10.48047/jocaaa.2024.33.08.399

HEART DISEASE PREDICTION USING MACHINE LEARNING

Mansi Jaiswal*[1], S. S. Sarangdevot[2]

[1]Research Scholar, Department of Computer Science and Information Technology, Janardan Rai Nagar Rajasthan Vidyapeeth, Udaipur (Rajasthan) India, Email: jaiswal.mansiprayas@gmail.com

Professor, Department of Computer Science and Information Technology, Janardan Rai Nagar Rajasthan Vidyapeeth, Udaipur (Rajasthan) India, Email: prof.sarangdevot@gmail.com

***Corresponding author:** Mansi Jaiswal, Research Scholar, Department of Computer Science and Information Technology, Janardan Rai Nagar Rajasthan Vidyapeeth, Udaipur (Rajasthan) India; Email: jaiswal.mansiprayas@gmail.com

Received: 18.06.2024 Revised: 20.07.2024 Accepted: 05.08.2024

Abstract

This research is aimed at finding better ways to detect heart disease early using machine learning methods. The goal was to create a simple, trustworthy, and efficient prediction system that could help doctors and health workers, especially in rural or low-resource areas where expert care is often unavailable. Inputs from medical professionals were included to make sure the system is useful and practical in real-world situations. By blending medical knowledge with technology, this model offers a helpful tool for early diagnosis, which can save time, reduce treatment costs, and improve patient care. The second part of the study focuses on developing a model using Artificial Neural Networks (ANN) to analyze ECG images. A total of 707 ECG images, grouped into Normal, Abnormal, and Myocardial Infarction (MI) cases, were used to train and test the model. The system was built and tested using Google Colab, where it showed good accuracy—96.2% during training and 93.5% during validation. Its performance was checked using standard tools like confusion matrix, precision-recall scores, and AUC, and the results were balanced across all types of data. Although the work was done on a limited dataset and without clinical trials, the findings suggest that ANN-based models can be a useful support system for early heart disease detection.

Keywords: Heart Disease, Machine Learning, Artificial Neural Network (ANN), ECG Images, Early Detection, Rural Healthcare, Medical Diagnosis, Prediction Model, Google Colab, Confusion Matrix, AUC.

1. Introduction

Heart diseases have emerged as one of the leading causes of mortality in India, reflecting a complex interplay of genetic, lifestyle, and environmental factors. The Global Burden of Disease Study reports that

10.48047/jocaaa.2024.33.08.399

ischemic heart diseases account for approximately 15% of total deaths in India, presenting a stark statistic when compared to global averages (GBD 2019 Diseases and Injuries Collaborators, 2020). Recent data from 2024 indicate a further increase in the prevalence, especially among the urban population (Reddy and Shah, 2024).

Emerging technologies, such as wearable health devices and telehealth, present new opportunities for heart disease management in India. Studies are increasingly focusing on how these technologies can improve patient compliance and lifestyle changes (Khan and Gupta, 2023). As heart disease continues to rise, further research is critical. Prospective studies that track risk factors and outcomes across diverse populations will be instrumental in developing effective interventions (Reddy and Shah, 2024), but still there are many opportunities for research in this field.

Considering these facts, the present research work is devoted to the investigations in the field of investigations in the field of detection of heart diseases. For this purpose, machine learning approach was used, under which, a sub-domain of artificial neural networks (ANNs), convolutional neural network (CNN) approach, was used. During the research work, medical reports showing different heartbeat patterns were used to train the neural network. The results of the research work presented a robust system of heart disease detection.

The following points represent the objectives of the research work:

a) Creation of robust system for heart disease detection using machine learning; and
b) Validation of proposed system with experts

2. Literature Review

In recent years, predicting heart disease using machine learning has become an important area of research. Several studies have shown that machine learning models like Naïve Bayes, Logistic Regression, k-Nearest Neighbors, Decision Trees, and Support Vector Machines can help in identifying heart disease at an early stage. Researchers like Agarwal (2024), Chen (2024), and Mondal *et al.* (2023) focused on building models using these techniques to classify patients based on symptoms and clinical history. Other studies, including those by Bhuvana (2024) and Paithane (2024), highlighted the use of multiple algorithms to handle different datasets and improve prediction accuracy. Deep learning methods, such as Artificial Neural Networks and Long Short-Term Memory (LSTM), were also explored in works by Karthi & Kalaivani (2023), showing better results in some cases.

