented log file processing not only reduces computational costs, but also supports or improves the quality of forecasting. These results confirmed the feasibility of including the log cleaning and optimisation step in the overall process of building machine learning models for automated testing. The results obtained can be integrated into automated pipelines for classifying defects and generating bug reports. This will help to reduce the amount of manual labour and increase the efficiency of teams

Keywords: regular expressions; lemmatisation; vectorisation; machine learning; test automation

Introduction

Automated testing of the latest software generates large volumes of unstructured log files, which are complicated by the growing number of tests and the complexity of systems. According to R. Peronto (2024), the annual growth in the volume of log files is 250%. Traditional log analysis methods are expensive and slow, which leads to delays in detecting defects and releasing software. Machine learning allows automating this process, but the effectiveness of models largely depends on the quality of preprocessing: redundant information slows down learning, and excessive filtering reduces accuracy. The balance between these extremes determines the performance of the system, especially in problems of classification of defect types, where it

is important not only to identify the problem, but also to understand its nature.

Recent international studies have mainly focused on detecting anomalies in logs. R.R. Abdalla & A.K. Jumaa (2022) in their study showed the effectiveness of ML (Machine Learning) when working with unstructured journals, but the classification of known defects was not considered. S. Ramachandran *et al.* (2023) proposed a model based on deep neural networks and custom “LogWord-2Vec” vectorisation for classifying errors in large log files, achieving high accuracy, although the impact of preprocessing on resource consumption was ignored. Log vectorisation is a key step in preparing for analysis. TF-IDF (Term

Suggested Citation:

Kaiafiuk, A. (2025). Impact of log file processing on learning speed and defect classification accuracy. *Technologies and Engineering*, 26(2), 27-36. doi: 10.30857/2786-5371.2025.2.2.

*Corresponding author



frequency – inverse document frequency) remains one of the most common approaches for highlighting informative words. A. Sandhu & S. Mohammed (2022) noted that frequently repeated terms reduce the quality of features, so limiting the number of duplicates helps reduce their impact while maintaining the semantic load of messages. In application systems, log analytics is increasingly used to detect states or types of errors. A. Brandão & P. Georgieva (2020) was one of the first to apply ML to network logs to detect attacks. P. Ryciak *et al.* (2022) adapted this approach to search for anomalies in system logs using sequential models. E. Shirzad & H. Saadatfar (2022) showed that analysis of patterns in unstructured logs allows predicting failures in a Hadoop cluster, which increases the reliability of the system. Y. Huangfu (2022) investigated the use of various machine learning approaches to diagnose log-based software failures, emphasising the importance of data preprocessing to improve model efficiency. Q. Qin *et al.* (2024) proposed a two-step approach to log file processing, using semi-supervised learning to speed up log classification in large online systems. Z.A. Khan *et al.* (2024) conducted an empirical study of the effect of log parsing on the accuracy of anomaly detection, emphasising that the distinguishing property in parsing results is key to achieving high accuracy. The problem of preprocessing log files was also actively investigated by Ukrainian researchers. In particular, M. Prodeus *et al.* (2024) proved that standardisation, normalisation, and feature selection significantly improve Random Forest results in network data anomaly detection problems. Despite the differences in domains, this confirmed the importance of filtering unnecessary technical information in logs to improve the efficiency of models. S. Kapitanets & G. Radelchuk (2022) emphasised the importance of orderly data logging for timely detection of failures in software systems, noting that chaotic logging practices significantly complicate error detection. O. Khil & V. Yakovina (2023) emphasised in their study that the success of the model largely depends on the quality of input data. The study by S.A. Hussein & S.R. Répás (2024) provided an overview of modern methods for detecting anomalies in log files using machine learning algorithms. The researchers analysed the effectiveness of statistical, ML, and DL approaches in the context of variable structure and large amounts of log data, highlighting the growing role of AI in strengthening cybersecurity.

The purpose of this study was to investigate the influence of log file preprocessing methods on the efficiency of machine learning in automatic defect classification. In particular, it was considered how noise removal using regular expressions, lemmatisation, and TF-IDF vectorisation affects the speed of converting text data to numeric vectors and model training.

Materials and Methods

To achieve this goal, a set of methods of scientific cognition was applied, including empirical and experimental approaches. In particular, a computational experiment was

conducted: a data processing and analysis pipeline was implemented, on which various scenarios for preprocessing log files were tested. The methods of cluster analysis (for preliminary detection of structures in data) and computer experiment were used to evaluate the performance of text data preprocessing, the learning time of the machine learning model, and the accuracy of model prediction. The results of various data preparation methods were compared with each other (comparative analysis method) to identify their impact on speed and accuracy. This approach allows drawing reasonable conclusions about the causal relationships between the applied data transformations and the obtained ML model metrics.

The open dataset HDFS_v3 (TraceBench), a collection of log files obtained in the Hadoop distributed file system (HDFS) during cluster operation (Zhou *et al.*, 2014). This set contains more than 370,000 trace records (log messages) from the real infrastructure-as-a-service (IaaS) environment. Below is the programme code for calculating the silhouette score based on a data set:

```
import matplotlib.cm as cm
import matplotlib.pyplot as plt
import numpy as np
from sklearn.metrics import silhouette_samples, silhouette_score
from sklearn.mixture import GaussianMixture

X = X.toarray()
range_n_clusters = [3, 4, 5, 6]

fig, axes = plt.subplots(2, 2, figsize=(15, 10))
axes = axes.flatten()

for idx, n_clusters in enumerate(range_n_clusters):
    gmm = GaussianMixture(n_components=n_clusters, random_state=42)
    cluster_labels = gmm.fit_predict(X)

    # Calculate avg silhouette coefficient
    silhouette_avg = silhouette_score(X, cluster_labels)
    print(
        f"n_clusters={n_clusters}, average silhouette coefficient={silhouette_avg:.4f}"
    )

    # Calculate silhouette coefficient for each sample
    sample_silhouette_values = silhouette_samples(X, cluster_labels)

    y_lower = 10
    ax = axes[idx]
    ax.set_title(
        f"Number of clusters: {n_clusters}\nAverage silhouette coefficient: {silhouette_avg:.4f}"
    )
    ax.set_xlabel("Silhouette coefficient")
    ax.set_ylabel("Cluster")

    ax.set_xlim([-0.1, 1])
    ax.set_ylim([0, len(X) + (n_clusters + 1) * 10])
    ax.axvline(x=silhouette_avg, color="red", linestyle="--")

    for i in range(n_clusters):
        ith_cluster_silhouette_values = sample_silhouette_values[cluster_labels==i]
        ith_cluster_silhouette_values.sort()
```

```

size_cluster_i = ith_cluster_silhouette_values.shape[0]
y_upper = y_lower + size_cluster_i

color = cm.nipy_spectral(float(i) / n_clusters)
ax.fill_betweenx(
    np.arange(y_lower, y_upper),
    0,
    ith_cluster_silhouette_values,
    facecolor=color,
    edgecolor=color,
    alpha=0.7,
)

ax.text(-0.05, y_lower + 0.5 * size_cluster_i, str(i))
y_lower = y_upper + 10 # 10 for space between clusters

ax.set_yticks([]) # Hide sticks for Y axis
ax.set_xticks(np.arange(-0.1, 1.1, 0.2))

plt.tight_layout()
plt.show()

```

Since the purpose of the study was to train the ML model to classify defect types, and the HDFS_v3_TraceBench dataset contains only log files without corresponding

labels, the first step was to create target classes (labels) for each log file. The first step was to read the log files and create a pandas.DataFrame structure with error messages; next, the data were vectorised using the TF-IDF method and clustered using Gaussian mixture; the optimal number of clusters was determined using the silhouette score, according to A. Géron (2022), and creating a pandas.DataFrame with target classes for further model training. Ultimately, the obtained clusters were analysed and compared with the types of defects given in the official documentation according to the HDFS architecture guide (n.d.).

The next step was data preprocessing. Problems identified in logs: unstructured, redundant information and noise. To solve these problems, the following pre-treatment steps were implemented (Fig. 1): using regular expressions, timestamps, logging levels (INFO, DEBUG, ERROR, WARN), ID (Identification Number), IP addresses, UUID (Universal Unique Identifier) and extra special characters were removed; words were returned to their original form to improve the quality of vectorisation and data consistency; the number of duplicate messages was reduced to reduce the amount of data.

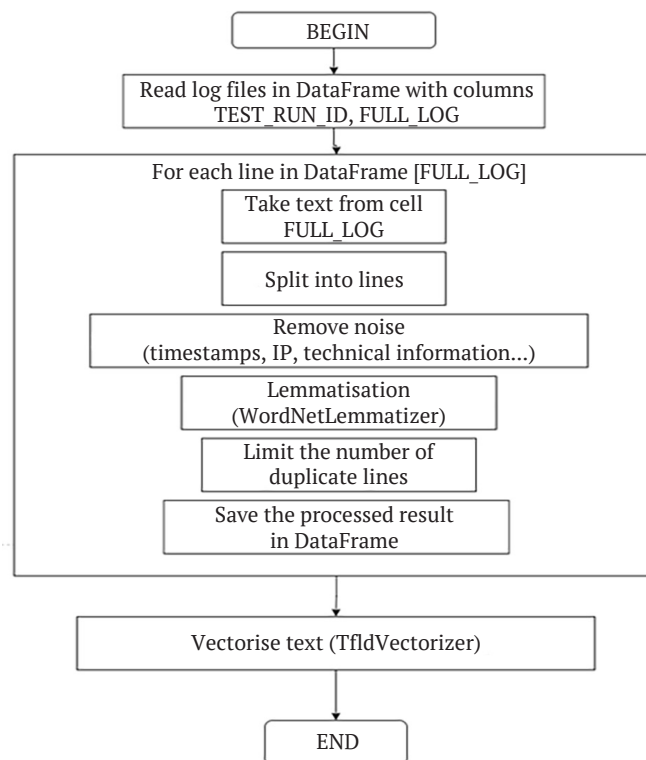


Figure 1. Block diagram of log file preprocessing

Source: developed by the author based on research

After preprocessing the data, the TF-IDF (Term frequency – inverse document frequency) method was used to vectorise the text, implemented using the TfidfVectorizer library from the scikit-learn package, according to the TfidfVectorizer (n.d.). This method was widely used for processing text data, as it allows determining

the weight of each word based on its frequency in the document and rarity throughout the corpus, according to G. Salton *et al.* (1975). TF-IDF helps to reduce the impact of noisy or overly frequent words by highlighting key terms that are unique or relevant to specific messages or errors.

Equation for calculating TF:

$$TF(t, d) = \frac{\text{Number of times the term } t \text{ appears in the document } d}{\text{Total number of terms in the document } d}, \quad (1)$$

where TF – frequency of the term, t – the term being analysed, d – the document for which the analysis is being performed.

Equation for calculating IDF:

$$IDF(t, D) = \frac{\text{Total number of documents in the corpus } D}{\text{Number of documents including the term } t}, \quad (2)$$

where IDF – inverse frequency of documents, t – the term being analysed, D – entire body of documents.

Equation for calculating TF-IDF:

$$TF-IDF(t, d, D) = TF(t, d) \times IDF(t, D), \quad (3)$$

where TF-IDF – combined metric that considers the frequency of the term in the document and its rarity throughout the body, t – the term being analysed, d – the document for which the analysis is performed, D – entire body of documents, TF – term frequency, IDF – inverted document frequency.

To implement this pipeline, the Pipeline interface was used, which inherits the BaseEstimator classes (BaseEstimator section in scikit-learn official documentation, n.d.) and TransformerMixin (TransformerMixin, n.d.), according to Developing scikit-learn estimators (n.d.):

```
import re
from typing import Union, Optional
from sklearn.base import BaseEstimator, TransformerMixin
from nltk.stem import WordNetLemmatizer

class RegexpCleaner(BaseEstimator, TransformerMixin):
    """Removes substrings for the given patterns."""

    def __init__(self, patterns: Union[str, list[str]]) -> None:
        """Initialize with a single pattern, or list of patterns to
        remove.

        :param patterns: e.g. r'^a-zA-Z\s', [r'^a-zA-Z\s'], ...
        """
        self.patterns: list[str] = [patterns] if isinstance(patterns, str)
        else patterns

    def fit(self, X: list[str], y: Optional[list] = None) -> "RegexpCleaner":
        """Dummy function to follow the interface of a
        transformer."""
        return self

    def transform(self, X: list[str], y: Optional[list] = None) -> list[str]:
        """For each line in X remove given pattern one by one.

        :param X: dataset to transform
        :param y: added to implement interface.
        :return: transformed dataset
        """
        cleaned: list[str] = []
        for text in X:
            for pattern in self.patterns:
                text = re.sub(pattern, "", text)
```

```
        cleaned.append(text)
        return cleaned

class Lemmatizer(BaseEstimator, TransformerMixin):
    """Lemmatizes data in a given dataset"""

    def __init__(self) -> None:
        self.lemmatizer = WordNetLemmatizer()

    def fit(self, X: list[str], y: Optional[list] = None) -> "Lemmatizer":
        """Dummy function to follow the interface of a
        transformer."""
        return self

    def transform(self, X: list[str], y: Optional[list] = None) -> list[str]:
        """
        Lemmatize the lines in the given dataset.
        """
        lemmatized: list[str] = []
        for text in X:
            words = text.split()
            lemmatized_words = [self.lemmatizer.lemmatize(word) for
            word in words]
            lemmatized.append(" ".join(lemmatized_words))
        return lemmatized

class DuplicateLimiter(BaseEstimator, TransformerMixin):
    """Limits number of duplicates in a given dataset"""

    def __init__(self, max_dups: int = 5) -> None:
        """Initialize with a number of duplicates to leave

        :param max_dups: number of duplicates to leave
        :raises ValueError: if max_dups has an incorrect type
        """
        if not isinstance(max_dups, int):
            raise ValueError(f"max_dups should be integer,
            {type(max_dups)} was given.")
        self.max_dups = max_dups

    def fit(self, X: list[str], y: Optional[list] = None) -> "DuplicateLimiter":
        """Dummy function to follow the interface of a
        transformer."""
        return self

    def transform(self, X: list[str], y: Optional[list] = None) -> list[str]:
        """Limit the duplicates in a dataset

        :param X: dataset to transform
        :param y: added to implement interface.
        :return: transformed dataset
        """
        line_counts = {}
        result = []
        for line in X:
            line_counts[line] = line_counts.get(line, 0) + 1
            if line_counts[line] <= self.max_dups:
                result.append(line)
        return result
```

Such a pipeline for data preprocessing has a concise structure and can be easily expanded if necessary:

```
from sklearn.pipeline import Pipeline

ID_PATTERN = r'\b[A-F0-9]{16},'
SPECIAL_CHAR_PATTERN = r'^a-zA-Z\s'
MULTIPLE_WHITESPACES_PATTERN = r'\s+'

pipeline = Pipeline([
    ('regex_cleaner', RegexpCleaner(patterns=[ID_PATTERN,
    SPECIAL_CHAR_PATTERN, MULTIPLE_WHITESPACES_
```

```
PATTERN])),
('lemmatizer', Lemmatizer()),
('dup_limiter', DuplicateLimiter(max_dupes = 5)),
])
```

```
X_preprocessed = pipeline.fit_transform(X)
```

The last step was to select the model and metrics for evaluation. The RandomForestClassifier algorithm, which is an ensemble machine learning method known for its high accuracy and resistance to retraining in classification

problems, was chosen to classify defect types. Previous studies confirm its effectiveness under similar conditions, according to M. Prodeus *et al.* (2024). Three metrics were used to evaluate the model quality: Log Loss, F1-score, and MCC (Table 1). Each of them measures different aspects of the algorithm's performance, which is especially important for working with unbalanced classes that are often found when classifying defects in log files, according to C. Cao *et al.* (2020).

Table 1. Metrics for evaluating the model

Metric	What it measures	Why it is used
Log Loss	How confident is the model in its predictions	Considers the probability of predictions
F1-score	Balance precision and recall	Important for unbalanced classes
MCC	General correlation of predictions with real classes	Best metric for unbalanced data

Source: developed by the author based on C. Cao *et al.* (2020)

The use of these metrics allows getting a comprehensive understanding of the quality of the classification model, ensuring the reliability and efficiency of automated analysis of defects in log files.

In this context, the programme code that was used to create and train the model is presented:

```
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size = 0.7, random_state = 42, stratify = y)
```

```
rf_clf = RandomForestClassifier(random_state = 42, n_estimators = 100)
start_time_train = time.time()
rf_clf.fit(X_train, y_train)
train_time = time.time() - start_time_train
```

```
y_pred = rf_clf.predict(X_test)
predict_proba = rf_clf.predict_proba(X_test)
```

```
log_loss = log_loss(y_test, predict_proba)
f1 = f1_score(y_test, y_pred, average="weighted")
matthews = matthews_corrcoef(y_test, y_pred)
print(f"log_loss: {log_loss:.4f}, f1: {f1}, matthews: {matthews}, train_time: {train_time}")
```

Results and Discussion

This section presents the results of an experimental study and their analysis. First, the obtained cluster structures and the formation of defect classes were considered, followed by the effect of pre-processing on the learning rate and model accuracy. In the end, a comparison is made with the latest developments of other researchers in the field of log file analysis. This approach ensured the integrity of the presentation and allowed interpreting the results qualitatively.

The obtained cluster analysis data confirmed the presence of five groups (clusters) in the data sample. The values of the silhouette coefficient for a different number of clusters were: for 3 clusters – 0.3758 (unsatisfactory separation); 4 – 0.4770 (improvement, but still weak); 5 – 0.5471 (optimal balance of cohesion and separation); 6 – 0.6088 (higher result, but there is an appearance of overtraining and blurring of borders between some clusters). Thus, the choice of 5 clusters as the optimal option for constructing

classification labels is justified. Figure 2 showed the results of the silhouette coefficient analysis for a different number of clusters, confirming the above values.

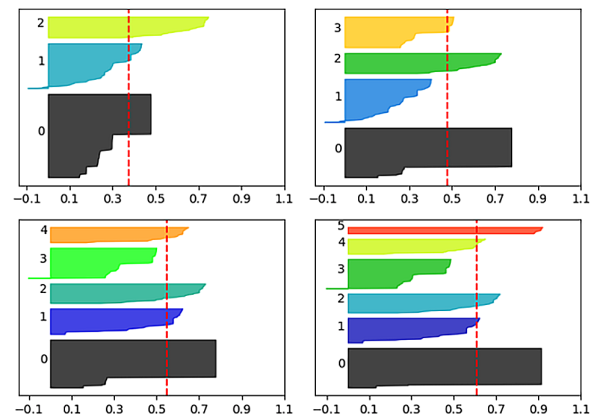


Figure 2. Silhouette factor analysis based on the number of clusters

Source: developed by the author based on research

Figure 3 showed the result of clustering using an uncontrolled Gaussian Mixture Model. For clarity, the data dimension was reduced to two main components using the PCA (Principal component analysis) method, which allows mapping clusters in two-dimensional space.

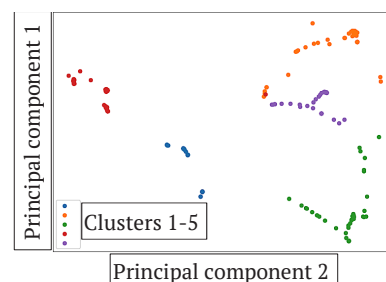


Figure 3. Visualisation of clusters obtained during the previous step

Source: developed by the author based on research

The resulting clusters look clearly separated, which indicates a successful choice of model parameters. The result corresponds to the previous silhouette analysis, according to which the optimal number of clusters is 5–6. In this case, the model also formed 5 clusters, which confirms the consistency between different analysis methods.

Next, a dataset was generated with labels (target values) for each log file from the dataset. Figure 4 showed the distribution of the sample by five classes of defects (codes 0–4) detected during automated testing. The X-axis shows the codes of defect classes, and the Y-axis shows the number of samples in each class. According to the HDFS architecture guide (n.d.), the resulting uneven distribution corresponds to the types of defects that were removed as part of the use of the system under study. It displays the actual types of failures, which include: data disk failure, heartbeats and re-replication, cluster rebalancing, data integrity, and metadata disk failure.

The distribution between classes was uneven, which indicates a more frequent occurrence of certain types of defects. In particular, class 0 (red) contains the largest number of samples, while classes 2 (purple) and 4 (pink) contain the least. This imbalance is typical for large distributed

systems and affects further training of models, because models can lean towards the dominant class.

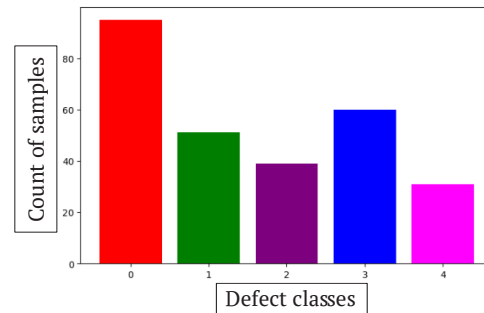


Figure 4. Distribution of the sample by five classes of defects
Source: developed by the author based on research

To illustrate the input data, Figure 5 showed a fragment of a raw log file from the HDFS_v3_TraceBench dataset. The file contains a significant amount of noise and duplicate service records. This format is typical for log files generated by software systems or automated tests, and confirms the presence of unstructured, redundant information in logs.

TaskID	TID	OpName	StartTime	EndTime	HostAddress	HostName	Agent
B076E6516B275ABB	E4F8152FB0F96806	getFileInfo	4119548017230538	4119548017995575	10.107.100.57	namenode	Namenode
B076E6516B275ABB	D7F390A0AF3856A0	RPC.getFileInfo	4113085176787665	4113085184877160	10.107.100.135	client024	RPC Client
B076E6516B275ABB	526C20ABFDA933F	getFileInfo	4119548029848716	4119548030243538	10.107.100.57	namenode	Namenode
B076E6516B275ABB	D7F390A0AF3856A0	RPC.getFileInfo	4113085190448805	4113085192061960	10.107.100.135	client024	RPC Client
B076E6516B275ABB	E024618C927914A1	getFileInfo	4119548042473974	4119548042701240	10.107.100.57	namenode	Namenode
B076E6516B275ABB	D7F390A0AF3856A0	RPC.getFileInfo	4113085203163602	4113085204571562	10.107.100.135	client024	RPC Client
B076E6516B275ABB	322C5157B58F8DC4	getFileInfo	4119548043862355	4119548044102628	10.107.100.57	namenode	Namenode
B076E6516B275ABB	D7F390A0AF3856A0	RPC.getFileInfo	4113085204730525	4113085205815934	10.107.100.135	client024	RPC Client
B076E6516B275ABB	F9CEE7766D3F80AE	getBlockLocations	4119548047519292	4119548048701722	10.107.100.57	namenode	Namenode

Description	
Success: returnCM[5]class=class org.apache.hadoop.hdfs.protocol.HdfsFileStatus	valien=org.apache.hadoop.hdfs.protocol.HdfsFileStatus@1d7a3413
Success: returnCM[5]class=class org.apache.hadoop.hdfs.protocol.HdfsFileStatus	valien=org.apache.hadoop.hdfs.protocol.HdfsFileStatus@2114e6f9
Success: returnCM[5]class=class org.apache.hadoop.hdfs.protocol.HdfsFileStatus	valien=org.apache.hadoop.hdfs.protocol.HdfsFileStatus@13c37e03
Success: returnCM[5]class=class org.apache.hadoop.hdfs.protocol.HdfsFileStatus	valien=org.apache.hadoop.hdfs.protocol.HdfsFileStatus@96b79959
Success: returnCM[5]class=class org.apache.hadoop.hdfs.protocol.HdfsFileStatus	valien=org.apache.hadoop.hdfs.protocol.HdfsFileStatus@550115e0
Success: returnCM[5]class=class org.apache.hadoop.hdfs.protocol.HdfsFileStatus	valien=org.apache.hadoop.hdfs.protocol.HdfsFileStatus@4807b8a13
Success: returnCM[5]class=class org.apache.hadoop.hdfs.protocol.HdfsFileStatus	valien=org.apache.hadoop.hdfs.protocol.HdfsFileStatus@4ac27a3b
Success: returnCM[5]class=class org.apache.hadoop.hdfs.protocol.HdfsFileStatus	valien=org.apache.hadoop.hdfs.protocol.HdfsFileStatus@4937e053
Success: returnCM[5]class=class org.apache.hadoop.hdfs.protocol.LocatedBlocks	valien=org.apache.hadoop.hdfs.protocol.LocatedBlocks@7023a603
Success: returnCM[5]class=class org.apache.hadoop.hdfs.protocol.LocatedBlocks	valien=org.apache.hadoop.hdfs.protocol.LocatedBlocks@141107103
Success: chosen bestNode = 10.107.100.96:50010 in nodes = 10.107.100.96:50010 10.107.100.86:50010 10.107.100.82:50010	

Figure 5. Log file data format from the HDFS_v3_TraceBench dataset

Source: J. Zhu et al. (2013)

The comparative analysis performed showed a clear advantage of the model trained on cleared logs over the option without preprocessing. First, the volume and structure of input data changed significantly: deleting duplicate and service records reduced the volume of log files by almost 80%, reducing the dimension of the feature space. Optimising the input data significantly reduced the training time of the model, as shown in Figure 6. The first column showed the time spent on training (red), and the second column showed the time spent on preprocessing (blue) and training. The results showed that although data preprocessing takes time, the total time spent processing data and training the model is almost five times less compared to training the model on raw data. In particular, the conversion of log files to TF-IDF vectors and further training of the model on cleaned data was much faster than on raw data. This is important from the standpoint of scaling, because it helps to

speed up the analysis of large volumes of log files without additional resources. In addition, the preprocessing process can take place separately from vectorisation and training, to save resources or save disk space spent on storing logs.

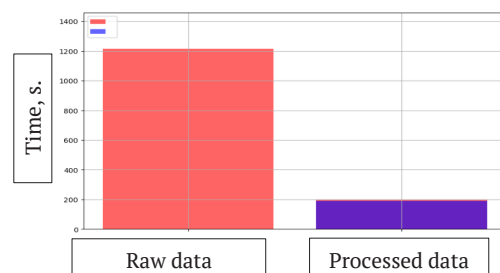


Figure 6. Impact of data quality on time spent on vectorisation

Source: developed by the author based on research

The simulation results showed that pretreatment of data contributed not only to the preservation, but also to the improvement of classification accuracy. Figure 7 compared the performance of the model trained on cleaned data with the model built on raw input data. There was a moderate increase in the values of the F1 measure and the Matthews correlation coefficient (MCC) after cleaning, which indicated a decrease in noise levels and the preservation of key informative features. A slight increase in the value of the Log Loss metric did not significantly affect the overall quality of classification, which indicates that the model remains confident in its forecasts. Thus, the model trained on cleared logs demonstrated both acceleration and high stability of accuracy metrics. Consequently, high-quality preprocessing of logs provided simultaneous acceleration of work and increased reliability of the model, which is a significant advantage from the standpoint of practical application in defect analysis.

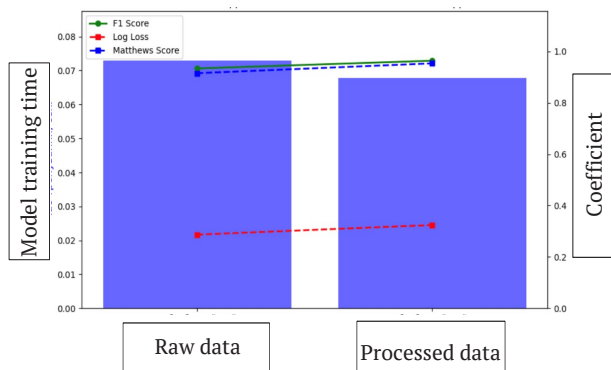


Figure 7. Influence of data quality on model accuracy
Source: developed by the author based on research

The results obtained in the framework of this study are consistent with the conclusions of modern works in the field of analysis of system journals. In particular, M. Aljabri *et al.* (2022) successfully applied Machine Learning and Deep Learning algorithms to classify network firewall records, achieving very high accuracy (more than 99% for Random Forest). This is consistent with the current study, where the Random Forest model also showed consistently high performance on cleaned data. This coincidence of results indicates the universality of the chosen approach to log classification, even if the data nature is different. Similar results were observed in the study by V. Koval *et al.* (2022), where the Random Forest Regression model was used to predict the quality of electricity in systems with renewable sources. Despite the different scope of application, the model demonstrated high stability and accuracy when working with a large array of synchronised data. This confirms the flexibility and reliability of Random Forest in tasks that require processing complex structured logs, and reinforces the argument for the universality of this approach in classification and forecasting problems. Other researchers have also emphasised the importance of effective preprocessing of journals. Thus, J. Wang *et al.* (2020) proposed

a LogEvent2Vec model that directly feeds log events to the input of a word2vec network for vectorisation. This helped to avoid multi-level transformations and speed up data analysis by about 30 times, while improving the accuracy of anomaly detection. Although the TF-IDF approach presented in this paper is simpler, it also provided approximately five-fold acceleration, which confirms the general trend: optimising the log representation significantly affects the performance of models and allows adapting the analysis to limited computing resources. S.A. Hussein & S.R. Répás (2024) noted the advantages of semantic features (based on word2vec) over purely statistical features (TF-IDF) for anomaly recognition tasks in system logs. However, using the word2vec technique itself requires significant computational resources and careful dictionary construction. In the case of the current study, even the use of a relatively simple TF-IDF proved effective due to preliminary clearing of logs, which reduced noise and highlighted meaningful terms, which allowed compensating for the lack of a deeper semantic context. Moreover, the results obtained are consistent with the findings of Z.A. Khan *et al.* (2024), who showed that it is not so much the absolute accuracy of log parsing as the ability to distinguish informative characteristics (“distinctiveness”) that determines the success of models. In this context, the experiment of the current study was aimed precisely at improving the selection of features through data cleaning, which ultimately allowed improving both the speed and accuracy of the model without complicating the architecture. Thus, the comparison confirmed the hypothesis that high-quality preprocessing of log files is a key factor in both speeding up calculations and improving the accuracy of defect prediction, which is consistent with trends in current research. U. Meteriz *et al.* (2020) presented a log classification system for telecom systems using a deep neural network. The researchers noted the success of advanced NLP (Natural Language Processing) for logs, which opens up prospects for automated analysis of complex technical messages. The results of the current study are consistent with their classification accuracy, although the approach used in the current study (TF-IDF + RandomForest) is significantly simpler and resource-saving compared to the deep CNN (Convolutional Neural Network) model. The effectiveness of data cleaning was also confirmed by the findings of O. Johnphill *et al.* (2024), according to which removing uninformative records and duplicate messages allows the model to focus on relevant signals, which increases learning efficiency. D.A. Bhanage & A.V. Pawar (2023) presented an approach to classifying system logs as natural language texts. The researchers tested a combination of several vectorisation techniques – TF-IDF, Word2Vec, and tonality analysis (polarity) – in combination with classical classifiers. According to the results, the most promising approach was to consider the “tone” of messages, which improved the recognition of complex irregular logs. D.A. Bhanage & A.V. Pawar’s approach demonstrated that additional semantic features (such as sentiment) can improve the quality of log analysis.

The method used in this study, on the contrary, is deliberately simplified (only TF-IDF without deep semantic analysis), which makes it less resource-intensive, although it is somewhat inferior in terms of contextual nuances. P. Marjai & A. Kiss (2024) pointed out in their paper that pre-structuring logs helps the model to better generalise error messages. The results obtained in the current study confirmed this trend: cleaning and normalising text data from log files helps to improve classification accuracy even without complicating the vectorisation approach. This is especially important in application environments where resources are limited and data processing speed requirements are critical.

The results showed that the quality and volume of log data have a complex relationship with the performance of machine learning models. In contrast to the common practice of accumulating and storing complete logs, the conducted experiment indicates the possibility of significantly reducing the amount of input information without negatively affecting the accuracy of classification. In particular, the removal of service records, repetitions, and technical labels helped not only to reduce processing time, but also to achieve better classification quality by key metrics. This effect can be explained by a decrease in the influence of information noise, which usually overloads the model and reduces its ability to distinguish relevant patterns in logs. The presence of a regular structure in messages after lemmatisation and cleaning can also help improve the vectorisation process, especially when using methods such as TF-IDF. However, the results indicate the potential adaptability of the proposed approach to other logging systems.

Conclusions

In the course of the study, it was found that log file preprocessing, which includes noise removal, lemmatisation, and limiting message duplication, significantly improves the efficiency of building machine learning models for defect classification. In particular, it has been shown that the use of regular expressions reduces the amount of unnecessary technical information, such as timestamps, identifiers, and log levels, which simplifies the structure of messages and facilitates further vectoring.

It has been proven that preprocessing reduces the time required for vectorisation and model training by almost five times without losing accuracy. The accuracy of the model is not only preserved, but also improved: an increase in the F1-score and Matthews correlation coefficient (MCC) was recorded, which indicates an improvement in the quality of classification even in the presence of unbalanced classes. In addition, it was found that the use of TF-IDF can

effectively emphasise meaningful words of diagnostic value, while suppressing repetitive terms. This has a positive effect on the model's noise resistance. It was also confirmed that the Log Loss value was decreasing, which indicates an increase in the model's confidence in its forecasts.

In the course of the study, a number of logically consistent and effective actions were identified that can be combined into a single pipeline for automatic processing of log files. Such a pipeline can be integrated into automated testing systems without significant losses for training the model, due to the effective allocation of the necessary information. This reduces the burden on analysts, ensures high classification accuracy, and reduces the time required to process large amounts of log data. It is important to note that the proposed approach does not depend on the use of complex or resource-intensive models and can be applied in real-world conditions without the need for high-performance equipment.

Thus, the results of the study confirmed the hypothesis that high-quality preprocessing of log files is a key factor in both speeding up calculations and improving the accuracy of models. The proposed approach can be successfully integrated into automated testing pipelines, helping to improve the efficiency of detecting and classifying software defects. In addition to its direct application value, the study highlights the potential of using simple word processing tools to solve complex software testing problems. Although the study was limited to one data set and one model type, the results provide prerequisites for further study of the method's stability to different domains and model architecture variants. The question of the extent to which other sources of information (e.g., contextual metadata) can influence the results of defect classification in logs remains relevant. In this context, the proposed solution is considered as a basis for future research.

Acknowledgements

The author is grateful to supervisor for support and valuable advice during the research process, as well as to colleagues who assisted in data analysis. Special thanks to the developers of the open data sets and libraries used in the study.

Funding

The study was not funded by any third-party organisation that could have influenced its results or conclusions.

Conflict of Interest

The author declares that there is no conflict of interest in the publication of this article.

References

- [1] Abdalla, R.R., & Jumaa, A.K. (2022). Log file analysis based on machine learning: A survey. *UHD Journal of Science and Technology*, 6(2), 77-84. [doi: 10.21928/uhdjst.v6n2y2022.pp77-84](https://doi.org/10.21928/uhdjst.v6n2y2022.pp77-84).
- [2] Aljabri, M., Alahmadi, A.A., Mohammad, R.M.A., Aboulmour, M., Alomari, D.M., & Almotiri, S.H. (2022). Classification of firewall log data using multiclass machine learning models. *Electronics*, 11(12), article number 1851. [doi: 10.3390/electronics11121851](https://doi.org/10.3390/electronics11121851).

- [3] BaseEstimator section in scikit-learn official documentation. (n.d.). Retrieved from <https://scikit-learn.org/stable/modules/generated/sklearn.base.BaseEstimator.html>.
- [4] Bhanage, D.A., & Pawar, A.V. (2023). Improving classification-based log analysis using vectorisation techniques. In A.B. Reddy, S. Nagini, V.E. Balas & K.S. Raju (Eds.), *Proceedings of 3rd international conference on advances in computer engineering and communication systems. Lecture notes in networks and systems* (Vol. 612, pp. 271-282). Singapore: Springer. doi: 10.1007/978-981-19-9228-5_24.
- [5] Brandão, A., & Georgieva, P. (2020). Log files analysis for network intrusion detection. In *Proceedings of the 2020 IEEE 10th international conference on intelligent systems (IS)* (pp. 328-333). Varna: IEEE. doi: 10.1109/IS48319.2020.9199976.
- [6] Cao, C., Chicco, D., & Hoffman, M.M. (2020). The MCC-F1 curve: A performance evaluation technique for binary classification. *ArXiv*. doi: 10.48550/arXiv.2006.11278.
- [7] Developing scikit-learn estimators. (n.d.). Retrieved from <https://scikit-learn.org/stable/developers/develop.html>.
- [8] Géron, A. (2022). *Hands-on machine learning with scikit-learn, keras, and tensorflow: Concepts, tools, and techniques to build intelligent systems* (3rd ed.). Sebastopol: O'Reilly Media.
- [9] HDFS architecture guide. (n.d.). Retrieved from https://hadoop.apache.org/docs/r1.2.1/hdfs_design.html.
- [10] Huangfu, Y. (2022). *Machine learning for log data analysis*. (PhD thesis, McMaster University, Hamilton, Canada).
- [11] Hussein, S.A., & Répás, S.R. (2024). Anomaly detection in log files based on machine learning techniques. *Journal of Electrical Systems*, 20(3), 1299-1311. doi: 10.52783/jes.1505.
- [12] Johnphill, O., Sadiq, A.S., Kaiwartya, O., & Aljaidi, M. (2024). An intelligent approach to automated operating systems log analysis for enhanced security. *Information*, 15(10), article number 657. doi: 10.3390/info15100657.
- [13] Kapitanets, S., & Radelchuk, G. (2022). Peculiarities of data logging and influence of application type on the choice of logging methodology. *Bulletin of Khmelnytskyi National University*, 6(315), 98-101. doi: 10.31891/2307-5732-2022-315-6-98-101.
- [14] Khan, Z.A., Shin, D., Bianculli, D., & Briand, L. (2024). Impact of log parsing on deep learning-based anomaly detection. *Empirical Software Engineering*, 29, article number 139. doi: 10.1007/s10664-024-10533-w.
- [15] Khil, O., & Yakovina, V. (2023). Analysis of the problems of applying machine learning methods for evaluating and predicting software defects. *Scientific Bulletin of UNFU*, 33(3), 110-116. doi: 10.36930/40330316.
- [16] Koval, V., Lysenko, V., Kiktev, N., Pylypenko, Yu., Samkov, O., Osinskiy, O., & Popov, I. (2022). Automated monitoring of time synchronisation devices and digital processing of vector measurements of dynamic characteristics of smart grid power systems. *Machinery & Energetics*, 13(2), 73-82. doi: 10.31548/machenergy.13(2).2022.73-82.
- [17] Marjai, P., & Kiss, A. (2024). The usage of template mining in log file classification. *IEEE Access*, 12, 96378-96386. doi: 10.1109/ACCESS.2024.3426959.
- [18] Meteriz, U., Yildiran, F.N., Kim, J., & Mohaisen, D. (2020). Understanding the potential risks of sharing elevation information on fitness applications. In *40th international conference on distributed computing systems (ICDCS)* (pp. 464-473). Singapore: IEEE. doi: 10.1109/ICDCS47774.2020.00063.
- [19] Peronto, R. (2024). The state of log data: 6 trends impacting observability and security. *Chronosphere Blog*. Retrieved from <https://chronosphere.io/learn/observability-log-data-trends/>.
- [20] Prodeus, M., Nichaporuk, A., & Ivanchenko, O. (2024). The influence of data preprocessing on the performance of the random forest model in detecting network attacks. *Measuring and Computing Devices in Technological Processes*, 4, 415-419. doi: 10.31891/2219-9365-2024-80-51.
- [21] Qin, Q., Aghili, R., Li, H., & Merlo, E. (2024). Preprocessing is all you need: Boosting the performance of log parsers with a general preprocessing framework. In *2025 IEEE international conference on software analysis, evolution, and reengineering (SANER)* (pp. 1-11). Singapore: IEEE. doi: 10.48550/arXiv.2412.05254.
- [22] Ramachandran, S., Agrahari, R., Mudgal, P., Bhilwaria, H., Long, G., & Kumar, A. (2023). Automated log classification using deep learning. *Procedia Computer Science*, 218, 1722-1732. doi: 10.1016/j.procs.2023.01.150.
- [23] Ryciak, P., Najgebauer, A., & Sołtysiak, M. (2022). A lightweight hybrid deep learning approach for log anomaly detection. *Applied Sciences*, 12(10), article number 5089. doi: 10.3390/app12105089.
- [24] Salton, G., Wong, A., & Yang, C.S. (1975). A vector space model for automatic indexing. *Communications of the ACM*, 18(11), 613-620. doi: 10.1145/361219.361220.
- [25] Sandhu, A., & Mohammed, S. (2022). Detecting anomalies in logs by combining NLP features with embedding or TF-IDF, 1-5. doi: 10.36227/techrxiv.19498769.v1.
- [26] Shirzad, E., & Saadatfar, H. (2022). Job failure prediction in Hadoop based on log file analysis. *International Journal of Computers and Applications*, 44(3), 260-269. doi: 10.1080/1206212X.2020.1732081.
- [27] TfidfVectorizer. (n.d.). Retrieved from https://scikit-learn.org/stable/modules/generated/sklearn.feature_extraction.text.TfidfVectorizer.html.
- [28] TransformerMixin. (n.d.). Retrieved from <https://scikit-learn.org/stable/modules/generated/sklearn.base.TransformerMixin.html>.

