International Scientific-Electronic Journal "Pioneering Studies and Theories" № 5
ISSN: 3060-5105 Volume 1
www.pstjournal.uz
April, May, June 2025

DOI: 10.5281/zenodo.15374475
Link: https://zenodo.org/records/15374475

A MACHINE-LEARNING APPROACH TO DETECT HEART DISEASE

Turaeva Makhliyo Shokir kizi[1]
[1] assistant teacher
[1,] "Tashkent University of Information Technologies", Tashkent, Uzbekistan.
max1209uz@gmail.com
Phone: + 998507740047

Abstract: Moreover, heart disease has kept on being the leading cause of death worldwide, hence the importance of coming up with different strategies for early detection and diagnosis. Machine learning algorithms have been found to be very efficient in diagnosing various cardiac diseases, through approaches such as Support Vector Machines (SVM), Random Forest, Neural Networks, and Logistic Regression. The materials of this research consist of a dataset that has various health markers, such as age, blood pressure, cholesterol level, and other clinical factors that are relevant to this issue. The Random Forest classifier was the only one with an incredible accuracy of 97.5% to outsmart the rest of the algorithms used. It also managed to record a 0.998 Area Under the Curve (AUC) score and F1, Precision, and Recall scores of 97.5%.

Keywords: Diagnosis, internal analysis, Dentistry, effective treatment, treatment plan, Dental diseases.

INTRODUCTION

Finding new approaches to early detection and diagnosis can be considered urgent, while cardiovascular disease remains one of the leading reasons for mortality internationally. A new revelation was just established, which applies the significance of machine learning (ML) as an innovative form of technology that can bring the most significant impact in the healthcare business, particularly the prediction and diagnosis of cardiovascular diseases. As a result, compared to top-down and traditional methods, machine learning can assess complex patient records characteristics based on large datasets and algorithms. This results in enhanced diagnostic accuracy and efficiency. Several machine learning algorithms are useful in predicting cardiac disease, with high accuracy rates, according to recent research reports. Certain examples of classifiers that have demonstrated promising outcomes are Support Vector Machines (SVM), Random Forests (RF), and Neural Networks. In certain instances, the accuracy rates of these classifiers have reached as high as 98.7% [1].

An individual's risk profile for cardiovascular disease may be evaluated completely thanks to the capability of machine learning to include many clinical parameters, such as age, blood pressure, cholesterol levels, and electrocardiogram (ECG) readings [2], [3]. A more comprehensive knowledge of a patient's health status is provided by this multifactorial approach, which also makes it possible to diagnose probable cardiac issues earlier and with more precision for the patient. Furthermore, machine learning helps in the diagnostic process and improves individualized therapy by identifying specific risk factors connected with individual patients with the disease. With prompt intervention, it is possible to considerably minimize the morbidity and mortality associated with cardiac diseases [4]. This skill is very beneficial in the healthcare industry. This allows healthcare practitioners to adjust therapies to each patient's particular needs, improving results and optimising resource allocation.

Machine learning algorithms guide personalized treatment techniques. For instance, machine learning may assist in identifying patients at a greater risk for particular forms of cardiac disease. This results in the implementation of focused preventative measures and the treatment of current problems in a more efficient manner. The scope of machine learning's potential to revolutionize cardiovascular disease diagnosis is becoming increasingly obvious as the amount of data about healthcare continues to develop. With the increasing access to electronic health records, wearable devices and other sources of data, it can be possible to build machine learning models that would provide higher levels of accuracy in diagnosing cardiac disease than is presently achievable. The change towards data-driven healthcare solutions represents a more effective, precise, and open path to care. This will, in turn, enhance the quality of patients and lower the prevalence of heart disease in health facilities across the globe [5]. Machine learning is set to become a significant component in cardiovascular healthcare in the future, promising the improved efficacy of avoiding and treating heart ailments. This is so since this area has ongoing research and development subdomains.

RELATED WORKS

Traditional machine learning methods may perform poorly when dealing with unbalanced datasets in heart disease prediction, as emphasized by Sharma et al.'s systematic literature review. This study highlights the need for more resilient algorithms to efficiently manage such data instabilities and synthesized results from 451 investigations. Veisi [6] et al. made another vital addition by using various models, such as Multilayer Perceptron (MLP) and XGBoost. The MLP model was remarkable in its accuracy, as seen by its 94.6% accuracy. Their work underlined the need for preprocessing measures, including outlier detection and normalization, to improve the predictive power of ML systems. Ramesh et al. conducted a thorough investigation and highlighted the revolutionary potential of machine learning (ML) in the healthcare industry, namely in heart disease identification. They looked at various machine learning approaches, such as ensemble methods and neural networks, that have been successfully used to analyze massive datasets and find intricate patterns linked to heart disease. The need for more research to increase these models' generalizability to a broader range of patient groups was highlighted by this study [7].