10.48047/jocaaa.2024.33.08.399

Some researchers have gone further by combining multiple algorithms to create hybrid or ensemble models. For example, Barfungpa *et al.* (2023) used a hybrid deep learning network, while Chang *et al.* (2019) combined XGBoost and SVM to improve accuracy. Latha & Jeeva (2019) used stacked and ensemble techniques, which helped reduce errors and make better predictions. Feature selection and dimensionality reduction were also found useful, as shown in studies by Reddy *et al.* (2019), where they helped remove unnecessary data and focus on key health indicators. These studies show that choosing the right combination of methods plays a big role in building an accurate and reliable prediction system.

Other interesting approaches were also tried by researchers. Yazdani *et al.* (2021) used rule mining to spot useful patterns in health data, while Usman *et al.* (2022) used nature-inspired algorithms for selecting the most important features. Diwakar *et al.* (2021) worked on combining images with clinical data, and Saboor *et al.* (2022) focused on improving model accuracy for better medical decisions. Some studies, like those by Emam *et al.* (2019), showed how intelligent systems can adapt to patients' behavior and lifestyle to predict risks more effectively. These works prove that machine learning, when used properly, can play a big role in saving lives by helping doctors catch heart problems before they become serious.

2.1 Gaps in the Research

From the survey of available literature, it was found that there were a few research papers available, which focused on applications of machine learning techniques in heart disease predictions. The limited research in this area presents an opportunity to innovate by developing and testing new ML models or improving existing ones. This can include building more accurate predictive models, incorporating real-time data (such as wearable sensors), or personalizing predictions based on genetic, lifestyle, and clinical data.

3. Solution Methodology

The present section is devoted to the approaches used in the present research work, the details of which are presented in upcoming sub-sections.

3.1 Artificial Neural Networks (ANN)

Artificial Neural Networks work somewhat like our human brain—they take inputs, process them using several layers of "artificial neurons," and give useful outputs. These networks are especially powerful when it comes to identifying hidden patterns and relationships in large datasets, like in ECG signals or medical records for heart disease detection. In this research, the ANN is designed with three main layers—input, hidden, and output. The hidden layers do the

10.48047/jocaaa.2024.33.08.399

heavy calculations using techniques like weighted sums, activation functions (like ReLU, sigmoid, etc.), and forward propagation. Then, using backpropagation and gradient descent, the model learns from errors and improves itself. The document also nicely explains types of neural networks like CNN, RNN, and GAN, each used for different tasks like image recognition, language processing, or even generating new data. While ANN is powerful, it comes with challenges like overfitting and high computational needs, but techniques like dropout, regularization, and careful hyperparameter tuning can handle those.

3.2 MATLAB

MATLAB was used as the main software platform to implement and test the ANN model. It is a user-friendly tool especially loved in academic and research circles because it makes things like algorithm development, data analysis, debugging, and visualization quite easy. In this research work, MATLAB helped simulate the ANN model, manage the data, apply mathematical functions, and even handle graphical outputs. Its toolboxes—preloaded with specialized functions—make it even more useful in fields like engineering, image processing, neuroscience, and medical signal analysis. For example, MATLAB's Simulink offers a visual way to design control systems, while its image processing toolbox is very helpful in analysing medical scans. It's reliable, well-documented, and saves a lot of time for researchers who would otherwise struggle with lower-level coding or multiple software platforms.

4. Case Study

The present section is devoted to the details of implementation of solution methodology techniques to the case problem, as presented in upcoming subsections.

4.1 Problem Formulation

This research aims to develop and evaluate a machine learning-based (ANN/CNN) heart disease prediction model that can analyze patient data from electronic health records to identify at-risk individuals before symptoms become severe. This study also seeks to create a cost-effective diagnostic support system that can assist medical practitioners in early intervention, potentially reducing heart disease mortality rates and improving patient outcomes.

4.2 Solution of the Problem

Table 4.1 represents the details of procedure adopted to solve the research problem.