- [29] Wang, J., Tang, Y., He, S., Zhao, C., Sharma, P.K., Alfarraj, O., & Tolba, A. (2020). LogEvent2vec: LogEvent-to-vector based anomaly detection for large-scale logs in Internet of Things. *Sensors*, 20(9), article number 2451. [doi: 10.3390/s20092451](https://doi.org/10.3390/s20092451).
- [30] Zhou, J., Chen, Z., Wang, J., Zheng, Z., & Lyu, M.R. (2014). Trace bench: An open data set for trace-oriented monitoring. In *6th IEEE international conference on cloud computing technology and science* (pp. 519-526). Singapore: IEEE. [doi: 10.1109/CloudCom.2014.79](https://doi.org/10.1109/CloudCom.2014.79).
- [31] Zhu, J., He, S., He, P., Liu, J., & Lyu, M.R. (2023). Loghub: A large collection of system log datasets for AI-driven log analytics. *ArXiv*. [doi: 10.48550/arXiv.2008.06448](https://doi.org/10.48550/arXiv.2008.06448).

Вплив обробки лог-файлів на швидкість навчання та точність класифікації дефектів

Антон Каяфюк

Аспірант

Київський національний університет технологій та дизайну

01011, вул. Мала Шияновська, 2, м. Київ, Україна

<https://orcid.org/0009-0003-9917-0834>

Анотація. Метою було дослідити вплив попередньої обробки лог-файлів автоматизованого тестування на швидкість векторизації та навчання моделей машинного навчання. Використано набір HDFS_v3_TraceBench, що містить понад 370 тисяч трасувань, зібраних у середовищі Hadoop Distributed File System. Обробка включала видалення шуму, лематизацію та зменшення дублікатів. Дані векторизовано методом Term frequency – inverse document frequency, після чого навчено модель RandomForestClassifier. Результати експериментів показали, що оптимізація вхідних даних дозволила зменшити загальний час обробки майже вп'ятеро. Час, необхідний для векторизації тексту та навчання моделі, скоротився, що дає змогу пришвидшити роботу з великими обсягами логів. При цьому точність класифікації не лише збереглася, а й продемонструвала незначне покращення: показники F1-score та коефіцієнта кореляції Метьюса залишилися стабільно високими. Також спостерігалось зниження значення Log Loss, що свідчило про підвищення впевненості моделі у власних прогнозах. Це особливо важливо в умовах незбалансованих класів, характерних для задач класифікації дефектів. Детальний аналіз виявив, що значна частина службової та повторюваної інформації в логах не є критичною для навчання моделі, а її видалення навпаки покращує якість підготовки даних. У ході роботи також було підтверджено, що отримані цільові мітки для логів відповідають типовим класам помилок. Реалізована обробка лог-файлів не лише скорочує обчислювальні витрати, але й підтримує або покращує якість прогнозування. Ці результати підтвердили доцільність включення етапу очищення та оптимізації логів у загальний процес побудови моделей машинного навчання для автоматизованого тестування. Отримані результати можуть бути інтегровані в автоматизовані пайплайни для класифікації дефектів і формування баг-репортів. Це сприятиме зменшенню обсягу ручної праці та підвищенню ефективності команд

Ключові слова: регулярні вирази; лематизація; векторизація; машинне навчання; автоматизація тестування



Gradient boosting as a tool for solving classification problems in data-constrained environments

Mykola Kyrychek*

Master

Taras Shevchenko National University of Kyiv

01033, 60 Volodymyrska Str., Kyiv, Ukraine

<https://orcid.org/0009-0009-4571-123X>

Abstract. In machine learning, the question of effective construction of classification models with insufficient amount of educational information has arisen. The purpose of the study was to analyse the possibilities of using gradient boosting to solve classification problems in data-constrained environments. The research methodology was based on a comprehensive analysis of the leading gradient boosting implementations: XGBoost, LightGBM, and HistGradientBoosting. The main focus was on investigating regularisation mechanisms, hyperparameter optimisation strategies, and adaptive learning techniques under small sample conditions. The research was aimed at identifying the architectural features of algorithms that can provide high classification accuracy with a minimum amount of data. It was established that the proposed algorithms have demonstrated a significant potential for effectively solving classification problems. It was found that the mechanisms of shrinkage and subsampling significantly increased the generalising ability of models. The results of the study expanded the theoretical understanding of ensemble machine learning methods and outlined promising areas for adapting algorithms to specific conditions of limited information resources. XGBoost, LightGBM, and HistGradientBoosting have been shown to have unique architectural features that allow working efficiently with different types of data. It was found that the internal regularisation mechanisms of these algorithms provided resistance to retraining and high prediction accuracy. The potential of gradient boosting for solving complex classification problems in medicine, finance, and other industries with limited information resources is shown. The practical significance of the study was to develop methodological recommendations for selecting and configuring gradient boosting algorithms for various types of classification problems. The results obtained will be useful for further development of machine learning methods

Keywords: machine learning; adaptive algorithms; model optimisation; XGBoost; hyperparameterisation

Introduction

Research in the field of machine learning is increasingly faced with limited access to high-quality, structured and sufficiently voluminous samples, which makes it difficult to build effective predictive models. In practical areas such as medicine, agricultural technology, market analytics, or cybersecurity, data is often fragmented, noisy, or presented in limited volumes due to ethical, financial, or technical barriers (García *et al.*, 2022). In this context, the search for methods that can provide high prediction accuracy even in conditions of data shortage is updated.

Gradient boosting as one of the leading strategies for ensemble learning shows a high ability to adapt in

conditions of limited or unevenly distributed samples. According to M. Maftoun *et al.* (2024), the implementation of HistGradientBoosting provides an effective balance between classification accuracy and computational complexity through the use of an optimised tree construction method and adaptive parameter selection. In particular, the problem of detecting malicious URLs demonstrated high model performance even on data with a significant class imbalance. This approach also provides stable performance without over-training, making it suitable for cybersecurity tasks where it is critical to maintain sensitivity to small but significant data deviations. Similarly, models based on

Suggested Citation:

Kyrychek, M. (2025). Gradient boosting as a tool for solving classification problems in data-constrained environments. *Technologies and Engineering*, 26(2), 37-47. doi: 10.30857/2786-5371.2025.2.3.

*Corresponding author



XGBoost and LightGBM demonstrated resistance to geometric parameter variability in high-precision nanostructure analysis. The study by G. Fan & K. Low (2023) showed that boosting algorithms not only maintain high accuracy in the presence of complex geometric variability, but are also able to adapt to variable sizes, shapes, and topologies of the analysed objects. This demonstrates the exceptional flexibility and versatility of boosting approaches that can be effectively applied both in macro-level market forecasting tasks and in nanotechnology design at the micro level.

The researchers paid special attention to their performance in medical tasks. Thus, A. Sun & F. Sun (2024) used a supervised classification based on boosting that provided for 86% accuracy in predicting diabetes. T.-H. Lin *et al.* (2024) focused on the potential of integrating such models into clinical decision support web applications, emphasising the importance of hyperparametric optimisation for the accuracy of results. The study by S. Ghimire *et al.* (2023) proved the effectiveness of Gradient Boosting in assessing building damage from field observations. Á. Baran *et al.* (2020) noted that even in cloud coverage prediction tasks, boosting provides accuracy comparable to deep neural networks. S. Guo & B. Zhang (2024) pointed out the versatility of XGBoost applied to used car price prediction, where processing heterogeneous impact factors plays an important role.

Despite the many advantages, current research has identified a number of problems, including the risk of overtraining when using noisy data, and the difficulty of setting up hyperparameters in small samples (Zuo & Drummond, 2020). C. Zhang *et al.* (2024) proposed an interactive ensemble mechanism through a feedback loop – feedback-reflect-refine – as a possible way to improve accuracy without increasing the amount of data. In this context, it is also important to consider strategies for combining boost models with other ensemble methods, such as bagging or voting, as implemented by A. Helmut & D.T. Murdiansyah (2023). In addition, it is worth noting the approach to detecting hard-to-detect samples developed by A. Okhrimenko & N. Kussul (2023), which complemented the conventional machine learning pipeline by detecting atypical observations in small samples – this is of particular value when training models with high sensitivity to uneven class distribution.

Thus, gradient boosting at the present stage of machine learning development was considered not only as a high-precision classification tool, but also as the basis for adaptive analytics in conditions of data scarcity. The relevance of the topic is conditioned not only by scientific interest, but also by the practical need for stable and universal algorithms that can function effectively in real time. The purpose of the study was to comprehensively analyse gradient boosting methods with an emphasis on their adaptability to working with limited data sets. The objectives of the study were to evaluate the effectiveness of various implementations of the algorithm (XGBoost, LightGBM, HistGradientBoosting) in solving classification problems, determining optimal strategies for setting

hyperparameters and applying data balancing methods, and developing recommendations for improving the performance of models in specific conditions.

Materials and Methods

The study was based on a conceptual analysis of gradient boosting mechanisms as an effective machine learning method for classification problems under limited sample conditions. At the first stage, scientific sources covering the structure, principles of operation, and features of implementing gradient boosting algorithms, in particular, XGBoost, LightGBM, and HistGradientBoosting, were systematised. Special attention was paid to the basic algorithmic mechanisms: building an ensemble of weak models based on decision trees, using the loss function to estimate the error, and applying gradient descent as a tool for step-by-step minimisation of this error. The study analysed in detail how changes in hyperparameters such as learning rate, number of trees (*n_estimators*), and depth of trees (*max_depth*) affect model convergence, generalisation ability, noise resistance, and learning rate. Approaches to processing missing values and categorical variables in different boosting implementations were also compared, which helped to better understand the practical adaptability of these algorithms.

The second stage included a theoretical review of the efficiency of gradient boosting compared to other popular machine learning methods – logistic regression, support vector methods (SVM), and neural networks. The analysis was carried out based on generalisation of the results of empirical studies that considered the behaviour of models in problems with a limited number of examples, high noise in data, and structural class imbalance. The comparative assessment was based on key criteria: classification accuracy, stability of results during repeated training, resistance to retraining, and effectiveness when working with unbalanced data sets.

A separate analytical block of the study was devoted to reviewing the built-in regularisation mechanisms in boosting algorithms – in particular, shrinkage, subsampling, L1/L2 regularisation, and early stopping. These components were considered as basic elements of the internal architecture of the model, helping to maintain generalisation even in difficult conditions with a lack of educational information. The analysis was based on a description of the theoretical foundations and practical conclusions of leading publications that focused on the role of regularisation in combating retraining and stabilising models at the training stage (Maftoun *et al.*, 2024; Zhang *et al.*, 2024).

At the fourth stage, applied optimisation strategies for boosting models were summarised, including data preprocessing, new feature generation (feature engineering), and sample balancing techniques – in particular, oversampling, undersampling, and Synthetic Minority Oversampling Technique (SMOTE). The source of practical substantiation was empirical research (Bej *et al.*, 2020; Liu *et al.*, 2022), which demonstrated the effectiveness of these approaches

in classification under complex data conditions. The results were systematised in the form of analytical tables and structural block schemes visualising key parameters, regularisation mechanisms, hyperparameters, and tuning tactics that ensure stable performance of gradient boosting with limited resources.

Results

Basic principles and methodology of gradient boosting

Gradient boosting is an advanced method of ensemble machine learning that provides high prediction accuracy by gradually forming a sequence of weak models, each of which eliminates the errors of the previous one. Unlike classical algorithms that build a single complex model, boosting creates a composition of simple solutions – most often decision trees with limited depth – and gradually improves their generalisation ability (Sun & Sun, 2024; Lin *et al.*, 2024). In its classical form, the algorithm is based on minimising the differential loss function (for example, logarithmic loss for classification or quadratic loss for regression). Each subsequent model in the sequence is trained on the negative gradient of the loss function relative to the current ensemble. Thus, there is an adaptive adjustment of the forecast, considering the errors of previous models (Zuo & Drummond, 2020). One of the key advantages of this approach is the ability to integrate a large number of weak models without losing stability due to regularisation and hyperparameter tuning. Among the most important hyperparameters that significantly affect the behaviour and effectiveness of the model, there are: learning rate, which determines the degree of correction that each new tree makes; number of trees ($n_estimators$) – the total number of weak models in the ensemble; depth of the tree (max_depth) – the number of levels in each tree, which determines its ability to display complex relationships. Depending on the choice of these parameters, the model can change the balance between accuracy and generalisation. Techniques such as subsampling (random subset training), regularisation (L1/L2), and early stopping (early termination of training when the loss function is stabilised) (Maftoun *et al.*, 2024). Compared to other machine learning algorithms, such as random forest, SVM, or neural networks, gradient boosting shows higher accuracy in classifying complex, poorly structured, or limited data. SVM often loses efficiency in the presence of noise, and neural networks have high requirements for the number of examples and computing resources (García *et al.*, 2022; Ghimire *et al.*, 2023). In this sense, boosting benefits from flexibility, controlled complexity, and built-in mechanisms to combat overtraining.

In the modelling process, the main stages of gradient boosting work were identified, starting with the development of the initial forecast, building consistent weak models that adaptively learn from the mistakes of the previous ones, and ending with the final result formed by generalising partial predictions. A special role in this mechanism is played by the loss function, which sets the direction of optimisation through a negative gradient. Figure 1 showed a

simplified logic of gradient boosting, which illustrates the key stages of building an ensemble of weak models in the iterative learning process.

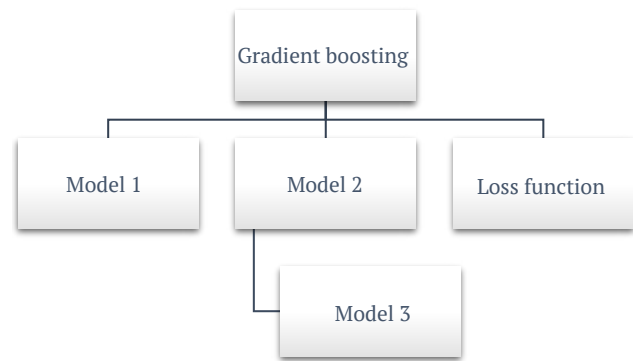


Figure 1. Algorithmic structure of gradient boosting operation

Source: compiled by the author based on the concept by Y. Zuo & T. Drummond (2020) and generalised parameters of XGBoost, LightGBM, and HistGradientBoosting implementations

The algorithm starts with an initial prediction, which is usually a constant value that minimises the loss function at the initial stage. Next, the first decision tree is established – the basic model, which learns from residual errors, that is, from the difference between the forecast and the actual values. Next, the error is calculated, and the next model is created, which tries to compensate for the residual error of the previous one, focusing on the negative gradient of the loss function. Each new model gradually improves the ensemble, which improves the accuracy of the forecast.

Classification problems with limited data

One of the most important problems encountered in classification problems is the limited amount of training data. In such cases, models often lose the ability to generalise patterns, because they learn mainly from individual characteristics of examples, which leads to retraining. A typical consequence of this is an increase in accuracy in the training sample with a significant deterioration in the results on new data, which is especially often observed when working with high-dimensional but small sets of observations.

Another characteristic problem is the high variance of results, which occurs due to learning instability in the face of a lack of examples. Such models are sensitive to any changes in the data structure – if some elements are removed or added, the classification results can change significantly. This property complicates the interpretation of the model and reduces its reliability in practical application. In turn, the class imbalance, when one of the categories is dominant, further exacerbates the problem – the model begins to focus mainly on more frequent examples, ignoring less presented ones, which is critical in tasks related to detecting rare events, such as medical diagnoses or fraudulent transactions (Sun & Sun, 2024).

Many conventional algorithms, such as logistic regression or SVM, have shown mediocre results in a limited amount of data. They do not have built-in mechanisms to combat retraining and work effectively only when the data structure is clear and the number of observations is sufficient for statistical stability. Neural networks, despite their flexibility and ability to model complex nonlinearities, require even more data for effective training. In cases where there are no such volumes, deep models usually either do not match, or form too complex dependencies that have no practical value (Lin *et al.*, 2024).

Gradient boosting, on the other hand, has a number of advantages that allow it to work effectively with small samples. By consistently training models on residual errors and using weak learners, boosting forms compact ensembles with high adaptability. Regularisation mechanisms, in

particular, shrinkage and subsampling, allow limiting the complexity of the model without significant loss of accuracy, while the early stop function provides an optimal balance between performance and generalisation.

Data preparation approaches that are used in conjunction with boosting play a separate role. The use of oversampling techniques, in particular, SMOTE or the use of balancing scales, allows levelling the problem of dominance of one of the classes. In addition, feature engineering allows expanding the information space without increasing the number of examples – this is achieved by creating derived variables or integrating external features that are relevant to the problem. Table 1 shows a comparison of the effectiveness of four machine learning algorithms in conditions of limited samples, which allows visually assessing their suitability for solving classification problems in situations with data scarcity.

Table 1. Comparison of model performance in small samples

Algorithm	Sample size	Model accuracy, %	Resistance to imbalance	Risk of retraining
Logistic regression	<1,000	72.4	Low	High
SVM	<1,000	78.1	Average	Average
Neural network	<1,000	80.3	Average	High
HistGradientBoosting	<1,000	85.6	High	Low

Source: created by the author based on summarising the findings by M. Maftoun *et al.* (2024), A. Sun & F. Sun (2024), T.-H. Lin *et al.* (2024)

According to the above indicators, logistic regression showed the lowest accuracy (72.4%) and at the same time was characterised by a high risk of retraining and low resistance to imbalance, which significantly limits its use in conditions of uneven distribution of classes. SVM showed slightly better results – an accuracy of 78.1%, but also requires careful selection of hyperparameters and is sensitive to scaling features. Neural networks showed higher accuracy (80.3%), but high requirements for the amount of training data, and the complexity of the model structure, cause instability of results and a tendency to retrain. The best results were shown by HistGradientBoosting – accuracy of 85.6%, high resistance to imbalance, minimal risk of overtraining. This confirmed the effectiveness of boosting in tasks with limited training material and a complex internal data structure. Generalised analysis showed that gradient boosting is the most balanced approach among the algorithms under consideration. Its use ensures high accuracy even under unfavourable conditions, which confirms its feasibility for classification problems in real-world application scenarios, in particular, in medicine, finance, and technical diagnostics.

Argumentation of the effectiveness of gradient boosting in conditions of limited data

Gradient boosting occupies a special place among machine learning algorithms due to its ability to adapt to complex data structures even in conditions of insufficient training data. Unlike models that require a large number of examples for

effective training, boosting can extract significant patterns from limited samples by gradually building an ensemble of weak models. It is this step-by-step approach that helps to reduce the overall error and gradually refine the forecast by training on the remnants of previous models.

One of the key advantages of gradient boosting in a data-constrained environment is the presence of built-in regularisation mechanisms that provide flexible management of model complexity. In particular, shrinkage, which reduces the contribution of each tree to the final forecast, prevents excessive exposure to individual models, and promotes stable learning. Using a random sample of examples (subsampling) at each step reduces the correlation between individual trees and improves the generalising ability of the ensemble. In addition, the regularisation parameters L1 and L2 control the weights involved in updating and help muffle noise variables that are particularly common in small or unfiltered datasets (Rismayati *et al.*, 2022).

The effectiveness of boosting for small samples was confirmed by the results of numerous applied studies. In particular, M. Maftoun *et al.* (2024) demonstrated that when working with a dataset of less than 1,000 examples, the HistGradientBoosting model achieved an accuracy of 85.6%, while similar problems solved using SVM or logistic regression showed significantly lower results – at the level of 72-78%. These figures are consistent with other studies where XGBoost has shown a steady superiority in medical and technical tasks, accompanied by a lack of balanced learning examples (Lin *et al.*, 2024; Sun & Sun, 2024).

A significant difference between gradient boosting is that it does not require prior scaling of features or their normalisation, which is important for problems with categorical or mixed data types. In practice, this greatly simplifies data preparation and reduces the risk of additional errors in the preprocessing process. Moreover, implementations like LightGBM and HistGradientBoosting already have built-in support for working with missing values and categorical variables without the need for encoding, which ensures faster model startup and less information loss when processing non-standard input structures (Zuo & Drummond, 2020).

Unlike neural networks, which are sensitive to architecture configuration, layer size, activation functions,

and computing resource requirements, boosting requires less training time and is less demanding on hyperparameters. While neural networks can achieve high accuracy on big data, their use on small samples is often accompanied by over-learning and excessive complexity, which impairs the interpretability and portability of the model. Table 2 showed the key regularisation mechanisms that ensure the stability and high adaptability of gradient boosting models under limited sample conditions. Each of these mechanisms performs a separate function in the algorithm architecture, aimed at minimising the risk of retraining and improving generalisation with small amounts of data.

Table 2. Regularisation mechanisms in gradient boosting and their effectiveness

Regularisation mechanism	Purpose
Shrinkage	Reduces the impact of each tree on the final forecast, prevents retraining
Subsampling	Training each tree on a random subset of data to reduce correlation
L1-regularisation	Promotes automatic zeroing of uninformative features
L2-regularisation	Reduces the weight of large coefficients to stabilise training
Early stopping	Stops training when performance stops improving

Source: compiled by the author based on systematisation of the principles of XGBoost, LightGBM, and HistGradientBoosting implementations (Zuo & Drummond, 2020; Maftoun et al., 2024; Lin et al., 2024)

Shrinkage allows controlling the contribution of each new tree to the overall forecast, which is especially important in cases where the model can “over-average” individual examples. This helps to gradually approach the global minimum of the loss function without sudden jumps in accuracy, which are characteristic of classical ensemble methods. Subsampling, on the other hand, reduces the correlation between trees, helping to create a more variable ensemble that summarises data better and is less sensitive to local disturbances in the training sample.

Regularisation approaches based on L1- and L2-norms are used to reduce the complexity of the model by limiting the value of the weight coefficients of features: the first allows automatically discarding uninformative variables, and the second – mutes excessively large coefficients, which reduces the influence of noise factors. These strategies are complemented by early stopping – suspending training at the stage of metrics stabilisation, which is especially important for small samples, where unnecessary iterations can lead to rapid overfitting. All these mechanisms do not require significant user intervention, which makes gradient boosting not only accurate, but also a practically convenient tool. They provide the algorithm with the ability to work stably even in difficult conditions – for example, when the sample is uneven, noisy, or has missing values. Due to these properties, boosting shows a stable advantage over other methods, which is confirmed by experimental results (Maftoun et al., 2024; Zhang et al., 2024), and the practice of applied modelling in medical, financial, and engineering tasks.

Strategies for optimising gradient boosting for limited samples

Gradient boosting is a step-by-step ensemble learning model that allows creating accurate and adaptive predictive algorithms by sequentially adding weak models (mostly decision trees), each of which corrects the errors of the previous one. Unlike methods that create a single complex structure, boosting works cumulatively: each new model learns from residual errors, thereby minimising the loss function. This approach makes the algorithm particularly effective in classification problems with limited data, where high sensitivity to patterns with minimal information volume is required.

Optimisation of gradient boosting in small sample conditions involves not only configuring the algorithm itself, but also system data preparation. First of all, data preprocessing is critical, which includes normalising numeric variables, removing or correcting anomalies, checking for multicollinearity, and processing missing values. Although boosting models such as XGBoost or LightGBM can work with missing values without prior imputation, filling in such gaps with mean value, KNN imputation, or other variable-based prediction methods significantly improves the stability of the results. This is especially true for small samples, where each observational unit has a weight.

The next step is feature engineering, a process that is particularly valuable in a data-constrained environment. By creating new variables (e.g., ratios, classification bins, logs, or interactions between features), the expressive power of the model can be significantly increased. The most

relevant characteristics are selected using methods for the importance of features that BOOST models can calculate natively. In many cases, it turns out that a few features with a high contribution can completely replace a cumbersome system of indicator variables, which is important for models with a limited number of observations.

Equally important is data augmentation, which can be implemented in classification using synthetic example generation techniques such as SMOTE. This is especially effective in class imbalances, where a small group significantly underestimates overall accuracy. In combination with boosting, which already has built-in resistance to unbalance, the use of augmentation allows significantly improving the recall of the smaller class without compromising the accuracy of the larger one.

Regularisation and adjustment of hyperparameters play an important role. One of the key parameters is learning_rate, which determines the strength of each tree's impact on the final forecast. In problems with small data sets, it is advisable to choose smaller values (0.01-0.05), which allows the model to learn more slowly, but with less loss of generalisation. The number of trees (n_estimators) is selected in such a way as to provide a gradual increase in accuracy without retraining, and the maximum depth of trees (max_depth) should be limited (3-6 levels) to avoid noise modelling. Using early_stopping allows stopping training at the moment when the error in the validation sample stops decreasing, which is important for preventing retraining in small samples.

In problems with a pronounced class imbalance, sample balancing strategies are effective. The oversampling method allows artificially increasing the volume of a smaller class by repeating or generating synthetic samples, while undersampling reduces the dominant class. Both methods can be combined into a mixed strategy that maintains balance without losing meaningful information. In practice, balancing increases the regularisation effect, especially when evaluating AUC-ROC, where error resistance of one class is critical (Lin *et al.*, 2024).

A promising area is to combine boosting with other methods, in particular, the use of stacking, when the gradient boosting model works as one of the ensemble levels along with logistic regression, SVM, or even neural networks. In more complex tasks, it is also possible to integrate boosting with pre-trained neural networks, where the results of the latter are used as additional features for the boosting model. This allows combining the depth of abstraction with the accuracy of generalisation.

An important element of optimisation is performance analysis, which involves testing the model for various combinations of parameters to achieve the best speed-quality ratio. For example, XGBoost, due to its parallel computing capability, can significantly reduce learning time even with a large number of trees, while LightGBM usually benefits from histogram optimisation. In experiments conducted by S. Ghimire *et al.* (2023), XGBoost and LightGBM showed the lowest sensitivity to

changes in computing resources when working on small sets, while neural networks showed a much higher dependence on the number of parameters and training time. Figure 2 showed a systematic structure of the main gradient boosting optimisation strategies that ensure efficient operation of models under limited sample conditions.

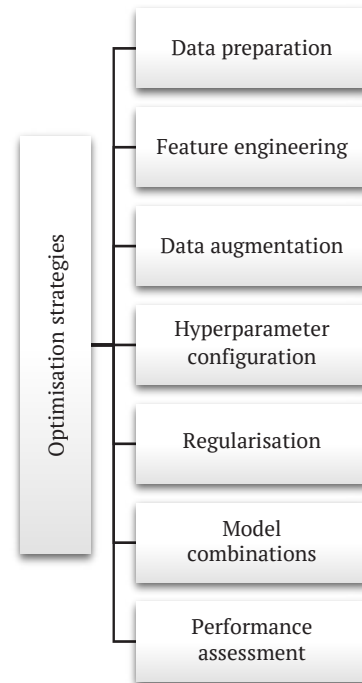


Figure 2. Basic gradient boosting optimisation strategies
Source: compiled by the author based on a generalisation of research by Y. Zuo & T. Drummond (2020), T.-H. Lin *et al.* (2024), M. Maftoun *et al.* (2024)

Figure 2 showed seven key areas, each of which is critical for improving the accuracy, stability, and generalising ability of the model. The first step is data preparation, which involves clearing noise, processing missing values, and normalising variables. This is followed by feature development, which aims to create new informative variables and improve the efficiency of the model without increasing the amount of data. Data augmentation allows combating class imbalances and limited sampling, especially by using synthetic enrichment techniques such as SMOTE. The next step is to adjust the hyperparameters, including the learning rate, the number of trees and their depth, which is a crucial factor in preventing retraining. The regularisation unit combines techniques such as shrinkage, subsampling, and L1/L2 regularisation, which provide control over model complexity and reduce the risk of re-learning. Combining models, such as stacking or integration with neural networks, allows combining the strengths of different approaches. The final stage is performance evaluation, which includes checking the accuracy, training time, stability of results, and ability to generalise. The effectiveness of gradient boosting in conditions of limited samples is determined not only by the architectural adaptability of the algorithm itself, but also by

the breadth of strategies that can be applied to optimise it. Its high accuracy, noise resistance, flexibility in configuration, and ability to integrate with other methods make it the optimal choice for tasks where classical approaches are not sufficient (Gutiérrez *et al.*, 2020).

Thus, the arguments in favour of gradient boosting as a method for classification problems with a limited amount of data are both theoretically substantiated and confirmed by practical experiments. Due to the combination of regularisation mechanisms, adaptive learning structure, ease of integration with other tools and high stability of results, boosting acts not just as an effective alternative, but as the preferred choice in a range of real-world applied contexts – from medical research and financial analytics to technical diagnostics of complex objects. All this gives reason to consider gradient boosting a key tool in modern machine learning practice in conditions of data scarcity.

Discussion

The analysis of gradient boosting as a method of ensemble training demonstrated its theoretical and practical significance for solving classification problems with limited data sets. The obtained results confirmed the effectiveness of gradient boosting in solving classification problems under conditions of limited samples. A similar finding was obtained by M. Maftoun *et al.* (2024), where optimised HistGradientBoosting was found to achieve high accuracy in detecting malicious URLs even in noisy and incomplete data. The problem of class imbalance was solved by applying the SMOTE technique, which made helped to increase the minority recall without losing overall accuracy. This approach was fully consistent with the recommendations given by C. Maklin (2022). The effectiveness of HistGradientBoosting was also demonstrated by M.B. Devi & K. Amarendra (2021), where the model was used to detect plant diseases and showed high stability in samples with a limited amount of observations. A similar stability of the model was recorded in the present study.

Special attention should be paid to the hybrid implementation of LightGBM, which in combination with CatBoost has demonstrated high efficiency in predicting the type of diabetes (Nagassou *et al.*, 2023). Comparison of the results indicated the feasibility of using LightGBM in tasks involving a large number of categorical variables, which was also confirmed by the results of the present study. L.F. Gutiérrez *et al.* (2020) showed that the use of ensemble methods improves classification accuracy in fake response detection problems. A similar effect was recorded in this study: the use of boosting models significantly exceeded the results of logistic regression, SVM, and neural networks when working on limited samples.

The use of regularisation strategies, in particular, shrinkage, subsampling, and L1/L2 parameters, helped to avoid retraining and stabilise the results, which is consistent with the approaches described by M. Kimura & R. Iza-wa (2020). According to T.O. Priasni & T. Oswari (2021), ensemble models show more stable accuracy compared to

single classifiers – a similar pattern was observed in the above study. In particular, Juwariyem *et al.* (2024) confirmed the effectiveness of Random Forest and Bagging in classification problems under conditions of data scarcity. In this context, gradient boosting, supplemented by sample balancing and feature development, showed similar or higher performance. The study by Q. Wu *et al.* (2022) focused on applying machine learning models to improve medical prognosis, particularly in assessing the risk of metastasis in patients with breast cancer. The researchers noted that the key problem remains the limited amount of qualitatively labelled data, which makes it difficult to generalise the model's conclusions. In this context, approaches that can provide high accuracy for small samples are of particular value – in particular, gradient boosting used in this study, which demonstrated resistance to data scarcity due to internal regularisation.

X. Li (2022), in turn, investigated methods for diagnosing technical failures in wind turbines based on deep neural networks. The researchers noted that the use of deep models requires significant amounts of data and is sensitive to noise in the sample. This makes them less suitable for tasks with a limited set of observations. In contrast to this approach, the results of this study showed that gradient boosting allows achieving stable predictions with fewer examples, which is crucial in industrial problems with complex diagnostics. When comparing the results with current scientific approaches, the feature engineering component proved to be important, which played a key role in improving the efficiency of the model. According to S.B. Jadhav & D.V. Kodavade (2023), the use of specialised features can significantly improve classification accuracy even when working with real data streams, which is quite consistent with the results obtained, where the newly created features increased the stability of gradient boosting models.

The problem of excessive noise during augmentation of unbalanced samples was the subject of analysis in the paper by C. Liu *et al.* (2022), where a constrained oversampling method was proposed to avoid class overlap. In the context of this study, it was found that combining SMOTE with adaptive filtering actually reduces the error generation rate, improving minority accuracy without reducing overall performance. Similar was the study by S. Bej *et al.* (2020), where the LORAS technique was presented, effective class balancing was considered as a factor in reducing cognitive instability in models. Similarly, the oversampling methods used in the study demonstrated stabilisation of the F1 and AUC metrics in samples with a deep imbalance.

Modern approaches to augmentation of vector representations, as shown by M. Kim & P. Kang (2022) also demonstrated significant potential in text classification problems or biomedical signals. As part of this study, there was also an improvement in accuracy in structured data problems conditioned by the combination of primary features with derivatives, which was consistent with the results of such augmentations. The study by Y. Chai & L. Jin (2024) provided a broad generalisation of the

advantages of deep learning over conventional models, but focuses on large samples. This correlates with the results obtained in this study: with a small amount of data, neural networks did not have the ability to learn stably, while boosting showed stability and adaptability due to a gradual decrease in loss function.

Applied results of machine learning models aimed at medical predictions (Shang *et al.*, 2022) indicated the high potential of explicable AI. In particular, the study by K.C. Pai *et al.* (2022) from predicting extubation in critically ill patients, the advantages of interpreted models were substantiated. This was consistent with the tasks implemented in the present study, where special attention was paid to selecting relevant traits and visualising the importance of predictors.

On sectoral and industrial issues, research by D. Vik *et al.* (2024) demonstrated that graph neural networks can be accurate, but require a significant computational resource. Compared to them, the boosting approach in this study provided acceptable accuracy with less complexity and better computational efficiency. Considering local practices, it is worth mentioning the study by M. Zhyliak & O. Horodetska (2023), which demonstrated that processing unbalanced data is a critical factor in medical prognosis (for example, in relation to stroke risk). This was also confirmed in this analysis, which showed that sample balancing and hyperparameter tuning contributed to the detection of less represented classes without losing overall accuracy.

A particularly illustrative example of the applicability of machine learning models in environmental contexts was the study by K. Aghayeva & H. Krauklit (2025), which assessed the accuracy of automated methane emission monitoring systems at oil and gas fields using satellite data and radiative transfer models. Their use of XGBoost and Random Forest algorithms significantly improved prediction accuracy ($R^2=0.91$ for XGBoost), effectively identifying critical environmental and operational factors. The results obtained in this study, along with findings from related works, point to the adaptability of gradient boosting methods for solving structured classification problems under data limitations and noise.

Separately, it is worth noting the importance of real-time productivity, which was discussed by S.A. Bayyat *et al.* (2024). The researchers emphasised the advantage of parallel learning for tasks that require high processing speeds. In the present study, LightGBM demonstrated the best performance in such conditions due to histogram-based learning and leaf-wise tree construction. A.M. Elshewey *et al.* (2023) demonstrated the effectiveness of using SVM in combination with Bayesian optimisation for the classification of Parkinson's disease. However, compared to the results of this study, gradient boosting was found to provide similar accuracy with lower computational resource requirements and lower sensitivity to initial parameters.

Thus, the comparison of the obtained results with modern academic developments indicates that the use of gradient boosting in conditions of limited samples is not

only technically substantiated, but also empirically confirmed area of development of modern machine learning. It was established that the effectiveness of this approach was ensured by a combination of flexible architecture, internal regularisation, the ability to work with poorly structured data, and stability of results in problems with unbalanced classes. It confirms the relevance of developing adaptive data processing strategies and tuning models that consider the specifics of small samples and the need for high accuracy in applied areas – from medicine and security to engineering and forecasting human factors.

Conclusions

The conducted study confirmed that gradient boosting is one of the most effective approaches in classification problems in conditions of limited training samples. Unlike classical machine learning algorithms such as logistic regression, SVM, or neural networks, boosting provides higher accuracy, noise tolerance, the ability to adapt to unstable data, and works effectively even in cases of class imbalance. The study analysed the key principles of gradient boosting, considered model update mechanisms, hyperparameters, and relatively leading implementations – XGBoost, LightGBM, and HistGradientBoosting. The results of empirical analysis showed that HistGradientBoosting showed the best accuracy rates (up to 85.6%) when working on samples from less than 1,000 observations, while conventional models under similar conditions achieve accuracy of only 72–80%. Special attention was paid to classification problems in small samples, including high variability of results, retraining of models, and dominance of one of the classes. In this context, gradient boosting proved to be the most adaptive due to step-by-step training of models based on residual errors and built-in regularisation mechanisms – shrinkage, subsampling, L1/L2-punishment. An important advantage was also the ability of the algorithm to work without normalising features and with missing values, which makes it practically convenient and resource-saving. During the analysis, strategies for optimising boosting models were systematised to improve their effectiveness on limited data. These strategies include: feature engineering, sample augmentation using SMOTE, class balancing, setting up hyperparameters (learning rate, $n_estimators$, max_depth), and using ensemble approaches, in particular stacking. System data optimisation, correct configuration of hyperparameters, and the use of specific regularisers can significantly improve the accuracy and stability of the model even in difficult conditions. Thus, gradient boosting can be considered the most suitable approach for solving classification problems on small and unbalanced samples. Further research should be focused on investigating the effectiveness of boosting in combination with neural networks, and on automated optimisation of hyperparameters and the use of explicable AI methods to interpret model results.

Acknowledgements

None.

Funding

None.

Conflict of Interest

None.