In addition, research by Ghosh et al. used feature selection methods like Relief and LASSO in conjunction with ML algorithms like Random Forest and Support Vector Machine (SVM) to improve prediction accuracy. According to their findings, the hybrid model considerably outperformed conventional techniques and had high prediction accuracy rates for cardiovascular events. This work demonstrates how crucial it is to combine sophisticated machine learning methods with careful feature selection to maximize predicted results [8]. Another noteworthy addition was a work that used eight different machine learning classifiers, such as Naïve Bayes and Learning Vector Quantization, to predict cardiac disease using the UCI repository dataset. Using the Learning Vector Quantization approach, the researchers obtained a fantastic accuracy of 98.7%, demonstrating the efficacy of several algorithms in finding essential variables that contribute to the prediction of heart disease [9]. Moreover, current research has investigated hybrid models that combine several machine-learning approaches.

Research has indicated that integrating models such as Gradient Boosting, Convolutional Neural Networks, and K-nearest neighbours might enhance the precision and dependability of heart disease prognosis. This method makes it possible to improve overall performance by utilizing the advantages of several different algorithms [10]. To sum up, research suggests that machine learning techniques have a great deal of potential to enhance heart disease prognosis. To improve the accuracy and usability of machine learning approaches in cardiovascular health, more innovation in model construction, data management, and integration of varied datasets is required, according to ongoing research in this field. These technologies will probably be crucial in revolutionizing the diagnosis and treatment of cardiac disease as they develop further. The usefulness of machine learning in image-based cardiac diagnostics has been demonstrated in recent research. Deep learning techniques have

shown better performance in analyzing computed tomography scans, cardiac magnetic resonance imaging, and echocardiograms. For example, a thorough analysis revealed that by automating the interpretation of imaging data, machine learning (ML) algorithms might greatly increase the diagnosis accuracy for diseases like heart failure (HF) and coronary artery disease (CAD) (Książek et al., 2019) [11].

By producing more accurate evaluations, machine learning (ML) integration in imaging expedites the diagnosis process and lessens the workload for medical practitioners. One of the most economical ways to diagnose heart disease is still using an ECG. However, reading ECG signals can be difficult since cardiac diseases vary and are so complicated. Interpretable machine learning (IML) approaches have recently advanced to provide models that not only categorize cardiac ailments but also provide insights into their decisionmaking processes in an attempt to solve these issues. Gaining the confidence of healthcare professionals and guaranteeing the practical application of ML models depend on this openness (Kresoja et al., 2023) [12]. Machine learning can also be used in agriculture and other domains [13] [14].

RESEARCH METHODOLOGY

A. Data Collection Collect relevant datasets, such as patient medical records and lab results.

B. Data Preprocessing Handle missing values, eliminate duplicates, and fix mistakes in your data. Finding and choosing pertinent features is the first step in feature selection (e.g., age, cholesterol levels, blood pressure). Scale the information so that each feature adds the same amount to the model.

C. Exploratory Data Analysis (EDA) Plots and charts can be used to visualize data linkages and distributions. Run statistical analyses to find essential features.

D. Model Selection Choose suitable machine learning algorithms, such as neural networks, random forests, and logistic regression.

E. Model Training Utilizing the training dataset, train the chosen models. Use methods such as Grid Search and Random Search to optimize model parameters.

F. Model Evaluation Metrics like F1-score, AUC-ROC, recall, accuracy, and precision are used to quantify models. Check the model's generalization by running it on a different validation dataset.

G. Model Deployment Introduce the trained model into an operational setting. Provide APIs so that other systems can communicate with the model.

Dataset available at https://www.kaggle.com/datasets/johnsmith88/heart-diseasedataset/data. Which consists of 1025 patient records with 13 attributes including age, chest pain type, blood pressure, cholesterol, and blood sugar.

