BULLETIN of Cherkasy State Technological University

Journal homepage: https://bulletin-chstu.com.ua/en

Vol. 29 No. 4. 2024

UDC 004.5:004.032.26
DOI: 10.62660/bcstu/4.2024.21

Optimisation of intelligent system algorithms for poorly structured data analysis

Mykola Demchyna*
PhD in Technical Sciences, Associate Professor
King Danylo University
76018, 35 Ye. Konovalets Str., Ivano-Frankivsk, Ukraine
https://orcid.org/0009-0002-9161-4843

Taras Styslo
Senior Lecturer
King Danylo University
76018, 35 Ye. Konovalets Str., Ivano-Frankivsk, Ukraine
https://orcid.org/0000-0002-2377-7985

Serhii Vashchyshak
PhD in Technical Sciences, Associate Professor
King Danylo University
76018, 35 Ye. Konovalets Str., Ivano-Frankivsk, Ukraine
https://orcid.org/0000-0002-1753-1540

Abstract. Integration of heterogeneous types of medical data using modern deep learning methods can improve the accuracy and efficiency of diagnosing complex diseases, such as cardiovascular diseases, which is relevant for personalised medicine and reducing the risk of medical errors. The study aimed to present the development of a decision support system for improving the diagnosis of cardiovascular diseases by integrating heterogeneous types of medical data. To create the knowledge base, data from real clinical scenarios were used, which underwent the stages of cleaning, standardisation, and semantic analysis using specialised medical dictionaries. The system demonstrated high efficiency due to its ability to integrate text, image and signal data into a single analysis process. The efficiency was evaluated by such metrics as accuracy, completeness, F1-score, and predictive values of positive and negative results. The introduction of transformers ensured a 15% increase in diagnostic accuracy compared to traditional methods, and the use of a hybrid computing approach reduced model training time by 30% and enabled the processing of up to 1 TB of data per day. Additionally, the integration of heterogeneous types of medical data into the system has improved the personalisation of diagnostics, accounting for individual patient characteristics such as medical history, genetic factors, or comorbidities. Transformer attention mechanisms improved resistance to noise and data gaps, which ensures reliable results even with incomplete or inaccurate information. Optimisation of the models reduced delays in data processing, which is critical for prompt clinical decision-making. In addition, transformers have proven their ability to dynamically scale to process new types of data without losing efficiency, opening opportunities

Article's History: Received: 16.07.2024; Revised: 11.10.2024; Accepted: 16.12.2024.

Suggested Citation:
Demchyna, M., Styslo, T., & Vashchyshak, S. (2024). Optimisation of intelligent system algorithms for poorly structured data analysis. *Bulletin of Cherkasy State Technological University*, 29(4), 21-31. doi: 10.62660/bcstu/4.2024.21.

*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/)

for further expansion of the system's functionality. The system has also increased the productivity of clinical specialists by automating routine tasks, allowing doctors to focus on more complex aspects of treatment

Keywords: artificial intelligence; knowledge base; decision support; neural networks; knowledge mining; classification algorithms; adaptive systems

INTRODUCTION

Modern medicine is facing increasing challenges in analysing large volumes of unstructured data, including text-based physician notes, medical imaging, and signal data. Traditional approaches to processing such information are often ineffective due to the limited ability to integrate heterogeneous data types into a single analysis process. This limits the accuracy of diagnosis and the speed of clinical decision-making, especially in the context of complex diseases such as cardiovascular disease.

The integration of artificial intelligence (AI) and neural networks in medicine creates new opportunities for improving the quality of diagnosis and treatment. Modern neural network architectures, such as multilayer perceptron (MLP), recurrent neural networks (RNN), convolutional neural networks (CNN), and transformers, improve the efficiency of various data format processing. The use of transformers with attention mechanisms enables complex pattern detection between text, image, and signal data, which is critical for medical research. The study by J. Meng & Z. Wang (2022) proposes a model that integrates transformer and RNN mechanisms to better understand the context of words and phrases, achieving an 18% increase in classification accuracy. Transformers provide a deeper analysis of long-term dependencies between words, while RNNs efficiently process text sequences, considering the local context. This combination not only increases the classification accuracy but also improves the model's ability to work with heterogeneous text data containing complex semantic relationships.

Z. Shen (2023) addressed the optimisation of data visualisation platforms based on Artificial Intelligence (AI). In particular, the study described approaches to interactive data analysis in real-time, which can reduce rendering time by 25%. The study emphasised the use of adaptive algorithms for processing large amounts of data, which ensures the smooth operation of platforms and their ability to scale. The study also emphasised the importance of integrating different data sources and applying effective methods to reduce computational costs, which is critical to ensuring the high performance of systems in real-time. D. Kushnir *et al.* (2021) analysed the use of deep neural networks for semantic text analysis. The main idea is to integrate vector representation of words (word embeddings) with transformers for natural language analysis. The results show that the model provides a 12% increase in text classification accuracy compared to classical machine learning methods.

The review analysis by C. Gambella *et al.* (2021) covered modern optimisation problems in machine learning. The authors reviewed methods for optimising hyperparameters, ways to avoid overfitting problems, and an approach to model selection. Improvement of the computational efficiency of algorithms, which can significantly reduce model training time without losing accuracy, was also prioritised. In addition, the authors analysed the role of hybrid computing approaches that combine local and cloud resources to ensure the scalability and adaptability of machine learning. The study emphasised the importance of integrating heterogeneous data and using transformative architectures to improve the accuracy of analysis and stability of models in different environments.