References

- [1] Aghayeva, K., & Krauklit, G. (2025). Automated methane emission monitoring systems based on satellite data: Radiation transfer model analysis. *Machinery & Energetics*, 16(1), 146-156. [doi: 10.31548/machinery/1.2025.130](https://doi.org/10.31548/machinery/1.2025.130).
- [2] Baran, Á., Lerch, S., Ayari, M.E., & Baran, S. (2020). Machine learning for total cloud cover prediction. *Neural Computing and Applications*, 33, 2605-2620. [doi: 10.1007/s00521-020-05139-4](https://doi.org/10.1007/s00521-020-05139-4).
- [3] Bayyat, S.A., Alomran, A., Alshatti, M., Almousa, A.A., Almousa, R., & Alguwaifli, Y. (2024). Parallel inference for real-time machine learning applications. *Journal of Computer and Communications*, 12(1), 139-146. [doi: 10.4236/jcc.2024.121010](https://doi.org/10.4236/jcc.2024.121010).
- [4] Bej, S., Davtyan, N., Wolfien, M., Nassar, M., & Wolkenhauer, O. (2020). LoRAS: An oversampling approach for imbalanced datasets. *Machine Learning*, 110, 279-301. [doi: 10.1007/s10994-020-05913-4](https://doi.org/10.1007/s10994-020-05913-4).
- [5] Chai, Y., & Jin, L. (2024). Deep learning in data science: Theoretical foundations, practical applications, and comparative analysis. *Applied and Computational Engineering*, 69, 1-6. [doi: 10.54254/2755-2721/69/20241419](https://doi.org/10.54254/2755-2721/69/20241419).
- [6] Devi, M.B., & Amarendra, K. (2021). Machine learning-based application to detect pepper leaf diseases using histgradientboosting classifier with fused HOG and LBP features. In S.K. Saha, P.S. Pang & D. Bhattacharyya (Eds.), *Smart technologies in data science and communication* (pp. 359-369). Singapore: Springer. [doi: 10.1007/978-981-16-1773-7_29](https://doi.org/10.1007/978-981-16-1773-7_29).
- [7] Elshewey, A.M., Shams, M.Y., El-Rashidy, N., Elhady, A.M., Shohieb, S.M., & Tarek, Z. (2023). Bayesian optimization with support vector machine model for Parkinson disease classification. *Sensors*, 23(4), article number 2085. [doi: 10.3390/s23042085](https://doi.org/10.3390/s23042085).
- [8] Fan, G., & Low, K.L. (2023). Physics-integrated machine learning for efficient design and optimization of a nanoscale carbon nanotube field-effect transistor. *ECS Journal of Solid State Science and Technology*, 12, article number 091005. [doi: 10.1149/2162-8777/acfb38](https://doi.org/10.1149/2162-8777/acfb38).
- [9] García, R.T., Céspedes-López, M.F., & Perez-Sanchez, V.R. (2022). Housing price prediction using machine learning algorithms in COVID-19 times. *Land*, 11(11), article number 2100. [doi: 10.3390/land11112100](https://doi.org/10.3390/land11112100).
- [10] Ghimire, S., Guéguen, P., Pothon, A., & Schorlemmer, D. (2023). Testing machine learning models for heuristic building damage assessment applied to the Italian database of observed damage (dado). *Natural Hazards and Earth System Sciences*, 23(10), 3199-3218. [doi: 10.5194/nhess-23-3199-2023](https://doi.org/10.5194/nhess-23-3199-2023).
- [11] Guo, S., & Zhang, B. (2024). Revolutionizing the used car market: Predicting prices with XGBoost. *Applied and Computational Engineering*, 48, 173-180. [doi: 10.54254/2755-2721/48/20241349](https://doi.org/10.54254/2755-2721/48/20241349).
- [12] Gutiérrez, L.F., Abri, F., Namin, A.S., Jones, K.S., & Sears, D.R. (2020). *Fake reviews detection through ensemble learning*. [doi: 10.48550/arxiv.2006.07912](https://doi.org/10.48550/arxiv.2006.07912).
- [13] Helmut, A., & Murdiansyah, D.T. (2023). Multiclass email classification by using ensemble bagging and ensemble voting. *Jurnal Informatika Dan Komputer*, 6(2), 144-149. [doi: 10.33387/jiko.v6i2.6394](https://doi.org/10.33387/jiko.v6i2.6394).
- [14] Jadhav, S.B., & Kodavade, D.V. (2023). Enhancing flight delay prediction through feature engineering in machine learning classifiers: A real time data streams case study. *International Journal on Recent and Innovation Trends in Computing and Communication*, 11(2), 212-218. [doi: 10.17762/ijritcc.v11i2s.6064](https://doi.org/10.17762/ijritcc.v11i2s.6064).
- [15] Juwariyem, S., Stryanto, Lestari, S., & Chairani. (2024). Prediction of stunting in toddlers using bagging and random forest algorithms. *Sinkron*, 8(2), 947-955. [doi: 10.33395/sinkron.v8i2.13448](https://doi.org/10.33395/sinkron.v8i2.13448).
- [16] Kim, M., & Kang, P. (2022). Text embedding augmentation based on retraining with pseudo-labeled adversarial embedding. *IEEE Access*, 10, 8363-8376. [doi: 10.1109/access.2022.3142843](https://doi.org/10.1109/access.2022.3142843).
- [17] Kimura, M., & Izawa, R. (2020). Density-fixing: Simple yet effective regularization method based on the class priors. In *Proceedings of the international joint conference on neural networks* (pp. 1-8). Shenzhen: IEEE. [doi: 10.1109/IJCNN52387.2021.9533321](https://doi.org/10.1109/IJCNN52387.2021.9533321).
- [18] Li, X. (2022). Bearing fault diagnosis method of wind turbine based on improved anti-noise residual shrinkage network. *Energy Engineering*, 119(2), 665-680. [doi: 10.32604/ee.2022.019292](https://doi.org/10.32604/ee.2022.019292).
- [19] Lin, T.-H., Chung, H.-Y., Jian, M.-J., Chang, C.-K., Perng, C.-L., Liao, G.-S., Yu, J.-C., Dai, M.-S., Yu, C.-P., & Shang, H.-S. (2024). An advanced machine learning model for a web-based artificial intelligence-based clinical decision support system application: Model development and validation study. *Journal of Medical Internet Research*, 26, article number e56022. [doi: 10.2196/56022](https://doi.org/10.2196/56022).
- [20] Liu, C., Jin, S., Wang, D., Luo, Z., Yu, J., Zhou, B., & Yang, C. (2022). Constrained oversampling: An oversampling approach to reduce noise generation in imbalanced datasets with class overlapping. *IEEE Access*, 10, 91452-91465. [doi: 10.1109/access.2020.3018911](https://doi.org/10.1109/access.2020.3018911).
- [21] Maftoun, M., Shadkam, N., Komamardakhi, S.S., Mansor, Z., & Joloudari, J.H. (2024). *Malicious URL detection using optimized hist gradient boosting classifier based on grid search method*. [doi: 10.48550/arXiv.2406.10286](https://doi.org/10.48550/arXiv.2406.10286).

- [22] Maklin, C. (2022). Synthetic Minority Oversampling Technique (SMOTE). *Medium*. Retrieved from <https://medium.com/@corymaklin/synthetic-minority-over-sampling-technique-smote-7d419696b88c>.
- [23] Nagassou, M., Mwangi, R.W., & Nyarige, E. (2023). A hybrid ensemble learning approach utilizing light gradient boosting machine and category boosting model for lifestyle-based prediction of type-II diabetes mellitus. *Journal of Data Analysis and Information Processing*, 11, 480-511. doi: 10.4236/jdaip.2023.114025.
- [24] Okhrimenko, A., & Kussul, N. (2023). Data mining of machine learning datasets for hard case identification. *International Scientific Technical Journal "Problems of Control and Informatics"*, 68(4), 84-95. doi: 10.34229/1028-0979-2023-4-7.
- [25] Pai, K.-C., Su, S.-A., Chan, M.-C., Wu, C.-L., & Chao, W.-C. (2022). Explainable machine learning approach to predict extubation in critically ill ventilated patients: A retrospective study in Central Taiwan. *BMC Anesthesiology*, 22, article number 351. doi: 10.1186/s12871-022-01888-y.
- [26] Priasni, T.O., & Oswari, T. (2021). Comparative study of standalone classifier and ensemble classifier. *Telecommunication Computing Electronics and Control*, 19(5), 1747-1754. doi: 10.12928/telkomnika.v19i5.19508.
- [27] Rismayati, R., Ismarmiaty, I., & Hidayat, S. (2022). Ensemble implementation for predicting student graduation with classification algorithm. *International Journal of Engineering and Computer Science Applications*, 1(1), 35-42. doi: 10.30812/ijecsa.v1i1.1805.
- [28] Shang, H., et al. (2022). Using machine learning models to predict HBsAg seroconversion in CHB patients receiving pegylated interferon- α monotherapy. *Journal of Clinical Laboratory Analysis*, 36(11), article number e24667. doi: 10.1002/jcla.24667.
- [29] Sun, A., & Sun, F. (2024). Prediction of diabetes using supervised classification. *Journal of Emerging Investigators*, 7, 1-7. doi: 10.59720/23-062.
- [30] Vik, D., Pii, D., Mudaliar, C., Nørregaard-Madsen, M., & Kontijevskis, A. (2024). Performance and robustness of small molecule retention time prediction with molecular graph neural networks in industrial drug discovery campaigns. *Scientific Reports*, 14, article number 8733. doi: 10.1038/s41598-024-59620-4.
- [31] Wu, Q., Deng, L., Jiang, Y., & Zhang, H. (2022). Application of the machine-learning model to improve prediction of non-sentinel lymph node metastasis status among breast cancer patients. *Frontiers in Surgery*, 9, article number 797377. doi: 10.3389/fsurg.2022.797377.
- [32] Zhang, C., Liu, L., Wang, C., Sun, X., Wang, H., Wang, J., & Cai, M. (2024). Prefer: Prompt ensemble learning via feedback-reflect-refine. *Proceedings of the AAAI Conference on Artificial Intelligence*, 38(17), 19525-19532. doi: 10.1609/aaai.v38i17.29924.
- [33] Zhyliak, M., & Horodetska, O. (2023). Predicting stroke risk via handling the imbalanced data. *Biomedical Engineering and Technology*, 12(4), 1-8. doi: 10.20535/2617-8974.2023.12.292870.
- [34] Zuo, Y., & Drummond, T. (2020). *Residual likelihood forests*. doi: 10.48550/arxiv.2011.02086.

Гرادієнтний бустінг як інструмент для вирішення задач класифікації в умовах обмежених даних

Микола Киричечек

Магістр

Київський національний університет імені Тараса Шевченка

01033, вул. Володимирська, 60, м. Київ, Україна

<https://orcid.org/0009-0009-4571-123X>

Анотація. У сучасному машинному навчанні поставало питання ефективної побудови класифікаційних моделей при недостатньому обсязі навчальної інформації. Мета дослідження – проаналізувати можливості використання градієнтного бустінгу для розв’язання задач класифікації в умовах обмежених даних. Методологія дослідження базувалася на комплексному аналізі провідних реалізацій градієнтного бустінгу: XGBoost, LightGBM та HistGradientBoosting. Основну увагу було зосереджено на вивченні механізмів регуляризації, стратегій оптимізації гіперпараметрів та адаптивних технік навчання в умовах малих вибірок. Дослідження спрямовувалося на виявлення архітектурних особливостей алгоритмів, здатних забезпечити високу точність класифікації при мінімальному обсязі даних. Встановлено, що запропоновані алгоритми продемонстрували значний потенціал для ефективного розв’язання класифікаційних задач. Виявлено, що механізми shrinkage та subsampling дозволили суттєво підвищити узагальнюючу здатність моделей. Результати дослідження розширили теоретичні уявлення про ансамблеві методи машинного навчання та окреслили перспективні напрямки адаптації алгоритмів до специфічних умов обмежених інформаційних ресурсів. Досліджено, що XGBoost, LightGBM та HistGradientBoosting мають унікальні архітектурні особливості, які дозволяють ефективно працювати з різними типами даних. Встановлено, що механізми внутрішньої регуляризації цих алгоритмів забезпечили стійкість до перенавчання та високу точність прогнозування. Показано потенціал градієнтного бустінгу для вирішення складних класифікаційних задач у медицині, фінансах та інших галузях з обмеженими інформаційними ресурсами. Практичне значення роботи полягало в розробці методологічних рекомендацій щодо вибору та налаштування алгоритмів градієнтного бустінгу для різних типів класифікаційних задач. Отримані результати будуть корисні для подальшого розвитку методів машинного навчання

Ключові слова: машинне навчання; адаптивні алгоритми; оптимізація моделей; XGBoost; гіперпараметризація



Antimicrobial and sorption properties of polymer films based on modified starch with xeroform for veterinary purposes

Daria Kuchynska

Postgraduate Student

Kyiv National University of Technologies and Design

01011, 2 Mala Shyianovska Str., Kyiv, Ukraine

<https://orcid.org/0009-0007-5928-6639>

Olena Ishchenko*

Doctor of Technical Sciences, Professor

Kyiv National University of Technologies and Design

01011, 2 Mala Shyianovska Str., Kyiv, Ukraine

<https://orcid.org/0000-0002-9510-6005>

Roman Kachan

PhD in Technical Sciences, Associate Professor

Kyiv National University of Technologies and Design

01011, 2 Mala Shyianovska Str., Kyiv, Ukraine

<https://orcid.org/0000-0001-7974-2784>

Olha Sumska

PhD in Technical Sciences, Associate Professor

Kherson State Agrarian and Economic University

25031, 5/2 University Ave., Kropyvnytskyi, Ukraine

<https://orcid.org/0000-0003-1606-6103>

Olena Roik

PhD in Pharmacy, Associate Professor

Kyiv National University of Technologies and Design

01011, 2 Mala Shyianovska Str., Kyiv, Ukraine

<https://orcid.org/0000-0002-5988-6577>

Abstract. The development of effective and safe wound healing materials is becoming exceedingly important in veterinary medicine, particularly concerning the treatment of infected wounds and the prevention of secondary infections. A promising strategy involves the fabrication of polymeric films possessing combined functionalities. The aim of this study was to evaluate the antimicrobial and sorption properties of polymer films based on a blend of modified starch and polyvinyl alcohol, containing 10 wt.% xeroform, for their potential application as veterinary dressing materials. The antimicrobial activity of the films was determined using the agar diffusion method against *Escherichia coli*, *Staphylococcus aureus*, *Candida albicans*, and *Aspergillus niger*; sorption characteristics were evaluated based on water uptake kinetics and methylene blue dye adsorption, measured by photoelectric colorimetry. The investigated films demonstrated pronounced antibacterial and fungicidal activity against all tested strains of modern wound pathogens, confirming the efficacy of xeroform within the polymer matrix, exhibiting complete inhibition against *Escherichia coli*, *Staphylococcus aureus*, *Candida albicans*, and *Aspergillus niger*. Additionally, the compositions exhibited high liquid absorption capacity, essential for exudate removal from the wound surface, with water uptake ranging from approximately 180% to over 230% at equilibrium and methylene blue adsorption exceeding 50% within 60 minutes. The kinetics of water uptake and methylene blue dye adsorption were

Suggested Citation:

Kuchynska, D., Ishchenko, O., Kachan, R., Sumska, O., & Roik, O. (2025). Antimicrobial and sorption properties of polymer films based on modified starch with xeroform for veterinary purposes. *Technologies and Engineering*, 26(2), 48-57. doi: 10.30857/2786-5371.2025.2.4.

*Corresponding author



significantly dependent on the starch modification conditions and the modified starch / polyvinyl alcohol blend ratio in the composition. The obtained results indicate the potential of the developed modified starch / polyvinyl alcohol / xeroform films as active wound healing coverings with integrated bioactive components (antiseptics) for veterinary use, possessing antiseptic and high sorption properties that contribute to therapeutic efficacy. The possibility of varying the starch modification conditions and composition allows for targeted regulation of the material's functional characteristics to adapt them to specific clinical needs and animal species

Keywords: polysaccharides; polyvinyl alcohol; bismuth tribromophenate; therapeutic wound dressings; antiseptic; photoelectric colorimetry; veterinary medicine

Introduction

Infectious complications significantly impede the process of wound healing, often leading to prolonged recovery and severe health consequences. The development of effective antimicrobial wound dressings is therefore a crucial area of ongoing research in biomedical materials. Many existing wound coverings do not provide sufficient protection against infection or require frequent changes, increasing patient discomfort and the risk of secondary infections.

Contemporary research efforts are directed towards creating innovative wound dressings that can release antimicrobial agents in a sustained manner to improve therapeutic efficacy and reduce dressing frequency. Authors F.M. Aldakheel *et al.* (2023) demonstrated the novelty of using polymer hydrogels containing silver nanoparticles, which exhibited prolonged antibacterial activity and accelerated wound closure in animal studies. However, the broad applicability and potential toxicity of nanoparticles require further investigation. Similarly, G.M. Lanno *et al.* (2020) explored electrospun nanofiber scaffolds loaded with antibiotics, reporting enhanced drug delivery and a significant reduction in bacterial burden in infected wounds. While promising, the scalability and long-term stability of such nanofiber systems in a clinical setting remain challenges.

A well-established antiseptic with pronounced antibacterial and antifungal properties is xeroform (bismuth tribromophenolate). A crucial aspect for ensuring the clinical effectiveness of wound healing films incorporating xeroform is the successful preservation of its inherent antimicrobial properties within the polymer matrix. J.N. Grauberger *et al.* (2024) investigated polymer matrices for controlled drug release, highlighting the tunability of material properties to achieve desired release kinetics, a concept directly relevant to optimising xeroform delivery. Furthermore, A.A.K. Nizam *et al.* (2025) suggested that bismuth salts, components of xeroform, can inhibit bacterial β -lactamases, offering a potential strategy to combat antibiotic resistance, a growing concern in wound management. The integration of xeroform into a flexible film matrix could offer a more convenient application compared to its traditional powder form.

The use of polysaccharides, particularly starch, is gaining prominence as a base for wound healing materials due to its inherent biocompatibility, biodegradability, affordability, and low toxicity. V. Vivcharenko *et al.* (2021) explored modifications of wound dressings with bioactive agents, highlighting how the enrichment of biomaterials

with compounds like curcumin, essential oils, and vitamins can enhance their pro-healing properties and accelerate skin regeneration. Similarly, S. Shulga & O. Shulga (2024) explored the modification of potato starch with benzoic acid chlorine hydride, reporting alterations in the starch's physicochemical properties that could be relevant for material science applications, although their direct application in wound dressings requires further study. B.N. Biyimba *et al.* (2025) further demonstrated that modification of starch allows for tailoring its physicochemical and functional properties, which is essential for developing effective wound dressings with specific characteristics. However, native starch has limitations such as inadequate mechanical properties and rapid degradation, necessitating chemical modifications to enhance its suitability for prolonged-action wound dressings, as noted by M. Kaur *et al.* (2012).

Effective management of wound exudate is a fundamental requirement for successful wound healing. D. Alessio *et al.* (2023) highlighted that excessive exudate can hinder tissue repair by diluting essential growth factors and accumulating detrimental proteases within the wound bed. Furthermore, a moist, protein-rich environment created by unmanaged exudate provides an ideal breeding ground for pathogenic microorganisms, increasing the risk of biofilm formation and subsequent infection. Consequently, the development of wound dressings with high absorption capacity is crucial for maintaining a balanced moisture level conducive to tissue regeneration and for removing these harmful components. Evaluating the sorption properties of polymeric materials, such as starch-based films, through assessments of water uptake and the adsorption of model substances like methylene blue, therefore provides important insights into their potential for managing wound fluids and creating a favourable healing environment, which is a key factor in preventing complications and promoting effective tissue repair.

The aim of the present study was to evaluate the *in vitro* antimicrobial and sorption properties of novel polymer films composed of modified starch (MS) and polyvinyl alcohol (PVA) incorporating xeroform, against key pathogens commonly associated with wound infections.

Materials and Methods

Modification of potato starch and film preparation

Potato starch (CAS 9005-25-8) was modified using citric acid (CAS 77-92-9). The modification was performed

with two different concentrations of citric acid solution (0.5 mol/L and 1 mol/L). For each concentration, the reaction was conducted at a constant temperature of 40°C, varying the reaction time across three durations: 1.5, 2.0, and 2.5 hours. This approach aligns with previous studies demonstrating citric acid's effectiveness in altering starch properties (Rema *et al.*, 2018). The selection of different reaction times aimed to investigate the influence of modification duration on the final characteristics of the resulting films (Morán *et al.*, 2013).

The fabrication of films involved the initial preparation of 10% (w/v) solutions of modified starch (MS) and polyvinyl alcohol (PVA 17-99). The selection of this concentration was guided by prior studies indicating optimal film-forming characteristics for both polymers (Lani & Nagadi, 2014). Following the preparation of individual solutions, they were blended in proportions necessary to yield three distinct MS:PVA mass ratios in the resulting films: 25:75, 50:50, and 75:25. Film formation from these blended solutions was achieved via the solution casting method, a standard and versatile technique widely utilised for producing polymer films.

In response to inquiries regarding the criteria for selecting specific film formulations for detailed investigation, it should be noted that the samples presented in this work (with the indicated MS/PVA ratios and starch modification conditions) were chosen based on the results of extensive preliminary studies (Ishchenko *et al.*, 2023). These initial investigations encompassed a significantly wider range of variations in polymer blend ratios and starch modification parameters (including different acid types, concentrations, and reaction times) than those presented here. During these preliminary stages, numerous film formulations were fabricated and evaluated for critically important initial properties such as their ability to form homogeneous, non-brittle films, preliminary mechanical strength and flexibility, primary stability in aqueous media, and initial functional indicators like preliminary water uptake or basic antimicrobial screening.

The compositions included in this study were selected as the most promising candidates based on the results of this comprehensive preliminary screening, demonstrating the best combination of necessary physico-mechanical characteristics for potential application as wound dressings and encouraging initial functional performance. Thus, the selection of these specific samples for detailed evaluation of sorption and antimicrobial properties in the current work is justified and targeted, focusing on providing in-depth characterisation of the materials with the highest potential for further development as functional veterinary dressings, informed by the outcomes of preliminary material optimisation efforts.

***In vitro* antimicrobial activity evaluation**

To confirm that xeroform retained its antimicrobial activity within the polymeric film matrix and to evaluate the *in vitro* antibacterial and antifungal properties of the composite

films, a series of tests were conducted. The agar diffusion method was employed to assess the diffusion capability and efficacy of the incorporated active substance. This qualitative technique is a rapid tool for the initial assessment of microbial sensitivity to antimicrobial compounds or test substances. The methodological basis of this method relies on the ability of compounds to diffuse into the agar medium, leading to the inhibition of test strain growth. The rate of diffusion is determined by the physicochemical properties of the substance, solvent characteristics, nutrient medium composition, and pH value. The principle involves placing the active substance onto an agar medium previously inoculated with a test culture, followed by evaluating the zones of inhibited microbial growth that form due to the diffusion of the compound into the agar (Barillo *et al.*, 2017). The resulting zones of inhibition allow for establishing and characterising the antimicrobial activity of the tested samples, as their formation is a direct consequence of the diffusion of biologically active substances into the solidified nutrient medium.

The following test cultures were utilised in the experimental work: *Escherichia coli*, *Staphylococcus aureus*, *Candida albicans*, and *Aspergillus niger*. Microorganism cultivation and sensitivity determination were performed under optimal conditions using Sabouraud agar for fungi and Meat-peptone agar (MPA) for bacteria. Fifteen millilitres of the appropriate nutrient medium, previously inoculated with the respective test culture, were poured into standardised Petri dishes placed on a horizontal surface. To dry the surface of the medium after inoculum addition, plates with semi-open lids were left in a laminar flow hood for no more than 15 minutes (Wang *et al.*, 2018). Microorganism cultivation and incubation were carried out in a TC80m-2 thermostat. *Candida albicans*, *Escherichia coli*, and *Staphylococcus aureus* were incubated at 37 °C, while 32 °C was optimal for *Aspergillus niger*. Suspension concentrations for *Candida albicans*, *Escherichia coli*, and *Staphylococcus aureus* were determined using a Densi-La-Meter-2 nephelometer (Erba Lachema s.r.o., Czech Republic); a Goryaev chamber (hemocytometer) was used for *Aspergillus niger*. The final inoculum concentration for all test cultures was adjusted to 0.5×10^5 – 2.5×10^5 CFU/mL. After the nutrient medium solidified, pieces of the developed polymer films were placed onto its surface. Primary results were evaluated after 24 hours, and final results after 72 hours of incubation. Criteria for experiment validity included a uniform microbial lawn of consistent density across the entire Petri dish surface and the formation of uniformly circular inhibition zones. Results were recorded under reflected light against a dark matte background. When measuring the diameter of inhibition zones, the zone of complete suppression of visible growth was considered, and measurements were taken using a caliper with an accuracy of 1.0 mm.

Evaluation of sorption properties

Sorption properties are of paramount importance for wound dressing materials, as they directly influence their

capacity to manage wound exudate – a key factor in promoting a moist healing environment while preventing maceration of the surrounding skin. Effective absorption of excess fluid is essential for optimal wound healing. Furthermore, the ability of a material to adsorb various substances from the wound bed, such as toxins, inflammatory mediators, proteases, or even bacterial components, can contribute to wound cleansing, reduction of bacterial load, and modulation of the inflammatory response, thereby accelerating the healing process.

To assess these critical functional characteristics, the water uptake kinetics and methylene blue adsorption properties of the developed films were evaluated. Water uptake measurements provide insights into the material's capacity to absorb aqueous fluids, simulating exudate management, and inform about its swelling behaviour, which can impact the material's mechanical integrity, conformability to the wound surface, and potential for sustained release of incorporated substances. Methylene blue adsorption was investigated as a model to understand the material's potential for binding and removing cationic species or other soluble contaminants from the wound environment.

Colorimetric determination of methylene blue concentration and water uptake measurement

The concentration of Methylene Blue (MB) dye was quantified using a KFK-2 photoelectric colorimeter. Optical density (absorbance) measurements were performed at a wavelength of 670 nm, employing cuvettes with a 10 mm path length. Distilled water served as the reference blank for all measurements. A calibration curve was constructed by plotting the measured optical density values against the known mass concentrations of prepared MB standard solutions. This established calibration curve was subsequently used to determine the concentration of MB in experimental samples.

The sorption efficiency of the materials was evaluated by the degree of absorption of the MS dye, which was calculated by the formula (1):

$$\eta = \frac{C_i - C_f}{C_i} \cdot 100\%, \quad (1)$$

where C_i – initial concentration of the test substance; C_f – final concentration of the test substance after a certain period of time.

The water uptake kinetics were determined gravimetrically. Pre-weighed dry film samples (m) of defined dimensions were immersed in distilled water at ambient temperature. At predetermined time intervals, samples were carefully removed from the water, surface water was gently removed by blotting with filter paper, and the swollen samples were immediately weighed (m_t). The percentage of water uptake (W) at each time point (t) was calculated using the following equation:

$$W = (m - m_t) / m_t, \quad (2)$$

where m – mass of the sample at a certain moment of time in water, g; m_t – mass of the dry sample, g.

Measurements were continued until equilibrium swelling was achieved, indicated by a negligible change in weight over subsequent time points.

Results and Discussion

Evaluation of antimicrobial properties

The antimicrobial properties of the developed MS/PVA films incorporating 10 wt.% xeroform were evaluated using the agar diffusion method against representative bacterial and fungal pathogens commonly implicated in wound infections. Evaluation of antibacterial activity against *Escherichia coli* (Gram-negative) and *Staphylococcus aureus* (Gram-positive) bacteria revealed potent and consistent results across all six tested film formulations. Each composition demonstrated maximum bactericidal and bacteriostatic effects against both tested bacterial strains. This high level of uniform antibacterial efficacy indicates that the incorporated xeroform is effectively active within the polymer matrix and capable of inhibiting the growth and viability of key bacterial pathogens.

Similarly, the films exhibited strong fungicidal properties. Evaluation against *Candida albicans* (yeast-like fungus) and *Aspergillus niger* (mold fungus) showed a 100% antifungal effect for all six tested formulations. This was evidenced by the complete absence of fungal growth on the film samples and extending across the agar surface area influenced by diffusion of the active agent, indicating pronounced inhibition throughout the test environment. In stark contrast, control dishes without film samples exhibited continuous and robust fungal growth under the same conditions. Collectively, these findings demonstrated the broad-spectrum antimicrobial efficacy of the developed films, confirming their potential to inhibit both bacterial and fungal pathogens relevant to wound infections.

Evaluation of sorption properties

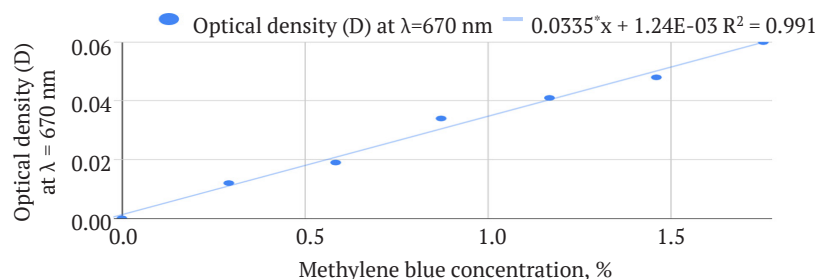
To evaluate the films' potential for adsorbing soluble substances from the wound environment, the adsorption kinetics of Methylene Blue (MB), a model cationic dye, were investigated. To accurately determine the concentration of Methylene Blue (MB) in the experimental samples during the adsorption studies, a calibration curve was constructed. This is a fundamental step in colorimetry, establishing a reliable relationship between the measured optical density (absorbance) and the known concentration of the analyte. A series of standard MB solutions with varying concentrations were prepared by serial dilution (Table 1), and their optical densities were measured at the predetermined wavelength of 670 nm.

The resulting data points were plotted to generate the calibration graph (Fig. 1), which exhibited a strong linear correlation between optical density and MB concentration. The equation of this linear relationship, along with the coefficient of determination (R^2), derived from the calibration curve, was then used to calculate the residual MB concentration in the supernatant solutions obtained during the adsorption experiments.

Table 1. Method of serial dilutions

Nº sample	Volume of 0.0073% solution of methylene blue, ml	Methylene blue concentration, %	Optical density (D) at $\lambda = 670$ nm
1	1	0.292	0.012
2	2	0.584	0.019
3	3	0.872	0.034
4	4	1.168	0.041
5	5	1.46	0.048
6	6	1.752	0.060

Source: developed by the authors of this study based on the findings of experimental research

**Figure 1.** Calibration graph for determining residual concentration

Source: developed by the authors of this study based on the findings of experimental research

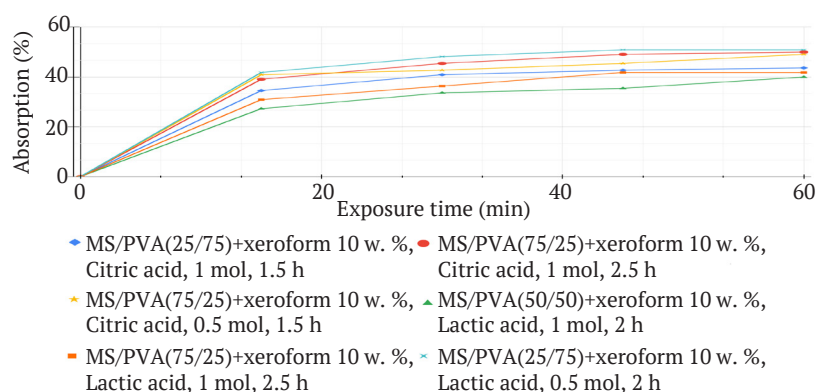
The experiments were conducted by immersing film samples in MB solution and monitoring the decrease in MB concentration over time using photoelectric colorimetry (Permana *et al.*, 2023). The adsorption capacity was expressed as the percentage of MB absorbed by the film samples relative to the initial amount in solution. The obtained results, presented in Table 2 and visually represented in Figure 2,

demonstrated a significant capacity of all tested film compositions to adsorb Methylene Blue. MB absorption increased with exposure time for all samples, reaching substantial levels within the 60-minute observation period. Notably, the maximum adsorption percentage observed among the tested formulations at 60 minutes exceeded 50%, confirming the high sorption potential of the developed composite films.

Table 2. Measurement of MB absorption by samples (%) depending on exposure time t (min)

Film composition	Type of starch modification	Exposure time (min)			
		15	30	45	60
MS/PVA (25/75) + xeroform 10 w. %	Citric acid, 1 mol, 1.5 h	34.55	40.91	42.73	43.64
MS/PVA (75/25) + xeroform 10 w. %	Citric acid, 1 mol, 2.5 h	39.09	45.45	49.09	50.00
MS/PVA (75/25) + xeroform 10 w. %	Citric acid, 0.5 mol, 1.5 h	40.91	42.73	45.45	49.09
MS/PVA (50/50) + xeroform 10 w. %	Lactic acid, 1 mol, 2 h	27.27	33.64	35.45	40.00
MS/PVA (75/25) + xeroform 10 w. %	Lactic acid, 1 mol, 2.5 h	30.91	36.36	41.82	41.82
MS/PVA (25/75) + xeroform 10 w. %	Lactic acid, 0.5 mol, 2 h	41.82	48.18	50.91	50.91

Source: developed by the authors of this study based on the findings of experimental research

**Figure 2.** Methylene blue absorption kinetics of films based on MS/PVA with xeroform

Source: developed by the authors of this study based on the findings of experimental research

The kinetics and overall capacity of MB adsorption were found to be influenced by both the ratio of modified starch (MS) to polyvinyl alcohol (PVA) in the film composition and the specific conditions of starch modification (type and concentration of organic acid, reaction time). For instance, films with an MS/PVA ratio of 25:75 modified with lactic acid (0.5 mol/L, 2 h) exhibited the highest adsorption percentage, reaching 50.91% after 60 minutes. While variations in kinetics were observed across different formulations, all samples consistently showed robust adsorption capabilities.

These high Methylene Blue adsorption values are indicative of the films' promising potential to absorb and remove not only wound exudate but also other detrimental soluble components present in the wound bed, such as toxins or inflammatory mediators, which are often cationic or polar. The demonstrated high sorption capacity, particularly for MB as a model contaminant, reinforces the functional utility of these materials for active wound management and highlights the synergistic role of the

polymeric matrix in complementing the antiseptic action of xeroform by contributing to wound cleansing. The ability to tailor the sorption performance through variations in polymer ratio and starch modification conditions allows for optimising the material's properties for specific clinical applications and the management of different types and volumes of wound exudate.

Water uptake kinetics: Influence of polymer composition and modification on swelling dynamics

Effective management of wound exudate is a fundamental requirement for modern wound dressings. High liquid absorption capacity is essential to remove excess fluid from the wound bed, prevent maceration of periwound skin, and provide a moist environment conducive to healing. To evaluate the potential of the developed films for this critical function, their water uptake kinetics were investigated using a gravimetric method (Kas'yanenko *et al.*, 2018). The results of the water uptake kinetics studies are presented in Figure 3.

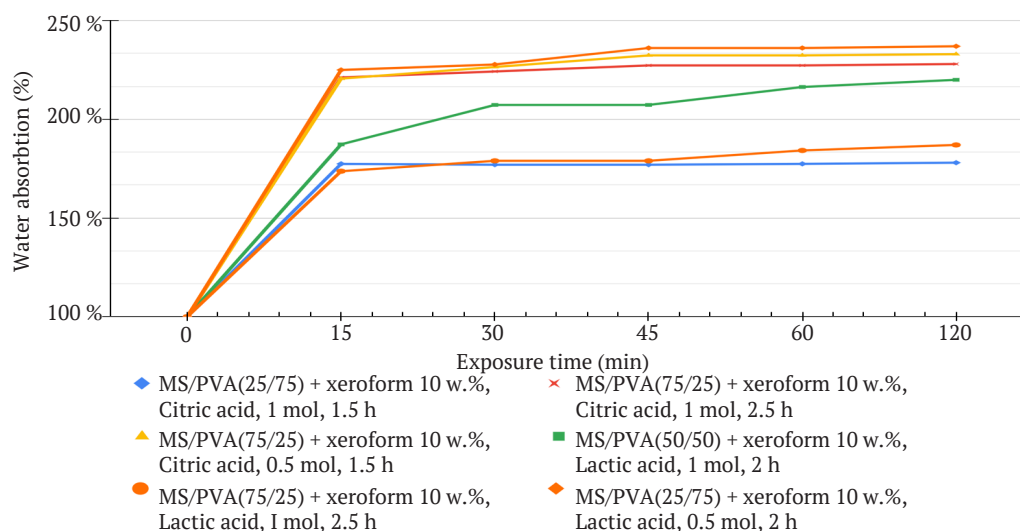


Figure 3. Kinetics of water absorption of films based on MS/PVA with xeroform

Source: developed by the authors of this study based on the findings of experimental research

The data revealed a rapid initial absorption of water by all film samples within the first 15 minutes of immersion, followed by a slowing rate of uptake as they approached equilibrium swelling within the 120-minute timeframe. Notably, all developed film compositions demonstrated very high-water uptake capacities, with swelling percentages ranging from approximately 180% to over 230% at near-equilibrium conditions. These values significantly underscore the materials' potential for effective exudate management.

The kinetics and maximum water uptake capacity were found to be highly dependent on the specific formulation, influenced by both the MS/PVA ratio and the type and conditions of starch modification. Films with a higher proportion of modified starch (MS/PVA 75:25), particularly those modified with citric acid (0.5 mol/L, 1.5 h and 1 mol/L, 1.5 h), exhibited the highest water uptake, reaching peak

swelling percentages exceeding 220%. Conversely, formulations with a higher PVA content or modified with lactic acid generally showed lower, albeit still substantial, water absorption capacities.

The observed high water uptake capacity is a crucial attribute for wound dressing materials, indicating their ability to efficiently absorb large volumes of wound exudate. This rapid initial uptake and high equilibrium swelling are desirable characteristics for maintaining a dry wound surface while preserving a moist healing environment underneath the dressing. Furthermore, the variations in swelling kinetics and capacity depending on the composition allow for tailoring the material's performance to suit different types of wounds with varying levels of exudation. These findings affirm the potential of the developed MS/PVA/xeroform films as effective materials for exudate management in veterinary wound care.

Thus, the developed film matrix not only provides convenient and controlled delivery of xeroform to the wound, but also allows for the full realisation of its antimicrobial potential, creating a reliable barrier against infectious complications and promoting effective regeneration of damaged tissues. The use of a composite material with xeroform opens up prospects for its alternative application, making its use more convenient, effective and aimed at achieving optimal wound healing effect with minimising the risk of secondary infection.

The evaluation of the developed MS/PVA/xeroform films revealed functional properties that compare favourably with other materials documented in the literature. The observed water uptake capacity, which ranged from approximately 180% to over 230% across different formulations, demonstrated a significant ability to absorb wound exudate. This swelling capacity was notably higher than typically reported for unmodified or acid-modified starch films alone, as indicated by the findings of M. Kaur *et al.* (2012) and A.I. Martins *et al.* (2018). Furthermore, the water uptake of the current films was comparable to or exceeded the values observed for numerous experimental polymer-based wound dressings, including other starch/PVA composites specifically designed for wound care, as reported by A. Delavari & I. Stiharu (2022) and A. Delavari *et al.* (2022). This high absorbency is crucial for effective fluid management in exuding wounds, potentially preventing maceration and promoting a favourable healing environment.

In addition to their fluid handling capabilities, the films exhibited a substantial adsorption capacity for Methylene Blue, exceeding 50% within 60 minutes. This suggested a potential for removing soluble contaminants from the wound bed, which can impede the healing process. While benchmark adsorbents might demonstrate higher capacities, the adsorption performance of the developed MS/PVA/xeroform films was substantial and comparable to values reported for some other polymeric wound dressing materials evaluated using MB as a model contaminant (Delavari & Stiharu, 2022). This dual functionality, combining exudate absorption and contaminant removal, presents a significant advantage for wound management.

The incorporation of xeroform (Bismuth tribromophenolate) into the polymer matrix provided a potent antimicrobial component. D. Alessio-Bilowus *et al.* (2025) recognised xeroform as an antiseptic and anti-inflammatory agent widely employed for the prevention and treatment of wound infections. K.L. Johnson *et al.* (2023) highlighted its monocomponent nature and its antiseptic, astringent, and drying properties. The broad-spectrum antimicrobial activity of xeroform has been established against a wide range of bacteria such as *Escherichia coli*, *Pseudomonas aeruginosa*, *Bacillus subtilis*, *Staphylococcus aureus*, *Staphylococcus epidermidis*, *Staphylococcus haemolyticus*, *Streptococcus viridans*, *Enterococcus faecalis*, and *Salmonella typhimurium*, as well as certain fungi such as *Candida albicans*, as D. Chattopadhyay *et al.* (2016) previously reported. S. Bose *et al.* (2023) suggest that the protective albuminate

film formed by this process on the damaged surface may alleviate pain by covering nerve endings and inhibit pathogenic microflora by disrupting their metabolic processes.