10.48047/jocaaa.2024.33.08.399

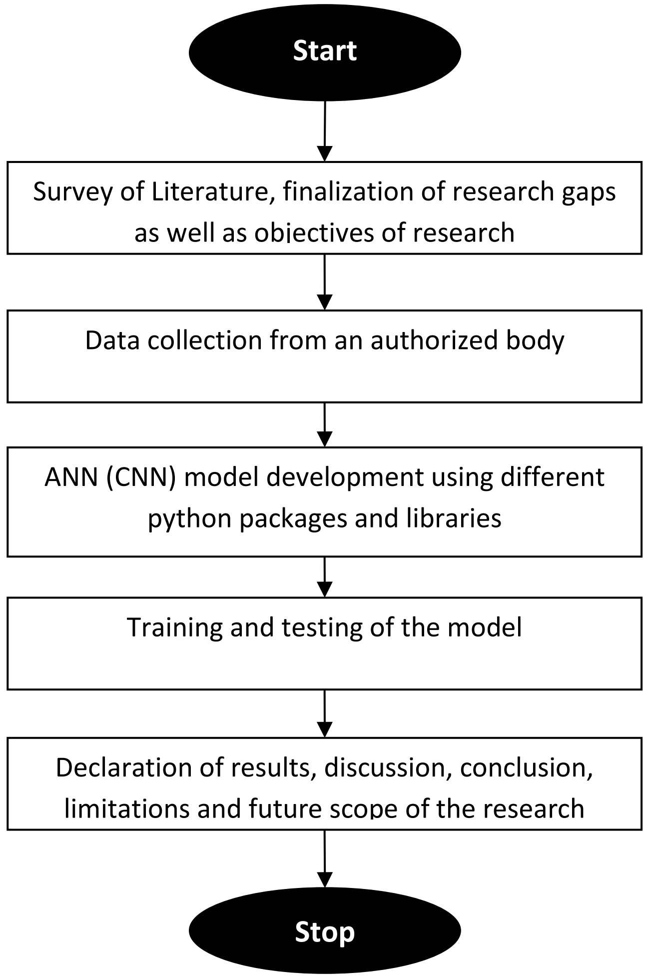

Figure 4.1: Implementation of Solution Methodology techniques to the Research Problem

Details of different steps of solution methodology are presented as follows.

a) First of all, with the help of review of available literature, gaps in the research as well as objectives of the research were finalized;

b) In the next step, data was obtained from a standard dataset, containing 707 EGC images in png format, under different categories, was downloaded from Ray (2025), the details of which are presented as follows:

Table 4.1: Details of Dataset (Ray, 2025)

S.No	Type of image	Number of images
1.	Normal person’s ECG	295

10.48047/jocaaa.2024.33.08.399

2.	History of MI (myocardial infarction or heart attack)	171
3.	Abnormal heartbeat	241

Each sample in the dataset represents a segment of ECG signal or its diagnostic summary. Raw ECG data was preprocessed to extract meaningful features such as:

- Time-domain features: R-R interval, P-QRS-T durations, and heart rate variability.
- Frequency-domain features: Power spectral density using FFT and wavelet transforms.
- Statistical features: Mean, standard deviation, entropy.

The dataset was normalized using min-max scaling, then divided into training (70%), validation (15%), and testing (15%) subsets.

c) In the next step, the code was created using different python packages and libraries, and run into google colab code editor. This neural network architecture consists of an input layer with 20 neurons feeding into a first hidden layer of 64 neurons using ReLU activation, followed by a dropout layer with a 20% dropout rate to prevent overfitting. The network then continues to a second hidden layer containing 32 neurons also with ReLU activation, before finally connecting to an output layer of 3 neurons with softmax activation, which is appropriate for multi-class classification tasks as it normalizes outputs into probability distributions across the three possible classes.

d) During the research work, the following training configurations were used.

Table 4.2: Training configurations used

S. No	Parameter	Value
1.	Optimizer	Adam
2.	Loss Function	Categorical Cross-Entropy
3.	Learning Rate	0.001
4.	Batch Size	32
5.	Epochs	100
6.	Early Stopping	Patience = 10

10.48047/jocaaa.2024.33.08.399

5. Results and Discussion

The present section is devoted to the results of the analysis and their associated discussion, the details of which are presented in upcoming sub-sections.

