

UDC 004.891.3

## **PREDICTIVE ANALYTICS FOR EARLY DISEASE DETECTION: HOW MACHINE LEARNING MODELS CAN HELP IN PREDICTING DISEASES**

*Madikhan M., 1st year Master's student  
Shaushenova A.G., Candidate of Technical Sciences, Associate Professor  
Nurpeisova A.A., PhD, Acting Associate Professor  
S. Seifullin Kazakh Agrotechnical Research University, Astana*

Early disease identification made possible by machine learning-powered predictive analytics is transforming the healthcare industry. Early risk identification has a major positive impact on treatment results and lowers medical expenses. Through the analysis of large and intricate healthcare datasets, machine learning models, including decision trees, random forests, and neural networks, provide effective methods for forecasting the beginning of diseases. Esteva et al. made significant progress in the classification of skin cancer, demonstrating the potential of deep learning. Their model, a convolutional neural network (CNN), demonstrated the effectiveness of AI in early detection by achieving dermatologist-level accuracy in recognising malignant tumours. Predictive models are being used in other fields, such as diabetes and cardiovascular illnesses, by utilizing imaging data and electronic health records (EHRs) to detect at-risk individuals prior to the development of diseases mentioned above. These developments are not just limited to cancer. However, despite the potential of these technologies, there are still a number of fields and questions we as researchers will have study further in the future, including issues with improper data quality, model biases, and data privacy violations in ethics. But there is hope for the potential of machine learning in healthcare since developments in predictive analytics will likely change how we approach patient care and illness prevention. This paper will explore how machine learning models are transforming predictive analytics, offering new possibilities for early disease detection and personalized healthcare [1].

The revolutionary significance of machine learning in early disease diagnosis is becoming more and more present in recent scientific papers on predictive analytics in healthcare. Deployment of deep learning models in medical image analysis has been studied, particularly in the classification of skin cancer, thanks to the work of Esteva et al. Their convolutional neural network (CNN) model demonstrated the potential of AI in giving accurate diagnoses by using filtered image pattern recognition layers within their CNNs. This work establishes a solid foundation for future AI applications in healthcare by highlighting the efficiency of neural networks in applications of vast medical data [1].

The paper written by Obulesu et al. makes another important addition by summarizing various machine learning applications in healthcare. Different

supervised learning algorithms, such as Random Forest and Support Vector Machines (SVM), are studied in the framework of predicting chronic diseases, including diabetes and cardiovascular problems, is covered in this work. These models use patient information, such as electronic health records (EHRs) and medical histories, to alert about possible health risks before actual symptoms appear [2].

Additionally, the survey on healthcare predictive analytics written by Badawy et al. looks at deep learning and machine learning methods, emphasising how well they can forecast diseases with the use of big datasets. The poll addresses important issues that can impact model accuracy, such as bias and poor data quality. Additionally, it examines moral issues pertaining to patient data privacy, emphasising the necessity of openness in AI-driven healthcare systems. Combined, these sources offer a thorough understanding of the developments and difficulties in using machine learning to disease prediction, highlighting the technology's potential to completely transform the healthcare industry [3].

Machine learning methods are currently at the forefront of illness prediction, providing tools to evaluate medical data and make very accurate predictions about health risks. In the field of medicine, supervised learning algorithms like Support Vector Machines (SVM) and Random Forests are frequently employed for tasks like chronic illness prediction. These models are ideal for early disease detection since they are trained on labelled data to find patterns and correlations between variables. This method is generally considered clearer and easier to use than other methods like unsupervised learning. In order to forecast the beginning of illnesses before symptoms appear, Random Forests, for instance, have been used to anticipate diseases like diabetes and heart disease. They achieve this by analysing sizable datasets from electronic health records (EHRs) [2].

In addition, deep learning models, in particular Convolutional Neural Networks (CNNs), are being employed more and more for disease prediction through image recognition. CNNs use more specialized neural network layers than regular machine learning techniques. CNNs have the ability to identify skin cancer with an accuracy level that is on par with experienced doctors, as demonstrated by Esteva et al. In order to determine the likelihood of having skin cancer, the model examined pictures using classification. Datasets consisting of images of healthy skin and skin having cancer were used to train CNN models. This showed how deep learning may transform medical diagnostics by producing quick and precise predictions from imaging data. In disease prediction, unsupervised learning methods such as clustering are very useful, especially when attempting to find risk factor-sharing subgroups in patient data. These models provide fresh perspectives on the course of disease by assisting in the identification of hidden patterns in unlabelled data. When combined, these machine learning approaches are increasing the precision, speed, and breadth of disease identification in the medical field, which is propelling predictive analytics forward [1][3].

Specialists have been using Machine Learning techniques in identification of various

diseases. Some of the most common ones will be studied in this paper. The identification of cancer, especially skin cancer classification, is one of the most important uses. Convolutional Neural Networks (CNNs) models' high degree of accuracy in image analysis creates new opportunities for skin cancer screening and diagnosis, guaranteeing early intervention and maybe enhancing patient outcomes. Other researchers have been using specific deep learning methods, like transfer learning, to achieve impressive results. Shames and Kamil used transfer learning to analyse CT scans in lung infection diagnosis, specifically COVID-19. This allowed them to use deep learning techniques with more specialized structure. Using pre-trained models to enhance performance with a smaller dataset, they trained a convolutional neural network (CNN) using a dataset of 2,482 CT scans. The model proved to be highly successful, correctly detecting patients with COVID-19 lung infections in 91.67% of cases [1][4].