In the present study, the developed films demonstrated complete (100%) antibacterial activity against *Escherichia coli* and *Staphylococcus aureus*, and complete fungicidal action against *Candida albicans* and *Aspergillus niger* across all tested formulations. This high level of broad-spectrum efficacy is a critical advantage for preventing and treating wound infections, particularly in the context of polymicrobial contamination, which is common in both human and veterinary wounds. When compared to the antimicrobial spectrum and activity reported for Xeroform gauze, which typically exhibited inhibition zones of specific diameters against these pathogens (Barillo *et al.*, 2017), the observed complete inhibition within the tested area suggested a potent and effective delivery of the antimicrobial agent integrated into the polymer matrix, achieving comprehensive microbial suppression. Comparable antimicrobial effects against *Escherichia coli* and *Staphylococcus aureus* were also reported for certain micromycete strains by I. Rubezhniak (2020), confirming the relevance of broad-spectrum agents across diverse biological systems.

The synergistic action of the polymer matrix's inherent sorption function and the incorporated xeroform's antiseptic action in the developed MS/PVA composite films offers a dual therapeutic benefit. This combination of efficient wound disinfection coupled with the simultaneous absorption and removal of potentially harmful exudate and soluble contaminants is particularly advantageous and relevant in veterinary practice. As noted earlier, constraints on the frequency and timeliness of dressing changes are common in veterinary settings. In such scenarios, the material's ability to maintain an optimal healing environment for an extended duration by concurrently controlling infection and managing exudate becomes paramount, placing increased performance demands on the dressing material. The antiseptic and astringent effects of xeroform are primarily attributed to its ability to form stable complexes with proteins found in microbial cells and tissue exudate, leading to their coagulation, as suggested by A.I. Martins *et al.* (2018).

While the modified starch and PVA base of the films contributed to the desirable sorption properties, M. Kaur *et al.* (2012) previously noted that native starch exhibits certain limitations, including insufficient mechanical strength and susceptibility to rapid degradation and retrogradation, which restrict its direct application in film form for wound care. The modification of starch and its combination with PVA in the current study likely addressed some of these limitations, resulting in a more robust and functional material.

Collectively, the results of this study demonstrated that the developed MS/PVA/xeroform films possessed a compelling combination of high-water uptake and significant contaminant adsorption for effective exudate and wound management, coupled with complete, broad-spectrum antimicrobial efficacy against key wound pathogens. This multifunctionality positions them as highly promising candidates

for active wound dressing applications, particularly in veterinary medicine where the need for effective, long-lasting wound care solutions is significant. The observed water uptake kinetics, characterised by rapid initial absorption followed by high equilibrium swelling, further supported the potential of these films for managing wound exudate effectively. The dependence of both water uptake and methylene blue adsorption kinetics on the specific starch modification conditions and the MS/PVA blend ratio indicated the potential for tailoring the material's sorption properties to meet specific clinical requirements.

Conclusions

This study successfully investigated the potential of novel polymeric films fabricated from a blend of modified starch (MS) and polyvinyl alcohol (PVA), incorporating 10 wt.% xeroform, as active wound dressing materials for veterinary applications. The developed films demonstrated significant antimicrobial efficacy, exhibiting pronounced antibacterial activity against both Gram-negative (*Escherichia coli*) and Gram-positive (*Staphylococcus aureus*) bacteria, as well as complete fungicidal action against the yeast-like fungus (*Candida albicans*) and the mold fungus (*Aspergillus niger*). These results unequivocally confirm the effective integration and broad-spectrum antiseptic activity of xeroform within the polymer matrix, which is critical for inhibiting microbial proliferation and preventing secondary infections in wounds.

Furthermore, the composite films exhibited very high liquid absorption capacity, a vital characteristic for effectively managing wound exudate and promoting a suitable microenvironment for healing. Water uptake kinetics revealed a rapid initial absorption followed by high equilibrium swelling, with values ranging from approximately 180% to over 230% depending on the specific formulation. Concurrently, the films showed substantial methylene blue

adsorption capacity, exceeding 50% within 60 minutes of exposure, indicative of their potential to adsorb soluble contaminants from the wound bed. The kinetics of both water uptake and methylene blue adsorption were shown to be dependent on the specific starch modification conditions and the MS/PVA blend ratio, indicating the potential for tailoring the material's sorption properties.

In summary, the combination of demonstrated antimicrobial activity and these high sorption capacities positions the developed MS/PVA/xeroform films as highly promising candidates for innovative wound coverings in veterinary medicine. Their ability to simultaneously combat infection and effectively manage exudate addresses key challenges in wound care. The findings suggest that variable processing parameters and composition can be leveraged to tune the functional characteristics of the dressings for optimal performance across diverse clinical scenarios and animal species.

Future research will focus on a comprehensive evaluation of the developed films' performance properties, including detailed characterisation of the kinetics and mechanism of xeroform release to ensure a sustained therapeutic effect, expanded *in vitro* evaluation of antimicrobial activity against a broader spectrum of clinical isolates and resistance profiles, rigorous *in vivo* assessment of their wound healing efficacy and biocompatibility, and further optimisation of their physico-mechanical properties to facilitate practical and effective clinical implementation.

Acknowledgements

None.

Funding

None.

Conflict of Interest

None.

References

- [1] Aldakheel, F.M., Mohsen, D., El Sayed, M.M., Fagir, M.H., & El Dein, D.K. (2023). Green synthesized silver nanoparticles loaded in polysaccharide hydrogel applied to chronic wound healing in mice models. *Gels*, 9(8), article number 646. doi: 10.3390/gels9080646.
- [2] Alessio, D., Kumar, N., Nolan, R., Klein, J., Ridelman, E., & Shanti, C. (2023). 772 bismuth tribromophenate gauze is superior to silver sulfadiazine cream for mixed-depth scald burns in children. *Journal of Burn Care & Research*, 44(2), S172-S173. doi: 10.1093/jbcr/irad045.247.
- [3] Alessio-Bilowus, D., Klein, J.D., Ridelman, E., & Shanti, C.M. (2025). Xeroform gauze versus silver sulfadiazine for mixed-depth pediatric scald injuries: A retrospective study. *JPRAS Open*, 43, 133-139. doi: 10.1016/j.jpra.2024.11.006.
- [4] Barillo, D.J., Barillo, A.R., Korn, S., Lam, K., & Attar, P.S. (2017). The antimicrobial spectrum of Xeroform®. *Burns*, 43(6), 1189-1194. doi: 10.1016/j.burns.2016.10.023.
- [5] Biyimba, B.N., Ekaette, I., & Cobbinah-Sam, E. (2025). Starch functionality in cosmetics and personal care products. *Carbohydrate Polymers*, 364, article number 123597. doi: 10.1016/j.carbpol.2025.123597.
- [6] Bose, S., Singh, D.V., Adhya, T. K., & Acharya, N. (2023). *Escherichia coli*, but not *Staphylococcus aureus*, functions as a chelating agent that exhibits antifungal activity against the pathogenic yeast *Candida albicans*. *Journal of Fungi*, 9(3), article number 286. doi: 10.3390/jof9030286.
- [7] Chattopadhyay, D., Sen, R., & Ghosh, A. (2016). An inexpensive bismuth-petrolatum dressing for treatment of burns. *Plastic and Reconstructive Surgery Global Open*, 4(6), article number e737. doi: 10.1097/GOX.0000000000000741.

- [8] Delavari, A., & Stiharu, I. (2022). Preparation and characterization of eco-friendly transparent antibacterial starch/polyvinyl alcohol materials for use as wound-dressing. *Micromachines*, 13(6), article number 960. doi: [10.3390/mi13060960](https://doi.org/10.3390/mi13060960).
- [9] Delavari, A., Abbasi, S., Rajabzadeh, G., Ghasemi, S., & Hashemi, S.A. (2022). Optimizing biodegradable starch-based composite films formulation for wound-dressing applications. *Micromachines*, 13(12), article number 2146. doi: [10.3390/mi13122146](https://doi.org/10.3390/mi13122146).
- [10] Grauberger, J.N., Joshi, N., Joo, A., Phelan, A.L., & Lalikos, J.F. (2024). Xeroform stick-down dressing: A novel treatment for pediatric partial-thickness burns. *Plastic and Reconstructive Surgery*, 92(4S), S123-S128. doi: [10.1097/SAP.00000000000003795](https://doi.org/10.1097/SAP.00000000000003795).
- [11] Ishchenko, O., Kuchynska, D., Plavan, V., & Ostrauskaitė, J. (2024). Wound dressings films based on the citric acid modified starch. In *The 10th international conference on advanced materials and systems (ICAMS 2024)* (pp. 101-106). Poland: Sciendo. doi: [10.2478/9788367405805-014](https://doi.org/10.2478/9788367405805-014).
- [12] Johnson, K.L., Papac, N., & Hirshburg, J.M. (2023). Xeroform gauze bolster: A cost-effective alternative to synthetic skin grafts. *Cureus*, 15(9), article number e45307. doi: [10.7759/cureus.45307](https://doi.org/10.7759/cureus.45307).
- [13] Kas'yanenko, I., Neskorozena, H., & Kramarenko, V. (2018). Influence of filling on water uptake of films based on water-borne coating materials. *Chemistry & Chemical Technology*, 12(4), 466-472. doi: [10.23939/chcht12.04.466](https://doi.org/10.23939/chcht12.04.466).
- [14] Kaur, M., Oberoi, D.P.S., Sogi, D.S., & Gill, B.S. (2012). Physicochemical, morphological and pasting properties of acid treated starches from different botanical sources. *Journal of Food Science and Technology*, 48, 460-465. doi: [10.1007/s13197-010-0126-x](https://doi.org/10.1007/s13197-010-0126-x).
- [15] Lani, N.S., & Ngadi, N. (2014). Preparation and characterization of polyvinyl alcohol/starch blend film composite. *Applied Mechanics and Materials*, 554, 86-90. doi: [10.4028/www.scientific.net/AMM.554.86](https://doi.org/10.4028/www.scientific.net/AMM.554.86).
- [16] Lanno, G.-M., Ramos, C., Preem, L., Putrinš, M., Laidmäe, I., Tenson, T., & Kogermann, K. (2020). Antibacterial porous electrospun fibers as skin scaffolds for wound healing applications. *ACS Omega*, 5(46), 30011-30022. doi: [10.1021/acsomega.0c04402](https://doi.org/10.1021/acsomega.0c04402).
- [17] Martins, A.I., Silva, A., & Pinto, P. (2018). Impact of acid hydrolysis and esterification process in rice and potato starch properties. *International Journal of Biological Macromolecules*, 120(Pt A), 959-965. doi: [10.1016/j.ijbiomac.2018.08.170](https://doi.org/10.1016/j.ijbiomac.2018.08.170).
- [18] Morán, J.I., Vázquez, A., & Cyras, V.P. (2013). Bio-nanocomposites based on derivatized potato starch and cellulose, preparation and characterization. *Journal of Materials Science*, 48, 7196-7203. doi: [10.1007/s10853-013-7536-x](https://doi.org/10.1007/s10853-013-7536-x).
- [19] Nizam, A.A.K., Masri, S., Fadilah, N.I.M., Maarof, M., & Fauzi, M.B. (2025). Current insight of peptide-based hydrogels for chronic wound healing applications: A concise review. *Pharmaceuticals*, 18(1), article number 58. doi: [10.3390/ph18010058](https://doi.org/10.3390/ph18010058).
- [20] Permana, M.D., Sakti, L.K., Luthfiah, A., Firdaus, M.L., Takei, T., Eddy, D.R., & Rahayu, I. (2023). A simple methods for determination of methylene blue using smartphone-based as digital image colorimetry. *Trends in Sciences*, 20(4), article number 5149. doi: [10.48048/tis.2023.5149](https://doi.org/10.48048/tis.2023.5149).
- [21] Rema, R., Alummoottil, N.J., & Janardanan, S. (2018). Effect of chemical modification with citric acid on the physicochemical properties and resistant starch formation in different starches. *Carbohydrate Polymers*, 202, 29-38. doi: [10.1016/j.carbpol.2018.08.128](https://doi.org/10.1016/j.carbpol.2018.08.128).
- [22] Rubezhniak, I. (2020). Antibacterial activities of cultural filtrates of some strains of micromycete. *Biological Systems: Theory and Innovation*, 11(2), 42-49. doi: [10.31548/biologiya2020.01.042](https://doi.org/10.31548/biologiya2020.01.042).
- [23] Shulga, S., & Shulga, O. (2024). Potato starch modification with benzoic acid chlorine hydride. *Ukrainian Chemistry Journal*, 89(12), 110-119. doi: [10.33609/2708-129X.89.12.2023.110-119](https://doi.org/10.33609/2708-129X.89.12.2023.110-119).
- [24] Vivcharenko, V., & Przekora, A. (2021). Modifications of wound dressings with bioactive agents to achieve improved pro-healing properties. *Applied Sciences*, 11(9), article number 4114. doi: [10.3390/app11094114](https://doi.org/10.3390/app11094114).
- [25] Wang, R., Lai, T.P., Gao, P., Zhuang, H., Ho, P.L., Woo, P.C.-Y., Ma, G., Kao, R.Y.-T., Li, H., & Sun, H. (2018). Bismuth antimicrobial drugs serve as broad-spectrum metallo- β -lactamase inhibitors. *Nature Communications*, 9, article number 439. doi: [10.1038/s41467-018-02828-6](https://doi.org/10.1038/s41467-018-02828-6).

Антимікробні та сорбційні властивості полімерних плівок на основі модифікованого крохмалю з ксероформом для ветеринарного призначення

Дар'я Кучинська

Аспірант

Київський національний університет технологій та дизайну

01011, вул. Мала Шияновська, 2, м. Київ, Україна

<https://orcid.org/0009-0007-5928-6639>

Олена Іщенко

Доктор технічних наук, професор

Київський національний університет технологій та дизайну

01011, вул. Мала Шияновська, 2, м. Київ, Україна

<https://orcid.org/0000-0002-9510-6005>

Роман Качан

Кандидат технічних наук, доцент

Київський національний університет технологій та дизайну

01011, вул. Мала Шияновська, 2, м. Київ, Україна

<https://orcid.org/0000-0001-7974-2784>

Ольга Сумська

Кандидат технічних наук, доцент

Херсонський державний аграрно-економічний університет

25031, просп. Університетський, 5/2, м. Кропивницький, Україна

<https://orcid.org/0000-0003-1606-6103>

Олена Роїк

Кандидат фармацевтичних наук, доцент

Київський національний університет технологій та дизайну,

01011, вул. Мала Шияновська, 2, м. Київ, Україна

<https://orcid.org/0000-0002-5988-6577>

Анотація. Розробка ефективних та безпечних ранозагоювальних матеріалів набуває надзвичайної важливості у ветеринарній медицині, особливо щодо лікування інфікованих ран та запобігання вторинним інфекціям. Перспективна стратегія передбачає виготовлення полімерних плівок, що мають комбіновані функціональні можливості. Метою цього дослідження було оцінити антимікробні та сорбційні властивості полімерних плівок на основі суміші модифікованого крохмалю та полівінілового спирту, що містять 10 мас. % ксероформу, для їхнього потенційного застосування як ветеринарних перев'язувальних матеріалів. Антимікробну активність плівок визначали методом дифузії в агаровому середовищі проти *Escherichia coli*, *Staphylococcus aureus*, *Candida albicans* та *Aspergillus niger*; сорбційні характеристики оцінювали на основі кінетики водопоглинання та адсорбції барвника метиленового синього, виміряних фотоелектроколориметрично. Досліджені плівки продемонстрували виражену антибактеріальну та фунгіцидну активність проти всіх протестованих штамів сучасних ранових патогенів, підтверджуючи ефективність ксероформу в межах полімерної матриці, зокрема, повне пригнічення росту *Escherichia coli*, *Staphylococcus aureus*, *Candida albicans* та *Aspergillus niger*. Крім того, композиції виявили високу рідинопоглинаючу здатність, необхідну для видалення ексудату з поверхні рани, з водопоглинанням у рівноважному стані від приблизно 180 % до понад 230 % та адсорбцією метиленового синього, що перевищувала 50 % протягом 60 хвилин. Кінетика водопоглинання та адсорбції барвника метиленового синього значною мірою залежала від умов модифікації крохмалю та співвідношення модифікованого крохмалю і полівінілового спирту в композиції. Отримані результати свідчать про потенціал розроблених плівок на основі модифікованого крохмалю, полівінілового спирту та ксероформу як активних ранозагоювальних покриттів з інтегрованими біоактивними компонентами (антисептиками) для ветеринарного застосування, що мають антисептичні та високі сорбційні властивості, які сприяють терапевтичній ефективності. Можливість варіювання умов модифікації крохмалю та складу дозволяє цілеспрямовано регулювати функціональні характеристики матеріалу для адаптації їх до специфічних клінічних потреб та видів тварин

Ключові слова: полісахариди; полівініловий спирт; вісмуту трибромфенолят; лікувальні ранові пов'язки; антисептик; фотоелектроколориметрія; ветеринарна медицина



Rheological properties of mucin from the snail *Achatina fulica*

Maryana Ovcharuk*

Postgraduate Student

National University of Food Technologies

01601, 68 Volodymyrska Str., Kyiv, Ukraine

<https://orcid.org/0009-0000-4922-6324>

Oksana Topchii

PhD in Technical Sciences, Associate Professor

National University of Food Technologies

01601, 68 Volodymyrska Str., Kyiv, Ukraine

<https://orcid.org/0000-0002-4832-2709>

Abstract. This article examines the functional role of various thickeners in cosmetic formulations. The study of their rheological properties and behaviour in aqueous solutions is essential for the development of stable and effective products within the skincare and dermocosmetic sectors. This study aimed to investigate the behaviour of a novel, naturally derived cosmetic ingredient – mucin from the snail *Achatina fulica* – in an aqueous medium under varying pH conditions, and to provide practical recommendations for its use. The study included an analysis of the physicochemical properties, mechanisms of action, and functional characteristics of five key thickeners: guar gum, Aristoflex, sodium alginate, gelatine, and carbomer. Their technological advantages, limitations, and applications in cosmetic products were outlined. In an aqueous solution, mucin exhibited a high swelling capacity: within 90 minutes, its volume increased ninefold, which is comparable to the performance of other hydrocolloids examined. The structural viscosity of mucin solutions within the concentration range of 0.1%-10.0% is characterised by a non-linear increase. A marked rise in viscosity is observed at concentrations between 1.5% and 3.0%, which is attributed to the attainment of maximum intermolecular interaction strength. It was found that the viscoelastic properties of aqueous mucin solutions depend not only on concentration but also on the pH of the medium, with mucin exhibiting non-Newtonian behaviour. Maximum structural stability was observed within the pH range of 5.0 to 6.5, where the structural viscosity of a 3.0% solution reached 127 mPa·s. Recommendations have been proposed regarding the inclusion of snail mucin in cosmetic formulations, taking into account the rheological and sensory characteristics of the final product. These insights may prove valuable to researchers and practitioners in the cosmetics industry

Keywords: thickeners; concentration; structure formation; medium acidity; cosmetic products

Introduction

Snail mucin, known for its moisturising, regenerative, and anti-inflammatory properties, is increasingly being used as a functional ingredient in skincare products. However, its behaviour as a thickener and stabiliser under varying conditions remains insufficiently studied. In particular, little is known about the impact of mucin on the rheological properties of low-concentration solutions, which is important for controlling the consistency and stability of cosmetic

formulations. In both cosmetic and biomedical fields, a range of polymolecular substances are employed to stabilise emulsions and suspensions, regulate viscosity, and control the release of active ingredients. These include polymers and hydrogels capable of forming stable structures that help preserve the efficacy of active compounds over time. Given the molecular structure of mucin, its ability to swell in aqueous solutions may be anticipated, suggesting

Suggested Citation:

Ovcharuk, M. & Topchii, O. (2025). Rheological properties of mucin from the snail *Achatina fulica*. *Technologies and Engineering*, 26(2), 58-65. doi: 10.30857/2786-5371.2025.2.5.

*Corresponding author



a potential influence on the consistency, stability, sensory characteristics, and bioavailability of final products.

The theoretical aspect of such research is linked to the development of polymer solution theory, while the practical focus lies in establishing empirical relationships between elastoplastic properties and technological parameters. From a practical standpoint, appropriate rheological properties are crucial in influencing consumer choice. For this reason, manufacturers often adjust the composition of the aqueous phase in cosmetic products by incorporating thickeners or structuring agents. Thickeners (structuring agents) are key ingredients used to increase viscosity and ensure the stability of dispersed systems in cosmetic formulations. Their application not only improves the rheological performance but also enhances the aesthetic qualities of the final product.

According to a study by E. Sikora (2019), hydrocolloids used as thickeners in cosmetics can be classified into the following categories: natural polymers – hydrocolloids of plant origin (mucilage, gums, resins, starch, dextrans, pectins, protein-based substances), and those of animal origin (polypeptides such as gelatine, casein, whey proteins, etc.); biosynthetic hydrocolloids – produced by microorganisms, e.g. dextran, xanthan gum; synthetic polymers – chemically or physically modified derivatives of cellulose, starch, or polyacrylic acid; inorganic compounds – minerals from the aluminosilicate group (such as bentonite), and colloidal silica.

P. Himashree *et al.* (2019) identified the primary functions of thickeners as moisture retention, structural modification, and the alteration of flow properties. Factors such as temperature, shear, pH, and ionic strength influence the functionality of these thickeners and must be carefully optimised by manufacturers during the production process. Proper optimisation of these conditions is critical to ensuring the stability and efficacy of the final products. Among the most effective thickeners are modified starches and proteins, used either individually or in combination with exudates, seed gums, seaweed extracts, and microbial polysaccharides.

The search for texturising agents and new combinations of known substances continues. For example, L. Rigano *et al.* (2019) proposed the use of *Caesalpinia spinosa* gum – a plant-derived polysaccharide obtained by milling the endosperm of *Caesalpinia spinosa* seeds – and investigated its rheological behaviour, compatibility with common cosmetic ingredients, and its use as a thickener in various types of cosmetic formulations. G. Tafuro *et al.* (2021) explored binary associations of sclerotium gum with iota-carrageenan, which imparted high viscosity, elasticity, and resilience to the system, as well as combinations of locust bean gum and pectin, which influenced the viscoelastic properties of the system, contributing to high adhesiveness and cohesiveness.

In the food industry, thickeners have traditionally been used to increase viscosity, stabilise texture, and regulate the rheological properties of products. For instance, R. Shkaraputa *et al.* (2024) highlighted the important role of structuring agents in developing the texture and consistency of food

products for vegans. These agents were used to replace collagen, preserve density and create a meatlike texture in meat substitutes, as well as to enhance the flavour and aroma of spices and aromatic plant-based additives.

There is also experience in applying polymolecular substances to achieve the desired rheological properties in technical systems. For example, R. Shilton *et al.* (2025) demonstrated that incorporating xanthan gum into alkali-activated material blends improved buildability during 3D printing and allowed the material to maintain its shape (structural integrity) while remaining flowable under stress – enabling pumping and extrusion.

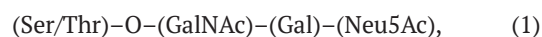
A deep understanding of the mechanisms behind the viscoelastic behaviour of hydrocolloids enables the optimisation of technological processes, the prediction of system behaviour during storage and use, and the development of new materials with tailored functional properties. This study aimed to investigate the rheological behaviour of *Achatina fulica* mucin in aqueous media under varying pH conditions, in order to provide practical recommendations for its use in cosmetic formulations.

Materials and Methods

The following ingredients were investigated: natural plant-based thickeners – guar and xanthan gum, sodium alginate, the animal-derived thickener mucin, and the synthetic agents Aristoflex and carbomer. Aqueous solutions of these gelling agents were prepared at concentrations ranging from 0.1% to 10.0%. To prepare the solutions, accurately weighed samples were dissolved in bidistilled water under stirring and made up to a final volume of 100 mL.

The characteristics of the selected gelling agents are outlined below. Mucin, a relatively novel and under-researched thickener derived from the mucus of giant snails (*Achatina fulica*), lacks a fixed chemical formula. It is a high-molecular-weight glycoprotein composed of a protein backbone with numerous attached oligosaccharide chains. The protein portion of mucins is rich in serine (Ser), threonine (Thr), and proline (Pro), while the carbohydrate side chains are linked to serine and threonine residues via O-glycosidic bonds (Ahmad *et al.*, 2021). The principal carbohydrates include N-acetylglucosamine (GlcNAc), galactose (Gal), fucose (Fuc), and sialic acid (Neu5Ac). Mucin exhibits high biocompatibility as it is a key component of mucus, which plays lubricating, protective, and barrier roles, particularly within the digestive, respiratory, and reproductive systems (Momoh *et al.*, 2019). Its inclusion in skincare products supports effective tissue repair, promotes healing of damaged skin, and offers protection against irritation – features of particular relevance in therapeutic and cosmetic creams designed for sensitive skin.

The general formula of the glycoprotein part (simplified):



where Ser/Thr is the amino acid residue (serine or threonine); O is the amino acid residue (serine or threonine);

GalNAc is the N-acetylgalactosamine; Gal is the galactose; Neu5Ac is the sialic acid.

Guar gum is a natural hydrocolloid that forms viscous colloidal solutions in aqueous media. It is extracted from the seeds of the guar plant (*Cyamopsis tetragonoloba*), a leguminous crop cultivated primarily in India and Pakistan. A notable advantage of guar gum in cosmetic formulations is its additional moisturising and conditioning properties (Mudgil *et al.*, 2014). Chemically, it is a polymer composed of mannose and galactose residues. Aristoflex is a polyacrylate derivative with high thickening efficiency at low concentrations. It demonstrates excellent compatibility with surfactants across a wide pH range (Baskar *et al.*, 2002). Its primary drawback is a relatively high production cost compared to natural alternatives. It is used in gels, serums, and lightweight emulsions for skincare. Sodium alginate is a natural polysaccharide that forms stable gels in the presence of calcium cations. It is derived from brown algae (Phaeophyceae) such as *Laminaria*, *Macrocystis*, and *Fucus*. Alginate is noted for its high hydrophilicity and moisture retention capacity; however, it is highly sensitive to changes in pH (Lee & Mooney, 2012). Gelatine is a protein obtained from collagen through the hydrolysis of animal bones, cartilage, and skin, most commonly from cattle or pigs. This natural protein polymer forms thermoreversible gels upon cooling (Chandra *et al.*, 2023); however, as of 2025, its use in cosmetic formulations is limited due to the rising popularity of vegan cosmetics.

Carbomer is a synthetic polymer based on acrylic acid, known for its high thickening efficiency in aqueous systems. It is synthesised by polymerising acrylic acid or its salts in the presence of initiators and stabilisers in an

aqueous solution. The resulting polymer is then neutralised (typically with alkalis), enabling it to form gels. Although effective at low concentrations, carbomer is sensitive to electrolytes and must be neutralised to achieve optimal performance (Khanlari & Dubé, 2015).

The degree of polymer swelling was calculated as the difference between the initial polymer volume and the swollen polymer volume using the formula:

$$\alpha = \frac{V - V_0}{V_0}, \quad (2)$$

where V_0 is the initial volume of the polymer, and V is the volume of the swollen polymer.

Viscosity was measured using an Ostwald viscometer, which operates by recording the flow rate of a liquid through a capillary tube of known radius and length. Density was determined using a pycnometer, with bidistilled water used as the reference liquid. The pH values were measured using a portable electronic pH meter with a measurement range of 0.00–14.

Results and Discussion

The effectiveness of thickeners can be indirectly assessed by examining their swelling capacity. The gel formation process is quantitatively characterised by the degree and rate of swelling. When preparing hydrogels with a low concentration of thickener, the gelation is primarily determined by the cooling phase and does not follow the kinetics of a first-order chemical reaction. Therefore, the rheological properties of the tested gelling agents were compared based on the degree of swelling, expressed as the amount of liquid absorbed per unit mass or volume of polymer, as illustrated in Figure 1.

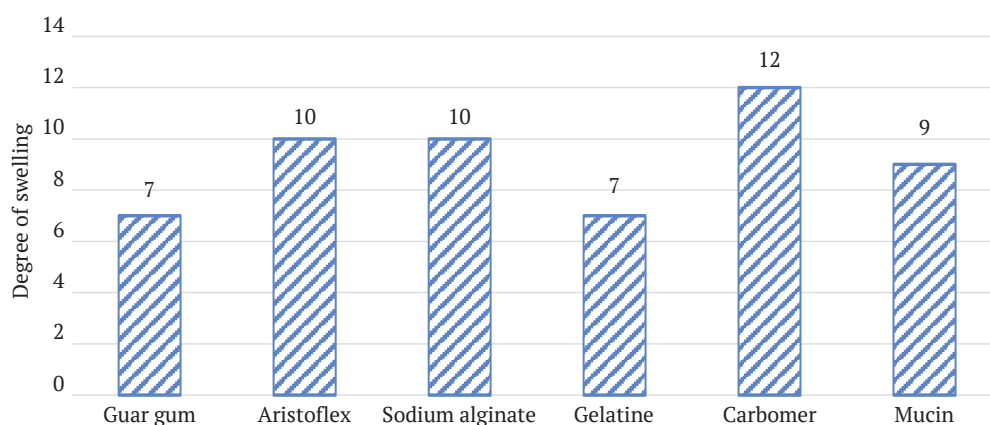


Figure 1. Degree of swelling after 90 minutes of observation

Source: compiled by the authors

Given that the molecular size of all the analysed thickeners is close to that of macromolecules, their swelling capacities are relatively similar. In terms of swelling ability, the synthetic thickeners Aristoflex and carbomer demonstrated superior performance, with sodium alginate performing nearly as well. Gelatine exhibited

moderate swelling, producing an opaque gel with a yellowish hue. The volume of guar gum increased sevenfold, while mucin showed a ninefold increase. As the polymer chains of the substances under investigation are flexible, complete dissolution by the end of the 90-minute experiment is expected.

Observation of the swelling process revealed that the dissolution of polymers differs significantly from the dissolution mechanism of low-molecular-weight compounds. In the latter case, molecules diffuse into the solvent. In contrast, when polymers dissolve, the polymer itself effectively acts as the solvent, while the low-molecular-weight liquid plays the role of the solute. This means that water molecules penetrate the polymer structure. Initially, they fill the spaces between macromolecular chains, resulting in a reduction in the overall volume of the system. Subsequently, the intermolecular bonds within the polymer weaken, allowing the solvent to penetrate further into the structure. As a result, the polymer's volume increases, although the total volume of the system remains constant.

According to F.K. Metze *et al.* (2023), the process of one-sided penetration of solvent molecules into a polymer is referred to as swelling. It has been observed that the studied hydrocolloids are prone to unlimited swelling,

which gradually transitions into dissolution, forming a homogeneous single-phase system. Overall, the behaviour of these substances in aqueous solution can be explained by the diffusion of molecules into the solvent and the formation of systems resembling colloids. Notably, all the resulting gels scatter sunlight, exhibiting the Tyndall effect characteristic of colloidal systems. In terms of swelling capacity, snail mucin performs on par with more well-known cosmetic thickeners and can be considered a competitive alternative.

As cosmetic formulations typically require low-concentration thickener solutions, the minimum gelation concentration of mucin was investigated. The effect of mucin concentration on the viscosity of aqueous solutions is presented graphically in Figure 2. Sample preparation included a swelling stage, as described in the previous experiment. The thickener concentration ranged from 0.1% to 10%, which falls within the recommended cosmetic usage levels.

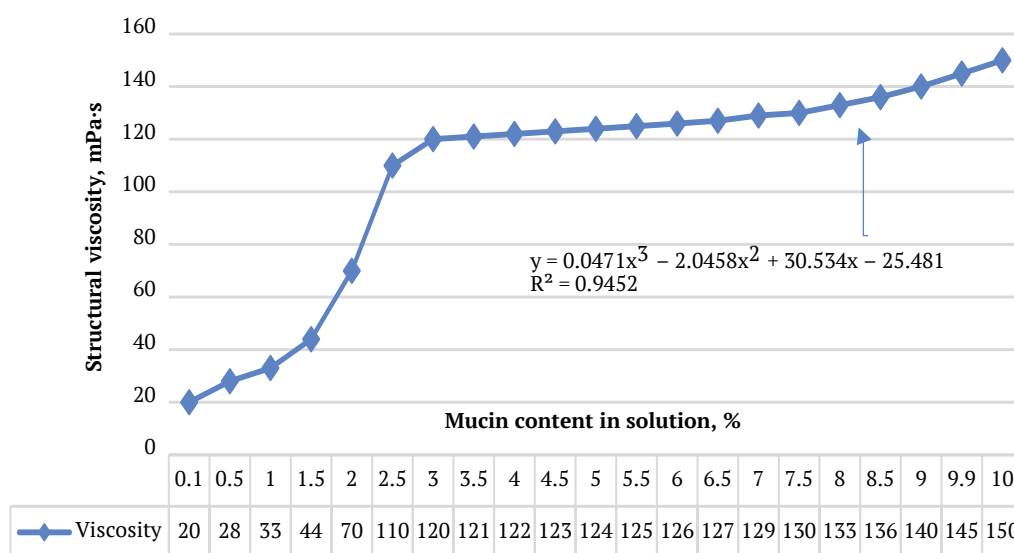


Figure 2. Mucin gelation concentration in aqueous solutions

Source: compiled by the authors

Figure 2 showed the relationship between the structural viscosity of the solution and the concentration of the mucin thickener. Within the concentration range of 1.5% to 3.0%, a sharp increase in viscosity is observed, indicating the formation of a structured system held together by intermolecular forces. Beyond this range, further increases in mucin concentration result in a linear rise in viscosity, which is characteristic of truly soluble substances. The trend line follows a third-degree polynomial equation. Based on the obtained results, the recommended dosage of mucin in cosmetic formulations may be narrowed to a maximum of 3.0%.

As established in the studies of B. Yuan *et al.* (2019) and several Ukrainian researchers, including O.P. Nekrasov & B.A. Veretenchenko (2018), the viscosity of colloidal solutions is influenced by the pH of the medium. Acidity is a critical characteristic of cosmetic products,

affecting both the selection of active ingredients and the overall efficacy of the formulation. Research conducted by O. Peredriy (2022) outlined regulatory safety requirements for cosmetic products within EU countries, which is particularly relevant when incorporating novel bio-active ingredients such as mucin derived from the snail *Achatina fulica*. The researcher emphasised that the inclusion of such components requires strict adherence to enhanced safety assessment procedures, monitoring of nanomaterials, and the provision of accurate product information. According to the Technical Regulations for Cosmetic Products (2021), products containing herbal extracts, fruit acids, and their derivatives must maintain a permissible pH range of 3.0-9.0. Therefore, the behaviour of mucin within this pH range was investigated. The relationship between the viscosity of mucin solution and pH value is presented in Figure 3.

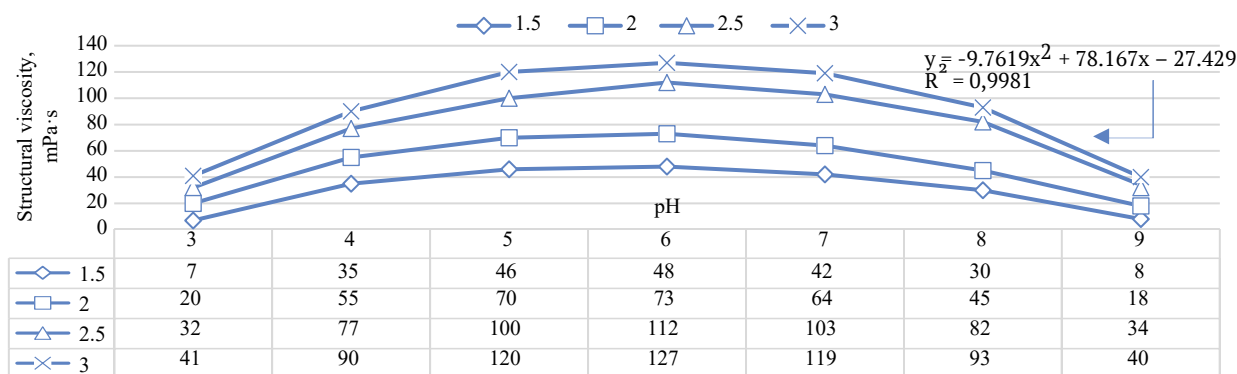


Figure 3. Dependence of mucin solution viscosity on pH

Source: compiled by the authors

Figure 3 illustrated that the viscosity of aqueous mucin solutions is significantly influenced by the pH of the medium. Viscosity fluctuations are observed across the entire investigated pH range, regardless of solution concentration. Based on the overall trend, a further decrease in viscosity can be anticipated as the pH deviates from neutrality. For all tested concentrations, peak viscosity is observed within the mildly acidic pH range of 5.0 to 6.5. This range should therefore be recommended as optimal, as it yields the maximum achievable viscosity. It was also noted that the general trend in structural formation within the solution remains consistent with the findings shown in Figure 2: higher mucin concentrations correspond to increased viscosity throughout the full pH range of 3.0 to 9.0. However, this difference becomes less pronounced at strongly acidic or strongly alkaline pH levels. All obtained viscosity profiles follow a similar pattern and can be described using a second-degree polynomial equation (one example of which is shown as the trendline equation of the obtained curves in Figure 3).

The results obtained can be explained by the behaviour of macromolecules in media with varying concentrations of H^+ and OH^- ions. The observed increase in viscosity is likely due to the unfolding of the molecule as a result of electrostatic repulsion, leading to a change in its conformation. In strongly acidic or strongly alkaline environments, the molecule tends to coil back into a compact structure, causing a reduction in viscosity. This occurs because an increase in ionic strength reduces electrostatic repulsion between charged groups within the molecule, thereby stabilising the compact, folded structure of the polymer.

These patterns align well with findings from other researchers studying the behaviour of low-concentration solutions of polymolecular substances in water. For instance, K.O. Lebedeva *et al.* (2022) investigated gelation and the rheological behaviour of hydrogels based on agar-agar at concentrations ranging from 25% to 50%, highlighting the critical role of temperature and cooling rate in the formation of spatial polymer networks. O.M. Kaminisky *et al.* (2024) worked with elastic gelatine-based gels at a 15% concentration, which exhibited high mechanical

stability and biocompatibility, making them suitable for applications in medicine and biotechnology. S.S. Stepanov & V.M. Mokrosnop (2021) studied the properties of sulphated polysaccharides from algae, which, owing to the presence of numerous hydroxyl and sulphated groups, form stable hydrogels and three-dimensional networks capable of retaining large volumes of water without losing elasticity. Thus, research into gelation has primarily focused on high-concentration solutions of conventional food thickeners, while non-food-grade polymeric substances remain less extensively studied. Notably, polymeric substances intended for cosmetic or pharmacological use continue to be a relatively underexplored area.

O.V. Maikovich *et al.* (2021) described the conditions for preparing a composite hydrogel based on polyacrylamide and gelatine with polypropylene microfibre. The authors identified the minimum component concentrations, as well as the temperature and pH ranges, required for the formation of a polyacrylamide-gelatine hydrogel. Their findings demonstrated that hydrogel formation occurs during the structuring reaction within a pH range of 2 to 4. Notably, no hydrogel was formed at pH values above 4, while at pH levels below 2, the samples were heterogeneous. This was attributed to the simultaneous hydrolysis of functional groups occurring alongside the structuring process. Therefore, the formation of the structured network is critically dependent on the acidity of the medium, and the stability of the resulting polymer network decreases outside the specified range. The pH range identified as optimal for the studied system lies in the strongly acidic region, whereas mucin has been shown to be more effective in mildly acidic conditions.