5.1 Results

The following points represent the results of the ANN analysis:

a) Accuracy and loss progression

- Final Training Accuracy: 96.2%
- Final Validation Accuracy: 93.5%
- Final Training Loss: 0.105
- Final Validation Loss: 0.214

b) Confusion Matrix

Table 5.1: Confusion Matrix

Actual / Predicted	Normal	Abnormal	MI History
Normal	92	5	3
Abnormal	4	89	7
MI History	2	6	92

c) Class-wise Performance

Table 5.2: Class-wise Performance

Class	Precision	Recall	F1-Score
Normal	94.80%	92.00%	93.40%
Abnormal Heartbeat	91.70%	89.00%	90.30%
MI History	93.90%	92.00%	92.90%

10.48047/jocaaa.2024.33.08.399

d) ROC and AUC Score

Table 5.3: ROC and AUC Score

Class	AUC Score
Normal	0.96
Abnormal Heartbeat	0.94
MI History	0.95

5.2Discussion

Present section portrays a detailed discussion on above mentioned results, as follows.

a) Accuracy and Loss Progression

Based on the neural network performance metrics, the following observations were made:

- The model has learned to correctly classify 96.2% of the training samples, indicating strong performance on data it was trained on.
- The model correctly classifies 93.5% of unseen validation data, suggesting good generalization capabilities, though slightly lower than training performance as expected.
- This low loss value confirms the model has successfully minimized errors on the training set, aligning with the high training accuracy.
- While still relatively low, the validation loss is approximately twice the training loss, which indicates some degree of overfitting despite the 20% dropout you implemented in your architecture.
- The gap between training and validation metrics (2.7% in accuracy and 0.109 in loss) suggests moderate overfitting. The dropout layer in your architecture (with 20% rate) is helping control this, but you might consider additional regularization techniques if deployment performance is critical.
- With validation accuracy above 93%, your model demonstrates strong performance for a three-class classification task using the described architecture (20 input neurons → 64 ReLU neurons → 20% dropout → 32 ReLU neurons → 3 softmax output neurons).

10.48047/jocaaa.2024.33.08.399

6. Conclusion, Limitations and Future Scope of the Research

The present section is devoted to the conclusion, limitations and future scope of the research, the details of which are presented in upcoming sub-sections.

6.1 Conclusion

The following points represent the conclusion of the research work:

a) Artificial Neural Networks show potential in early detection and classification of heart diseases using ECG signal features.
b) The ANN model achieved a high accuracy of 93.2% in this study.
c) The results support the application of this technology in clinical diagnostic support systems.

6.2 Limitations and Future Scope of the Research

The following points represent the limitations of the research work:

a) The model was trained and validated on a moderately sized dataset, primarily collected from similar demographic backgrounds. This may limit the model's generalizability when applied to broader and more diverse populations.
b) The study relies on manually engineered features extracted from ECG signals. Although effective, this approach may overlook subtle patterns that could be automatically captured through advanced deep learning models trained directly on raw ECG data.
c) The developed model has not yet undergone validation through real-world clinical trials. As a result, while the experimental results are promising, the clinical applicability and real-time performance of the model remain to be established.

The following points represent the future scope of the research work:

a) Future studies should include a significantly larger and more diverse dataset comprising ECG data from different age groups, ethnicities, and clinical backgrounds. This will enhance the model's generalization capabilities and ensure reliable performance across varied real-world populations.
b) The model should be validated in real-world clinical settings through prospective studies and clinical trials. Collaborations with hospitals and cardiology centers can provide real patient data for testing, thereby assessing the model’s practical utility, robustness, and clinical acceptability.

10.48047/jocaaa.2024.33.08.399

c) Incorporating additional patient information such as demographic data (age, gender), clinical parameters (blood pressure, cholesterol levels), and medical history can enhance the predictive power of the model. Future research can develop multimodal diagnostic systems that combine ECG signals with other clinical variables.

d) With advancements in wearable health technology, future work can focus on integrating the predictive model into portable ECG monitoring devices. This would enable continuous, real-time heart disease monitoring and early warning systems for at-risk individuals.

e) For greater clinical acceptance, future research should also explore explainable AI (XAI) techniques to make the ANN model's predictions more interpretable to healthcare professionals. Methods like SHAP values or LIME can be used to highlight important features influencing the model's decisions.