A. Wilson & M.R. Anwar (2024) discussed adaptive machine learning algorithms for processing multivariate data. A new approach to data dimensionality reduction using stochastic methods was proposed, which significantly reduces computation time without loss of accuracy. The process of integrating neural networks for analysing poorly structured data was emphasised. Three types of architectures were compared: multilayer perceptron (MLP), recurrent neural networks (RNN), and transformers. MLPs were used to process structured data, such as test results, RNNs were used to analyse text records, while transformers provided image processing and integration with other data. The results showed that transformers had the highest accuracy in diagnosis, outperforming traditional methods by 15%, while RNNs demonstrated significantly better text processing compared to other approaches.

N. Abbas & S. Gasmi (2024) analysed in detail the approaches to optimising machine learning algorithms, in particular for natural language processing and text analytics. The authors addressed the development of efficient methods for processing streaming data in real-time, which is critical in modern dynamic information systems. The main goal of the study was to improve the performance of algorithms by reducing delays in data processing and optimising the use of computing resources. T.B. Brown *et al.* (2020) analysed natural language processing models that can perform tasks with a minimum number of examples (few-shot learning). The authors demonstrated how large language models, such as GPT-3, can adapt to new tasks without the need for large amounts of training data. The key task was to optimise the resources required to process large amounts of data without losing the accuracy of the results. For this purpose, a hybrid learning methodology that combines local data processing with cloud computing was used. A detailed analysis of the system performance was conducted: the model training time was reduced by 30% due to parameter optimisation, and the use of cloud computing resources allowed processing up to 1 terabyte of data per day.

The study aimed to develop and evaluate the effectiveness of a decision support system (DSS) for medical diagnoses that integrated heterogeneous data types to improve the accuracy of cardiovascular disease diagnosis. As part of the study, a knowledge base was created that included data from real clinical scenarios that went through the stages of cleaning, standardisation and semantic analysis. The development of a methodology for data integration and neural network optimisation was emphasised.

MATERIALS AND METHODS

The research was based on the development of a decision support system (DSS) for medicine that can integrate heterogeneous types of medical data. For this purpose, a knowledge base was created that includes information from various sources: textual records of doctors, laboratory data, computed tomography (CT) images and electrocardiographic (ECG) signals. Ethical approval covered all aspects of the study, including procedures for data collection, processing, analysis and storage. The Ethics Committee reviewed whether the study met ethical standards, including the informed consent of patients (if required) or the appropriate legal basis for using anonymised data. This correlated with the principles stipulated in the World Medical Association's Declaration of Helsinki (1964), the Law of Ukraine No. 2297-VI (2010), and the European General Data Protection Regulation (GDPR) (2016). All the steps were designed to comply with the principles of transparency, responsibility and security of data processing. Particular attention was devoted to standardisation of text data, which was processed using natural language processing (NLP) methods such as tokenisation, text normalisation and semantic analysis using specialised medical dictionaries.

Different neural network architectures were used to integrate the data into the system, with appropriate optimisation algorithms to ensure that each model was trained efficiently. Multilayer Perceptron (MLP) was used to analyse structured data such as laboratory parameters. The Adaptive Moment Estimation (AdamW) algorithm was used to train this model, which provides fast convergence and stability when dealing with structured numerical data. Recurrent neural networks (RNNs) were used to process sequential text records, such as medical reports, to allow for context. For this model, the RMSprop algorithm was used, which works well with sequential data and avoids the problem of exploding or disappearing gradients typical of RNNs. Transformers were used to integrate text, image, and signal data. Attention mechanisms detected complex patterns between heterogeneous data. They were trained using the AdamW algorithm (adaptive optimisation with weight decay), which reduces the risk of over-computing gradients and improves regularisation, which is critical when working with large datasets and complex architectures. Convolutional Neural Networks (CNNs) were used to analyse medical images, in particular CT images. For their optimisation, the SGD (Stochastic Gradient Descent) algorithm with momentum was used, which provides high efficiency in computer vision tasks such as image classification and segmentation. Both RNNs and transformers were used to process ECG signals. The RNNs used the RMSprop algorithm, which improved the efficiency and consistency of physiological data, while the transformers were trained using AdamW, ensuring stable convergence and reducing the risk of overfitting.

These optimisation algorithms were carefully selected following the specifics of the data being processed and the architecture of the neural network. This was used to achieve high efficiency and accuracy of the system, ensuring high-quality integration of text, image and signal data within a single analytical process. The study results were evaluated using a sample from the control set of patients whose data were not used during the model training phase. This limited overtraining and ensured the independence of the evaluation. The sample included cases with varying degrees of complexity, covering mild, moderate and severe forms of disease, to represent the full range of clinical scenarios. To increase the objectivity of the assessment, heterogeneity of the data, including textual records of doctors, images of medical examinations (e.g., CT scans), electrocardiographic (ECG) signals and laboratory indicators, were prioritised. Additionally, proportionality in terms of age and gender characteristics of patients was addressed to avoid bias in statistical results, as well as comorbidities that could complicate diagnosis. All patient data was anonymised, in compliance with ethical principles and legal standards, such as the Law of Ukraine No. 2297-VI (2010) and the European General Data Protection Regulation (GDPR) (2016). This ensured the legality and transparency of the information collection and analysis process. This approach ensured the most objective assessment of the effectiveness of the developed system.

To evaluate the system's performance, the models were benchmarked using metrics such as accuracy, recall, F1-score, and clinically relevant indicators such as the positive predictive value (PPV) and negative predictive value (NPV). The models were trained on large volumes of labelled data, which ensured high diagnostic accuracy. In addition, the computational efficiency of the developed system was evaluated, including model training time and performance when processing large data sets. To optimise computational costs, a hybrid approach was used, combining local computing and scaling in the cloud environment.