The study by K. Belinska & T. Marusei (2022) provided insights into the swelling kinetics of extrudates, which may contribute to a deeper understanding of biopolymer behaviour in various media. The research established that the rate and extent of swelling are dependent on the type of dispersing medium – an important factor in assessing the hydration properties of mucin, which in turn influence its rheological characteristics and potential application in cosmetic or pharmaceutical systems. K. Joyner *et al.* (2019) reported the development of a mucin-based hydrogel produced via

a thiol-based cross-linking strategy. Mucin hydrogels with viscoelastic properties resembling those of natural mucus were synthesised using a cross-linking reagent (the specific compound was not disclosed). The authors found that the rate of gel formation could be effectively controlled by adjusting the ratio of mucin to cross-linker. Structural formation within the gel network was attributed to hydrogen bonding and disulphide linkages. This approach enables the mechanical properties of the material to be tailored according to its intended use – ranging from soft coatings to structured gels suitable for the delivery of bioactive compounds. The findings support the fabrication of viscoelastic microstructures with physiologically relevant properties based on highly concentrated mucin solutions.

The present study focused on low-concentration solutions suitable for practical use in cosmetic manufacturing and cosmetic chemistry. Based on the rheological properties and stability of *Achatina fulica* mucin in aqueous media, its suitability for inclusion in various types of cosmetic formulations was established. The results support the following recommended concentrations of snail mucin in cosmetic products: in facial creams and serums – 1.0%-3.0%; in masks and patches – 2.0%-3.0%; in toners and essences – 0.5%-1.0%; and in cleansing products – 0.1%-0.5%. These concentrations allow for the development of conventional textures while ensuring minimal use of the valuable active ingredient. The inclusion of mucin is not advised in acid-based peels, gels, or neutralising sprays, nor products intended for carboxytherapy, as these formulations typically involve strongly acidic environments in which the gel-forming ability of mucin is reduced. The findings of this study thus contribute to a broader understanding of the potential uses of *Achatina fulica* mucin in cosmetic compositions. It has been demonstrated that, even at low concentrations, mucin exhibits stable rheological properties suitable for the formulation of a variety of product types.

Conclusions

Thickeners are essential components of cosmetic products, as they determine the rheological properties, stability, and sensory characteristics of the final formulation. Their use enables control over viscosity, enhances texture, and ensures the even distribution of active ingredients. As the properties of thickeners vary depending on their chemical structure and origin, it is important

to select them rationally per the specific requirements of the cosmetic formula. The optimal choice of thickener depends on the desired rheological behaviour and the final texture of the product. Natural thickeners, such as guar gum, sodium alginate, and gelatine, are suitable for the formulation of environmentally friendly cosmetics. Meanwhile, synthetic thickeners like Aristoflex and carbomer offer high stability and allow for the creation of unique textures. Optimal product performance can often be achieved through the combination of different types of thickeners within a single formulation. The rheological behaviour of a cosmetic thickener – mucin from *Achatina fulica*. In an aqueous solution, mucin demonstrated the ability to swell, increasing its volume ninefold over a 90-minute observation period, showing a swelling capacity comparable to that of other commonly studied hydrocolloids. The structural viscosity of mucin solutions at concentrations ranging from 0.1% to 10.0% increased non-linearly across the entire studied range. A marked rise in viscosity was observed between 1.5% and 3.0%, corresponding to the point of maximum intermolecular interaction. It was established that the viscoelastic properties of mucin depend not only on concentration but also on the pH of the medium, with the substance exhibiting non-Newtonian behaviour. The structure reached maximum stability within a pH range of 5.0 to 6.5, where structural viscosity values were recorded as follows: 48 mPa · s for the 1.5% solution; 73 mPa · s for the 2.0% solution; 112 mPa · s for the 2.5% solution; and 127 mPa · s for the 3.0% solution. These findings should be taken into account when developing cosmetic formulations. The results obtained may be applied to optimise the composition of cosmetic and pharmaceutical products containing mucin, and also provide a basis for further research into its functional properties in biomaterials and medical applications.

Acknowledgements

None.

Funding

The study received no funding.

Conflict of Interest

None.

References

- [1] Ahmad, M., Ritzoulis, C., Chen, J., Meigui, H., Bushra, R., Jin, Y., & Xiao, H. (2021). Xanthan gum – mucin complexation: Molecular interactions, thermodynamics, and rheological analysis. *Food Hydrocolloids*, 114, article number 106579. doi: 10.1016/j.foodhyd.2020.106579.
- [2] Baskar, G., Ramya, S., & Mandal, A. (2002). Synthesis and solution properties of comblike polymers from octadecyl methacrylate and acrylic acid. *Colloid & Polymer Science*, 280, 886–891. doi: 10.1007/s00396-002-0697-9.
- [3] Belinska, K., & Marusei, T. (2022). Investigation of kinetics of extrudate swelling in different types of renewed milk. *Commodity Bulletin*, 15(1), 103–112. doi: 10.36910/6775-2310-5283-2022-15-9.
- [4] Chandra, J., Sasi, S., & Bindu Sharmila, T.K. (2023). Material applications of gelatin. In S.A. Thomas, C. Jose Chirayil, & B. Thomas. (Eds.), *Handbook of biopolymers* (pp.749–782). Singapore: Springer. doi: 10.1007/978-981-19-0710-4_28.

- [5] Himashree, P., Animesh Singh Sengar, & Sunil, C.K. (2022). Food thickening agents: Sources, chemistry, properties and applications – a review. *International Journal of Gastronomy and Food Science*, 27, article number 100468, doi: [10.1016/j.ijgfs.2022.100468](https://doi.org/10.1016/j.ijgfs.2022.100468).
- [6] Joyner, K., Song, D., Hawkins, R.F., Silcott, R.D., & Duncan, G.A. (2019). A rational approach to form disulfide linked mucin hydrogels. *Soft Matter*, 15, 9632–9639. doi: [10.1039/C9SM01715A](https://doi.org/10.1039/C9SM01715A).
- [7] Kaminsky, O.M., Denisyuk, R.O., Chaika, M.V., Pisarenko, S.V., Yevdochenko, O.S., & Panasyuk, D.Y. (2024). Kinetics of the growth of Liesegang rings in elastic gelatin gels. *Ukrainian Journal of Natural Sciences*, 8, 131–139. doi: [10.32782/naturaljournal.8.2024.14](https://doi.org/10.32782/naturaljournal.8.2024.14).
- [8] Khanlari, S., & Dubé, M.A. (2015). Effect of pH on poly (acrylic acid) solution polymerization. *Journal of Macromolecular Science*, 52(8), 587–592. doi: [10.1080/10601325.2015.1050628](https://doi.org/10.1080/10601325.2015.1050628).
- [9] Lebedeva, K.O., Cherkashina, G.M., Savchenko, D.O., & Lebedev, V.V. (2022). Study of the features of gelation and rheological properties of agar-agar-based hydrogels. *Integrated Technologies and Energy Saving*, 3, 42–51. doi: [10.20998/2078-5364.2022.3.05](https://doi.org/10.20998/2078-5364.2022.3.05).
- [10] Lee, K.Y., & Mooney, D.J. (2012). Alginate: Properties and biomedical applications. *Progress in Polymer Science*, 37(1), 106–126. doi: [10.1016/j.progpolymsci.2011.06.003](https://doi.org/10.1016/j.progpolymsci.2011.06.003).
- [11] Maikovich, O.V., Nosova, N.G., Yakoviv, M.V., Dron, I.À., Stasiuk, A.V., Samaryk, V.Y., Varvarenko, S.M., & Voronov, S.A. (2021). Composite materials based on polyacrylamide and gelatin reinforced with polypropylene microfibre. *Issues of Chemistry and Chemical Technology*, 1, 45–54. doi: [10.32434/0321-4095-2021-134-1-45-54](https://doi.org/10.32434/0321-4095-2021-134-1-45-54).
- [12] Metze, F.K., Sant, S., Meng, Z., Klok, H.-A., & Kaur, K. (2023) Swelling-activated, soft mechanochemistry in polymer materials. *Langmuir*, 39(10), 3546–3557. doi: [10.1021/acs.langmuir.2c02801](https://doi.org/10.1021/acs.langmuir.2c02801).
- [13] Momoh, M.A., Chime, S.A., Ogbodo, D.U., Akudike, P.K., Udochukwu, S.U., Ossai, E.C., Kenechukwu, F.C., Ofokansi, K.C., & Attama, A.A. (2019). Biochemical, rheological and hydrophile-lipophile balance (HLB) evaluation of *Archachatina marginata* (snail) mucin extract for possible nutraceutical and nano biopharmaceutical applications. *Tropical Journal of Pharmaceutical Research*, 18(5), 927–934. doi: [10.4314/tjpr.v18i5.3](https://doi.org/10.4314/tjpr.v18i5.3).
- [14] Mudgil, D., Barak, S., & Khatkar, B.S. (2014). Guar gum: Processing, properties and food applications – a review. *Journal of Food Science and Technology*, 51, 409–418. doi: [10.1007/s13197-011-0522-x](https://doi.org/10.1007/s13197-011-0522-x).
- [15] Nekrasov, O.P., & Veretenchenko, B.A. (2018). *Surface phenomena and dispersed systems*. Kharkiv: National Technical University 'Kharkiv Polytechnic Institute'.
- [16] Peredriy, O. (2022). Regulations of safety of cosmetics goods in EU countries and Ukraine. *Commodity Bulletin*, 15(1), 252–263. doi: [10.36910/6775-2310-5283-2022-15-22](https://doi.org/10.36910/6775-2310-5283-2022-15-22).
- [17] Rigano, L., Deola, M., Zaccariotto, F., Colleoni, T., & Lionetti, N. (2019). A new gelling agent and rheology modifier in cosmetics: *Caesalpinia spinosa* gum. *Cosmetics*, 6(2), article number 34. doi: [10.3390/cosmetics6020034](https://doi.org/10.3390/cosmetics6020034).
- [18] Shilton, R., Wang, S., & Banthia, N. (2025). Use of polysaccharides as a rheology modifying admixture for alkali activated materials for 3D printing. *Construction and Building Materials*, 458(5), article number 139661. doi: [10.1016/j.conbuildmat.2024.139661](https://doi.org/10.1016/j.conbuildmat.2024.139661).
- [19] Shkaraputa, R., Melnyk, O., & Spytskiy, V. (2024). Analysis of raw materials for the production of vegetable meat. *Herald of Khmelnytskyi National University. Technical Sciences*, 337(3(2)), 305–309. doi: [10.31891/2307-5732-2024-337-3-46](https://doi.org/10.31891/2307-5732-2024-337-3-46).
- [20] Sikora, E. (2019). *Cosmetic emulsions: Monograph*. Kraków: Cracow University of Technology Press.
- [21] Stepanov, S.S., & Mokrosnop, V.M. (2021). *Metabolic processes and valuable substances of algae*. Kyiv: Scientific Thought.
- [22] Tafuro, G., Costantini, A., Baratto, G., Francescato, S., Busata, L., & Semenzato, A. (2021). Characterization of polysaccharidic associations for cosmetic use: Rheology and texture analysis. *Cosmetics*, 8(3), article number 62. doi: [10.3390/cosmetics8030062](https://doi.org/10.3390/cosmetics8030062).
- [23] Technical regulations for cosmetic products. (January, 2021). Retrieved from <https://zakon.rada.gov.ua/laws/show/65-2021-%D0%BF#Text>.
- [24] Yuan, B., Ritzoulis, C., Wang, X., Pan, W., & Chen, J. (2019), Interactions between mucin and okra gum during pH cycling. *Food Hydrocolloids*, 95, 1–9, doi: [10.1016/j.foodhyd.2019.03.050](https://doi.org/10.1016/j.foodhyd.2019.03.050).

Реологічні властивості муцину равлика *Achatina fulica*

Мар'яна Овчарук

Аспірант

Національний університет харчових технологій

01601, вул. Володимирська, 68, м. Київ, Україна

<https://orcid.org/0009-0000-4922-6324>

Оксана Топчій

Кандидат технічних наук, доцент

Національний університет харчових технологій

01601, вул. Володимирська, 68, м. Київ, Україна

<https://orcid.org/0000-0002-4832-2709>

Анотація. У статті розглянуто функціональну роль ряду згущувачів у косметичних рецептурах. Дослідження їх реологічних властивостей та поведінки у водних розчинах є важливим для розробки стабільних і ефективних косметичних засобів в сегментах доглядової та дерматокосметичної продукції. Метою було вивчити поведінку нового косметичного інгредієнту природного походження – муцину равлика *Achatina fulica* – у водному середовищі зі змінним значенням рН та надати практичні рекомендації по його застосуванню. Проведено аналіз фізико-хімічних властивостей, механізмів дії та функціональних характеристик п'яти ключових згущувачів: гуарової камеді, Арістофлексу, альгінату натрію, желатину та Карбомеру. Описано їхні технологічні переваги, обмеження та сфери застосування в косметичних засобах. У водному розчині муцин продемонстрував високу здатність до набухання: впродовж 90 хвилин спостереження його об'єм збільшився у 9 разів, що є співставним із показниками інших досліджуваних гідрокоолідів. Структурна в'язкість розчинів муцину в концентраційному діапазоні 0,1-10,0 % характеризується нелінійним зростанням. Значне підвищення в'язкості реєструється при концентраціях від 1,5 % до 3,0 %, що обумовлено досягненням максимальної сили міжмолекулярних взаємодій. Встановлено, що в'язкісно-пружні характеристики водних розчинів муцину залежать не лише від концентрації, а також від рН середовища, коли муцин демонстрував неньютонівську поведінку. Максимальна стабільність його структури спостерігалася в діапазоні рН від 5,0 до 6,5, коли структурна в'язкість 3,0 %-го розчину досягає 127 мПа·с. Запропоновано рекомендації щодо вмісту равликового муцину в косметичних складах з урахуванням реологічних та сенсорних характеристик кінцевого продукту, які будуть корисні для науковців та фахівців-практиків косметичної індустрії

Ключові слова: згущувачі; концентрація; структуроутворення; кислотність середовища; косметичні засоби



Analysis of the key models, methods, and means of data collection in the Internet of Things

Maksym Savka*

Postgraduate Student

National Technical University of Ukraine "Igor Sikorsky Kyiv Polytechnic Institute"

03056, 37 Beresteyskyi Ave., Kyiv, Ukraine

<https://orcid.org/0000-0003-2049-9438>

Abstract. The rapid development of mobile technologies and the technological revolution are contributing to the active implementation of the Internet of Things, leading to an increase in the number of physical devices connected to the Internet. However, along with the advantages of the Internet of Things, there are also growing disadvantages, including cyber threats, privacy violations, and risks of unauthorised data collection. In this regard, research into the challenges and security measures in the field of the Internet of Things is necessary to ensure the safe implementation of these technologies. The purpose of the present study was to analyse key models, methods, and tools for data collection in the Internet of Things environment. The methodological framework of this study included a systematic approach, which allowed presenting public policy in the field of the Internet of Things as a complex, structurally organised system that requires improvement and development. The study employed general scientific and specific methods (systemic, abstraction and comparison, formalisation, institutional analysis, historical-genetic, logical-semantic, classification, generalisation, analysis, formalisation, and comparison), as well as statistical data, which was used extensively and consistently. The study found that among the existing data collection technologies, sensor networks and cloud computing are the most widely used, but they do not cover all the requirements of mobile Internet of Things systems, specifically in terms of energy efficiency, throughput, and security. It was found that modern architectural solutions (including social architectures of the Internet of Things) allow scaling the system and connecting various devices, but have limitations related to deployment complexity and bandwidth. The study also found that the major security threats arise at all levels of mobile Internet of Things networks, from devices to cloud infrastructure, confirming the need for a comprehensive approach to their protection. It was shown that most existing solutions do not simultaneously meet the requirements for low power consumption, speed of data collection and processing in networks with dynamic topology, as well as the requirements for reliability and confidentiality. The proposed developments complemented existing methods, enhancing the flexibility and resilience of the data collection process and laying the foundation for the development of new energy-efficient and secure tools in the Internet of Things

Keywords: cloud computing; IoT architecture; data security; CoAP protocol; fog computing

Introduction

Changes in network structure complicate the determination of the level of trust between devices. The primary reason for this situation is the complex structure of the network. Mobility forces devices to consume more energy, but their resources stay unchanged, and therefore data collection in dynamic Internet of Things (IoT) networks with complex structures becomes a pressing issue. Data encryption increases the overhead costs of data transmission, which contributes to increased network traffic, higher

network energy consumption, and a shorter network lifetime. Therefore, fresh solutions are necessary to ensure that the basic requirements of the Internet of Things are met when collecting data from mobile device networks. The complexity of data collection tasks in the Internet of Things is increasing considerably due to the transition from the use of simple static IoT devices to the use of a wide range of smart mobile devices, such as mobile phones, smart watches, smart vehicles, and mobile devices in smart

Suggested Citation:

Savka, M. (2025). Analysis of the key models, methods, and means of data collection in the Internet of Things. *Technologies and Engineering*, 26(2), 66-78. doi: 10.30857/2786-5371.2025.2.6.

*Corresponding author



cities. Modern mobile device networks are heterogeneous networks with complex dynamic structures. Data collection in such networks poses new challenges, such as new opportunities for attacks (Mishra & Pandya, 2021).

W.E.H. Youssef *et al.* (2023) addressed the problem of secure data transmission between end devices and servers in the Internet of Things using cryptographic mechanisms. The essence of their study was to present a novel approach to solving IoT security problems by introducing a secure lightweight cryptosystem based on Chaos. The findings of the study confirmed the effectiveness of the approach in terms of computational complexity, memory requirements, and energy consumption. The researchers noted that after reviewing and comparing their findings with the existing reports, the advantage of their cryptosystem became evident. Modelling results and performance analysis revealed that the proposed secure chaos-based lightweight cryptosystem (SCBLC) can be considered an ideal choice for protecting communications in resource-constrained IoT devices, taking major steps towards improving IoT network security. However, some drawbacks included increased computational overhead on IoT devices caused by their complexity, the need for reliable key management protocols, and the probability of susceptibility to advanced cryptanalytic attacks. The problems of data collection and transmission delays and high energy consumption have been solved with the ESR (energy-efficient and secure routing protocol); the ESR protocol for the Internet of Things was developed by K. Haseeb *et al.* (2019). As the researchers noted, experimental findings using a network simulator demonstrated that the proposed routing protocol achieved a 37% improvement in network lifetime, a 24% improvement in average end-to-end delay, a 30% improvement in packet delivery rate, average communication cost by 29%, network overhead by 28%, and route reopening frequency by 38% compared to existing work in dynamic network topologies. The proposed protocol created various energy-efficient clusters based on the intrinsic qualities of nodes. Based on the (k, n) threshold scheme of Shamir's secret sharing, reliability and security of sensor information between the base station (BS) and the cluster head were achieved. The proposed security scheme provided an easy solution to combat intrusions created by malicious nodes. This protocol can reduce energy consumption and data transmission delays. Notably, this only provided data protection from nodes to the cluster head and further to the base station from malicious threats. This protocol was not extended to factor in the network communication with multiple hops. These studies did not fully address the issue of data transmission security along with energy efficiency in highly resource-constrained environments.

The basic concept of the Internet of Things envisages the ubiquity of objects (things, devices) that can interact to achieve shared goals using wireless and wired connections and unique addressing schemes. Internet of Things networks generate multiple data streams, each with its unique characteristics (Khare & Totaro, 2019). D. Honcharenko *et*

al. (2023) offered a detailed analysis of modern approaches to building an information system for monitoring physical indicators based on Internet of Things technologies, with a particular focus on the selection of network solutions and IoT platforms according to performance and energy efficiency requirements. Thus, there is currently no solution that guarantees the simultaneous fulfilment of the three basic requirements of the Internet of Things: security of data collection and transmission, low network response time, and minimal energy consumption. Therefore, the purpose of the present study was to comprehensively compare models, methods, and protocols for data collection in mobile IoT systems in terms of energy efficiency, delays, and security.

Materials and Methods

The study employed a comprehensive scientific and methodological approach combining systematic analysis, comparison and classification methods, as well as the use of relevant statistical data and literary sources. To obtain objective and comprehensive findings, each stage of the study was conducted sequentially, ensuring verifiability and reproducibility. The first step was to select scientific publications, monographs, conference materials, and official documents (international standards, recommendations, patents) on the topic of the Internet of Things (IoT). The focus was on recent studies reflecting the current state of technology in the field of mobile IoT systems, the specific features of social IoT solution architectures, as well as various data collection methods.

The study analysed open statistical materials (specifically, reports of analytical agencies, data available on the websites of international organisations), covering indicators of IoT solution implementation, data transfer volumes, the dynamics of the mobile device market, and cyberattack risk assessments. The study was conducted considering the multi-level structure of IoT systems (devices, network, cloud infrastructure), as well as the factors affecting the security and efficiency of data collection at each level. The use of a systematic approach helped to identify internal and external interrelationships between components of IoT architectures.

Abstraction and comparison methods were employed to identify key parameters and characteristics of data collection models and methods, which allowed generalising and systematising the information obtained from various sources. The descriptions of existing data collection technologies were formalised to simplify further analysis and comparison of technical characteristics (throughput, response time, energy consumption, scalability, etc.). Within the framework of the classification, analysis, and generalisation methods based on the collected data, existing technical solutions were systematised and structured according to their architecture type and technological features, and their relevance for mobile IoT systems was determined.

At the stage of sampling and initial analysis, a search, selection, and analysis of scientific papers and statistical

materials on the topic was performed. The results were systematised in a database according to key parameters (year of publication, description of technologies, areas of application, security features). The study determined the criteria for evaluating the effectiveness and security of data collection (e.g., energy consumption, performance, security complexity, scalability). Based on general scientific methods of comparison and classification, several groups of technological solutions were formed (e.g., sensor networks, cloud and fog computing, social IoT architectures), and their features and possible areas for development were identified. Particular attention was paid to the study of threats at various levels of the IoT: from devices to central (cloud) infrastructure, as well as protection mechanisms that can be applied during data collection.

Data analysis was performed using publicly available software (Microsoft Excel) and specialised libraries for processing, visualising, and comparing the technical characteristics of solutions. To assess security and performance indicators, the results of experimental studies with network load modelling and review tests of security tools described in scientific literature were used. The study only considered solutions with up-to-date technical documentation and fairly widespread practical application. Highly specialised products or protocols with limited scope or at an early stage of development were not considered. It was also assumed that the conditions for integrating IoT solutions into mobile networks may vary

depending on the concrete region and regulatory environment, yet the general technological approaches, data collection and processing methods were still comparable on a global scale. Overall, this approach and set of methods allowed covering a variety of data collection models and analysing their compliance with the requirements of mobile IoT systems in terms of energy consumption, throughput, security, and adaptability to dynamic environments. This allowed identifying the advantages and disadvantages of various solutions and helped to formulate proposals for their improvement.

Results and Discussion

General architecture of the Internet of Things

Based on a review of the literature on Internet of Things (IoT) system architectures, the study found that IoT system architecture consists of three layers (Fig. 1): physical, network, and applied (application or service layer). The applied layer includes services such as smart grids and smart transport services. IoT devices collect data about events occurring in the physical environment. This data is sent to the network layer; the primary task of the IoT network layer is to process the collected data. Wired or wireless networks are used to transfer data to higher levels: LAN (local area network), WAN (wide area network), Wi-Fi (wireless fidelity), LTE (long-term evolution), 3G (3rd Generation). Various network technologies can be used for data transfer, including Bluetooth and ZigBee.

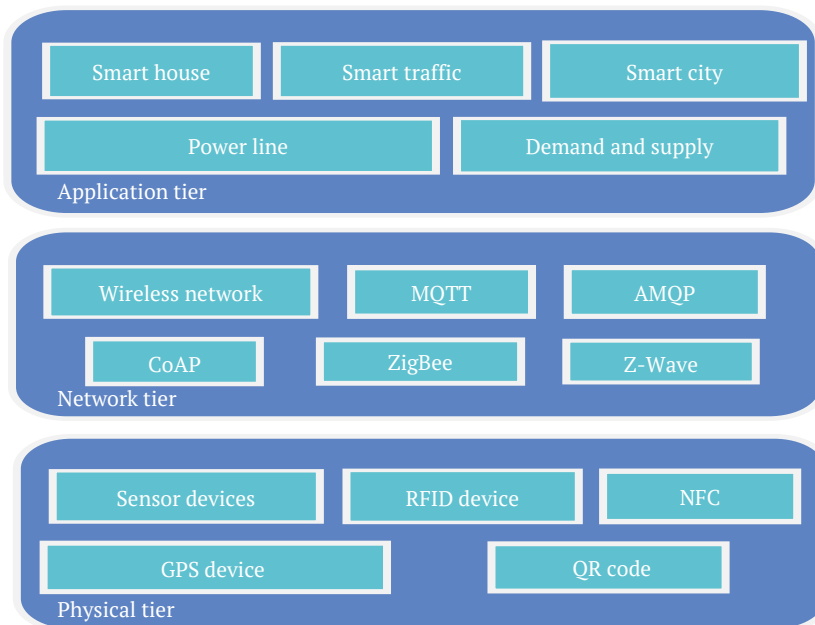


Figure 1. Schematic representation of the three-tier architecture of the Internet of Things system

Source: developed by the author of this study based on the experimental research results

The three-tier architecture ensures direct interaction between the device and the application layer. No intermediate nodes are required to convert or interpret data; building a system with a three-tier architecture allows using

the existing equipment to create a unified Internet of Things infrastructure that collects and transmits data from end devices to applications. The advantage of this is that services are independent of each other and of the

physical hardware. Abstraction of the physical tier facilitates access to network resources from the application tier (Chanal & Kakkasageri, 2020).

Architecture of IoT systems with middleware platforms (Fig. 2). Middleware platform technology in IoT systems emerged in parallel with the concept of cloud computing. Middleware platforms enable the integration of many systems with different functionalities on their own basis. The platform

ensures requirements such as enhanced data security and resource sharing. Middleware platforms are designed to provide access to data collected about environmental objects and their interactions anytime, anywhere, and from any device. Middleware platforms provide services to software applications hosted in the cloud. The platform can be built to create an infrastructure for running applications and use general tools for programming physical devices (Waheed *et al.*, 2024).

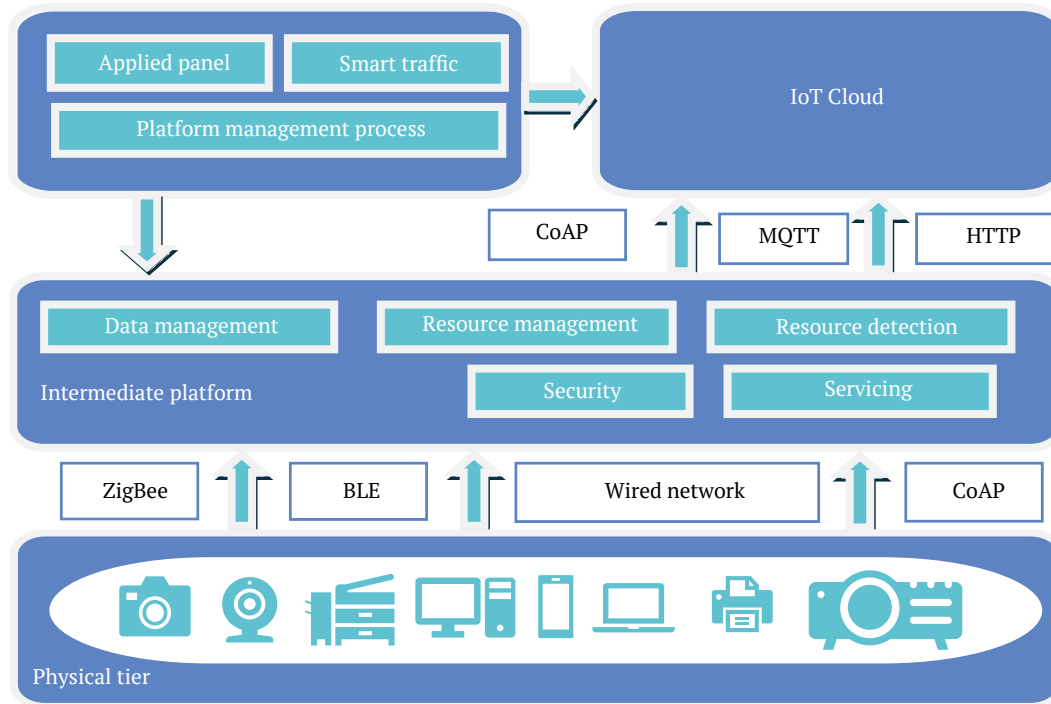


Figure 2. Architecture of IP systems with an intermediate software platform

Source: developed by the author of this study based on the experimental research results

In a conventional three-tier architecture, data processing and support tiers, also known as the interface tier or middleware tier, are created and used. In this architecture, the data processing tier is located between the applied and network tiers. The primary tasks of this tier are to receive user requests, collect the data necessary to fulfil them, and provide responses. Various services must be used to process requests. Therewith, the necessary level of security must be ensured when collecting and transmitting data. Cloud computing architectures focus on the distributed processing of raw data collected from a device. Various data analysis methods, including machine learning methods, can be used for processing. It may be necessary to process query history, information about previous actions of mobile devices, and connections between devices, or to process operational data in real time.

Data processing in cloud nodes extends the processing capabilities offered by cloud services. Processing is performed in close proximity to the data source. System performance is improved by reducing the amount of data transferred to the cloud for processing, analysis, and long-term storage. Cloud computing is currently the most

popular option because it allows protecting and managing data in an IP environment and building systems with sufficiently high performance. The study (Ivanov *et al.*, 2023) proposed an architectural model for IP that included the use of cloud computing and described existing standards for the collection, processing, and transmission of data from mobile devices. IoT environments are based on the collection, storage, and processing of data. IoT devices such as sensors, RFID tags, and Bluetooth devices generate massive amounts of data that is transferred to cloud environments for storage and processing, which users can subsequently access. These standards form the basis for the interconnection of the Internet of Things and Big Data. The problematic of the cited study was the increasing complexity of processing and analysing the collected data. As conclusions, further research was proposed to develop new methods for collecting and processing data from IoT devices in the context of Big Data, especially in the area of developing effective algorithms for optimising the processing of large amounts of data. S. Kasiviswanathan *et al.* (2024) enabled the validation of IoT system deployments by providing dynamic deployment of distributed IP applications,

computing infrastructure management, and automation of visualisation systems.

S. Supreeth & K. Patil (2022) proposed an approach (VMS-EDMVM) to distributed virtual machine management in a cloud computing environment focused on real-time data processing. This method uses only cloud data centre resources and does not consider intermediate (edge/fog) nodes, focusing on dynamic migration and scaling of virtual machines to ensure continuity and efficient resource utilisation. Y. Huang *et al.* (2024) proposed a cloud-based model called E-norm. This model provides centralised management of applications and nodes. Proximity to the end user is factored in when distributing workloads between nodes. Furthermore, an automatic scaling mechanism is implemented to adjust resource allocation according to network delays and task execution times.

U. Khadam *et al.* (2024) proposed a cloud computing-based data aggregation model. This model performs two-dimensional data aggregation, which reduces data transmission costs and increases network throughput. V.K. Prasad *et al.* (2023) proposed using a cloud computing layer to minimise server resource consumption and determine the resources required for data collection and processing. M. Peyman *et al.* (2021) proposed an IoT solution for data collection and analysis focused on the efficient use of cloud resources and optimisation of energy consumption. The model involves step-by-step processing: initial signal processing is performed directly at the nodes, while deeper analysis and long-term storage are

performed in the cloud. The researchers pointed out that this approach reduced the load on the local infrastructure and improved the scalability of the system as the number of sensors increased.

Cloud computing provides virtually unlimited scalable processing and storage resources. However, routing data to remote data-centres typically increases the round-trip time (RTT) for mobile and IoT devices. For latency-sensitive workloads (≤ 50 ms), architects therefore employ edge- or fog-computing approaches, which shift computation closer to the data source and, in practice, lower end-to-end latency by a factor of 2-4. Such responsiveness is essential, for instance, when developing software for autonomous vehicles. Thus, the edge (fog) computing tier acts as an intermediate layer between the device tier and the centralised cloud, optimising data processing and storage while improving the quality of service delivered to end users.

Registration of new social objects, creation of new device groups, determination of “social connections” between devices, management of established connections. A socially oriented approach to defining IP networks allows extending the conventional IP model by adding social elements. Social IP architecture was developed as part of Software as a Service (SaaS), which offers heterogeneous sensor data. This service ensured data compatibility and reusability. As presented in Figure 3, the architecture of a social IS system includes a physical tier, a network tier, and an application tier. The physical and network tier allow collecting data from devices using various data transmission networks.

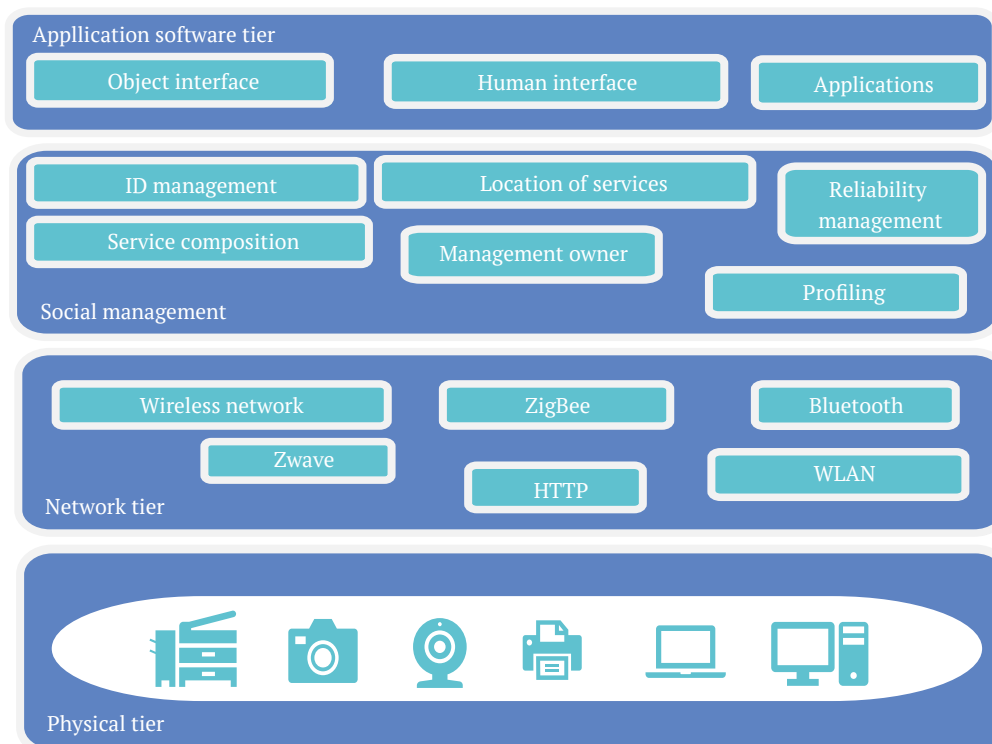


Figure 3. Architecture of social IP systems

Source: developed by the author of this study based on the experimental research results

At the application level, IP applications are deployed together with services, the functionality of which is enabled by social connections between objects. Such services form an independent sub-tier of the application tier. These include identity management services for identifying objects, profiling services for semi-automatic collection of data about objects, and management services for determining the actions that objects can perform. The relationship management service is one of the basic services for establishing relationships between objects. Reliability management services are used to create a database of reliable objects based on the analysis of object behaviour. The reliability management service is closely related to the relationship management service.

Communication technologies and data transfer protocols in IoT

Communication technologies for data collection in IoT define the rules for transmitting digital messages, including packet size parameters, data transfer rates, synchronisation methods, types of possible errors, address mapping methods, as well as rules for packet delivery control and data flow management. In Internet of Things systems, data is transmitted using various protocols.

Bluetooth LE (Low Energy), Wi-Fi, ZigBee, and LoRaWAN (Long Range Wide Area Network) are used for device-to-device communication at the channel level. At the application level, DDS (Data Distribution Service), MQTT (Message Queue Telemetry Transport), HTTP (Hypertext Transfer Protocol), CoAP (Constrained Application Protocol), and XMPP (Extensible Messaging and Presence Protocol) protocols are used to transfer data from devices to servers, while AMQP (Advanced Message Queuing Protocol) is used to transfer data from servers. The Data Distribution Service (DDS) architecture has two tiers of information transfer to the destination and an additional tier of data flow management. It is an open standard that defines the mechanism for transferring information in a publish-subscribe paradigm. Unlike the conventional client-server model, DDS is focused on distributed, decentralised interaction between nodes, where data is published and consumed by various applications without the need for a central broker or server for routing (Internal documentation..., n.d.). The key advantage is decentralisation, which eliminates a single point of failure, as DDS does not require

a central broker. Data is published and received directly between nodes. The disadvantages of DDS include the complexity of configuration and compatibility issues between implementations from different vendors.

The MQTT (Message Queuing Telemetry Transport) protocol is designed to collect data from devices and send it to a server (MQTT version 5.0, n.d.). This protocol can collect data from different devices and works on top of the Transmission Control Protocol (TCP) to ensure reliable data transfer and prevent data loss. The key disadvantage of this protocol is that it is not suitable for devices with limited resources. The CoAP (Constrained Application Protocol) protocol is analogous to the HTTP (Hypertext Transfer Protocol) protocol. However, HTTP has greater computational complexity, lower data transfer rates, and higher energy consumption than CoAP. CoAP was developed as a special data transfer protocol for web networks. CoAP offers the possibility of use by nodes with limited resources (Shelby *et al.*, 2014). Although HTTP is versatile, it has higher overhead and longer connection establishment times compared to lightweight IoT protocols such as CoAP. It provides independence and scalability of network components; HTTP is a suitable option for organising interaction between nodes in a local network. AMQP (Advanced Message Queuing Protocol). This protocol is designed to organise interaction between servers. AMQP is implemented by two key components. The data exchange component routes messages and stores them in a queue. The message queue is used to store messages until they are sent to their destination; the second component, the TCP protocol, is used by providers to interact with subscribers and provides a reliable point-to-point connection.

The study considered the most widespread applied-tier protocols. Table 1 presents a comparative overview of key characteristics and communication approaches (client-server or publisher-subscriber architecture), allowing their suitability for various data collection and processing scenarios to be assessed. DDS provides decentralised interaction without a broker, while MQTT, AMQP, and XMPP are based on the presence of a server or broker. HTTP and CoAP offer a client-server (request-response) model, but CoAP is designed for substantially limited devices and typically runs on top of UDP (User Datagram Protocol), providing lower bandwidth costs.

Table 1. Comparative review of applied-tier protocols

Protocol	Interaction model	Transport	Advantages	Disadvantages
DDS	Decentralised "pub-sub"	UDP/TCP (RTPS)	No single point of failure. Rich set of QoS. Real-time support.	Complexity of configuration. Different versions may be incompatible.
MQTT	Broker-oriented "pub-sub"	TCP	Ease of implementation. Widespread popularity in IoT.	Not sufficiently optimised for devices with ultra-low resources. Requires a broker.
CoAP	Lightweight client-server model (REST)	UDP (often)	Low overhead costs. Optimised for limited resources.	Fewer delivery guarantees. Some security features require additional extensions.

Table 1. Continued

Protocol	Interaction model	Transport	Advantages	Disadvantages
HTTP	"Client-server" (request-response)	TCP	Universal web standard. Easy to integrate.	Relatively high load.
AMQP	"Pub-sub", queues, advanced routing	TCP	Powerful shift model. Flexible route settings.	Complex protocol. Requires more resources. More frequently used at the server-to-server level.