RESULTS AND DISCUSSION

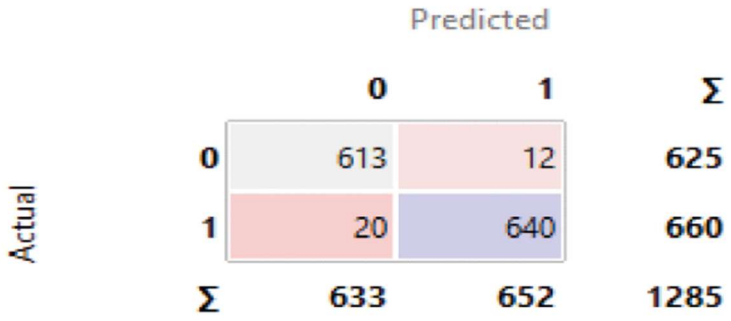

Figure 1. (a) Random Forest confusion matrix

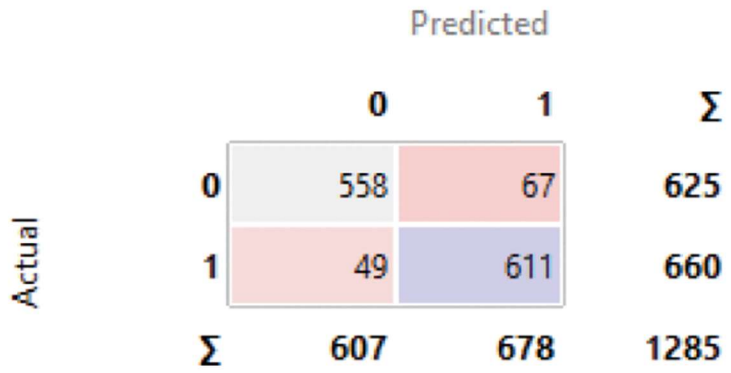

Figure 2. (b) SVM confusion matrix

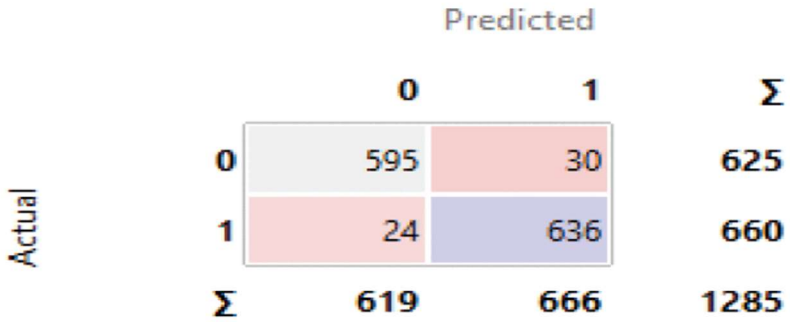

Figure 3. (c) Neural Network confusion matrix

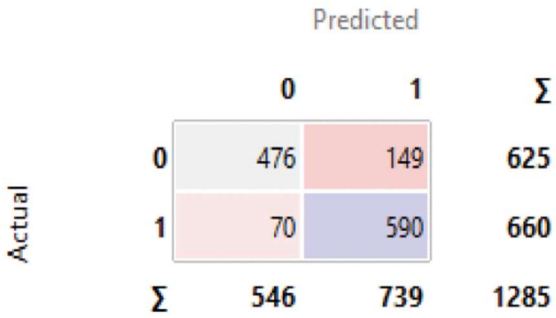

Figure 4. Logistic Regression confusion matrix

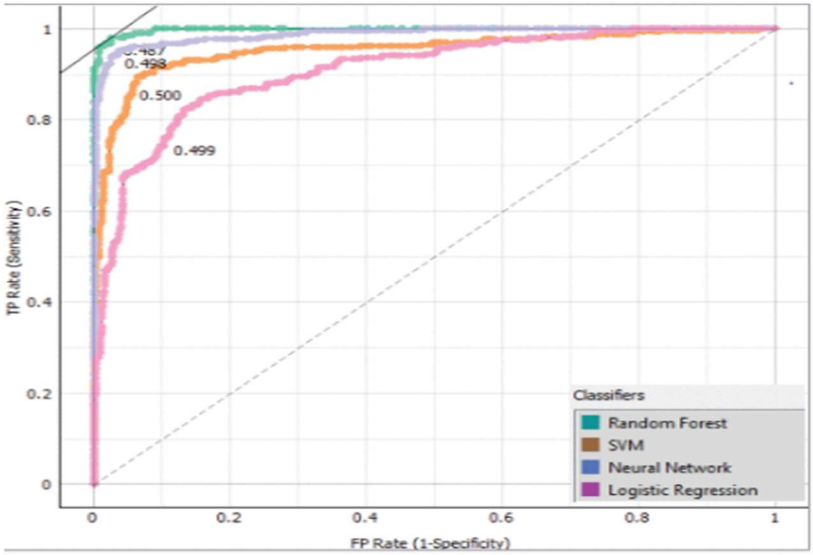

Figure 5. (e) ROC Curve for target class 0 no heart disease

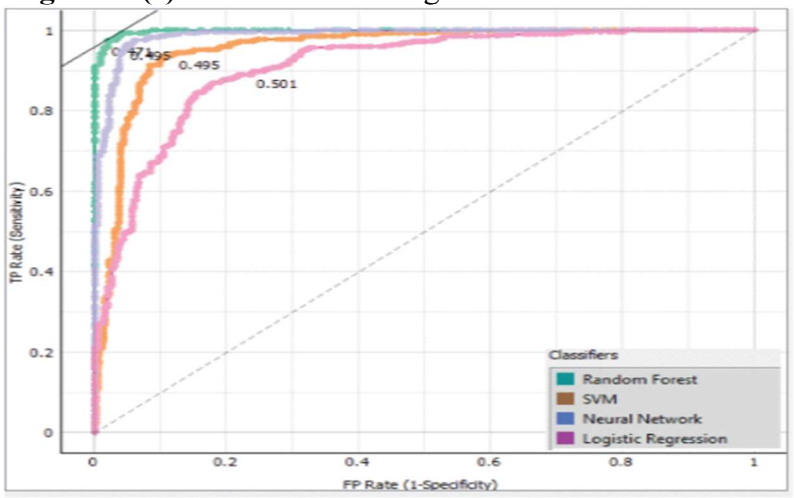

Figure 6. (f) ROC Curve for target class 1 heart disease.

Table 1. Results of Various Used Classifiers

Model	AUC	CA	F1	Precision	Recall
SVM	0.954	0.910	0.910	0.910	0.910
Random Forest	0.998	0.975	0.975	0.975	0.975
Neural Network	0.988	0.958	0.958	0.958	0.958
Logistic Regression	0.909	0.830	0.829	0.834	0.830

Based on the following critical performance measures, the table compares four machine learning models: Random Forest, Neural Network, SVM (Support Vector Machine), and Logistic Regression. These metrics include AUC (Area Under the ROC Curve), F1 Score, Precision, and Recall. The Random Forest model performs better than the other models in every metric. With an AUC of 0.998, it attains the greatest level and shows good discrimination between positive and negative classes. Additionally, this model has a Classification Accuracy of 97.5%, which indicates that 97.5% of occurrences are accurately predicted. Furthermore, the Random Forest model exhibits

a balanced performance in identifying true positives and avoiding false positives, as seen by its F1 Score, Precision, and Recall of 0.975.

Conclusion. This study shows how machine learning algorithms have a great deal of potential for early diagnosis and detection of heart disease, one of the world's leading causes of death. Out of all the tested algorithms, the Random Forest classifier performed exceptionally well, attaining an astounding 97.5% accuracy and an almost perfect AUC score of 0.998. These findings highlight how machine learning models can improve diagnostic accuracy and assist healthcare providers in making better decisions about patient care. This study adds to the expanding corpus of information on using sophisticated computational approaches in healthcare settings by utilizing machine learning. According to the research, using machine learning in standard clinical practice may help identify cardiac problems earlier, eventually improving patient outcomes and saving medical expenses.