RESULTS

The first step was to evaluate the effectiveness of traditional approaches such as individual MLP architectures that provided intermediate results when analysing structured data, such as laboratory results. Such approaches proved to be useful in cases where the information was clearly structured and easily formalised. However, during the processing of contextual factors (e.g., the sequence of textual notes from doctors describing a patient's medical history) or the complex structure of image data, the accuracy of MLP models was not high enough. RNN architectures were more efficient at analysing text containing descriptions of symptoms, doctors' opinions, and recommendations. Due to the memory mechanism and the ability to account for the temporal sequence, RNN models demonstrated better completeness, and F1-score compared to MLPs when processing text data. However, their ability to integrate different types of data and consider images or signals remained limited, as RNNs mainly focused on sequential text input.

The use of transformers has become a key step towards achieving the highest performance. Due to their attention mechanisms, transformers can identify relevant pieces of information in different types of data, compare them with each other, and build complex links between medical records, CT images, and ECG signals (Zhang *et al.*, 2024). Experiments have shown that transformative architectures can achieve 15% higher accuracy than traditional methods that work with only one type of data or combine data less efficiently. This result suggests that transformers can detect complex patterns that other models have failed to capture. This is especially important for medical diagnostics, where even small correlations between different indicators can be critical. Table 1 illustrates the accuracy, precision, recall, and F1-score scores for the three main types of neural models: MLPs, RNNs, and Transformers.

Table 1. Comparison of the efficiency of different neural network architectures for the diagnosis of cardiovascular diseases

Architecture	Accuracy	Precision	Recall	F1-score
MLP	0.82	0.79	0.80	0.795
RNN	0.87	0.85	0.84	0.845
Transformers	0.95	0.93	0.92	0.925

Source: compiled by the authors

Following Table 1, the MLP architecture has the lowest results among the models considered. This can be attributed to the fact that multilayer perceptrons are effective with structured data but lose efficiency in situations where it is necessary to integrate information from different sources. RNN architectures have significantly improved the accuracy and completeness rates compared to MLPs by processing textual data in a consistent form, which allows for a better understanding of the contextual information available in medical history. However, the limitations of RNNs in simultaneously integrating complex visual and signal characteristics did not allow them to achieve transformer-like performance.

Transformer architectures proved to be the leaders across the board. Their ability to accommodate different data formats and identify complex relationships between them allowed them to significantly exceed the accuracy achieved by other models. In addition, transformers ensured that the results remained stable even when new data types were included or updated. This approach is especially important for medical applications where examination protocols may change periodically, or new research methods may appear. Given the above results, it is possible to argue that transformers are best suited to the needs of complex medical diagnostics, as they provide the most accurate and reliable results when working with heterogeneous, poorly structured data.

A critical aspect of medical diagnostics is not only the accuracy but also the speed of results. In situations where a clinical decision needs to be made within a limited time, it is also important to evaluate the performance of the models, i.e., their ability to learn quickly and process large amounts of information. To this end, a detailed analysis of training time and computational costs for different architectural approaches was conducted. The experiment considered both local computing resources and the cloud infrastructure, which made it possible to scale the load and ensure the processing of up to 1 TB of data per day. As a result of optimising hyperparameters and applying a hybrid computing approach, model training time was reduced by 30% compared to standard methods that relied solely on local resources or did not use optimisation strategies. Table 2 illustrates the average model training time on a benchmark dataset (approximately 500 GB), as well as the amount of data that can be processed per day after optimisation. Table 2 shows the results for MLP, RNN, and transformers, addressing the use of hybrid infrastructure.

Table 2. Computational efficiency of different neural network architectures in medical data processing

Model	Learning time (hours)	Processing volume per day (GB)
MLP	10	600
RNN	12	700
Transformers	7	1,000

Source: compiled by the authors

Table 2 shows that MLP models were trained relatively quickly, but their limited ability to integrate complex data negatively affected the final accuracy and usefulness in a clinical context. RNNs require more time to learn due to the processing of sequential information and context-preservation mechanisms. However, even with longer training times, RNNs did not reach the accuracy of transformers. In turn, the transformers proved to be optimal not only in terms of accuracy but also in terms of computational efficiency, as their training time was only 7 hours on average, and the implemented hybrid infrastructure allowed processing up to 1 TB of data per day, which is 30% faster than previous approaches. Such computational optimisation is extremely important as modern medical systems generate large amounts of data, and their prompt processing is one of the key conditions for timely diagnosis.

The quick large amounts of data processing, combined with the high accuracy of transformational models, means that the developed DSS can provide reliable results in real-time or close to it. This significantly increases the value of the system for medical practice, as clinicians can make complex diagnostic decisions quickly and confidently. For example, in the case of suspected acute cardiac conditions such as myocardial infarction, a timely and accurate diagnosis can save a patient's life by allowing them to start appropriate treatment immediately. While traditional methods require several hours to analyse a complex data set, the new system, thanks to transformers and a hybrid computing approach, can provide a preliminary decision much faster. In addition to time and computational aspects, an important criterion for the system's effectiveness is the ability to integrate and analyse different types of medical data simultaneously. To confirm this, additional experiments were conducted to evaluate the diagnostic accuracy when new sources of information were added sequentially. Initially, only textual records of medical histories were used, then laboratory values were added, and then CT scans and ECG signals were connected. The results showed that the transformers demonstrated a steady improvement in accuracy and F1-score with each new type of data. This means that the model is not "overloaded" with additional information, but rather successfully forms interconnections, improving the quality of its predictions. For a medical diagnostic system, this result is extremely valuable, as it confirms that the introduction of new data sources (e.g., new types of visual examinations or more accurate sensory signals) will allow the model to dynamically improve its estimates without the need for a fundamental rebuilding of the entire architecture.