Source: developed by the author of this study based on the experimental research results

The information in the table about DDS, MQTT, CoAP, HTTP, and AMQP protocols reflects various approaches to organising the application tier of interaction in IoT systems and distributed environments. DDS is characterised by a decentralised pub-sub model, which avoids a single point of failure and ensures real-time operation thanks to flexible QoS mechanisms. MQTT, on the other hand, relies on a broker-oriented pub-sub scheme, where all messages pass through a central broker. This greatly simplifies implementation and administration but can create a single point of failure.

CoAP represents a client-server architecture adapted for resource-constrained environments. CoAP operates primarily over UDP, using acknowledged and unacknowledged messages, which reduces network overhead. HTTP, compared to CoAP, is more versatile and understandable, but has high overhead due to TCP connections and a more complex transmission scheme, which can be critical for microcontroller devices with minimal computing resources. AMQP is mostly used in enterprise systems that require advanced routing and queuing mechanisms. It supports both the pub-sub model and queuing but requires more resources and is typically not used directly on low-power devices, but rather on servers. AMQP is often chosen for systems where finely tuned message distribution is critical.

The choice between DDS, MQTT, CoAP, HTTP, or AMQP depends on several key factors: performance requirements, reliability level, available network resources, and security and scalability needs. DDS is best suited for real-time industrial applications, MQTT is a versatile solution for most IoT scenarios, while CoAP is an option for resource-constrained sensor networks. HTTP is the basic protocol for web-oriented applications, but it is somewhat excessive in resource-constrained environments. AMQP is the best solution when detailed routing and queue management are required.

Internet of Things data exchange and collection technologies

The technology for exchanging information between machines on the Internet of Things is called 'M2M' (Machine

to Machine). It allows devices such as mobile phones to send and receive data via cellular networks. Satellites are mostly used for communication in industrial environments where small amounts of data must be transmitted. The Wi-Fi protocol is used to provide Internet access to devices within the coverage area of a wireless network. Wi-Fi is widely used in Internet of Things systems. The use of Wi-Fi offers advantages such as security, integrity, flexibility, and scalability in data transmission, and this connection is based on the use of IP (Internet Protocol) protocols (Farhan *et al.*, 2021).

Bluetooth technology allows transferring data between two or more devices in close proximity to each other in conditions that do not require high network bandwidth. It also provides service discovery and configuration to regulate the interaction between devices. Bluetooth provides network addresses and sets permissions for data access. Secure data transmission is guaranteed by the Bluetooth protocol. Radio frequency communication is one of the simplest forms of communication between devices; both ZigBee and ZWave use low-power radio frequency receivers embedded in electronic devices and systems. In the Internet of Things, these technologies offer considerable advantages such as low power consumption, high security, and network scalability.

RFID technology uses electromagnetic fields to identify objects. In this case, a device is installed that reads the data stored in an RFID tag, which can be used for many applications. NFC (near field communication) technology uses magnetic field induction for communication, where two frame antennas are placed in a magnetic field close to each other, effectively forming an air-core transformer. The key disadvantage of this technology is the relatively low security of data transmission and the possibility of unauthorised connections (Onumadu & Abroshan, 2024). Below is a comparative analysis of the key wireless technologies and protocols used in the Internet of Things, including Wi-Fi, Bluetooth, ZigBee, LoRaWAN, RFID, NFC, etc. (Table 2).

Table 2. Comparison table of wireless technologies for IoT

Technology	Operating range/radius	Throughput	Area of application	Key benefits	Key disadvantages
M2M	Variable (satellite, cellular).	Depends on the operator (2G/3G/4G/5G, etc.).	Industrial systems, remote monitoring.	Global coverage. Autonomous data exchange.	Cost depends on tariffs. Often requires licensed bands.

Table 2. Continued

Technology	Operating range/radius	Throughput	Area of application	Key benefits	Key disadvantages
Wi-Fi	~30-100 m (depending on standard and obstacles).	High (tens to hundreds of Mbps).	Home/office IoT systems, smart homes, video surveillance.	High speed. Widespread infrastructure.	Relatively high energy consumption. Shorter range compared to some LPWANs.
Bluetooth	~1-100 m (depending on class and version).	Low/medium (< 24 Mbps for Bluetooth 5).	Wearable devices, personal gadgets, audio devices.	Low power consumption (especially BLE). Easy connectivity.	Limited range. Only suitable for local applications.
ZigBee / Z-Wave	~10-100 m (mesh-network).	Low (up to ~250 Kbps for ZigBee).	Smart homes, industrial sensor networks, lighting.	Low energy consumption. Mesh topology (self-organising nodes).	Low bandwidth. Requires a gateway to connect to the Internet.
LoRaWAN	Up to 2-5 km in urban conditions; up to 15 km in rural areas.	Very low (0.3-50 Kbps).	Distributed sensor networks, smart meters, agriculture.	Very low power consumption. Wide coverage area. Operates in the free frequency band (ISM).	Low transmission speed. Not suitable for large amounts of data. Dependence on the density of base stations.
RFID	Several centimetres to several metres.	Very low (only identification data).	Logistics, access control, inventory.	Affordable tags. Passive operation (no battery required).	Very short distance. Small amount of data stored.
NFC	Up to ~10 cm	Very low (identification data).	Contactless payments, fast identification.	Easy interaction with smartphones. Quick device pairing.	Extremely short distance. Potential vulnerability (possibility of unauthorised connections).
Satellite communications	Global (orbit coverage).	Low/medium.	Remote industrial facilities, transport, sea vessels.	Large coverage area. Access to hard-to-reach regions.	High cost. Signal delays. Low speed compared to terrestrial networks.

Source: developed by the author of this study based on the experimental research results

The study examined Internet of Things data collection technologies, including cloud computing and middleware technologies. Cloud computing is a web technology that provides end-user applications with access to distributed computing resources. It ensures data storage and processing capabilities, enabling access to resources as a local service. As a result, computing resources, data storage, and other services are located between the user's device and the cloud server. A prominent advantage is that cloud computing nodes are closer to the end user, which substantially reduces data transfer delays. Furthermore, the spatial distribution of resources allows using these technologies effectively in mobile device networks.

Another significant technology is middleware platforms, which act as an interface between the hardware and application levels. They enable interaction with devices, manage information, and provide a unified model for working with various devices. Such platforms also help solve the problem of device distribution that arises when end users access them. Many sensors, which form an integral part of smart objects, are widely used to collect data in Internet of Things applications. They are affordable, consume minimal energy, and are characterised by battery capacity and ease of deployment. Since data collection, transmission, and

storage require a strong level of security, encryption methods play a critical role. Their key purpose is to prevent unauthorised monitoring, access, misuse, and modification of information. At the same time, conventional cryptographic algorithms are not always suitable for resource-constrained devices. For this purpose, lightweight encryption methods are used, which enable data protection even on devices with low computing power (He *et al.*, 2024).

Apart from encryption, social network analysis methods are an effective tool for analysing information in the Internet of Things. They allow establishing connections between objects based on human interactions in the digital environment. These "social connections" can be used to search for and collect data on the Internet. A major challenge in the Internet of Things is the transmission of massive data flows. In this context, routing methods play a vital role, ensuring the secure and efficient transport of large amounts of information from devices to storage facilities, thereby ensuring the reliable operation of the entire system.

Security and data management issues in IP systems

Many security problems in IP systems are related to the limited computing and energy resources of the devices used. Many of these devices are mobile, which determines the dynamic nature of the network. Data transmission

delays often occur in networks with a dynamic structure. Therefore, it is necessary to ensure network security, considering the possibility of data transmission delays and energy resource limitations. Security issues can be addressed through node authentication and identification and data access restrictions. Data management issues include data routing problems during data transmission and data storage problems. The complexity of solving these problems is determined by the large amounts of data coming from devices in modern IP networks. The increase in data volume is achieved by expanding the components of the data collection system. Threats can arise at different architectural tiers of the IP system, such as threats to the security of the IS system at the physical tier, which is the base tier of the IS system. At this tier, there are physical and virtual objects. Physical objects include various sensors, including wireless ones. At the physical tier, various data transmission technologies and protocols are used, such as the ZigBee protocol, RFID tags, QR codes, and GPS systems for positioning mobile devices. The key function of this tier is to collect data about environmental objects and the environment itself. For instance, parameters such as wind speed, acidity, humidity, and vibration can be measured.

With the increase in the number of devices connected to a dynamic IP network, new data management challenges arise, such as data collection, classification, compression, and archiving. In a dynamic network, the configuration of devices is constantly changing. New elements are connected to the network, while others leave it. Devices generate large amounts of data that must be processed and stored. These tasks are handled by data management systems. The need to provide real-time access to data and high requirements for response times to user requests impose limitations on the amount of data that can be processed and stored in modern systems. To achieve network scalability, lightweight infrastructures must be created and used. When designing such an infrastructure, it is necessary to factor in the security requirements for data collection and storage, as well as current resource constraints, including energy.

The primary cause of threats is the dynamic change in the network structure caused by the changing location of mobile devices that form part of the network. Attackers exploit changes in the network structure to organise their attacks. Attacks on the network at the physical level can include denial-of-service (DDoS) attacks and network eavesdropping. Attacks may also include the capture of nodes or the addition of malicious nodes. Attacks via side channels are also possible. Attacks may be aimed at gaining unauthorised access to data, disrupting data routing, or disrupting normal customer service, which may result in data integrity violations or network overload. There are also threats to the security of IP systems at the network level. At this level, data is transmitted over wired and wireless networks, where data routing tasks are performed. IEEE 802.15 (n.d.) is used for data transmission, which is a technical standard that defines the operation of low-speed wireless personal area networks (LR-WPAN), 3G, 4G,

LTE, Wi-Fi, ZigBee, and MQTT, but the choice of a concrete technology depends on the devices used and the type of data to be transmitted.

At the network tier, secure connection of IP devices is ensured by various communication protocols, network interfaces, and communication channels. There are various network management mechanisms. As a result, secure data transmission from the physical tier to the data processing tier must be guaranteed. This tier is crucial for data management, network intrusion detection, and network traffic coordination in IP systems. Attacks by malicious actors at the network tier that can compromise data security include denial of service (Al-Hadhrani & Hussain, 2021), man-in-the-middle attacks, attacks on gateways, attacks on data stores, as well as attacks based on traffic analysis and spoofing attacks. DDoS/DoS attacks can cause destructive changes to network components and threaten the security of Wi-Fi connections. In a man-in-the-middle attack, an attacker transmits data or alters the connection between two nodes, believing that they are communicating directly.

As a result, the nodes lose access to the necessary information. Attackers often target gateways. Gateways are a crucial element of IP systems. If an attack on a gateway is successful, the attacker gains control over the part of the IP system where the gateway acts as a bridge between the device and the cloud. Therefore, network-level security consists of ensuring the authenticity, confidentiality, and integrity of data as it is transmitted over the network. Encryption algorithms based on cryptographic keys, node authentication mechanisms, and intrusion detection tools can be used to protect the network from attacks. The data processing layer is part of the network layer. This level hosts the middleware platform. Functions at this level include the implementation of cloud computing, the distribution of data processing between nodes and cloud services, and data storage using various database management systems (DBMS).

The data management tier processes and stores heterogeneous data from devices. This includes web data, documents, sensor data, and emails. This data is irrelevant, and various formats are used to represent it. Processing and analysing such data require the use of intelligent data analysis and machine learning methods (Mahdaviinejad *et al.*, 2018). These methods can be employed to extract the data needed to solve an application problem from the original information flow. The resulting data can be placed in temporary storage on a node or transferred to the cloud for long-term storage. Cloud nodes can process device data in real time, and processing on nodes involves data collection. Cloud services allow processing, analysing, and providing stored data to end users on demand (Abdulkaareem *et al.*, 2021). Thus, cloud nodes in combination with cloud services enable short-term and long-term data storage, processing, and analysis of stored data in real time.

Attacks on cloud services and nodes are among the most widespread types of attacks. Data transmitted and stored in the cloud is at the greatest risk (Somani *et al.*, 2017). At the

processing tier, the primary types of attacks include denial-of-service (DoS) attacks and man-in-the-middle attacks. In conclusion, attackers gain control over shared network resources and can gain unauthorised access to data through hidden paths. Most services provided in the cloud on demand are accessible via the Internet (Younis *et al.*, 2024). In this case, attackers can use the Internet to penetrate IP networks. Once inside the network, they can perform actions aimed at disrupting the operation of the network or its elements. Therefore, the problem of protecting data stored in the cloud can be considered one of the key security challenges for IP systems. There are also threats to the security of IP systems at the application level. Various applications for creating an intelligent environment are located at the application level. At this level, intelligent applications are managed and accessible to end users.

Data management at the application tier must ensure that data is provided to end users. Management is based on the requirements of the application running in real time. For example, healthcare and transportation applications must have access to operational data and process it in real time. Delays in data delivery can have severe negative consequences. Applications that collect information about people's daily lives, such as travel routes, shopping habits, and daily energy consumption, do not have strict processing time requirements. However, because the data is sensitive, security requirements for its collection, transmission, and processing still apply. At the application tier, security issues are caused by the creation and use of computer viruses, worms, Trojan horses, spyware, and other malicious software. Attacks at this level can lead to data interception, incorrect data exchange, and denial of service. Successful attacks also result in the attacker being recognised by the system as a legitimate user and gaining unrestricted access to applications and data containing confidential information. Data and application security at the application level requires preventing unauthorised access and use of data, as well as restoring the functionality of applications after an attack.

To ensure security, vulnerabilities that may exist in applications should be identified and access rights to data should be restricted. Data encryption methods can also be used, but their use leads to an extensive increase in data transfer delays. Other security mechanisms include firewalls, intrusion detection systems (IDS), and antivirus software. As a rule, when choosing a method of protecting applications and data at the application tier, the risks of various threats and the probability of successful attacks by malicious actors should be assessed.

The reviews by M.F. Usmani (2021) and D. Dinculeană & X. Cheng (2019) are most relevant to the subject under study, where the researchers also analysed various application-tier protocols in IoT networks. Specifically, M.F. Usmani (2021) emphasised that MQTT is most often used to collect telemetry from many nodes due to its simplicity and low overhead. The researchers' findings confirmed these conclusions, as MQTT truly provides effective

data collection in device-to-cloud scenarios. At the same time, analogously to the comments of D. Dinculeană & X. Cheng (2019), the present study also noted that MQTT is not always suitable for devices with extremely limited resources, as it still requires a persistent connection (TCP) and the presence of a broker, as well as the use of TLS protocols for data protection.

G. Sciangula *et al.* (2023) or M.-H. Ho *et al.* (2022) offered a slightly distinct perspective, where the use of DDS in real-time mode was evaluated. The researchers concluded that DDS has advantages in decentralisation and provides lower latency compared to the conventional client-server model. The present review confirmed this thesis: according to (Internal documentation..., n.d.), DDS allows publishers and subscribers to interact without a central broker, eliminating the single point of failure. However, the complexity of DDS configuration and possible differences in implementations by different vendors (RTI, OpenDDS, etc.) may be a serious obstacle to mass adoption. S.K. Routray *et al.* (2023) focused on CoAP and demonstrated that this protocol is the optimum solution for resource-constrained devices due to its simpler structure (UDP) and lower power consumption compared to HTTP. The researchers confirmed that CoAP does have lower overhead costs, which is consistent with the official specifications of Z. Shelby *et al.* (2014). The issue of scalability was also considered. The researchers found that CoAP works better in small networks, while in large-scale environments, delays may occur due to large amounts of broadcast traffic.

M. Saban *et al.* (2021) and Z. Liu *et al.* (2022) addressed low-power protocols (LoRaWAN, ZigBee). They suggested that LoRa and ZigBee were among the best options for remote sensor nodes due to their wide range and very low power consumption. The presented review confirmed this opinion: Wi-Fi is suitable where greater throughput is required and a constant power supply is available, while LoRa allows collecting small amounts of data over long distances. However, the findings revealed that LoRa is not always capable of providing low latency, which can be a problem for some mobile applications.

In terms of security, A. Alrawais (2020) highlighted the potential risks of using NFC, particularly the danger of unauthorised connections. The researcher's analysis found that NFC does have a short range, but insufficient protection against 'over-the-shoulder' attacks (especially without additional encryption measures). This position is consistent with that of P. Onumadu & H. Abroshan (2024), who, in their study of technologies for IoT, also emphasised the significance of security protocols for data transmission in the near field. Most researchers agree on the absence of a single 'universal' technology for data collection and transmission in IoT. The conclusion remains that the choice of protocol or technology (MQTT, DDS, CoAP, Wi-Fi, LoRaWAN, etc.) depends on the use case, among which the following are decisive: access to energy resources, requirements for data speed and volume, permissible delay and real-time criticality, and the level of security and safety.

Thus, the literature review and comparison with previous studies showed that the presented findings were consistent with the findings of other researchers (Usmani, 2021; Ho *et al.*, 2022; Sciangula *et al.*, 2023) regarding the effectiveness of DDS in real time and the suitability of MQTT for telemetry. They also confirmed the findings of M. Saban *et al.* (2021) and Z. Liu *et al.* (2022) on the advantages of LoRaWAN and ZigBee in energy-efficient sensor networks. The views of A. Alrawais (2020) and the position of P. Onumadu & H. Abroshan (2024) on NFC security risks coincided with the reservations presented in the current study on the low security of near-field technologies. The difference in emphasis (QoS, energy consumption, security) indicated that the subject of IoT communication requires further investigation, particularly in the context of integrated hybrid solutions. This will enable developers to tailor data collection systems to the concrete requirements of a particular industry, increasing their flexibility, scalability, and reliability.

Conclusions

The study analysed five key application-level protocols (DDS, MQTT, CoAP, HTTP, and AMQP). It found that decentralised DDS eliminates a single point of failure but is characterised by complex configuration. MQTT simplifies telemetry but requires a broker and a persistent TCP connection. CoAP is optimised for resource-constrained nodes but provides fewer delivery guarantees. HTTP continues to be versatile but creates additional overhead. AMQP provides advanced queuing mechanisms but requires more hardware resources. It was found that the choice of protocol depends on the use case and the acceptable energy consumption, data transfer volumes and delays. It was established that Wi-Fi, Bluetooth, ZigBee, LoRaWAN, RFID, and NFC wireless technologies cover different ranges, bandwidths, and energy consumption levels. None of them is universal for all applications. This was confirmed by the feasibility of developing and using hybrid approaches that combine the advantages of several protocols and technologies depending on the environmental conditions. The study also analysed the concept of M2M (machine communication via cellular and satellite networks). The comparison showed that none of the existing methods and protocols for collecting data in the Internet of Things (including DDS, MQTT, CoAP, LoRaWAN, ZigBee, etc.) would be capable of simultaneously meeting all the key requirements of mobile IoT systems by 2025: high security, minimal

delays and energy efficiency. In real dynamic environments, a compromise must be found between data exchange speed, energy consumption, and security level, which complicates the development of a universal solution.

The conclusions obtained suggest the need for a comprehensive, multi-stage approach to the implementation of mobile IoT networks: protocols must be optimised depending on the physical environment, specialised tools must be used to reduce energy consumption, and mechanisms to counter unauthorised actions must be put in place. Specifically, for nodes with critically low resources and a need for energy-saving modes, it is advisable to use protocols such as CoAP or LoRaWAN, while in cases where fast response and scalability are critical, DDS-based approaches with decentralised interaction may be justified. For telemetry with many sensors and reliance on a broker-oriented “publisher-subscriber” scheme (MQTT), server performance and connection security requirements (TLS, key management, etc.) should be factored in. It is also advisable to provide additional security mechanisms (authentication, encryption, intrusion detection), as security measures have substantially influence latency and energy consumption. Further research should focus on three inter-related tasks. First, empirical verification of the effectiveness of DDS, MQTT, and CoAP protocols in mobile IoT networks with dynamic topology is required, with a particular focus on complex security mechanisms, which are currently largely unexplored. Second, the development of hybrid architectures combining publish-subscribe protocols (DDS, MQTT) with lightweight transport solutions (CoAP) and low-power LPWAN technologies (e.g., LoRaWAN) is promising; this approach will simultaneously ensure decentralisation, scalability, and energy efficiency in large sensor networks. Third, these integrated solutions must be tested in real-world dynamic environments to better understand the trade-offs between security, latency, and resource efficiency, thereby increasing the flexibility of IoT deployment in various application areas.

Acknowledgements

None.

Funding

None.

Conflict of Interest

None.

References

- [1] Abdulkareem, N.M., Zeebaree, S.R.M., Sadeeq, M.A.M., Ahmed, D.M., Sami, A.S., & Zebari, R.R. (2021). IoT and cloud computing issues, challenges and opportunities: A review. *Qubahan Academic Journal*, 1(2), 1-7. doi: 10.48161/QAJ.V1N2A36.
- [2] Al-Hadhrani, Y., & Hussain, F.K. (2021). DDoS attacks in IoT networks: A comprehensive systematic literature review. *World Wide Web*, 24, 971-1001. doi: 10.1007/s11280-020-00855-2.
- [3] Alrawais, A. (2020). Security issues in near field communications (NFC). *International Journal of Advanced Computer Science and Applications*, 11(11), 621-628. doi: 10.14569/IJACSA.2020.0111176.
- [4] Chanal, P.M., & Kakkasageri, M.S. (2020). Security and privacy in IoT: A survey. *Wireless Personal Communications*, 115, 1667-1693. doi: 10.1007/S11277-020-07649-9.

- [5] Dinculeană, D., & Cheng, X. (2019). Vulnerabilities and limitations of MQTT protocol used between IoT devices. *Applied Sciences*, 9(5), article number 848. doi: [10.3390/APP9050848](https://doi.org/10.3390/APP9050848).
- [6] Farhan, L., Hameed, R.S., Ahmed, A.S., Fadel, A.H., Gheth, W., Alzubaidi, L., Fadhel, M.A., & Al-Amidie, M. (2021). Energy efficiency for green Internet of Things (IoT) networks: A survey. *Network*, 1(3), 279-314. doi: [10.3390/network1030017](https://doi.org/10.3390/network1030017).
- [7] Haseeb, K., Almogren, A., Islam, N., Din, I.U., & Jan, Z. (2019). An energy-efficient and secure routing protocol for intrusion avoidance in IoT-based WSN. *Energies*, 12(21), article number 4174. doi: [10.3390/EN12214174](https://doi.org/10.3390/EN12214174).
- [8] He, P., Zhou, Y., & Qin, X. (2024). A survey on energy-aware security mechanisms for the Internet of Things. *Future Internet*, 16(4), article number 128. doi: [10.3390/FI16040128](https://doi.org/10.3390/FI16040128).
- [9] Ho, M.-H., Lai, M.-Y., & Liu, Y.-T. (2022). Implementation of DDS cloud platform for real-time data acquisition of sensors for a legacy machine. *Electronics*, 11(13), article number 2096. doi: [10.3390/ELECTRONICS11132096](https://doi.org/10.3390/ELECTRONICS11132096).
- [10] Honcharenko, D., Mokin, V., & Protsenko, D. (2023). Building an information system for monitoring physical indicators based on the internet of things technology. *Information Technologies and Computer Engineering*, 52(2), 99-108. doi: [10.31649/1999-9941-2023-57-2-99-108](https://doi.org/10.31649/1999-9941-2023-57-2-99-108).
- [11] Huang, Y., Liu, Y., Song, J., & Meng, W. (2024). A lightweight and efficient raw data collection scheme for IoT systems. *Journal of Information and Intelligence*, 2(3), 209-223. doi: [10.1016/J.JIIXD.2024.03.004](https://doi.org/10.1016/J.JIIXD.2024.03.004).
- [12] IEEE 802.15.4. (n.d.). Retrieved from <https://docs.nordicsemi.com/bundle/ncs-latest/page/zephyr/connectivity/networking/api/ieee802154.html>.
- [13] Internal documentation – OpenDDS 3.28.0. (n.d.). Retrieved from <https://opendds.readthedocs.io/en/dds-3.28/internal/index.html>.
- [14] Ivanov, D., Alekseienco, V., & Yarmolenko, T. (2023). Features of the interdependence of Big Data and Internet of Things technologies. *Bulletin of Priazovsky State Technical University*, 46, 13-27. doi: [10.31498/2225-6733.46.2023.288096](https://doi.org/10.31498/2225-6733.46.2023.288096).
- [15] Kasiviswanathan, S., Gnanasekaran, S., Thangamuthu, M., & Rakkiyannan, J. (2024). Machine-learning- and Internet-of-Things-driven techniques for monitoring tool wear in machining process: A comprehensive review. *Journal of Sensor and Actuator Networks*, 13(5), article number 53. doi: [10.3390/JSAN13050053](https://doi.org/10.3390/JSAN13050053).
- [16] Khadam, U., Davidsson, P., & Spalazzese, R. (2024). Exploring the role of Artificial Intelligence in Internet of Things systems: A systematic mapping study. *Sensors*, 24(20), article number 6511. doi: [10.3390/S24206511](https://doi.org/10.3390/S24206511).
- [17] Khare, S., & Totaro, M. (2019). Big Data in IoT. In *2019 10th international conference on computing, communication and networking technologies, ICCCNT* (pp. 1-7). Kanpur: India. doi: [10.1109/ICCCNT45670.2019.8944495](https://doi.org/10.1109/ICCCNT45670.2019.8944495).
- [18] Liu, Z., Li, Y., Zhao, L., Liang, R., & Wang, P. (2022). Comparative evaluation of the performance of ZigBee and LoRa wireless networks in building environment. *Electronics*, 11(21), article number 3560. doi: [10.3390/ELECTRONICS11213560](https://doi.org/10.3390/ELECTRONICS11213560).
- [19] Mahdavinejad, M.S., Rezvan, M., Barekatain, M., Adibi, P., Barnaghi, P., & Sheth, A.P. (2018). Machine learning for Internet of Things data analysis: A survey. *Digital Communications and Networks*, 4(3), 161-175. doi: [10.1016/J.DCAN.2017.10.002](https://doi.org/10.1016/J.DCAN.2017.10.002).
- [20] Mishra, N., & Pandya, S. (2021). Internet of Things applications, security challenges, attacks, intrusion detection, and future visions: A systematic review. *IEEE Access*, 9, 59353-59377. doi: [10.1109/ACCESS.2021.3073408](https://doi.org/10.1109/ACCESS.2021.3073408).
- [21] MQTT version 5.0. (n.d.). Retrieved from <https://docs.oasis-open.org/mqtt/mqtt/v5.0/mqtt-v5.0.html>.
- [22] Onumadu, P., & Abroshan, H. (2024). Near-Field Communication (NFC) cyber threats and mitigation solutions in payment transactions: A review. *Sensors*, 24(23), article number 7423. doi: [10.3390/S24237423](https://doi.org/10.3390/S24237423).
- [23] Peyman, M., Copado, P.J., Tordecilla, R.D., Martins, L.D.C., Xhafa, F., & Juan, A.A. (2021). Edge computing and IoT analytics for agile optimization in intelligent transportation systems. *Energies*, 14(19), article number 6309. doi: [10.3390/EN14196309](https://doi.org/10.3390/EN14196309).
- [24] Prasad, V.K., Dansana, D., Bhavsar, M.D., Acharya, B., Gerogiannis, V.C., & Kanavos, A. (2023). Efficient resource utilization in IoT and cloud computing. *Information*, 14(11), article number 619. doi: [10.3390/INFO14110619](https://doi.org/10.3390/INFO14110619).
- [25] Routray, S.K., Pappa, M., Jha, M.K., & Sharmila, K.P. (2023). Performance of CoAP in the limited resource environment: An empirical study. In *5th international conference on inventive research in computing applications, ICIRCA* (pp. 510-515). Coimbatore: India. doi: [10.1109/ICIRCA57980.2023.10220891](https://doi.org/10.1109/ICIRCA57980.2023.10220891).
- [26] Saban, M., Aghzout, O., Medus, L.D., & Rosado, A. (2021). Experimental analysis of IoT networks based on LoRa/LoRaWAN under indoor and outdoor environments: Performance and limitations. *IFAC-PapersOnLine*, 54(4), 159-164. doi: [10.1016/J.IFACOL.2021.10.027](https://doi.org/10.1016/J.IFACOL.2021.10.027).
- [27] Sciangula, G., Casini, D., Biondi, A., Scordino, C., & Di Natale, M. (2023). Bounding the data-delivery latency of DDS messages in real-time applications. In *35th Euromicro conference on real-time systems (ECRTS 2023). Leibniz international proceedings in informatics (LIPIcs)* (pp. 9:1-9:26). Vienna: Austria. doi: [10.4230/LIPICS.ECRTS.2023.9](https://doi.org/10.4230/LIPICS.ECRTS.2023.9).
- [28] Shelby, Z., Hartke, K., & Bormann, C. (2014). *The Constrained Application Protocol (CoAP)*. doi: [10.17487/RFC7252](https://doi.org/10.17487/RFC7252).
- [29] Somani, G., Gaur, M.S., Sanghi, D., Conti, M., & Buyya, R. (2017). DDoS attacks in cloud computing: Issues, taxonomy, and future directions. *Computer Communications*, 107, 30-48. doi: [10.1016/J.COMCOM.2017.03.010](https://doi.org/10.1016/J.COMCOM.2017.03.010).

- [30] Supreeth, S., & Patil, K. (2022). VM Scheduling for Efficient Dynamically Migrated Virtual Machines (VMS-EDMVM) in cloud computing environment. *KSII Transactions on Internet and Information Systems*, 16(6), 1892-1912. [doi: 10.3837/TIIS.2022.06.007](https://doi.org/10.3837/TIIS.2022.06.007).
- [31] Usmani, M.F. (2021). *MQTT protocol for the IoT – review paper*. Frankfurt: Frankfurt University of Applied Sciences. [doi: 10.13140/RG.2.2.26065.10088](https://doi.org/10.13140/RG.2.2.26065.10088).
- [32] Waheed, S., Hanif, S., Hafeez, R., Sharif, M.I., Siddique, K., & Akhtar, Z. (2024). A comprehensive survey on applications, challenges, threats and solutions in IoT environment and architecture. *Journal of Computer Science*, 20(3), 310-332. [doi: 10.3844/JCSSP.2024.310.332](https://doi.org/10.3844/JCSSP.2024.310.332).
- [33] Younis, R., Iqbal, M., Munir, K., Javed, M.A., Haris, M., & Alahmari, S. (2024). A comprehensive analysis of cloud service models: IaaS, PaaS, and SaaS in the context of emerging technologies and trend. In *5th international conference on electrical, communication and computer engineering, ICECCE* (pp. 1-6). Kuala Lumpur: Malaysia. [doi: 10.1109/ICECCE63537.2024.10823401](https://doi.org/10.1109/ICECCE63537.2024.10823401).
- [34] Youssef, W.E.H., Abdelli, A., Kharroubi, F., Dridi, F., Khriji, L., Ahshan, R., Machhout, M., Nengroo, S.H., & Lee, S. (2023). A secure chaos-based lightweight cryptosystem for the Internet of Things. *IEEE Access*, 11, 123279-123294. [doi: 10.1109/ACCESS.2023.3326476](https://doi.org/10.1109/ACCESS.2023.3326476).

Аналіз основних моделей, методів та засобів збору даних в Інтернеті речей

Максим Савка

Аспірант

Національний технічний університет України «Київський політехнічний інститут імені Ігоря Сікорського»
03056, просп. Берестейський, 37, м. Київ, Україна
<https://orcid.org/0000-0003-2049-9438>

Анотація. Швидкий розвиток мобільних технологій та технологічна революція сприяють активному впровадженню Інтернету речей, що веде до зростання кількості фізичних пристроїв, підключених до Інтернету. Однак поряд із перевагами Інтернету речей зростають і недоліки, зокрема кіберзагроза, порушення приватності та ризики несанкціонованого збору даних. У зв'язку з цим дослідження викликів та заходів захисту в сфері Інтернету речей є необхідним для забезпечення безпечного впровадження цих технологій. Метою дослідження було проаналізувати ключові моделі, методи та інструменти збору даних у середовищі Інтернету речей. Методологічною основою роботи був системний підхід, який дозволив представити державну політику у сфері Інтернету речей як складну, структурно організовану систему, що потребує вдосконалення та розвитку. В роботі застосовувались загальнонаукові та специфічні методи (системний, абстрагування та порівняння, формалізації, інституціонального аналізу, історико-генетичний, логіко-семантичний, класифікації, узагальнення, аналізу, формалізації та порівняння), а також широко і послідовно використовувалися статистичні дані. Встановлено, що серед існуючих технологій збору даних найширше застосовуються сенсорні мережі й хмарні обчислення, але вони не покривають усіх вимог мобільних систем Інтернету речей, зокрема щодо енергоефективності, пропускну здатності й безпеки. Виявлено, що сучасні архітектурні рішення (у тому числі соціальні архітектури Інтернету речей) дають змогу масштабувати систему та підключати різноманітні пристрої, однак мають обмеження, пов'язані зі складністю розгортання та пропускну здатністю. З'ясовано, що основні загрози безпеці виникають на всіх рівнях мобільних мереж Інтернету речей – від пристроїв до хмарної інфраструктури, що підтверджує необхідність комплексного підходу до їх захисту. Показано, що переважна більшість наявних рішень не задовольняє одночасно вимоги до низького енергоспоживання, швидкості збору й обробки даних у мережах із динамічною топологією, а також вимоги до надійності й конфіденційності. Запропоновані напрацювання доповнюють існуючі методи, посилюючи гнучкість і стійкість процесу збирання даних та створюючи підґрунтя для розробки нових енергоефективних і безпечних засобів в Інтернеті речей

Ключові слова: хмарні обчислення; архітектура IoT; безпека даних; протокол CoAP; fog-обчислення



Study of the effect of excipients on the pharmacotechnological attributes of ribavirin capsules using the random balance method

Olena Saliy

PhD in Pharmaceutical Sciences, Associate Professor
Kyiv National University of Technologies and Design
01011, 2 Mala Shyianovska Str., Kyiv, Ukraine
<https://orcid.org/0000-0001-7103-2083>

Andrii Bublyk

Master
Kyiv National University of Technologies and Design
01011, 2 Mala Shyianovska Str., Kyiv, Ukraine
<https://orcid.org/0009-0002-6495-9969>

Mariia Popova*

Postgraduate Student
Kyiv National University of Technologies and Design
01011, 2 Mala Shyianovska Str., Kyiv, Ukraine
<https://orcid.org/0000-0002-2579-0331>

Larysa Fukleva

PhD in Pharmaceutical Sciences, Senior Lecturer
Zaporizhzhia State Medical Pharmaceutical University
69035, 26 Maria Prymachenko Ave., Zaporizhzhia, Ukraine
<https://orcid.org/0000-0002-2930-0619>

Abstract. One of the main stages in the creation of solid pharmaceutical forms is pharmaceutical development, during which the quality, efficacy, and safety of the future drug are established. The selection of biopharmaceutical factors when planning technological research is a critical task in pharmaceutical development, complicated by the need for a comprehensive consideration of all influencing factors. The aim of this work was to investigate and analyse the impact of the quantitative characteristics of excipients on the quality indicators of powder masses and hard capsules containing ribavirin. Mathematical experimental planning was applied, enabling the creation of an optimal composition by identifying the relationships between factors and technological parameters that must ensure the pharmacological action of ribavirin. The random balance method was employed to select optimal excipients for inclusion in the composition of gelatin capsules. Based on the obtained statistical data, scatter diagrams were constructed, where the median magnitude indicated the significance of the factors' influence on the corresponding quality characteristics. The study examined the effect of quantitative factors on the properties of capsule masses and hard capsules with ribavirin, manufactured using wet granulation. It was found that quantitative characteristics at both "high" and "low" levels did not demonstrate a significant effect on indicators such as "Bulk density", "Tapped density" for capsule masses, and "Disintegration" for hard capsules. The greatest influence of excipients on the Hausner ratio (flowability) for powder masses and the uniformity of the capsule content mass was exerted by binding agents such as hydroxypropyl cellulose and povidone. Increased amounts of these binders improved these characteristics. Scatter diagrams showed that the loss in mass during the drying of capsule masses significantly improved with an increased amount of microcrystalline

Suggested Citation:

Saliy, O., Bublyk, A., Popova, M., & Fukleva, L. (2025). Study of the effect of excipients on the pharmacotechnological attributes of ribavirin capsules using the random balance method. *Technologies and Engineering*, 26(2), 79-88. doi: 10.30857/2786-5371.2025.2.7.

*Corresponding author



cellulose and a reduced content of croscopovidone. Factors affecting the “Dissolution” parameter were identified, including an increased amount of croscarmellose sodium and a reduced content of magnesium stearate and povidone

Keywords: direct-acting antiviral drugs; Design of Experiments; bulk density; Hausner ratio; capsule disintegration; capsule dissolution

Introduction

Modern pharmaceutical science is focused on improving the technology for the production of medicinal forms to enhance their pharmacotechnological properties, bio-availability, and stability. Ribavirin, as an effective agent for the treatment of chronic hepatitis C, is widely used in combination therapy regimens; however, the impact of excipients on the quality of its encapsulated forms has not been sufficiently studied. Optimising the composition and technology for the production of ribavirin capsules will improve their technological parameters, increase treatment efficacy, and reduce side effects.

The authors A.V. Volkova & A.A. Nozdrina (2020) noted that, as of 2019, the WHO (World Health Organization) reported 58 million people suffering from the chronic form of HCV (Hepatitis C Virus) HCV, which poses serious medical and social problems and is a global health threat that does not receive sufficient attention. Researcher T.L. Hrydina (2022) highlighted that one of the effective DAAs (Data Authentication Algorithm) DAAs for treating chronic HCV is ribavirin. As noted by V.I. Trykhlil (2020), ribavirin is effective in ribavirin/sofosbuvir combinations (with an Sustained virological response (SVR) rate of 91% in patients without liver cirrhosis).

N.S. Behei & O.V. Tryhubchak (2020) noted that one of the primary stages in creating solid dosage forms is pharmaceutical development, which lays the groundwork for the drug's quality, effectiveness, and safety. A. Singh & C.S. Chauhan (2022) pointed out that pharmaceutical drug development faces numerous regulatory challenges due to inconsistent product quality and the absence of certain product characteristics, leading regulatory authorities worldwide to continually compel pharmaceutical manufacturers to apply systematic tools. I.M. Fukuda *et al.* (2018) highlighted that in this regard, Design of Experiments (DoE) is considered an indispensable tool for enhancing product quality through rational planning and meaningful execution of experiments.

The random balance method is one of the screening designs most frequently used by Ukrainian researchers to study the influence of quantitative characteristics of excipients on the composition being developed. For instance, N. Behei & O. Tryhubchak (2020) investigated the influence of seven factors on amlodipine tablets with enalapril, conducting ten experiments with varying quantitative ratios of active pharmaceutical ingredients. In a study by B.V. Pavliuk *et al.* (2022), the effect of various amounts of 12 excipients on the pharmacotechnological parameters of tablets with lyophilised drone homogenate powder was examined, with 16 experimental series performed using the direct compression method. N.P. Darzuli & T.A. Hroshovyi (2018)

analysed the influence of seven factors on the pharmacotechnological properties of powder masses and tablets of sharp-leaved wintergreen extract, conducting eight experiments using direct compression. In the work of M. Vasenda *et al.* (2023), the quantitative effects of six factors on the pharmacotechnological properties of tablets based on walnut septum phytosubstances were examined, with eight experimental series performed using the direct compression method. O.I. Hordiienko & T.A. Hroshovyi (2021) explored the impact of different amounts of 10 excipients on the pharmacotechnological parameters of masses for tableting and tablets containing plant extracts and essential oils, conducting 16 experiments via direct compression.

A review of the scientific literature demonstrated that when using mathematical experiment planning, the random balance method is widely employed to investigate the effects of multiple excipients on the properties of medicinal products. This approach provides a scientific basis for optimising formulations. The aim of the study was to investigate the effect of excipients and their quantity on the pharmacological and technological parameters of capsule masses and hard gelatin capsules containing ribavirin using the random balance method.