REFERENCES

1. H. Ge, Y. Shi, M. Zhang, Y. Wei, H. Zhang and X. Cao, "YOLO: An Improved High-Accuracy Method for PCB Defect Detection," 2024 IEEE 12th International Conference on Computer Science and Network Technology (ICCSNT), Dalian, China, 2024, pp. 159-165, doi: 10.1109/ICCSNT62291.2024.10776686.
2. Mekhriddin Rakhimov, Dilnoza Zaripova, Shakhzod Javliev, Jakhongir Karimberdiyev; Deep learning parallel approach using CUDA technology. AIP Conf. Proc. 27 November 2024; 3244 (1): 030003. https://doi.org/10.1063/5.0241439.
3. M. Rakhimov, R. Akhmadjonov and S. Javliev, "Artificial Intelligence in Medicine for Chronic Disease Classification Using Machine Learning," 2022 IEEE 16th International Conference on Application of Information and Communication Technologies (AICT), Washington DC, DC, USA, 2022, pp. 1-6, doi: 10.1109/AICT55583.2022.10013587.
4. Rakhimov, M., Karimberdiyev, J., Javliev, S. (2024). Artificial Intelligence in Medicine: Enhancing Pneumonia Detection Using Wavelet Transform. In: Choi, B.J., Singh, D., Tiwary, U.S., Chung, WY. (eds) Intelligent Human Computer Interaction. IHCI 2023. Lecture Notes in Computer Science, vol 14531. Springer, Cham. https://doi.org/10.1007/978-3-031-53827-8_16
5. Goran Oreski. 2023. YOLO*C — Adding context improves YOLO performance. Neurocomput. 555, C (Oct 2023). https://doi.org/10.1016/j.neucom.2023.126655
6. M. Rakhimov, J. Elov, U. Khamdamov, S. Aminov and S. Javliev, "Parallel Implementation of Real-Time Object Detection using OpenMP," 2021 International Conference on Information Science and Communications Technologies (ICISCT), Tashkent, Uzbekistan, 2021, pp. 1-4, doi: 10.1109/ICISCT52966.2021.9670146.
7. Nasimov, R., Rakhimov, M., Javliev, S., Abdullaeva, M. (2024). Parallel Approaches to Accelerate Deep Learning Processes Using Heterogeneous Computing. In: Koucheryavy, Y., Aziz, A. (eds) Internet of Things, Smart Spaces, and Next Generation Networks and Systems. NEW2AN ruSMART 2023 2023. Lecture Notes in Computer Science, vol 14543. Springer, Cham. https://doi.org/10.1007/978-3-031-60997-8_4.
8. Mekhriddin Rakhimov, Shakhzod Javliev, and Rashid Nasimov. 2024. Parallel Approaches in Deep Learning: Use Parallel Computing. In Proceedings of the 7th International Conference on Future Networks and Distributed Systems (ICFNDS '23). Association for Computing Machinery, New York, NY, USA, 192–201. https://doi.org/10.1145/3644713.3644738
9. A. Thulaseedharan and L. P. P. S, "Deep Learning based Object Detection Algorithm for the Detection of Dental Diseases and Differential Treatments," 2022 IEEE 19th India Council International Conference (INDICON), Kochi, India, 2022, pp. 1-7, doi: 10.1109/INDICON56171.2022.10040109.
10. Wang, M.; Yang, B.; Wang, X.; Yang, C.; Xu, J.; Mu, B.; Xiong, K.; Li, Y. YOLO-T: Multitarget Intelligent Recognition Method for X-ray Images Based on the YOLO and Transformer Models. Appl. Sci. 2022, 12, 11848. https://doi.org/10.3390/app122211848.

11. Otabek Ismailov, Xosiyat Temirova; Tooth square detection using artificial intelligence. AIP Conf. Proc. 27 November 2024; 3244 (1): 030030. https://doi.org/10.1063/5.0242591
12. Terven, J.; Córdova-Esparza, D.-M.; Romero-González, J.-A. A Comprehensive Review of YOLO Architectures in Computer Vision: From YOLOv1 to YOLOv8 and YOLO-NAS. Mach. Learn. Knowl. Extr. 2023, 5, 1680-1716. https://doi.org/10.3390/make5040083.
13. Davron Ziyadullaev, Dildora Muhamediyeva, Sholpan Ziyaeva, Umirzoq Xoliyorov, Khasanturdi Kayumov, Otabek Ismailov. “Development of a traditional transport system based on the bee colony algorithm”. E3S Web of Conf. 365 01017 (2023). DOI: 10.1051/e3sconf/202336501017.
14. R. Mohan, R. Arunmozhi and V. Rajinikanth, "Deep-Learning Segmentation and Recognition of Tooth in Thresholded Panoramic X-ray," *2023 Winter Summit on Smart Computing and Networks (WiSSCoN)*, Chennai, India, 2023, pp. 1-5, doi: 10.1109/WiSSCoN56857.2023.10133861.
15. Suryani, D & Shoumi, M & Wakhidah, Rokhimatul. (2021). Object detection on dental x-ray images using deep learning method. IOP Conference Series: Materials Science and Engineering. 1073. 012058. 10.1088/1757-899X/1073/1/012058.
16. Abdusalomov, A.B.; Nasimov, R.; Nasimova, N.; Muminov, B.; Whangbo, T.K. Evaluating Synthetic Medical Images Using Artificial Intelligence with the GAN Algorithm. Sensors 2023, 23, 3440. https://doi.org/10.3390/s23073440
J. George, T. S. Hemanth, J. Raju, J. G. Mattapallil and N. Naveen, "Dental Radiography Analysis and Diagnosis using YOLOv8," 2023 9th International Conference on Smart Computing and Communications (ICSCC), Kochi, Kerala, India, 2023, pp. 102-107, doi: 10.1109/ICSCC59169.2023.10335023.