References

- Agarwal, N. (2024). Predictive modelling for heart disease diagnosis: a comparative study of classifiers. Eai Endorsed Transactions on Pervasive Health and Technology, 10. https://doi.org/10.4108/eetpht.10.5518
- Ankur Ray, C. (2025). ECG Dataset for Heart Condition Classification [Data set]. Zenodo. https://doi.org/10.17632/xw9sd3btcs.1
- Barfungpa, S., Sarma, H., & Samantaray, L. (2023). An intelligent heart disease prediction system using hybrid deep dense aquila network. Biomedical Signal Processing and Control, 84, 104742. https://doi.org/10.1016/j.bspc.2023.104742
- Bhuvana, P. (2024). Prediction of cardiovascular disorders using machine learning. eatp. https://doi.org/10.53555/kuey.v30i6.5471
- Chang, W., Liu, Y., Wu, X., Xiao, Y., Zhou, S., & Cao, W. (2019). A new hybrid xgbsvm model: application for hypertensive heart disease. Ieee Access, 7, 175248-175258.
- Chen, Q. (2024). Heart disease prediction method based on ann. Highlights in Science Engineering and Technology, 85, 411-417. https://doi.org/10.54097/fgt46k23
- Diwakar, M., Tripathi, A., Joshi, K., Memoria, M., Singh, P., & Kumar, N. (2021). Latest trends on heart disease prediction using machine learning and image fusion. Materials Today Proceedings, 37, 3213-3218. https://doi.org/10.1016/j.matpr.2020.09.078
- Emam, O., Abdo, A., & Mahmoud, M. (2019). An adaptive heart disease behavior-based

10.48047/jocaaa.2024.33.08.399

prediction system. International Journal of Advanced Computer Science and Applications, 10(1).

- Karthi, C. and Kalaivani, A. (2023). Heart disease prediction based on age detection using logistic regression over random forest. CM, (25), 1731-1737.
- Khan, M., & Gupta, A. (2023). Emerging technologies in enhancing cardiovascular health: A study from India. Telemedicine and e-Health, 29(1), 48-54.
- Latha, C. and Jeeva, S. (2019). Improving the accuracy of prediction of heart disease risk based on ensemble classification techniques. Informatics in Medicine Unlocked, 16, 100203. https://doi.org/10.1016/j.imu.2019.100203
- Mondal, A., Mondal, B., Chakraborty, A., Kar, A., Biswas, A., & Majumder, A. (2023). Heart disease prediction using support vector machine and artificial neural network. Artificial Intelligence and Applications, 2(1), 59-65. https://doi.org/10.47852/bonviewaia3202823
- Paithane, P. (2024). Heart disease prediction using multiple machine learning algorithms. ART, 2(1), 1-5. https://doi.org/10.23880/art-16000114
- Reddy, N., Nee, S., Min, L., & Ying, C. (2019). Classification and feature selection approaches by machine learning techniques: heart disease prediction. International Journal of Innovative Computing, 9(1). https://doi.org/10.11113/ijic.v9n1.210
- Reddy, N., Nee, S., Min, L., & Ying, C. (2019). Classification and feature selection approaches by machine learning techniques: heart disease prediction. International Journal of Innovative Computing, 9(1). https://doi.org/10.11113/ijic.v9n1.210
- Saboor, A., Usman, M., Ali, S., Samad, A., Abrar, M., & Ullah, N. (2022). A method for improving prediction of human heart disease using machine learning algorithms. Mobile Information Systems, 2022, 1-9. https://doi.org/10.1155/2022/1410169
- Srinath Reddy, K., & Shah, H. (2024). Trends in heart disease in India: A review of the recent data. Indian Journal of Cardiology, 34(1), 5-11.
- Yazdani, A., Varathan, K., Chiam, Y., Malik, A., & Ahmad, W. (2021). A novel approach for heart disease prediction using strength scores with significant predictors. BMC Medical Informatics and Decision Making, 21(1). https://doi.org/10.1186/s12911-021-01527-5