In medical practice, it is necessary to account for individual patient characteristics, such as pre-existing conditions, and reactions to certain drugs or genetic markers. With transformers, the model was able to focus on key segments of medical data relevant to a particular patient. This became apparent when analysing cases with atypical symptoms or comorbid conditions when traditional methods or simpler neural network models could not detect hidden patterns. Transformational architectures, on the contrary, can identify unique patterns and form more accurate and personalised diagnostic recommendations based on them. This confirms the importance of this approach for the development of personalised medicine, as more accurate diagnostics pave the way for better and more individually tailored treatment regimens, reducing the risk of inadequate therapy and side effects. In addition, the analysis of the results shows that the introduction of transformers contributes to greater resilience of the system to noise and data errors. Medical data often has a heterogeneous structure, containing not only different formats but also different quality. For instance, doctors' text records may contain spelling errors, abbreviations or acronyms, and images may contain shooting artefacts. ECG signals can be interfered with by patient movement or electrical interference. Traditional processing methods are often vulnerable to such distortions and can either ignore or misinterpret some of the information. Transformers, on the other hand, can dynamically determine which parts of the data should be given more attention and which parts can be partially ignored as noise, thanks to their attention mechanisms. This increases the reliability of the system and provides better results even with incomplete or poorly structured data.

In a real-world clinical setting, a medical decision support system must not only be accurate and efficient but also easy to use. The study concentrated mainly on quantitative evaluation of the results but also considered qualitative aspects. For instance, the time taken to complete a diagnostic decision is reduced not only due to faster data processing but also due to the preparation of more generalised and understandable recommendations for doctors. The transformer-based model provided summary information in the form of concise text messages or annotations on images. This allowed clinicians to quickly evaluate the proposed diagnoses and recommendations. This functionality proved to be extremely useful in situations where doctors were faced with many patients and limited time for each of them. Another aspect of practical importance is the scalability of the system. The results showed that the implemented hybrid infrastructure and the use of transformers allow the system to easily adapt to the growth of data or the emergence of new sources of information (Meng & Wang, 2022). Combining transformers with CNN increases the efficiency of integrating text, numeric, and visual data, which is especially important for systems that work with multimodal data in real-time (Zhang *et al.*, 2024).

In modern healthcare facilities, data volumes are growing rapidly as new examination methods are introduced (e.g., fractional imaging analysis, biometric sensors, patient data from smart devices). The ability to scale quickly without losing accuracy is a crucial factor in ensuring the system's long-term relevance and efficiency. The results confirmed that the proposed architecture is well suited for such conditions, as the models can be updated or retrained on new data sets without significantly reducing the quality of diagnosis. Notably, the introduction of transformers and hybrid computing approaches contributes to better resource utilisation. The optimisation of hyperparameters and adaptive load balancing between local and cloud resources reduced the time and cost of processing without compromising the quality of the results (Nedosnovanyi *et al.*, 2023). This can have significant economic value for healthcare facilities that want to improve diagnostic efficiency without the need for massive investments in new equipment or expensive computer clusters. Instead, by using cloud computing and adapting the system architecture, data processing can be scaled to meet the needs and resources available.

Another important aspect is the system's ability to work with unstructured data, which includes textual notes from doctors, laboratory values, images, and signals. The study results confirmed that the use of deep neural networks significantly improves the analysis of such data. The use of neural networks for semantic text analysis can significantly improve the classification accuracy of poorly structured information (Kushnir *et al.*, 2021). Cognitive computing is also used to effectively analyse unstructured data (Chen *et al.*, 2020). Another aspect of practical importance is the scalability of the system. The implemented hybrid infrastructure makes it easy to adapt to the growth of data volumes or the emergence of new sources of information. The use of AI strategies allows for optimising big data processing in dynamic environments (Oza & Domadiya, 2023; Smetaniuk & Tsisar, 2024).

An important consequence of improved accuracy and speed of data processing is increased confidence in the decision support system. Doctors and other medical professionals need to be confident that the proposed diagnosis is reliable and justified. If the model is highly accurate and stable across different data sets, they will be more willing to use the system in their daily practice. This, in turn, can lead to wider adoption of DSS in the clinical process, increased standardisation of diagnostic decisions, and ultimately improved medical outcomes for patients.

The use of transformers improves the accuracy and speed of data processing but also lays the foundation for further development of medical DSS in the direction of integrating new types of data, improving the interpretation of results and increasing the trust of the medical community. The results obtained confirm that the system can be used in real clinical settings and can contribute to improving the quality of medical services and the processes of diagnosing cardiovascular diseases. Thus, the results of the experiments indicate the high efficiency and prospects of using transformable architectures and hybrid computing approaches in decision support systems for medical diagnostics. Comparison with traditional approaches (MLP, RNN) has demonstrated a significant advantage of transformers in terms of accuracy, completeness, prediction accuracy and F1-score, as well as better computational efficiency and the ability to quickly process large volumes of heterogeneous data. The practical benefit is the ability to significantly reduce diagnostic time, increase the level of treatment personalisation, and provide more reliable and scalable tools for medical practice. The data obtained confirm that the integration of various types of data – text, image, signal and structured – within a single decision support system is quite achievable and brings tangible benefits.

DISCUSSION

The obtained results demonstrate the effectiveness of using transformable architectures and hybrid computing approaches for medical decision support systems (DSS) capable of processing poorly structured data of various natures (text, images, signals). A 15% increase in the accuracy of cardiovascular disease diagnosis compared to traditional methods and a 30% reduction in model training time demonstrate the feasibility of the chosen approach. These results are in line with the general trend in artificial intelligence (AI), machine learning (ML), and natural language processing (NLP) research, where transformers, multitasking learning (MTL), hybrid computing platforms, and hyperparameter optimisation have become key elements in building effective systems.