Deep learning models have also been used as deployment for devices with medical use. In order to forecast cardiac illness, S. Sabarunisha Begum and colleagues used the ThingSpeak platform to integrate deep learning with IoT-based data. IoT sensors collect health data from patients, which is instantly processed and examined by a cloud-hosted deep learning model. By continuously tracking health parameters, this integration makes it possible to make accurate forecasts about heart disease and assists clinicians in identifying any problems earlier and more effectively. This study shows that neural networks can be go beyond usual software implications and can be deployed on devices with a wider range of uses [5].

Machine learning has also demonstrated potential in the early identification of kidney disease. Decision trees and random forests are used in the study "Early Prediction of Healthcare Diseases Using Machine Learning and Deep Learning Techniques" to predict kidney-related illnesses. By evaluating factors like age, blood pressure, and lifestyle choices, these models allow for an early diagnosis even in the absence of symptoms. Healthcare practitioners can intervene with preventative strategies, improving patient outcomes and slowing the progression of disease, via early identification of patients who are at-risk. This strategy improves overall healthcare precision by giving alternative applications in the prediction of cardiovascular diseases [2].

Advances in the prediction of cardiovascular disease (CVD) are also being made by machine learning algorithms. In order to identify those at risk for heart problems, predictive models, such as Random Forest and Support Vector Machines (SVM), have been used to evaluate patient data, including medical histories and lifestyle characteristics. Research has demonstrated that these models are more accurate than conventional techniques at predicting the risk of CVD, particularly when they are used to find patterns and correlations in patient data from Electronic Health Records (EHRs). In addition, machine learning is essential for controlling and alerting possible the course of diabetes. Models like logistic regression and decision trees have been used to analyse data points like blood glucose levels, age, and lifestyle characteristics in order to predict which people

are most likely to develop diabetes. Early diagnosis enables prompt lifestyle modifications that can either postpone or stop the disease's start. By examining brain imaging data and cognitive test results, machine learning has been used in the field of neurological illnesses to anticipate the beginning of diseases like Alzheimer's. Early brain alterations can be identified by deep learning models, which may enable earlier diagnosis and treatment plans [2]. Summing up, the use of machine learning to a variety of diseases, including neurological disorders, diabetes, cancer, and cardiovascular ailments, is radically altering patient care and disease management by providing better accuracy and earlier detection.

Machine learning is enabling earlier and more accurate disease identification, greatly assisting the development of predictive analytics in the healthcare industry. Models have a promising potential to change how diseases will be treated in the near future. Ongoing research, such as works of Esteva et al., who pioneered the use of Convolutional Neural Networks (CNNs) to classify skin cancer, and works of other researchers show that lives and health issues of many patient will be positively affected by using AI models. Increasing number of research in the fields of Random Forests and Support Vector Machines (SVM) in diabetes and cardiovascular disease classification is also encouraging. Current capabilities of AI engineers have proven that they can cooperate closely with medical professionals. Large volumes of medical data, such as electronic health records (EHRs) and medical imaging, can be processed by these techniques, which allow them to find patterns that other methods would not be able to process [1]. Even though the developments are encouraging, more study is required to enhance the models' scalability and dependability across a range of demographics. Enhancing the fairness and accuracy of prediction models will require ongoing innovation in deep learning and unsupervised learning approaches, which will ultimately result in improved patient outcomes and more individualised treatment. In addition, further research in ethical considerations of using AI in sensitive patient data might yield valuable results

Данная научная статья подготовлена в рамках проекта № AP23486538

«Исследование и разработка системы распознавания изображений в видеопотоках на основе искусственного интеллекта» по программе 217 «Развитие науки», подпрограмме 102 «Грантовое финансирование научных исследований».

## References

1. Esteva, A., Kuprel, B., Novoa, R. *et al.* (2017). Dermatologist-level classification of skin cancer with deep neural networks. *Nature*, 542, 115–118.
2. Obulesu, O., Venkateswarulu, N., Sri Vidya, M., Manasa, S., Pranavi, K., Brahmani, C. (2023). Early Prediction of Healthcare Diseases Using Machine Learning and Deep Learning Techniques. In: Seetha, M., Peddoju, S.K., Pendyala, V., Chakravarthy, V.V.S.S.S. (eds) *Intelligent Computing and Communication. ICICC 2022. Advances in Intelligent Systems and Computing*, 1447.

3. Badawy, M., Ramadan, N., Hefny, HA. (2023). Healthcare predictive analytics using machine learning and deep learning techniques: a survey. *Journal of Electrical Systems and Information Technology*, 10(1), 40.
4. Shames, M. A., Kamil, M. (2024). Lung Infection Detection via CT Images and Transfer Learning Techniques in Deep Learning. *Journal of Advanced Research in Applied Sciences and Engineering Technology*, 47(1), 206-218.
5. Begum, SS, Karthikeyan, B., Alanazi, FZ. (2024). A Prediction of Heart Disease using IoT based ThingSpeak Basis and Deep Learning Method. *Journal of Advanced Research in Applied Sciences and Engineering Technology*, 47(1), 166-179.