Materials and Methods

The study was conducted at Kyiv National University of Technologies and Design, Ukraine. The composition of hard gelatin capsules was developed using ribavirin samples manufactured by Jinan Mingxin Pharmaceutical Co., Ltd., China. The following excipients were used: microcrystalline cellulose (MCC), type 101 (Comprecel® M101D +), polyvinylpyrrolidone (PVPP, Kollidon® CL), croscarmellose sodium (Solutab® A), povidone (PVP, Plasdone™ K29/32), hydroxypropyl cellulose (HPC, type L), colloidal anhydrous silica (Aerosil® 200), and magnesium stearate (type LIGAMED MF-2-V). MCC 200 was used as a compensator.

To determine the significance of the quantitative factors, scatter diagrams were constructed, where the x-axis represents the influencing factor and the y-axis represents the evaluation response variable (Revyatskyy & Barchuk, 2019). Each evaluated attribute was visually represented in separate charts as scatter plots based on statistical data processing. The median value visually indicated the dominance of one factor over another, i.e., the significance of the quantitative characteristics of excipients in relation to the critical attributes of ribavirin formulations. The median orientation indicated the direction of this effect.

Physicochemical and pharmacotechnological attributes of the capsule mass and the finished dosage form were

investigated in accordance with the methodology of the State Pharmacopoeia of Ukraine (2015). Hard capsules containing the active pharmaceutical ingredient (API) ribavirin were produced using wet granulation with high-shear

mechanical processing (Dipy & Wildfong, 2017). The effect of the quantitative attributes of excipients was investigated. These attributes (excipient quantities) were examined at “low” (–) and “high” (+) levels (Table 1).

Table 1. Quantitative factors and their levels studied in the development of hard gelatin capsules with ribavirin

Factor	Factor levels	
	“Low” (–)	“High” (+)
x_1 – quantity of MCC (Comprecel® M101D+), mg	69.0	74.0
x_2 – quantity of Crospovidone (Kollidon® CL), mg	3.0	6.0
x_3 – quantity of Croscarmellose sodium (Solutab® A), mg	3.0	6.0
x_4 – quantity of Povidone (Plasdone™ K29/32), mg	3.5	7.0
x_5 – quantity of Hydroxypropyl cellulose (HPC, type L), mg	3.0	6.0
x_6 – quantity of Silica, colloidal anhydrous (Aerosil® 200), mg	1.0	2.0
x_7 – quantity of Magnesium stearate (type LIGAMED MF-2-V), mg	1.5	3.0

Source: developed by the authors based on their own research

The influence of excipients was determined by constructing scatter plots based on the data of pharmacotechnological parameters of the capsule mass and hard gelatin capsules. The random balance method enabled the selection of certain excipients in optimal amounts and the exclusion of those that did not fully meet the pharmacotechnological quality attributes of the capsules (Darzuli & Hroshovyi, 2018). This approach allowed for a reduction in the number of test series.

Results and Discussion

To obtain hard gelatin capsules of ribavirin, the attributes of 25 excipients were preliminarily studied. Based on the results, excipients with the most significant effect on the pharmacotechnological quality attributes of capsules were selected (Shah & Shahzad, 2015). The optimal quality parameters for the capsule mass based on ribavirin as follows:

“Bulk density before and after tapping” – 0.57-0.62 g/ml and 0.62-0.70 g/ml, respectively; “Hausner ratio” – 1.12-1.15 (good flowability); “Loss on drying” – 2-3%. The optimal quality parameters for hard gelatin capsules containing the active pharmaceutical ingredient ribavirin are: “Uniformity of capsule content mass” – $\pm 5.0\%$; “Disintegration” – no more than 15 minutes; “Dissolution” – at least 85 % of the active substance must dissolve within 15 minutes.

Eight batches were prepared, differing in the quantitative composition ratios of excipients. To achieve a capsule content mass of 300 mg, MCC 200 was used as a compensator during the mixing and powdering operations. It acted as an independent quantitative factor without significant effect on the studied characteristics. The produced powder masses and capsules were analysed for pharmacotechnological attributes in accordance with the requirements of the State Pharmacopoeia of Ukraine (2015). The experimental design and analysis results are presented in Table 2.

Table 2. Experimental design and results of studies on hard gelatin capsules with ribavirin

Batch	x_1	x_2	x_3	x_4	x_5	x_6	x_7	y_1	y_2	y_3	y_4	y_5	y_6	y_7
1	–	+	–	–	+	–	+	0.589	0.689	1.17	3.0	4.6	93	92.8
2	–	+	+	+	+	–	+	0.568	0.647	1.14	3.7	3.6	95	93.2
3	–	+	+	–	–	–	–	0.549	0.659	1.20	4.1	5.2	91	99.6
4	+	–	+	–	–	+	–	0.612	0.704	1.15	2.7	4.2	100	98.5
5	+	–	–	+	–	–	–	0.550	0.655	1.19	2.4	5.0	99	83.7
6	–	+	–	–	–	+	+	0.531	0.627	1.18	3.1	4.8	102	91.1
7	+	–	+	+	+	+	–	0.563	0.636	1.13	2.5	3.3	104	96.0
8	+	–	–	+	+	+	+	0.573	0.653	1.14	2.2	3.5	92	86.4

Note: y_1 – Bulk density, g/mL; y_2 – Tapped density, g/mL; y_3 – Hausner ratio (flowability); y_4 – Loss on drying, %; y_5 – Uniformity of capsule content mass, $\pm \%$; y_6 – Disintegration time, s; y_7 – Dissolution, % in 15 minutes

Source: developed by the authors' based on their own research

The results of experimental studies of eight experiments on pharmacotechnological indicators, shown in Table 2, demonstrated the following values: bulk density ranges from 0.531 g/ml to 0.612 g/ml (the best results were obtained in experiments No. 1, No. 4 and No. 8); bulk density after shrinkage – from 0.627 g/ml to 0.704 g/ml (the best results were experiments No. 1, No. 3 and No. 4);

Hausner coefficient (characterising the fluidity of the mass for encapsulation) – from 1.13 to 1.20 (the best results were experiments No. 2, No. 7 and No. 8); mass loss during drying – from 2.2% to 4.1% (the best results were experiments No. 5, No. 7 and No. 8); mass uniformity of the capsule contents – from $\pm 3.3\%$ to 5.2% (the best results were experiments No. 2, No. 7 and No. 8); disintegration – from 91 s to

104 s, which is not a significant difference in the results, making it difficult to single out the best experiments; dissolution – from 83.7% to 99.6% (the best results are experiments No. 3, No. 4 and No. 7). If the experiments are

analysed by all indicators together, then experiments No. 4 and No. 7 showed the most optimal results. The effect of quantitative factors on the “Bulk density” attribute (y_1) is shown in the scatter diagram in Figure 1.

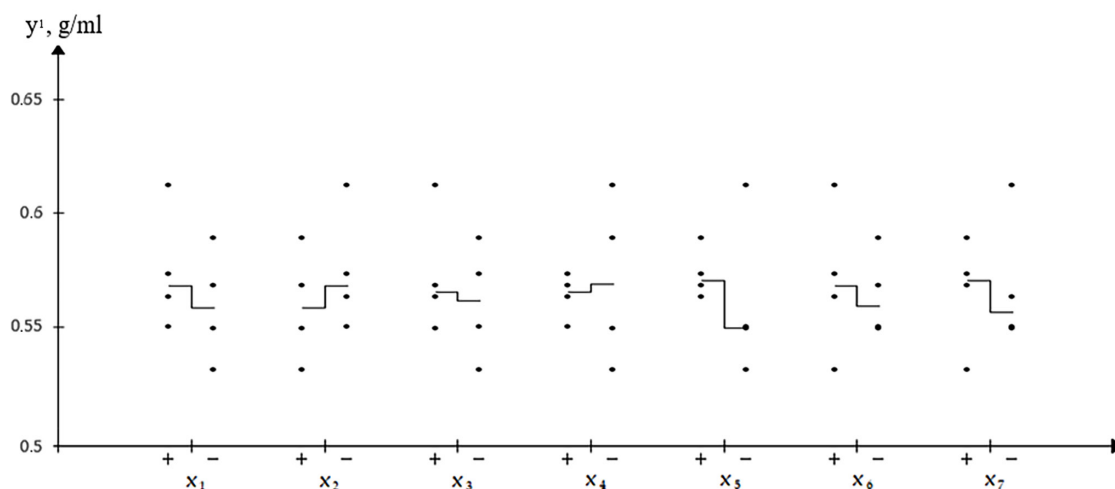


Figure 1. Effect of quantitative factors on “Bulk Density” of ribavirin capsule mass

Source: developed by the authors' based on their own research

The bulk density of the encapsulation mass, determined using the graduated cylinder pouring method, characterises fluidity and flowability and should ensure uniform dosing on the capsule filling machine. Based on the analysis of the scatter diagram of bulk density results, were

determined that the binding agent HPC-L (x_5) had a minor effect, while the effects of other excipients were negligible and could be disregarded. When examining the effect of quantitative factors on the “Tapped density” attribute (y_2), the results are presented in the scatter diagram in Figure 2.

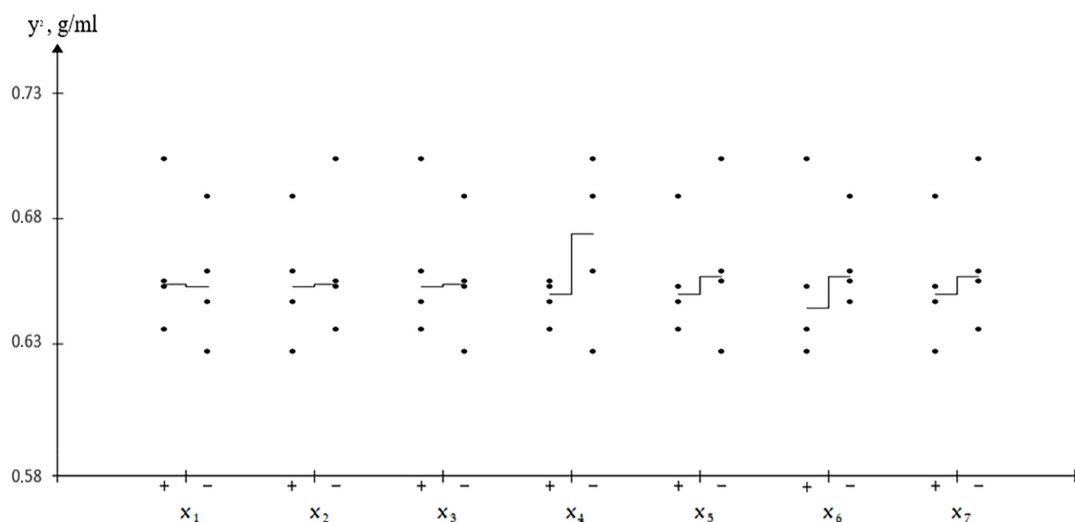


Figure 2. Effect of quantitative factors on “Tapped Density” of ribavirin capsule mass

Source: developed by the authors' based on their own research

Tapped Density is used to calculate the Hausner Ratio that for capsule filling. The results indicated that the binding agent Plasdone™ K29/32 (x_4) and the glidant Aerosil® 200 (x_6) had a minimal effect on this attribute. Other factors did not affect “Tapped density” at all. The scatter diagram in Figure 3 demonstrated that the Hausner Ratio (flowability) (y_3) of ribavirin capsule masses

was significantly affected by HPC-L (x_5), with a slightly lower effect from Plasdone™ K29/32 (x_4). Quantitative characteristics of the following excipients also showed a significant effect: Comprecel® M101D + (x_1), Kollidon® CL (x_2), Solutab® A (x_3), and Aerosil® 200 (x_6). The lubricant magnesium stearate (MF-2-V) (x_7) exhibited a negligible effect.

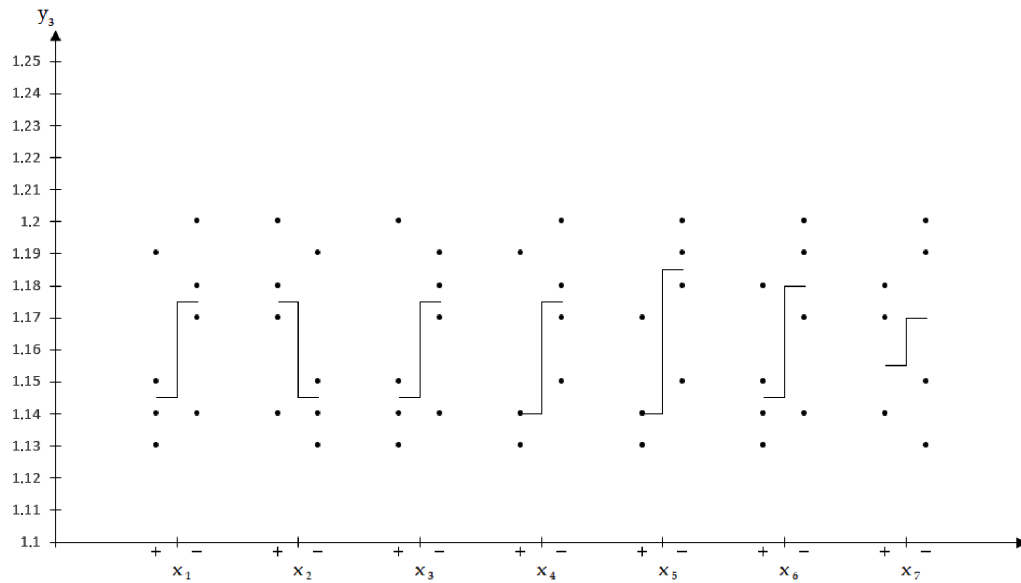


Figure 3. Effect of Quantitative Factors on “Hausner Ratio” (flowability) of Ribavirin Capsule Mass

Source: developed by the authors’ based on their own research

Based on the statistical data and median directions, the following results were obtained: increasing the amounts of Comprcel® M101D+ (x_1), Solutab® A (x_3), Plasdone™ K29/32 (x_4), HPC-L (x_5), Aerosil® 200 (x_6) and magnesium stearate (MF-2-V) (x_7) decreased the Hausner Ratio (y_3), improving the flowability of powder masses containing ribavirin; increasing the disintegrant Kollidon® CL (x_2) increased the Hausner Ratio (y_3), reducing the flowability of capsule masses.

The scatter plot data in Figure 4 demonstrated that the “Loss on Drying” attribute (y_4) of the mass for encapsulation is significantly affected by the factors Comprcel® M101D+ (x_1) and Kollidon® CL (x_2). Less pronounced effects are observed from Solutab® A (x_3), Plasdone™ K29/32 (x_4), Aerosil® 200 (x_6), and magnesium stearate (MF-2-V) (x_7). The quantitative factor HPC-L (x_5) has almost no effect on the studied attribute y_4 .

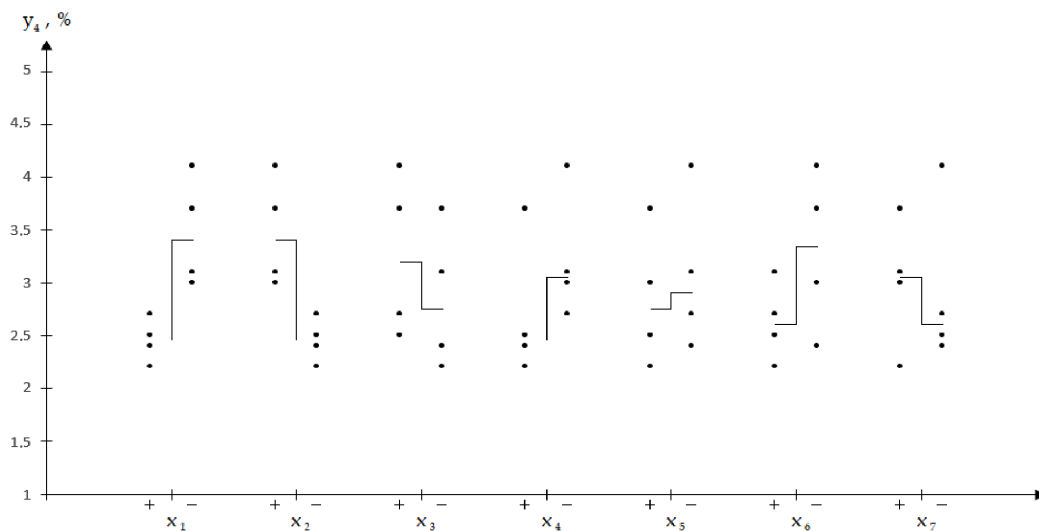


Figure 4. Effect of quantitative factors on “Loss on Drying” of ribavirin capsule mass

Source: developed by the authors based on their own research

Based on the statistical data and the positioning of the medians: increasing the amounts of factors x_2 , x_3 and x_7 increased the “Loss on Drying” attribute (y_4), which raised the moisture content in the powder masses containing ribavirin; increasing the amounts of factors x_1 , x_4 , x_5 , and x_6 decreased the “Loss on Drying” attribute (y_4), reducing

the moisture content in the capsule masses. Granules were dried to a moisture content within the range of 1.8-2.0%. A moisture content of more than 3.0% in the capsule mass led to material “sticking” in the capsule-filling machine hopper and partial powder adherence to the hopper walls. Therefore, the optimal “Loss on Drying” attribute (y_4)

for capsule masses should range between 2.0% and 3.0%. Based on these data, it is better to avoid combining disintegrants (Kollidon® CL and Solutab® A) and to use smaller amounts of one disintegrant. Additionally, higher quantities of the diluent Comprcel® M101D+, the binder Plasdone™ K29/32, and the glidant Aerosil® 200 should be used.

The effect of quantitative factors on the “Uniformity of Capsule Content Mass” attribute (y_5) is shown in the scatter diagram in Figure 5. According to the requirements of the State Pharmacopoeia of Ukraine, for a capsule content mass of 300 mg with ribavirin, deviation of this attribute should be within $\pm 7.5\%$.

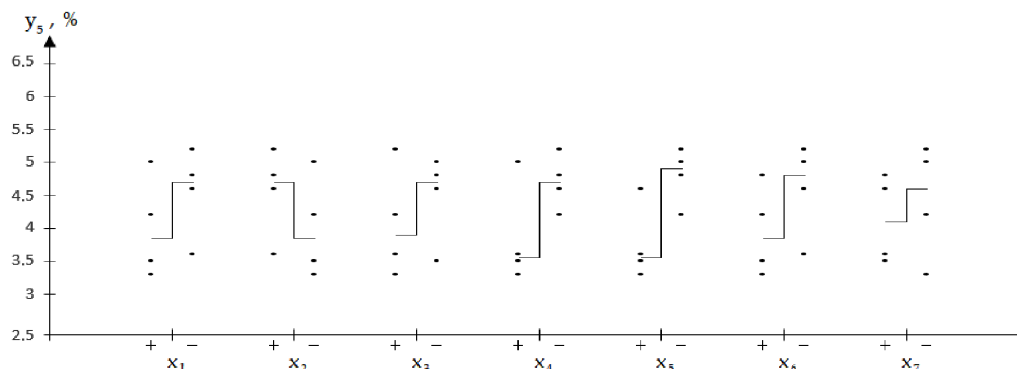


Figure 5. Effect of quantitative factors on “Uniformity of Capsule Content Mass” with API ribavirin

Source: developed by the authors based on their own research

When studying the effect of quantitative factors on the y_5 factor, it was found that the binding agents Plasdone™ K29/32 (x_4) and HPC-L (x_5) had the greatest effect. Significant but slightly lesser effects were observed for the following excipients: Comprcel® M101D+ (x_1), Kollidon® CL (x_2), Solutab® A (x_3), and Aerosil® 200 (x_6). The lubricant magnesium stearate (MF-2-V) (x_7) had an insignificant effect on the uniformity of capsule content mass with ribavirin.

Based on the statistical data and the orientation of medians, the following results were obtained: increasing the amounts of Comprcel® M101D+ (x_1), Solutab® A (x_3), Plasdone™ K29/32 (x_4), HPC-L (x_5), Aerosil® 200 (x_6) and magnesium stearate (MF-2-V) (x_7) reduced the deviations between the minimum and maximum values of capsule content mass, thus improving the “Uniformity of Capsule Content Mass” (y_5) attribute for ribavirin. Conversely increasing the disintegrant Kollidon® CL (x_2) increased the deviations between the minimum and maximum values of capsule content mass, thereby deteriorating the “Uniformity of Capsule Content Mass” (y_5).

To reduce deviations, i.e., improve the uniformity of capsule content mass, enhancing the flowability of the capsule mass is necessary. Based on the above data, this can be achieved by increasing the quantities of binding agents (x_4 or x_5), diluent (x_1), and glidants (x_6). Based on the research results none of the selected factors significantly affected the “Disintegration” attribute (y_6), as this attribute depends not on the quantitative characteristics of excipients but on the rate at which the hard gelatin capsule shell breaks down to release its contents.

The graphical data from the scatter diagram in Figure 6 showed that the “Dissolution” attribute (y_7) of hard ribavirin capsules is significantly affected by the disintegrant Solutab® A (x_3). A lesser, but still notable, effect is exerted by Plasdone™ K29/32 (x_4) and the lubricant magnesium stearate (MF-2-V) (x_7). Insignificant effects were observed for the excipients Comprcel® M101D+ (x_1), Kollidon® CL (x_2), and HPC-L (x_5), while the quantitative factor Aerosil® 200 (x_6) had almost no effect on the “Dissolution” attribute (y_7).

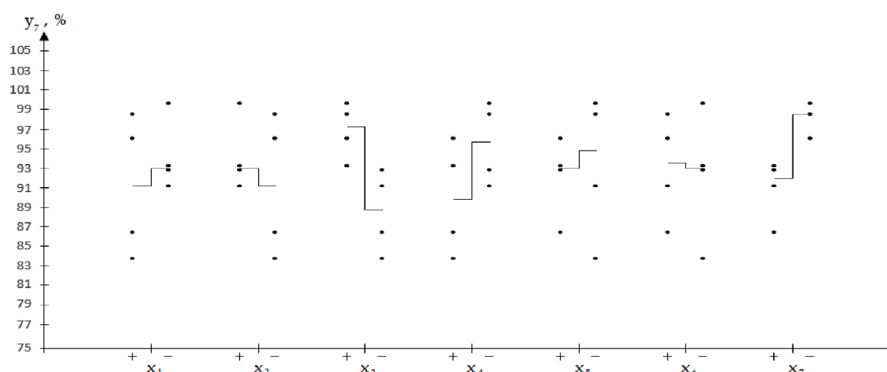


Figure 6. Effect of quantitative factors on “Dissolution” of hard capsules with API ribavirin

Source: developed by the authors based on their own research

Based on statistical data and the positioning of medians: increasing the amounts of x_2 , x_3 , and x_6 enhanced the “Dissolution” attribute (y_7), accelerating the release of the active substance from the dosage form; Conversely increasing the amounts of x_1 , x_4 , x_5 , and x_7 reduced the “Dissolution” attribute (y_7), slowing down the release of the active substance. V. Loustaud-Ratti *et al.* (2016) highlighted that adding peginterferon to this combination resulted in an SVR rate ranging from 86% to 91% in patients with liver cirrhosis, etc. Ribavirin is primarily used in solid oral dosage forms, including Rebetol®, hard capsules of 200 mg; Ribasphere®, hard capsules of 200 mg, 400 mg, and 600 mg; Copegus®, film-coated tablets of 200 mg, and others.

As noted by R. Samineni *et al.* (2022) ribavirin belongs to Class I of the Biopharmaceutical Classification System (BCS), meaning it refers to highly soluble substances with a high degree of permeability. G.F. Plöger *et al.* (2018) noted that, according to the State Pharmacopoeia of Ukraine, at least 85% of the API must be released within 15 minutes. Researchers O.O. Salii *et al.* (2010) studied that to achieve the best results for the “Dissolution” attribute of hard ribavirin capsules, higher amounts of the disintegrant (x_3) should be used while reducing the quantities of the binding agent (x_4) and the lubricant (x_7), ensuring no significant adverse effects on the “Hausner Ratio” (flowability), “Loss on Drying”, or “Uniformity of Capsule Content Mass” attributes. Authors H. Pham-The *et al.* (2013) highlighted that the main difficulty in applying the random balance method lies in the fact that experiments must be planned so that the sum of all components does not exceed 100%.

Author O.I. Hordiienko & T.A. Hroshovyi (2021) demonstrated the best way to select the number of studied factors according to the formula $N-1$, where N is the number of experiments, in order to study the influence of a larger number of factors with the fewest possible number of experiments. Researchers selected critical quality indicators for tablets as the main feedback: loss on drying, flowability, angle of repose, bulk density, tapped density, stability, abrasion, disintegration, etc. In this work, in addition to the pharmacotechnological parameters adapted for hard gelatin capsules, there is a difference in the choice of indicator for evaluating flowability, namely, the “Hausner ratio” was preferred, as the authors believe it best illustrates the flowability of the capsule mass (good, satisfactory, etc.).

Scientific publications provide data on the use of DoE design by authors L.A. Fukleva *et al.* (2024) for studying the most commonly used types of mathematical planning of the experiment, and selecting the most appropriate model for developing the optimal composition of a vaginal ointment. Authors O.A. Ruban *et al.* (2018) identified key features and capabilities application of experimental design method in optimisation of composition and technology for matrix tablets containing *Vaccinium myrtillus* leaf dry extract. The random balance method has been applied by M. Demchuk *et al.* (2019) to study the effect of excipient quantities on the technological attributes of metformin

orodispersible tablets. The random balance method has been used by S. Chernetska *et al.* (2021) to develop formulations and investigate the effect of excipient quantities on the technological attributes of dry extract-based mixtures and tablets derived from *Origanum vulgare* L. herb. The method was employed by O. Panysheva (2022) to study the effects of 27 excipients from five functional groups and to create of powder technology for oral solutions in sachets.

According to A.S. Deyneka & I.O. Zhuravel (2022), DoE enables a reduction in the number of scientific studies, streamlining, optimising, and intensifying research while minimising errors and delivering highly reliable and scientifically validated results at minimal cost. The authors B. Kovács *et al.* (2021) noted that DoE is a core component of the statistical toolkit for implementing the systematic “Quality by Design” (QbD) approach in both research and industrial settings, allowing for continuous improvement throughout the entire lifecycle of a drug. Researchers S. Beg *et al.* (2019) studied the trend since the late 20th century, with numerous DoE methods having been developed for application in pharmaceutical research. T. Waghule *et al.* (2021) noted that experimental designs can be classified as screening designs, while S. Beg (2021) highlighted optimisation designs.

In scientific articles authors M. Vasenda *et al.* (2023) highlighted that the direct compression method is used for tablet manufacturing, which allows for a better study of the impact of quantitative factors on the pharmacotechnological properties of powder masses. In this work, the wet granulation method was used, which differs from the sources mentioned above. This technology did not fully allow the analysis of the impact of quantitative factors on the indicators “Bulk density and tapped density” of capsule mass, as almost all excipients had a negligible effect, which was difficult to evaluate.

It should also be noted that in some scientific publications the desirability function (D) is used for the general evaluation of results for each experiment. This indicator is not mandatory when using the random balance method, but it facilitates the evaluation of results when the overall impact of factors in a given experiment on all the studied quality indicators needs to be assessed, rather than each separately. Scientific publications authors B.V. Pavliuk *et al.* (2022) provide data on the following compensators are used: sucrose, MCC 102, MCC 200, etc. Also, the number of experiments must be a multiple of 4, which is due to the peculiarity of constructing scatter diagrams. But the desirability function was not used in their scientific articles.

The obtained results indicate that the use of the random balance method allows for the effective optimisation of the composition of hard gelatin capsules with ribavirin. Comparison with the works of other authors confirmed the significance of selecting the quantitative characteristics of excipients to ensure pharmacotechnological parameters. Thus, the obtained results can serve as the basis for further research in the development of antiviral drugs with direct activity and other active pharmaceutical ingredients.

Conclusions

The use of the random balance method made it possible to determine the most significant quantitative factors and stabilise them at optimal levels in the composition of the developed ribavirin capsules. Based on statistical data and the magnitude of medians in the scatter diagrams, it was established that quantitative factors at “high” and “low” levels did not significantly affect the following attributes: “Bulk Density”, “Tapped Density”, “Disintegration”. Thus, for selecting more significant excipients, the focus was placed on the following attributes: “Hausner Ratio”, “Loss on Drying”, “Uniformity of Capsule Content Mass”, “Dissolution”.

For the optimal formulation of the drug containing ribavirin as the active pharmaceutical ingredient (API), which ensures appropriate pharmacotherapeutic attributes, the following factors and their levels were selected, the following excipients were used: microcrystalline cellulose (Comprecel® M101D+) as a compensator; sodium croscarmellose (Solutab® A) at 6 mg per capsule; povidone (Plasdone™ K29/32) at 7 mg; colloidal anhydrous silica (Aerosil® 200) at 2 mg; magnesium stearate (LIGAMED MF-2-V) at 1.5 mg.

The random balance method in experimental studies allowed for the scientific justification of the selection of quantitative ratios of excipients in the development of solid capsule formulations and achieving target parameters

of the drug profile with ribavirin according to the “Disintegration” and “Dissolution” tests. Therefore, determining the most significant quantitative influencing factors at the pharmaceutical development stage enables the identification of the most critical functional excipients, which will affect the drug release profiles, especially in the case of poorly soluble substances.

The randomisation method has numerous advantages, as it prevents random bias in the experiment and ensures that comparable groups are created across all factors. Further research could be aimed at improving randomisation strategies to optimise the design of randomised clinical trials of medicines. Of particular interest is the application of this approach in small and medium-sized clinical trials ($n < 100$), where the use of randomisation helps to achieve a balance in treatment. This can provide more reliable and valid trial results, which, in turn, will increase the efficiency of the development and evaluation of new medicines.

Acknowledgements

None.

Funding

None.

Conflict of Interest

None.

References

- [1] Beg, S. (Ed.). (2021). *Design of experiments for pharmaceutical product development*. Singapore: Springer. doi: 10.1007/978-981-33-4717-5.
- [2] Beg, S., Swain, S., Rahman, M., Hasnain, M.S., & Imam, S.S. (2019). Application of design of experiments (DoE) in pharmaceutical product and process optimization. In *Pharmaceutical quality by design* (pp. 43-64). Cambridge: Academic Press. doi: 10.1016/B978-0-12-815799-2.00003-4.
- [3] Behei, N.S., & Tryhubchak, O.V. (2020). Research of the excipients' quantitative characteristics influence on the basic quality indicators of amlodipine with enalapril tablets. *Ukrainian Biopharmaceutical Journal*, 2(63), 24-29. doi: 10.24959/ubphj.20.263.
- [4] Chernetska, S., Beley, N., & Demchuk, M. (2021). The method of random balance for studying the influence of excipients' quantities on technological parameters of tablets based on *Origanum vulgare* L. dry extract. *EUREKA: Health Sciences*, 2, 73-81. doi: 10.21303/2504-5679.2021.001664.
- [5] Darzuli, N.P., & Hroshovyi, T.A. (2018). Study of quantitative factors on technological properties of compression mixture and tablets of round-leaved wintergreen extract. *Pharmaceutical Review*, 3, 45-51. doi: 10.11603/2312-0967.2018.3.9377.
- [6] Demchuk, M., Chubka, M., & Hroshovyi, T. (2019). The method of random balance for studying the influence of excipients' quantities on technological parameters of metformin orodispersible tablets. *International Journal of Applied Pharmaceutics*, 11(3), 168-175. doi: 10.22159/ijap.2019v11i3.32792.
- [7] Deyneka, A.S., & Zhuravel, I.O. (2022). *Development of technology for obtaining an extract from the flowers of Celosia cristata (L.) Kuntze by applying the method of mathematical planning of the experiment*. *Annals of Mechnikov's Institute*, 2, 41-46.
- [8] Dipy, D.M., & Wildfong, P.D. (2017). Solid-state transformations of ribavirin as a result of high-shear mechanical processing. *International Journal of Pharmaceutics*, 524(1-2), 339-350. doi: 10.1016/j.ijpharm.2017.04.002.
- [9] Fukleva, L.A., Saliy, O.O., Baula, O.P., Puchkan, L.A., & Hrechana, O.V. (2024). Methods of mathematical planning of the experiment and their use in the development of the composition of an ointment with the essential oil *Thyme vulgaris* L. For vaginal application. *News of Pharmacy*, 107(1), 74-79. doi: 10.24959/nphj.24.130.
- [10] Fukuda, I.M., Pinto, C.F.F., Moreira, C.D.S., Saviano, A.M., & Lourenço, F.R. (2018). Design of experiments (DoE) applied to pharmaceutical and analytical quality by design (QbD). *Brazilian Journal of Pharmaceutical Sciences*, 54,

article number e01006. doi: [10.1590/s2175-97902018000001006](https://doi.org/10.1590/s2175-97902018000001006).

- [11] Hordiienko, O.I., & Hroshovyi, T.A. (2021). Study of the quantitative factors influence on pharmacotechnological properties of powder masses and tablets with plant extracts and essential oil. *Pharmaceutical Journal*, 1(1), 35-42. doi: [10.32352/0367-3057.1.21.04](https://doi.org/10.32352/0367-3057.1.21.04).
- [12] Hrydina, T.L. (2022). Peculiarities of treatment and prevention of infection caused by hepatitis C virus. Review. *Medical Science of Ukraine*, 18(1), 64-72. doi: [10.32345/2664-4738.1.2022.09](https://doi.org/10.32345/2664-4738.1.2022.09).
- [13] Kovács, B., Kovacs-Deak, B., Szekely-Szentmiklosi, I., Fulop, I., Baba, L.I., Boda, F., & Peterfi, O. (2021). Quality-by-design in pharmaceutical development: From current perspectives to practical applications. *Acta Pharmaceutica*, 71(4), 497-526. doi: [10.2478/acph-2021-0039](https://doi.org/10.2478/acph-2021-0039).
- [14] Loustaud-Ratti, V., et al. (2016). Ribavirin: Past, present and future. *World Journal of Hepatology*, 8(2), 123-130. doi: [10.4254/wjh.v8.i2.123](https://doi.org/10.4254/wjh.v8.i2.123).
- [15] Panysheva, O. (2022). [Development of formulation and technology of combined generic powder for oral solution in sachets based on the quality-by-design approach](https://doi.org/10.32345/2664-4738.1.2022.09). *Global Journal of Medical Research: B Pharma, Drug Discovery, Toxicology & Medicine*, 22(B2), 77-98.
- [16] Pavliuk, B.V., Chubka, M.B., Hroshovyi, T.A., & Dobrynychuk, M.M. (2022). Application of the method of random balance to study the effect of the amounts of excipients on the technological parameters of tablets with lyophilized powder of drone brood homogenate. *Pharmaceutical Review Farmaceutičnij Časopis*, 1, 58-65. doi: [10.11603/2312-0967.2022.1.13056](https://doi.org/10.11603/2312-0967.2022.1.13056).
- [17] Pham-The, H., Garrigues, T., Bermejo, M., González-Álvarez, I., Monteagudo, M.C., & Cabrera-Pérez, M.A. (2013). Provisional classification and in silico study of biopharmaceutical system based on caco-2 cell permeability and dose number. *Molecular Pharmaceutics*, 10(6), 2445-2461. doi: [10.1021/mp4000585](https://doi.org/10.1021/mp4000585).
- [18] Plöger, G.F., Hofsäuss, M.A., & Dressman, J.B. (2018). Solubility determination of active pharmaceutical ingredients which have been recently added to the list of essential medicines in the context of the biopharmaceutics classification system-biowaiver. *Journal of Pharmaceutical Sciences*, 107(6), 1478-1488. doi: [10.1016/j.xphs.2018.01.025](https://doi.org/10.1016/j.xphs.2018.01.025).
- [19] Revyatsky, I., & Barchuk, O. (2019). Presentation optimization of research results using the random balance method and influence factors that have different equivalents of measurement. *Pharmaceutical Journal*, 1, 76-85. doi: [10.32352/0367-3057.1.19.07](https://doi.org/10.32352/0367-3057.1.19.07).
- [20] Ruban, O.A., Kolisnyk, T.Ye., & Slipchenko, G.D. (2018). Application of experimental design method in optimization of composition and technology for matrix tablets containing *Vaccinium myrtillus* leaf dry extract. *Pharmaceutical Review*, 2, 41-47. doi: [10.11603/2312-0967.2018.2.9003](https://doi.org/10.11603/2312-0967.2018.2.9003).
- [21] Saliy, O.O., Litka, V.V., & Los, O.V. (2010). [Manufacturing process development and process procedure scaling of drug capsules ribavirin](https://doi.org/10.32345/2664-4738.1.2010.09). *Ukrainian Biopharmaceutical Journal*, 1(6), 4-9.
- [22] Samineni, R., Chimakurthy, J., & Konidala, S. (2022). Emerging role of biopharmaceutical classification and biopharmaceutical drug disposition system in dosage form development: A systematic review. *Turkish Journal of Pharmaceutical Sciences*, 19(6), 706-713. doi: [10.4274/tjps.galenos.2021.73554](https://doi.org/10.4274/tjps.galenos.2021.73554).
- [23] Shah, S.N., & Shahzad, W. (2015). Manufacturing of formulation of ribavirin 200 mg capsules. *Canadian Journal of Applied Science*, 5(2), 13-18. doi: [10.21065/19257430.5.13](https://doi.org/10.21065/19257430.5.13).
- [24] Singh, A., & Chauhan, C.S. (2022). Factorial designing – an essential tool in pharmaceutical optimization. *ResearchGate*, 21(7), 519-525. doi: [10.37896/YMER21.07/41](https://doi.org/10.37896/YMER21.07/41).
- [25] State Pharmacopoeia of Ukraine. (2015). Retrieved from <https://sphu.org/en/state-pharmacopoeia-of-ukraine>.
- [26] Trykhlil, V.I. (2020). Chronic viral hepatitis C. Effectiveness of antiviral therapy. The WHO methodological recommendations (literature review). *Actual Infectology*, 8(1), 1-7. doi: [10.22141/2312-413x.8.1.2020.196165](https://doi.org/10.22141/2312-413x.8.1.2020.196165).
- [27] Vasenda, M., Budniak, L., Kramar, H., & Kravchuk, L. (2023). Study of the effect of quantitative factors on the pharmacotechnological properties of tablets based on the phytosubstance of *Juglans regia* L. membranes. *Phytotherapy Journal*, 3, 95-100. doi: [10.32782/2522-9680-2023-3-95](https://doi.org/10.32782/2522-9680-2023-3-95).
- [28] Volkova, A.V., & Nozdrina, A.A. (2020). [Research on the availability of drugs for the treatment of hepatitis C](https://doi.org/10.32345/2664-4738.1.2020.09). In *Scientific and practical distance conference with international participation "Modern directions of improvement of pharmaceutical ensure of the population: From development to application of drugs of natural and synthetic origin"* (pp. 16-18). Ivano-Frankivsk: Ivano-Frankivsk National Medical University.
- [29] Waghule, T., Dabholkar, N., Gorantla, S., Rapalli, V.K., Saha, R.N., & Singhvi, G. (2021). Quality by design (QbD) in the formulation and optimization of liquid crystalline nanoparticles (LCNPs): A risk based industrial approach. *Biomedicine & Pharmacotherapy*, 141, article number 111940. doi: [10.1016/j.biopha.2021.111940](https://doi.org/10.1016/j.biopha.2021.111940).