Numerous studies emphasise the importance of using modern neural architectures and optimisation strategies for processing low-structured data. For instance, I. Malyha & S. Shmatkov (2022) combined word embeddings with multitasking learning, increasing the accuracy of text contextualisation by 15%. This approach is particularly promising for clinical case analysis, where semantic aspects of the text – including contextual nuances and terminological variations – are critical. Considering such works inspires the integration of MTL into the developed system, which can further improve the accuracy of medical record analysis. Similar ideas were expressed by Y. Zhang & Q. Yang (2022), who emphasised the potential of MTL to simultaneously solve several interrelated tasks, such as disease classification, symptom analysis, and complication prediction. The 20-30% increase in model accuracy due to MTL is a strong argument in favour of implementing this approach for medical DSS. This will allow the system to account for the diversity of data, including textual descriptions, images, and signals, by jointly training several complementary objective functions.

G. Zhang *et al.* (2024) demonstrated the benefits of combining transformers and CNNs to integrate textual, visual, and numerical data, which increased the accuracy of analysis by 25%. This integration is particularly relevant for systems that consider CT images of the cardiovascular system, ECG signals, and text records.

The combination of CNNs for image processing with transformers that can integrate text and signal data opens new opportunities for creating complex models that consider all aspects of medical data. A review by R. Rueda *et al.* (2024) analysed the application of machine learning methods for exacerbation detection and clustering in patients with chronic obstructive pulmonary disease (COPD). The authors reviewed approaches to analysing time series and multivariate data to identify precursors of exacerbations and automatically group patients according to the nature of the disease. The paper emphasises the importance of using machine learning to improve the quality of medical care for patients with COPD, focusing on the possibility of early intervention and a personalised approach to treatment. S.V. Mahadevkar *et al.* (2024) addressed the capabilities of transformers and NLP algorithms to analyse poorly structured documents, reducing analysis time by 30%. This is especially important in real-world clinical settings when doctors need prompt recommendations and diagnostic conclusions. Integration of such solutions will ensure faster detection of critical patterns in medical information and timely response to changes in the patient's condition.

The effectiveness of model training depends not only on the choice of architecture but also on the optimal setting of hyperparameters. H. Fatima & S. Gasmi (2023) demonstrated that dynamic tuning of hyperparameters can speed up the learning process by 40% without losing accuracy. The use of such strategies in DSS facilitates faster adaptation of models to new data while reducing overall computational costs. The relevance of this aspect is also confirmed by N.L. Rane *et al.* (2024), where optimised algorithms such as AdamW and Ranger can reduce training time by 15-25% without compromising accuracy.

A critical part of the system's effective operation is scalability and the ability to process large amounts of data in real-time. S. Kumar Rachakatla *et al.* (2023) emphasise that the introduction of AI-oriented approaches and cloud technologies can reduce data processing time by 35% and increase analysis efficiency by 20%. This is of particular importance for the medical DSS, as healthcare systems constantly generate large amounts of information that need to be analysed quickly. Cloud-based infrastructure provides flexibility and scalability, can be used to quickly adapt to the growth of data or the connection of new sources. A hybrid approach to the integration of structured and unstructured data is considered extremely promising. J.G. Turet & A.P.C.S. Costa (2022) proposed an effective combination of different sources of information for public safety systems. Similar methods can be adapted to DSS, including additional clinical databases or new types of signals. D. Baviskar *et al.* (2021) demonstrated the possibilities of automated document processing in the legal and medical fields, which can reduce data processing time by 30%. This highlights the versatility of the proposed approaches and confirms the potential for their scaling to other industries.

The expediency of improving approaches to processing poorly structured data was confirmed by the results of the research by S. Singh & S. Hooda (2023). The study proposes strategies for pre-processing and model adaptation that reduce classification errors by 20%. Optimisation of data analysis workflows can improve the performance of models in real-world conditions. The proposed approaches demonstrate efficiency in working with large amounts of information and ensure the stability of results when integrating new data sources.

The review paper by R. Cong *et al.* (2024) addressed the multidimensional selection of features based on an optimisation strategy for the causal analysis of medical data. The authors proposed an innovative approach to the selection of relevant characteristics that allow for the efficient identification of key variables for analysing relationships in complex medical systems. Particular attention was devoted to the use of optimisation methods to overcome the problem of high-dimensional data, which is typical for medical research. The authors also analysed the impact of various types of noise and missing data on the accuracy of models and proposed strategies to minimise them. In addition, the paper emphasises the importance of interpreting the results for decision-making in clinical practice. The inclusion of adaptive optimisation algorithms not only improved the quality of the analysis but also ensured the flexibility of the system for use in environments with limited computing resources. The study emphasises the potential of a multidimensional approach to improve diagnostic accuracy and treatment effectiveness based on causal models. S. Chen *et al.* (2020) propose cognitive computing to analyse customer information and improve the effectiveness of marketing strategies, which increases by 20%. While this study has a different focus, similar approaches can be applied to analyse patient feedback, optimise clinical services, and improve the quality of care. Lastly, tools that improve the understanding of context and intentions in data have universal value.

The adaptive control systems mentioned by B. Zohuri (2024), which use AI to adapt to changing environmental conditions, can be useful for healthcare facilities. Their application can provide dynamic resource optimisation, improve the stability of medical systems, and allow flexible responses to changing clinical protocols or the emergence of new diagnostic methods. S. Jain *et al.* (2021) investigated deep learning approaches for processing unstructured text data. The proposed transformer architecture for integrating different text formats increases the extraction of key data by 28%, which is especially useful for analysing complex clinical records where important details may be scattered across different parts of the document. This increase in the efficiency of information extraction contributes to faster diagnostic recommendations. Z. An *et al.* (2023) examined the intelligent design of complex products and the reduction of their creation time by 20%. Similar principles can be used to create and improve medical protocols, personalised treatment plans, or new algorithms for DSS. Optimising processes with AI make systems more flexible, adaptive, and focused on rapid innovation.

Acceleration of model training without loss of accuracy was addressed by K. Karthick (2024) in an analysis of optimisers and mechanisms for their adaptation. Reducing training time by 30% through the right choice of optimisers was identified as a key aspect for systems that manage medical data that are frequently updated and expanded. Ensuring that new models can be trained quickly, or existing models can be retrained with up-to-date data is critical to maintaining the accuracy and relevance of recommendations.

The potential of machine learning methods for optimisation tasks in communication networks was demonstrated by H. Dahrouj *et al*. (2021). Although the focus of this study is on network infrastructure, the proposed algorithms for improving bandwidth, reducing latency, and integrating different types of data are relevant to the medical field. For instance, accelerating the transfer and processing of information between hospital subsystems (laboratory results, CT scans, ECG data) will allow the DSS to respond more quickly to clinical changes. H. Lu & Y. Li (2020) examined cognitive computing and the processing of semi-structured information. According to the results, the use of deep neural networks and NLP techniques can reduce data analysis time by 20% and increase decision-making accuracy by 18%. Similar approaches can be integrated into a DSS for detailed analysis of mixed types of information, including data from medical devices, laboratory systems, and patient records. Lastly, D. Van De Berg *et al*. (2022) demonstrated data-driven optimisation techniques for managing engineering processes, increasing system performance by 25%. Although the field of engineering is different from the medical field, the general approaches to optimisation, streaming data analysis, and automated decision-making are universal. The ability to quickly adapt systems to new conditions, update models, and test different information processing options will allow DSS to support doctors in their daily work even more effectively.

Given the prospects for further development, it is possible to conclude that the integration of MTL will increase the accuracy and flexibility of the DSS, allowing the model to simultaneously solve several clinically important tasks. The use of optimised learning algorithms, adaptive optimisers, dynamic hyperparameter tuning mechanisms and cloud technologies will help reduce data processing time and improve the overall efficiency of the system. Expanding the range of data types, including non-standard medical formats, as well as the use of explanatory artificial intelligence (XAI) techniques will increase clinicians' confidence in the results of the DSS and facilitate the system's implementation in real medical practice. Thus, the use of transformers and related modern data processing methods creates the preconditions for the emergence of new generations of medical DSS capable of quickly and accurately analysing large amounts of diverse information, providing personalised recommendations for each patient, and effectively scaling to the needs of modern medicine. Comparison with the results of other studies confirms the high relevance and scientific novelty of the chosen approach.

CONCLUSIONS

A study on the development and evaluation of the effectiveness of a decision support system for medical diagnosis of cardiovascular diseases that integrates heterogeneous types of data using neural networks, in particular transformers, shows the significant potential of this approach for modern medicine. The proposed architecture of the DSS, which processes text records, images, and signals, demonstrated the possibility of effective integration of poorly structured data of different natures in a single analytical process. The Transformers significantly improved the system's ability to detect key patterns and relationships between data using an attention mechanism, which increased the diagnostic accuracy by 15% compared to traditional methods.

The achieved level of accuracy and stable results when using new data sources are the result of an optimal combination of different types of neural networks. The use of multilayer perceptrons, recurrent and convolutional neural networks in combination with transformers allowed each of the subsystems to focus on specific aspects of analysis: MLP – on structured indicators, RNN – on sequential text records, CNN – on visual data, and transformers – on the complex integration of these formats. Such a multimodal model better considers the contextual and individual characteristics of patients, which is important in the context of personalised medicine and the provision of quality healthcare services. An equally important achievement was the introduction of a hybrid computing approach that combined local and cloud resources. This made it possible to reduce model training time by 30% and process up to 1 TB of data per day. The use of optimisation techniques and adaptive optimisers ensured the speed of analysis, which is critical in situations where rapid clinical decision-making is required. Thus, computational efficiency and scalability have become key factors in implementing a system suitable for real-world clinical settings where the amount of data is constantly growing.

The results of the study confirmed that transformers not only improve the accuracy and speed of diagnostics but also contribute to increasing the system's resistance to noise and incomplete data. This helps to ensure the reliability and predictability of decisions made by the decision support system and reduces the risk of medical errors. In this regard, the introduction of such DSS can significantly improve the quality of diagnostic procedures, optimise resource allocation in healthcare facilities, and support doctors in their daily practice.

The findings correlate with current trends in artificial intelligence and machine learning, with special attention paid to multimodal data, multitasking, hyperparameter optimisation, scalability, and the use of cloud technologies. A promising area for further research is the introduction of explanatory artificial intelligence (XAI) approaches, which will increase the transparency of the decision-making process and the trust of doctors in the system's results. It is also possible to expand the list of data types (e.g., genetic or wearable device signals), which will contribute to even greater personalisation and accuracy of diagnosis. Thus, the developed DSS has all the prerequisites to become the basis for future generations of intelligent medical systems, increasing the effectiveness of treatment, reducing time and resources, and improving the quality of life of patients.

ACKNOWLEDGEMENTS

None.

CONFLICT OF INTEREST

None.