Дослідження впливу допоміжних речовин на фармако-технологічні параметри капсул з рибавірином методом випадкового балансу

Олена Салій

Кандидат фармацевтичних наук, доцент
Київський національний університет технологій та дизайну
01011, вул. Мала Шияновська, 2, м. Київ, Україна
<https://orcid.org/0000-0001-7103-2083>

Андрій Бублик

Магістр
Київський національний університет технологій та дизайну
01011, вул. Мала Шияновська, 2, м. Київ, Україна
<https://orcid.org/0009-0002-6495-9969>

Марія Попова

Аспірант
Київський національний університет технологій та дизайну
01011, вул. Мала Шияновська, 2, м. Київ, Україна
<https://orcid.org/0000-0002-2579-0331>

Лариса Фуклева

Кандидат фармацевтичних наук, старший викладач
Запорізький державний медико-фармацевтичний університет
69035, бульв. Марії Примаченко, 26, м. Запоріжжя, Україна
<https://orcid.org/0000-0002-2930-0619>

Анотація. Одним із основних етапів створення твердих лікарських форм є фармацевтична розробка, коли закладаються якість, ефективність та безпека майбутнього препарату. Вибір біофармацевтичних факторів у плануванні технологічних досліджень є важливим завданням фармацевтичної розробки, що ускладнюється необхідністю комплексного врахування усіх факторів. Метою роботи було дослідження та аналіз впливу кількісних характеристик допоміжних речовин на показники якості порошкових мас та капсул твердих з рибавірином. Застосовано математичне планування експерименту, що дозволило створити оптимальний склад шляхом встановлення взаємозв'язків впливу факторів на технологічні параметри, які мають забезпечувати фармакологічну дію рибавіріну. У дослідженнях використано метод випадкового балансу для відбору оптимальних допоміжних речовин для включення у склад желатинових капсул. На основі отриманих статистичних даних побудовано діаграми розсіювання, де величина медіани показувала значущість впливу досліджуваних факторів на відповідну характеристику якості. Вивчено вплив кількісних факторів на властивості мас для капсулювання і твердих капсул з рибавірином, виготовлених методом вологої грануляції. Виявлено, що кількісні характеристики на вищих та нижніх рівнях не показують істотної сили дії на показники: «Насипна густина», «Насипна густина після усадки» для капсульних мас та «Розпадання» для капсул твердих. Найбільший вплив допоміжних речовин на коефіцієнт Гауснера для порошкових мас та однорідність маси вмісту капсул здійснювали зв'язуючі речовини, такі як гідроксипропілцелюлоза і повідон. З діаграми розсіювання виявлено, що найбільший вплив на поліпшення втрат у масі при висушуванні маси для капсулювання мали збільшена кількість мікрокристалічної целюлози та зменшений вміст кросповідону. Визначено фактори, що впливають на показник «Розчинення»: збільшення кількості натрію кроскармелози та зменшення вмісту магнію стеарату та повідону.

Ключові слова: противірусні препарати прямої дії; математичне планування експерименту; насипна густина; коефіцієнт Гауснера; розпадання капсул; розчинення капсул



The use of artificial intelligence for information campaigns in wartime: Visual tools for countering disinformation

Vladyslav Shchuka*

Master

Kyiv National University of Culture and Arts

01601, 36 Yevhen Konovalets Str., Kyiv, Ukraine

<https://orcid.org/0009-0004-1550-4727>

Abstract. In the context of the modern information war, the development of effective methods to counter disinformation and hostile propaganda has become critically important. The amount of false information and the speed of its dissemination necessitated the implementation of automated systems involving artificial intelligence to optimise the processes of creating visual counter-propaganda content. This research aimed to develop a methodology for the integration of artificial intelligence technologies into the processes of creating effective visual tools for countering disinformation, taking into account the principles of graphic design and the psychology of visual information perception. The research was based on a comprehensive approach that combined theoretical analysis of scientific literature, comparative analysis of the neural networks Midjourney, Stable Diffusion, and DALL-E, semiotic analysis of visual materials, as well as experimental implementation of the developed system for the automated creation of counter-propaganda visual content. A comprehensive approach has been developed for the creation of visual tools to counter disinformation, which combines the capabilities of automated information collection systems, artificial intelligence algorithms for generating graphic content, and the principles of effective graphic design. It has been found that the use of artificial intelligence in graphic design has optimised up to 20% of routine tasks in the creation of visual content, allowing designers to focus on the strategic and creative aspects of development. The developed recommendations for the use of artificial intelligence in graphic design may be implemented by state institutions, media, and public organisations to respond promptly to information threats

Keywords: graphic design; visual communication; counter-propaganda content; information warfare; generative design; perception psychology; compositional solutions

Introduction

In the context of the full-scale war of the Russian Federation against Ukraine, the information front has become one of the key elements of resistance. Russian propaganda actively utilised modern technologies, particularly artificial intelligence (AI), to create and disseminate disinformation, which posed a serious threat to Ukraine's national security. The Centre for Democracy and Rule of Law emphasised the ability of AI to generate texts that supported conspiracy theories and included real events, as well as to form disinformation messages based on propaganda narratives (Center for Democracy and Rule of Law, 2023).

In the collective monograph by M. Kolosnichenko *et al.* (2023), it was established that visual communications played a key role in the modern information space. Graphic

design was used for aesthetic decoration, as well as for effectively conveying key messages and refuting misinformation. Researcher O. Hrozna (2024) noted that global digital transformation has significantly impacted journalism and design, altering them through the implementation of cutting-edge technologies. Artificial intelligence and virtual reality have revolutionised these fields, fundamentally changing how content is created, consumed, and interacted with. The use of AI in visual communications allowed for the prompt processing and analysis of massive volumes of data, enhancing the accuracy and timeliness of information presentation.

K. Kertysova (2019) examined artificial intelligence as a bilateral tool in the context of disinformation. On the

Suggested Citation:

Shchuka, V. (2025). The use of artificial intelligence for information campaigns in wartime: Visual tools for countering disinformation. *Technologies and Engineering*, 26(2), 89-98. doi: 10.30857/2786-5371.2025.2.7.

*Corresponding author



one hand, it provided powerful, scalable, and cost-effective solutions to prevent the distortion of information online, but on the other hand, it had its own limitations and unpredictable consequences. The researcher emphasised the importance of developing critical thinking for restoring healthy relationships with information and enhancing the resilience of democracies to disinformation. The analysis of ethical aspects conducted by K. Kertysova (2019) revealed that the use of AI in combating fake news necessitated the development of clear ethical standards to prevent potential negative consequences for freedom of speech and privacy. The results of the research determined the necessity of the parallel development of media literacy among the population alongside the implementation of technological solutions.

Y. Hlushchuk (2023) defined fakes as deliberately false or misleading information that transcended objective reality and was aimed at manipulating public consciousness. The analysis revealed that in the context of the current war of the Russian Federation against Ukraine, the dissemination of fake news through social networks, online platforms, and other channels has influenced society, undermined trust in reliable sources of information, and led to manipulation of public opinion and socio-political upheavals. In the analytical materials of the Alliance for the Defence of Democracy (Salvo *et al.*, 2018), the necessity of a comprehensive approach to countering disinformation was substantiated, which combined the efforts of government structures, the private sector, technology companies, and civil society. Experts have formulated recommendations regarding the transparency of the activities of technology companies, providing information about technologies, privacy policies, and business models to the relevant authorities and the public. It also emphasised the importance of supporting independent research aimed at protecting platforms from disinformation campaigns and cyberattacks.

N. Derevyanko & O. Zalevska (2023) conducted a comparative analysis of the neural networks Midjourney, Stable Diffusion, and DALL-E, which were used in the design. Researchers have discovered that each of these neural networks had a specific purpose and architecture, which allowed them to effectively perform various tasks in the field of design. The use of these tools in the educational process of future designers contributed to the development of creative potential and expanded the possibilities of visual expression. The research by T. Bozhko & V. Arefiev (2023) demonstrated that artificial intelligence was actively employed in graphic design – from the creation of identification and navigation signs to the development of fonts, graphics, posters, animations, and infographics. Technologies have helped to transform the learning experience for students in design disciplines and have contributed to the development of individual approaches and creativity.

A. Borji (2022) conducted a quantitative comparison of Stable Diffusion, Midjourney, and DALL-E regarding

their ability to generate photorealistic faces. The research highlighted the potential of these tools for use in the design of visual materials aimed at countering disinformation. C. Marsden & T. Meyer (2019) noted that fully automated fact-checking remained a distant goal. Social media continued to rely on a combination of AI for the most repetitive tasks and human oversight for more nuanced cases. The authors emphasised the importance of keeping the individual “in the loop” of AI systems to ensure ethical use and correct interpretation of data. N. Dufour *et al.* (2024) conducted a two-year study on online disinformation based on media content, during which a detailed typology of images used in disinformation was developed and the dynamics of their dissemination were analysed. Research has shown that despite widespread concern since the late 2010s, AI-generated content was rare until the spring of 2023 when its presence in fake news sharply increased.

This research aimed to develop and implement methodological principles for the use of artificial intelligence in creating effective visual tools to counter disinformation, taking into account contemporary achievements in graphic design and perception psychology.

Materials and Methods

The research on the use of artificial intelligence for the creation of visual tools to counter disinformation in wartime was conducted from January 2023 to April 2024 at the Digital Media Laboratory of Taras Shevchenko National University of Kyiv. To achieve the set aim, a comprehensive approach was employed, combining general scientific and special research methods adapted to the specifics of graphic design and visual communications. The methodological basis of the research was formed by transdisciplinary approaches, which allowed for the examination of the problem of creating visual tools to counter disinformation from the perspectives of various fields of knowledge. The study employed theoretical, empirical, and mathematical methods.

During the research, observations were made of the work of designers who utilise artificial intelligence, which allowed for the calculation of the approximate proportion of time spent on routine tasks and the assessment of their suitability for automation. The methodological justification was based on the working processes of graphic designers and the identification of typical operations that are subject to automation. In particular, this includes searching for references, processing images, generating composition variants, and creating basic design elements. The theoretical and informational basis of the research consisted of fundamental studies by leading Ukrainian and international scholars, materials from periodical publications, analytical reports from specialised organisations, as well as empirical data from online media and social networks. Particular attention was paid to the analysis of publications from research centres specialising in the study of

mechanisms for countering disinformation in the context of the Russian-Ukrainian war.

To investigate the psychological and cognitive factors influencing the impact of graphic design on information perception, a method of content analysis of visual materials and semiotic analysis was employed. Samples of graphic content aimed at countering disinformation were examined using a specially developed matrix of criteria, which included assessments of emotional impact, informational value, aesthetic appeal, and psychological accessibility. The methodology of analysis was based on the concepts of visual semantics proposed by S.V. Pryshchenko (2020), who considered the artistic imagery of the poster as a complex semiotic phenomenon, and the research on psychological aspects of perception by H.V. Mykytiv & Y. Kondratenko (2016), who studied extratextual elements as a means of shaping media culture.

A comparative analysis of the neural networks Midjourney, Stable Diffusion, and DALL-E was conducted using a methodology adapted from the research of N. Derevyanko & O. Zalevska (2023). The aim of the analysis was to determine optimal technological solutions for the creation of visual tools to counter disinformation. Each neural network was evaluated according to the following parameters: accuracy of visual concept reproduction, flexibility in stylistic design, speed of content generation, ease of integration into workflows, and adaptability for specific tasks in countering disinformation. The analytical part of the research included a comparative analysis of the results of using various neural networks for the creation of visual materials to counter disinformation. A series of test queries with identical parameters have been created for each neural network to objectively compare the quality and relevance of the results, which has allowed for the identification of optimal tools for various types of visual tasks.

Semiotic analysis was applied to study the sign systems in visual materials aimed at countering disinformation. The methodology of semiotic analysis was adapted from the research of H.V. Mykytiv & Y. Kondratenko (2016), who considered extratextual elements as a means of shaping media culture. This approach enabled the identification of effective visual codes and symbolic systems that contributed to a better understanding and retention of counter-propaganda messages.

For the analysis and interpretation of the obtained results, qualitative comparative analysis and visual analytics methods were employed. The evaluation of the effectiveness of visual tools was conducted according to a developed system of criteria, which included compositional expressiveness, colour harmony, informational richness, and emotional response. The comparative analysis was conducted using the visual analytics methods proposed by O.V. Kolisnyk *et al.* (2023), which allow for the identification of structural and stylistic features of graphic content. Visual modelling and testing were conducted using

Adobe Creative Cloud, including Photoshop and Illustrator, with the integration of artificial intelligence algorithms for the analysis of compositional solutions. Graphical experiments with neural networks were conducted through the web interfaces of Midjourney, Stable Diffusion, and DALL-E using the methods described by T. Bozhko & V. Arefiev (2023). The ethical aspects of the research included adherence to the rules for the use of generative AI models, proper citation of the data sources used for training the models and ensuring the confidentiality of the personal data of the experts who participated in the surveys. All experiments were conducted for the purpose of scientific research and did not involve the creation or dissemination of disinformation materials.

Results and Discussion

The analysis of contemporary technological solutions for the creation of visual content has revealed a variety of tools that could be employed to counter disinformation in the context of information warfare. Research has shown that artificial intelligence systems have significantly optimised the work of graphic designers, automating up to 20% of routine tasks, however, they have not completely replaced the designer. Platforms such as Midjourney, StyleGAN, and Adobe Generative AI have enabled the rapid creation of high-quality images, the generation of design variations, and experimentation with styles, significantly accelerating workflows and expanding the creative possibilities for designers. Systematic observation of the process of creating visual materials has shown that these aspects of work occupy approximately one-fifth of the total time of designers and can be effectively automated using modern AI tools.

The study by R. Rombach *et al.* (2022) demonstrated that latent diffusion models (LDM) provided a significant improvement in both the training process and the sampling efficiency of diffusion noise models without compromising quality. This was identified as particularly important for creating high-quality visual content used in countering disinformation. The study found that LDM achieved new FID (Fréchet Inception Distance) benchmarks for image synthesis with conditional classification and high performance for various tasks, including text-to-image synthesis, unconditional image generation, and super-resolution. As a result of the comparative analysis of the neural networks Midjourney, Stable Diffusion, and DALL-E, specific advantages for various tasks of visual communication were identified (Table 1).

The semiotic analysis of 10 visual materials aimed at countering disinformation has revealed the most effective visual codes and sign systems. It has been established that the use of contrasting visual comparisons of “truth-falseness”, fact-checking symbols, visual metaphors of revelation or exposure, as well as elements that emphasise transparency and credibility, significantly enhanced the effectiveness of visual materials (Table 2).

Table 1. Comparison of neural network characteristics for graphic content creation

Assessment criterion	Midjourney	Stable Diffusion	DALL-E
Visual language	Utilises a visual metaphor with clean, minimalist solutions.	Diversity of styles within a single concept.	Official, informative style with elements of infographics.
Interactivity	High level of user engagement through testing.	Predominantly static formats.	Intermediate level with a focus on video content.
Adaptability	High adaptability for various languages and cultural contexts. The system allows for fine-tuning of visual styles by the cultural characteristics of the target audience through the modification of prompts and the use of specific style references.	Average adaptability. It provides opportunities for content localisation but requires a more detailed configuration to achieve cultural relevance and accurate compliance with regional characteristics.	High adaptability for various platforms. The system integrates with various publication formats and optimises visual content for diverse digital platforms with minimal adjustments.
Emotional impact	Strong emotional impact due to personal involvement. The created images evoke an immediate emotional response due to the refined combination of colours, shapes, and compositional solutions that resonate with the cultural and personal experiences of the viewer.	Extremely high emotional impact. Demonstrates an outstanding ability to generate visual content with a powerful emotional charge, particularly in genres that require dramatic expressiveness and imagery.	Predominantly rational approach. It focuses on informativeness, structure, and logical organisation of data, prioritising clarity over emotional expressiveness.
Informative nature	High informativeness while maintaining simplicity. Visual solutions effectively convey complex ideas and concepts through thoughtful metaphors and symbols, maintaining a balance between informational richness and accessibility of perception.	Moderate to high informativeness. It provides a sufficient level of informational value, but sometimes sacrifices detail for the sake of stylistic expressiveness.	Maximum informativeness. Specialises in creating structured, analytically rich visual content with an optimal balance of textual and graphic elements.

Source: adapted from the research of N. Derevyanko & O. Zalevska (2023)

Table 2. A comparison of the semiotic features of disinformation and counter-propaganda materials

Assessment criterion	Disinformation materials	Counter-propaganda materials
Dissemination speed	High, due to simplicity and emotionality	Moderate, depends on institutional support
Emotional impact	Powerful, often based on negative emotions	Varies from emotional (Russia's Hunger Games) to rational (NAD)
Resilience to critical analysis	Low, easily refuted upon detailed examination	High, especially in the presence of a factual basis
Resource intensity	Low, especially with the use of AI	High, requires professional design and research
Adaptability	High, responsive to changes in the information environment	Moderate, requires time to develop strategies
Long-term impact	Low, forms short-term reactions	High, contributes to the formation of resilient narratives
Media literacy development	Absence or negative impact	Positive, especially in educational formats
Plausibility for the target audience	High among vulnerable audiences, low among media-literate audiences	High when supported institutionally
Semiotic complexity	Low, employs simple binary oppositions	High, especially in metaphorical campaigns
Technological innovativeness	High, actively employs AI	Moderate, employs comprehensive design solutions

Source: author's development

The analysis of the practice of Brand Ukraine (n.d.), which is engaged in countering disinformation, has revealed effective strategies for the use of visual means. In particular, the campaign Russia's Hunger Games was analysed, which employed metaphorical visual solutions to expose manipulations regarding the food crisis (Fig. 1).

The experience of using dramatic visual contrasts to demonstrate the consequences of information manipulation proved to be particularly valuable (Fig. 2). The Nations

Against Disinformation campaign utilised an interactive "Distortion Test" that visualised the process of distorting reality under the influence of disinformation, allowing users to experience this effect firsthand.

A comparative analysis of the disinformation materials disseminated on social media was also conducted during the research (Fig. 3). It has been found that such materials often exploited emotional triggers, employed simplified binary oppositions, and utilised manipulative visual techniques.



Figure 1. First sample of counter-propaganda material
Source: Brand Ukraine (n.d.)



Figure 2. Second sample of counter-propaganda material
Source: Center for Countering Disinformation (n.d.)

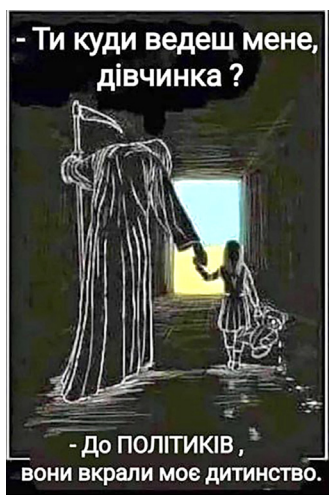


Figure 3. First sample of disinformation material
Source: taken from the Facebook platform

Particularly dangerous are materials created using AI that appear sufficiently convincing for uncritical perception but contain characteristic signs of artificial generation (anatomical anomalies, perspective distortions) (Fig. 4).



Figure 4. Second sample of disinformation material
Source: taken from the Facebook platform

The experimental implementation of the developed system for the automated creation of visual materials to counter disinformation has significantly improved the efficiency of graphic designers compared to traditional approaches. The quality of the materials, assessed according to the criteria of informativeness, visual appeal, and persuasiveness, remained at a high level.

The proposed technological conveyor for the automated creation of visual content to counter disinformation included the following stages:

1. Collection of information from specified media resources and social networks through RSS aggregation (Really Simple Syndication).
2. Analysis of the collected data using artificial intelligence algorithms to detect potential disinformation and its key narratives.
3. Generation of counter-propaganda content using large language models.
4. Creation of visual content using generative AI models under the guidance of a designer.
5. The dissemination of created material through social networks using automated mailing systems.

An important outcome of the research was the development of recommendations regarding optimal approaches to prompt formulation for generative AI models aimed at creating effective visual content to counter misinformation in wartime conditions.

The author's research has shown that the effectiveness of AI-generated visual materials directly depends on the quality and structure of the prompts. The most successful prompts followed a clear structure: a precise description of the genre and format (e.g. informational poster, explanatory infographic, counter-propaganda poster); specification of the visual style (minimalist design, photorealistic

collage, graphic illustration); description of the colour scheme (with particular attention to the psychological impact of colours); compositional layout (placement of elements and focal points); the core message (a clearly articulated counter-propaganda statement); and factual basis (real data used to refute disinformation).

For instance, an effective prompt for creating material aimed at exposing manipulation around the food crisis might read as follows: Create an information poster in A3 format in a minimalist style using a contrasting colour scheme (yellow and blue as the main colours, red for accents) The composition should divide the poster into two parts: the left shows the actual statistics of Ukrainian grain exports, while the right displays the manipulative statements of Russian propaganda with a clear visual marking of "FAKE". Include a visual metaphor of weighing scales, with the factual side outweighing the disinformation. Include the actual export figures for 2023. In contrast, less effective prompts were marked by a lack of clear structure, overly vague language, and the absence of a factual foundation, such as: Create a nice poster about the fakes regarding Ukrainian wheat. Brightly and convincingly.

The analysis of the results of working with various types of prompts showed that the more details and structural elements the prompt contained, the more accurate and effective the generated visual material proved to be. In particular, several elements were found to be especially important in prompts intended for the creation of counter-propaganda content: emotional tone of the content – specifying in the prompt the need for a balance between emotional impact and rational argumentation yielded better results than focusing solely on either emotional appeal or factual information; cultural context – adapting the content to the cultural background of the target audience by including instructions in the prompt to use recognisable cultural symbols and visual codes significantly improved relevance and resonance; technical specifications – including details such as size, format, and intended distribution channel (e.g. social media, print materials, outdoor advertising) in the prompt facilitated better adaptation of the content to the specific medium; duration of impact – indicating in the prompt whether the material was intended for short-term or long-term impact influenced the selection of expressive techniques and the level of content detail.

An additional significant outcome was the identification of a synergistic effect from the combination of various neural networks for the creation of complex visual materials. Indeed, the combination of Midjourney's capabilities for creating the visual foundation, Stable Diffusion for stylistic adaptation, and DALL-E for informational elements allowed for the production of more layered and effective visual solutions. Based on the conducted analysis, a comprehensive methodology for the integration of artificial intelligence technologies into the processes of creating counter-propaganda materials has been developed. The methodology includes:

1. The algorithm for automated detection of potential disinformation is a systematic monitoring of the information space using machine learning to identify typical patterns of manipulative content.

2. The model for generating counter-narratives involves the development of strategies for refutation that are specific to each type of disinformation, using large language models.

3. The system of automated visualisation of counter-narratives – the transformation of textual rebuttals into visual content using generative AI models, taking into account the psychological aspects of perception and the semiotic characteristics of visual communication.

4. The mechanisms of evaluation and optimisation are analytical tools for assessing the effectiveness of created materials and enhancing the quality of future content.

An important component of the methodology was the development of principles for the interaction between artificial intelligence and the designer, where AI acts as an extension of creative capabilities rather than a replacement for human creativity. Within the framework of semiotic analysis, three key strategies of visual resistance to disinformation were identified, all of which can be effectively implemented using AI technologies: the strategy of contrast – visually opposing truthful information and disinformation with clear labelling (e.g. verified/refuted, fact/fake); the strategy of narrative deconstruction – visually breaking down manipulative narratives into their constituent parts to expose their falsity or manipulative nature; the strategy of emotional recoding – transforming negative emotions triggered by disinformation into constructive responses through the use of visual metaphors and symbols.

The obtained results confirm the conclusions of A. Borji (2022) conducted a comparative analysis of generative models and identified various capabilities for creating visual content. This research has expanded these conclusions by focusing on the specifics of applying these technologies to counter disinformation, particularly in the context of the graphic design of informational materials. Graphic design and visual communications are particularly important tools in information confrontation, as they allow for the rapid and effective delivery of key messages to the target audience. As noted by T.H. Dibrova *et al.* (2018), the primary advantage of visual communication is the speed of their perception and the immediate response of consumers to communication signals. While textual information is perceived sequentially and requires significant cognitive resources for processing, visual messages are processed by the brain holistically and considerably faster. According to a study by MIT (Potter *et al.*, 2014), the time taken to perceive a quality graphic image is approximately 13 milliseconds, which is 60,000 times faster than the perception of a text message with equivalent informational content. In the context of information overload for the modern media consumer, such speed of perception becomes a decisive factor for effective communication.

The identified optimisation of designers' workflows by 20% due to the use of AI correlates with the data presented in the study by S.S. Gerenko (2023), which also noted the ability of artificial intelligence to enhance the effectiveness of design solutions while maintaining the need for human oversight and creative direction. The author's conclusions regarding the structure of effective visual materials for countering disinformation align with the research of O.V. Kolisnyk *et al.* (2023), which identified that Midjourney is a powerful tool for generating design graphics. However, unlike this study, the author has found that the most effective are not individual generative models, but their combined use with subsequent professional refinement.

The results of the analysis conducted by Y.I. Siuiva (2024) indicated that machine learning algorithms trained on large datasets identified fake news based on linguistic, semantic, and contextual features of the text with high accuracy. The main approaches identified were naïve Bayes classifiers, logistic regression, decision trees, random forests, support vector machines, and deep learning (Siuiva, 2024). According to the research by L. Graves (2018), fully automated fact-checking has remained a distant aim. Social media continued to employ a combination of AI for the most repetitive operations and human oversight for the analysis of more complex cases. Particularly emphasised was the necessity of maintaining human control in AI systems to ensure ethical use and correct interpretation of data (Graves, 2018). The study by J. Bullock & M. Luengo-Oroz (2019) demonstrated that using the pre-trained AWD-LSTM model, it is possible to easily generate texts in the style of speeches by political leaders at the UN's General Assembly, which confirmed the potential risks of abusing such technologies.

M. Kryzhnia (2023) in her research emphasised the dual nature of AI, which can both assist in countering disinformation and be used to create it. The authors' results confirm this thesis, demonstrating that the ethical responsibility for the use of AI in the design of counter-propaganda materials lies with the designer. However, unlike M. Kryzhnia, the author has developed specific recommendations for the ethical use of AI in creating effective visual materials. The articles of I.M. Tsidylo & C.E. Sendra (2023) examined the use of Midjourney as a methodological innovation in the training of future designers. This research complements the findings, demonstrating the practical application of this and other generative models in the context of information warfare and countering disinformation. C. Marsden & T. Meyer (2019) emphasised the importance of keeping the human "in the loop" of AI systems to ensure ethical use and correct interpretation of data. These results confirm this thesis, demonstrating that the most effective materials were those created as a result of collaboration between the designer and AI systems, rather than a fully automated process. In contrast, W. Liu & R. Bezuhla (2024) demonstrated that artificial intelligence plays a key role in image generation in design, as it is capable of not only automating individual stages of visual design but also

refining outcomes, supporting the creative process, and influencing aesthetic decisions.

Particular interest lies in comparing current findings with the research of K. Kertysova (2019), who considered AI as a bilateral tool in the context of disinformation. This outcome confirms her thesis regarding the potential of AI for both the creation and countering of disinformation; however, it focuses specifically on developing methodological approaches for harnessing this potential within graphic design. N. Dufour *et al.* (2024) noted the rapid increase in the amount of AI-generated content in fake news since spring 2023. The current study confirms this trend and proposes specific approaches to countering this threat through visual communication methods. O. Hrozna (2024) in her research emphasised that global digital transformation has significantly impacted journalism and design. The present study expands its conclusions, demonstrating how these transformations can be utilised to counter disinformation through visual tools.

Unlike the studies considered, the author's study offers a comprehensive methodology for the integration of artificial intelligence technologies into the processes of creating visual tools for countering disinformation, which takes into account technological aspects and psychological factors of perception of visual information, as well as semiotic features of communication in the context of information warfare. The results support the conclusion that the optimal use of AI technologies for designing visual tools to counter disinformation requires: a clear understanding of the psychological mechanisms of visual information perception; the ability to formulate effective queries to generative models; professional evaluation and refinement of the generated content; integration of the created materials into comprehensive information campaigns.

Conclusions

The conducted research on the use of artificial intelligence for the creation of information campaigns in wartime, particularly visual tools for countering disinformation, has led to the following conclusions. It has been established that modern artificial intelligence technologies significantly expand the capabilities of graphic design in the context of creating counter-propaganda materials. Generative models, such as Midjourney, Stable Diffusion, and DALL-E, have enabled the optimisation of up to 20% of routine tasks in the creation of visual content, allowing designers to focus on the strategic and creative aspects of development. However, as the research showed, the most effective approach turned out to be a hybrid one that combined automation through AI with professional refinement of materials by designers. A comprehensive methodology for the integration of artificial intelligence technologies into the processes of creating visual tools for countering disinformation has been developed. The proposed methodology included algorithms for the automatic detection of potential disinformation, the generation of relevant counter-narratives, and their visualisation using

generative models. An important element of the developed methodology was the recommendations for the formation of effective prompts for generative models, which took into account the psychological aspects of the perception of visual information and the semiotic characteristics of communication in the context of information warfare.

The semiotic analysis of visual materials has revealed the most effective visual codes and sign systems for countering disinformation. These include: contrasting visual comparisons of “truth/falsehood”, symbols of fact-checking, visual metaphors of revelation or exposure, as well as elements that emphasise transparency and credibility. The findings have practical value for the development of a design system of visual tools for countering disinformation. The experimental implementation of the developed system for the automated creation of visual materials for countering disinformation has demonstrated a significant improvement in operational efficiency compared to traditional approaches while maintaining a high quality of materials, which confirms the effectiveness of the proposed methodology and its potential for widespread adoption in the practice of information campaigns. The conducted research has both theoretical and practical significance. From a theoretical perspective, the research expands the understanding of the role of artificial intelligence in contemporary design practices and its potential to counter

misinformation. From a practical standpoint, the developed recommendations and methodological approaches can be directly implemented in the work of design studios, media organisations, and state institutions engaged in countering disinformation.

The prospects for further research lie in the development of specialised software solutions that integrate the identified approaches to the automated creation of visual tools for countering disinformation; the study of the long-term effects of visual counter-propaganda campaigns; and the investigation of adaptive design systems that automatically adjust to different target audiences and cultural contexts. The development of educational programmes for designers, aimed at the effective use of artificial intelligence to counter disinformation, deserves special attention, taking into account both the technical and ethical aspects of this activity.

Acknowledgements

None.

Funding

None.

Conflict of Interest

None.

References

- [1] Borji, A. (2022). *Generated faces in the wild: Quantitative comparison of Stable Diffusion, Midjourney and DALL-E 2*. doi: 10.48550/arXiv.2210.00586.
- [2] Bozhko, T., & Arefiev, V. (2023). Neural networks as a graphic design tool. *Bulletin of KNUKIM. Series in Arts*, 48, 125-135. doi: 10.31866/2410-1176.48.2023.282475.
- [3] Brand Ukraine. (n.d.). *Russia's famine games*. Retrieved from <https://brandukraine.org.ua/uk/projects/golodni-igri-rosiyi/>.
- [4] Bullock, J., & Luengo-Oroz, M. (2019). Automated speech generation from UN General Assembly statements: Mapping risks in AI generated texts. In *International conference on machine learning AI for social good workshop* (pp. 1-5). Long Beach. doi: 10.48550/arXiv.1906.01946.
- [5] Center for Countering Disinformation. (n.d.). Retrieved from <https://cpd.gov.ua>.
- [6] Center for Democracy and Rule of Law. (2023). *Disinformation and artificial intelligence: Invisible threat to modernity*. Retrieved from <https://cedem.org.ua/analytics/dezinformatsiya-shtuchnyi-intelekt/>.
- [7] Derevyanko, N., & Zalevska, O. (2023). Comparative analysis of neural networks Midjourney, Stable Diffusion, and DALL-E and ways of their implementation in the educational process of students of design specialties. *Scientific Bulletin of Mukachevo State University. Series "Pedagogy and Psychology"*, 9(3), 36-44. doi: 10.52534/msu-pp3.2023.36.
- [8] Dibrova, T.H., Solntsev, S.O., & Bazherina, K.V. (2018). *Advertising management: Theory and practice*. Kyiv: Kyiv Polytechnic Institute named after Igor Sikorsky.
- [9] Dufour, N., et al. (2024). *AMMeBa: A large-scale survey and dataset of media-based misinformation in-the-wild*. doi: 10.48550/arXiv.2405.11697.
- [10] Gerenko, S.S. (2023). *Artificial intelligence in graphic design: Challenges and perspectives*. (Doctoral dissertation, Kyiv National University of Culture and Arts, Kyiv, Ukraine).
- [11] Graves, L. (2018). Understanding the promise and limits of automated fact-checking. *Factsheet*, 1-8. doi: 10.60625/risj-nqnx-bg89.
- [12] Hlushchuk, Y. (2023). Fakes as a pressure tool in war conditions: Specifics of application and perception. *Scientific Papers of Vernadsky National Library of Ukraine*, 67, 96-107. doi: 10.15407/np.67.096.
- [13] Hrozna, O. (2024). Technological innovations in online media: The role of artificial intelligence and virtual reality in content transformation. *Information Technologies and Learning Tools*, 1(15), 102-112. doi: 10.20535/2522-1078.2024.1(15).302843.

- [14] Kertysova, K. (2019). Artificial intelligence and disinformation: How AI changes the way disinformation is produced, disseminated, and can be countered. *Security and Human Rights*, 29, 55-81. doi: [10.1163/18750230-02901005](https://doi.org/10.1163/18750230-02901005).
- [15] Kolisnyk, O.V., Mykhailova, R.D., Beregovyi, O.S., Vlasjuk, V.V., & Kurovska, D.V. (2023). Midjourney Neural Network as a tool for generating design graphics. *Art and Design*, 1(21), 106-115. doi: [10.30857/2617-0272.2023.1.10](https://doi.org/10.30857/2617-0272.2023.1.10).
- [16] Kolosnichenko, M.V., et al. (2023). *Graphic design in information and visual space*. Riga: Baltija Publishing.
- [17] Kryzhnia, M. (2023). Artificial intelligence in creativity and at war. About opportunities, risks and copyright. *Suspilne Culture*. Retrieved from <https://suspilne.media/culture/426576-stucnij-intelekt-u-tvorcosti-ta-na-vijni-pro-mozlivosti-riziki-ta-avtorske-pravo/>.
- [18] Liu, W., & Bezuhla, R. (2024). Comparative analysis and collaborative innovation of parametric generation and aigc. *Art and Design*, 4, 22-32. doi: [10.30857/2617-0272.2024.4.2](https://doi.org/10.30857/2617-0272.2024.4.2).
- [19] Marsden, C., & Meyer, T. (2019). *Regulating disinformation with artificial intelligence: Effects of disinformation initiatives on freedom of expression and media pluralism*. Brussels: European Parliamentary Research Service (EPRS). doi: [10.2861/003689](https://doi.org/10.2861/003689).
- [20] Mykytiv, H.V., & Kondratenko, Y. (2016). [Non-textual elements as a means of forming media culture of readers of popular science journals](#). In *Actual problems of media education in Ukraine and the world: Collection of abstracts of reports of the international scientific-practical conference* (pp. 50-53). Zaporizhzhia: Zaporizhzhya National University.
- [21] Potter, M.C., Wyble, B., Hagmann, C.E., & McCourt, E.S. (2014). Detecting meaning in RSVP at 13 ms per picture. *Attention, Perception, & Psychophysics*, 76, 270-279. doi: [10.3758/s13414-013-0605-z](https://doi.org/10.3758/s13414-013-0605-z).
- [22] Pryshchenko, S.V. (2020). Visual semantics and artistic imagery of the poster. *Culture and Modernity*, 1, 97-103. doi: [10.32461/2226-0285.1.2020.221344](https://doi.org/10.32461/2226-0285.1.2020.221344).
- [23] Rombach, R., Blattmann, A., Lorenz, D., Esser, P., & Ommer, B. (2022). *High-resolution image synthesis with latent diffusion models*. doi: [10.48550/arXiv.2112.10752](https://doi.org/10.48550/arXiv.2112.10752).
- [24] Salvo, D., Fly, J., & Rosenberger, L. (2018). Policy blueprint for countering authoritarian interference in democracies. *The German Marshall Fund of the United States (GMF)*. Retrieved from <https://www.gmfus.org/publications/asd-policy-blueprint-countering-authoritarian-interference-democracies>.
- [25] Siuiva, Y.I. (2024). [Artificial intelligence technologies in the practice of detecting fake news](#). (Master's thesis, National University of Kyiv-Mohyla Academy, Kyiv, Ukraine).
- [26] Tsidylo, I.M., & Sendra, C.E. (2023). Artificial intelligence as a methodological innovation in the training of future designers: Midjourney tools. *Information Technologies and Learning Tools*, 97(5), 203-214. doi: [10.33407/itlt.v97i5.5338](https://doi.org/10.33407/itlt.v97i5.5338).

Використання штучного інтелекту для створення інформаційних кампаній в умовах війни: візуальні інструменти протидії дезінформації

Владислав Щука

Магістр

Київський національний університет культури і мистецтв

01601, вул. Євгена Коновальця, 36, м. Київ, Україна

<https://orcid.org/0009-0004-1550-4727>

Анотація. В умовах сучасної інформаційної війни критично важливим став розвиток ефективних методів протидії дезінформації та ворожій пропаганді. Кількість неправдивої інформації та швидкість її поширення вимагали впровадження автоматизованих систем із залученням штучного інтелекту для оптимізації процесів створення візуального контрпропагандистського контенту. Мета полягала у розробці методології інтеграції технологій штучного інтелекту в процеси створення ефективних візуальних інструментів протидії дезінформації з урахуванням принципів графічного дизайну та психології сприйняття візуальної інформації. Дослідження базувалося на комплексному підході, що поєднував теоретичний аналіз наукової літератури, порівняльний аналіз нейромереж Midjourney, Stable Diffusion та DALL-E, семіотичний аналіз візуальних матеріалів, а також експериментальне впровадження розробленої системи автоматизованого створення контрпропагандистського візуального контенту. Розроблено комплексний підхід до створення візуальних інструментів протидії дезінформації, який поєднав можливості систем автоматизованого збору інформації, алгоритмів штучного інтелекту для генерації графічного контенту та принципів ефективного графічного дизайну. Виявлено, що використання штучного інтелекту в графічному дизайні оптимізувало до 20 % рутинних завдань при створенні візуального контенту, дозволяючи дизайнерам зосередитись на стратегічних і творчих аспектах розробки. Розроблені рекомендації щодо використання ШІ у графічному дизайні можуть бути впроваджені державними інституціями, медіа та громадськими організаціями для оперативного реагування на інформаційні загрози

Ключові слова: графічний дизайн; візуальна комунікація; контрпропагандистський контент; інформаційна війна; генеративний дизайн; психологія сприйняття; композиційні рішення

ТЕХНОЛОГІЇ ТА ІНЖИНІРИНГ

Науковий журнал

Том 26, № 2, 2025

Оригінал-макет видання виготовлено
у редакційно-видавничому відділі Київського національного університету технологій та дизайну

Відповідальний редактор:

О. Кривонос

Підписано до друку 30.04.2025 р. Формат 60*84/8
Умовн. друк. арк. 11,7
Наклад 100 примірників

Адреса видавництва:

Київський національний університет технологій та дизайну
01011, вул. Мала Шияновська, 2, м. Київ, Україна
тел/факс: +38 (044) 256-84-27
E-mail: info@technologies-engineering.com.ua
<https://technologies-engineering.com.ua/uk>

TECHNOLOGIES AND ENGINEERING

Scientific Journal

Vol. 26, No. 2, 2025

The original layout of the publication is made
in the publishing department of Kyiv National University of Technologies and Design

Managing editor:

O. Kryvonos

Signed for print 30.04.2025. Format 60*84/8
Conventional printed pages 11.7
Circulation 100 copies

Publishing Address:

Kyiv National University of Technologies and Design
01011, 2 Mala Shyianovska Str., Kyiv, Ukraine
tel/fax: +38 (044) 256-84-27
E-mail: info@technologies-engineering.com.ua
<https://technologies-engineering.com.ua/en>