REFERENCES

[1] Abbas, N., & Gasmi, S. (2024). Optimizing machine learning techniques for big data analysis in natural language processing and text analytics. *ResearchGate*. doi: 10.13140/RG.2.2.15547.84002.

[2] An, Z., Bu, W., Wu, Z., & Li, D. (2023). Intelligent design of complex products based on extraction and reconstruction of key dimensions. In *2023 China automation congress (CAC)* (pp. 7966-7970). Chongqing: IEEE. doi: 10.1109/cac59555.2023.10450557.

[3] Baviskar, D., Ahirrao, S., Potdar, V., & Kotecha, K. (2021). Efficient automated processing of the unstructured documents using artificial intelligence: A systematic literature review and future directions. *IEEE Access*, 9, 72894-72936. doi: 10.1109/ACCESS.2021.3072900.

[4] Brown, T.B., *et al.* (2020). Language models are few-shot learners. *ArXiv*. doi: 10.48550/arXiv.2005.14165.

[5] Chen, S., Kang, J., Liu, S., & Sun, Y. (2020). Cognitive computing on unstructured data for customer co-innovation. *European Journal of Marketing*, 54(3), 570-593. doi: 10.1108/ejm-01-2019-0092.

[6] Cong, R., Deng, O., Nishimura, S., Ogihara, A., & Jin, Q. (2024). Multiple feature selection based on an optimization strategy for causal analysis of health data. *Health Information Science and Systems*, 12, article number 52. doi: 10.1007/s13755-024-00312-8.

[7] Dahrouj, H., *et al.* (2021). An overview of machine learning-based techniques for solving optimization problems in communications and signal processing. *IEEE Access*, 9, 74908-74938. doi: 10.1109/access.2021.3079639.

[8] European General Data Protection Regulation (GDPR). (2016, April). Retrieved from https://surl.li/wwepce.

[9] Fatima, H., & Gasmi, S. (2024). Optimization strategies in machine learning for improved big data analysis and natural language processing. *ResearchGate*. doi: 10.13140/RG.2.2.28130.75208.

[10] Gambella, C., Ghaddar, B., & Naoum-Sawaya, J. (2021). Optimization problems for machine learning: A survey. *European Journal of Operational Research*, 290(3), 807-828. doi: 10.1016/j.ejor.2020.08.045.

[11] Jain, S., Jain, S., & Jain, A.K. (2021). Deep learning approach towards unstructured text data utilization: Development, opportunities, and challenges. In A. Sharaff, G. Sinha & S. Bhatia (Eds.), *New opportunities for sentiment analysis and information processing* (pp. 29-49). New York: IGI Global Scientific Publishing. doi: 10.4018/978-1-7998-8061-5.ch002.

[12] Karthick, K. (2024). Comprehensive overview of optimization techniques in machine learning training. *Control Systems and Optimization Letters*, 2(1), 23-27. doi: 10.59247/csol.v2i1.69.

[13] Kumar Rachakatla, S., Ravichandran, P., & Reddy Machireddy, J. (2023). Advanced data science techniques for optimizing machine learning models in cloud-based data warehousing systems. *Australian Journal of Machine Learning Research & Applications*, 3(1), 396-419.

[14] Kushnir, D., Ocherklevich, O., & Paramud, Ya. (2021). Deep neural network model for text semantic analysis based on word embeddings. In *Proceedings of the 11th international conference on advanced computer information technologies (ACIT)* (pp. 718-721). Deggendorf: IEEE. doi: 10.1109/acit52158.2021.9548393.

[15] Law of Ukraine No. 2297-VI "On the Protection of Personal Data". (2010, June). Retrieved from https://www.president.gov.ua/documents/2297vi-11567.

[16] Lu, H., & Li, Y. (2020). Cognitive computing for intelligence systems. *Mobile Networks and Applications*, 25(4), 1434-1435. doi: 10.1007/s11036-019-01428-y.

[17] Mahadevkar, S.V., Patil, S., Kotecha, K., Soong, L.W., & Choudhury, T. (2024). Exploring AI-driven approaches for unstructured document analysis and future horizons. *Journal of Big Data*, 11, article number 92. doi: 10.1186/s40537-024-00948-z.

[18] Malyha, I., & Shmatkov, S. (2022). Machine learning methods for solving semantics and context problems in processing textual data. *Bulletin of V.N. Karazin Kharkiv National University, Series "Mathematical Modeling. Information Technology. Automated Control Systems"*, 56, 35-42. doi: 10.26565/2304-6201-2022-56-03.

[19] Meng, J., & Wang, Z. (2022). Intelligent algorithms of English semantic analysis based on deep learning technology. In *2022 IEEE Asia-Pacific conference on image processing, electronics and computers (IPEC)* (pp. 1530-1533). Dalian: IEEE. doi: 10.1109/IPEC54454.2022.9777363.
[20] Nedosnovanyi, O., Cherniak, O., & Golinko, V. (2023). Comparative analysis of cloud services for geoinformation data processing. *Information Technologies and Computer Engineering*, 20(2), 50-57. doi: 10.31649/1999-9941-2023-57-2-50-57.
[21] Oza, R.R., & Domadiya, D.H. (2023). Analysis of unstructured data using artificial intelligence. *International Journal of Creative Research Thoughts*, 11(5), 969-973.
[22] Rane, N.L., Mallick, S.K., Kaya, Ö., & Rane, J. (2024). Techniques and optimization algorithms in deep learning: A review. In *Applied machine learning and deep learning: Architectures and techniques* (pp. 59-79). Yakutiye: Deep Science Publishing. doi: 10.70593/978-81-981271-4-3_3.
[23] Rueda, R., Fabello, E., Silva, T., Genzor, S., Mizera, J., & Stanke, L. (2024). Machine learning approach to flare-up detection and clustering in chronic obstructive pulmonary disease (COPD) patients. *Health Information Science and Systems*, 12, article number 50. doi: 10.1007/s13755-024-00308-4.
[24] Shen, Z. (2023). Algorithm optimization and performance improvement of data visualization analysis platform based on artificial intelligence. *Frontiers in Computing and Intelligent Systems*, 5(3), 14-17. doi: 10.54097/fcis.v5i3.13836.
[25] Singh, S., & Hooda, S. (2023). A study of challenges and limitations to applying machine learning to highly unstructured data. In *2023 7th international conference on computing, communication, control and automation (ICCUBEA)* (pp. 1-6). Pune: IEEE. doi: 10.1109/ICCUBEA58933.2023.10392115.
[26] Smetaniuk, O., & Tsisar, D. (2024). Digital economy as a foundation for creating digital strategies of enterprises. *Innovation and Sustainability*, 4(3), 68-75. doi: 10.31649/ins.2024.3.68.75.
[27] Turet, J.G., & Costa, A.P.C.S. (2022). Hybrid methodology for analysis of structured and unstructured data to support decision-making in public security. *Data & Knowledge Engineering*, 141, article number 102056. doi: 10.1016/j.datak.2022.102056.
[28] Van De Berg, D., Savage, T., Petsagkourakis, P., Zhang, D., Shah, N., & del Rio-Chanona, E.A. (2022). Data-driven optimization for process systems engineering applications. *Chemical Engineering Science*, 248(B), article number 117135. doi: 10.1016/j.ces.2021.117135.
[29] Wilson, A., & Anwar, M.R. (2024). The future of adaptive machine learning algorithms in high-dimensional data processing. *International Transactions on Artificial Intelligence*, 3(1), 97-107. doi: 10.33050/italic.v3i1.656.
[30] World Medical Association's Declaration of Helsinki. (1964, June). Retrieved from https://www.wma.net/policies-post/wma-declaration-of-helsinki/.
[31] Zhang, G., Fu, C., & Zhou, H. (2024). Research on key technologies of deep learning techniques in unstructured data processing. *Applied Mathematics and Nonlinear Sciences*, 9(1). doi: 10.2478/amns-2024-3175.
[32] Zhang, Y., & Yang, Q. (2022). A survey on multi-task learning. *IEEE Transactions on Knowledge and Data Engineering*, 34(12), 5586-5609. doi: 10.1109/TKDE.2021.3070203.
[33] Zohuri, B. (2024). Artificial intelligence and machine learning driven adaptive control applications. *Journal of Material Sciences and Engineering Technology*, 2(4). doi: 10.61440/JMSET.2024.v2.27.

Оптимізація алгоритмів інтелектуальних систем для аналізу слабоструктурованих даних

Микола Демчина

Кандидат технічних наук, доцент
Університет Короля Данила
76018, вул. Є. Коновальця, 35, м. Івано-Франківськ, Україна
https://orcid.org/0009-0002-9161-4843

Тарас Стисло

Старший викладач
Університет Короля Данила
76018, вул. Є. Коновальця, 35, м. Івано-Франківськ, Україна
https://orcid.org/0000-0002-2377-7985

Сергій Ващишак

Кандидат технічних наук, доцент
Університет Короля Данила
76018, вул. Є. Коновальця, 35, м. Івано-Франківськ, Україна
https://orcid.org/0000-0002-1753-1540

Анотація. Інтеграція різнорідних типів медичних даних із використанням сучасних методів глибокого навчання дозволяє підвищити точність та ефективність діагностики складних захворювань, таких як серцево-судинні, що має вирішальне значення для персоналізованої медицини та зниження ризику медичних помилок. Метою роботи було представити розробку системи підтримки прийняття рішень, спрямованої на покращення діагностики серцево-судинних захворювань шляхом інтеграції різнорідних типів медичних даних. Для створення бази знань були використані дані реальних клінічних сценаріїв, що пройшли етапи очистки, стандартизації та семантичного аналізу за допомогою спеціалізованих медичних словників. Система продемонструвала високу ефективність завдяки здатності інтегрувати текстові, зображувальні та сигнальні дані в єдиний процес аналізу. Ефективність оцінювалася за такими метриками, як точність, повнота, F1-score, а також прогностичні значення позитивних і негативних результатів. Впровадження трансформерів забезпечило підвищення точності діагностики на 15 % порівняно з традиційними методами, а використання гібридного підходу до обчислень дозволило скоротити час навчання моделей на 30 % і обробляти до 1 ТБ даних на добу. Додатково, інтеграція різнорідних типів медичних даних у системі дала змогу покращити персоналізацію діагностики, враховуючи індивідуальні особливості пацієнтів, такі як анамнез, генетичні фактори чи супутні захворювання. Завдяки механізмам уваги трансформерів система демонструє стійкість до шуму та пропусків у даних, що забезпечує надійність результатів навіть за умови неповної або неточної інформації. Оптимізація моделей сприяла зменшенню затримок в обробці даних, що є критично важливим для оперативного прийняття клінічних рішень. Крім того, трансформери довели свою здатність динамічно масштабуватися для обробки нових типів даних без втрати ефективності, відкриваючи можливості для подальшого розширення функціоналу системи. Система також підвищила продуктивність клінічних фахівців завдяки автоматизації рутинних завдань, що дозволило лікарям зосередитися на складніших аспектах лікування

Ключові слова: штучний інтелект; база знань; підтримка прийняття рішень; нейромережі; видобування знань; алгоритми класифікації; адаптивні системи