Volume (1) No (11), 2025
Journal of Global Research in Multidisciplinary Studies (JGRMS)
Review Paper/Research Paper
Available online at https://saanvipublications.com/journals/index.php/jgrms/index

Machine Learning Approaches for Early Identifictaions of Heart Disease in Healthcare Applications

Dr Manish Saraswat
Associate Professor
(CSE) and Controller of Examinations
Faculty of Science and Technology
The ICFAI University, Himachal Pradesh
manish.saraswat@iuhimachal.edu.in

***Abstract*—Early detection and accurate diagnosis are crucial for improving patient outcomes and prescribing appropriate therapy for heart disease, which is still one of the leading causes of death globally. It can be extremely costly, time-consuming, and intrusive to use conventional diagnostic methods in settings with limited resources. Automated, data-driven prediction of cardiovascular problems has shown promising potential thanks to new ML and DL advances. This research presents a strong artificial neural network (ANN) model for the prediction of cardiovascular disease using a large-scale Kaggle dataset on the subject. The dataset includes more than 300,000 samples with 17 health-related variables. Complete data pre-processing, including missing value filling, duplicate removal, feature selection via the Chi-Square test, and min-max normalization, is required by the method. Two sets of data, one for training and one for testing, are used to teach an artificial neural network (ANN) to correctly identify cases of cardiac diseases. The experimental findings indicate that the proposed ANN is more effective, with an accuracy (ACC) of 98.75%, precision (PRE), recall (REC) and F1-score (F1) of 99, which is better than the baseline models including MLP, DNN, and Random Forest. Remarkably, the model is stable, can be generalized and is reliable, which is evidenced by ROC and loss curve analyses. The results point out the potential of the proposed method as a high-precision, cost-effective and clinically appropriate tool of early heart disease detection, which may help achieve timely interventions and better healthcare decisions.**

***Keywords*—Heart Disease Prediction, Artificial Neural Network (ANN), Data Normalization, Cardiovascular Disease, Healthcare Analytics, ROC Curve.**

I. INTRODUCTION

The heart is important in keeping one alive as it pumps blood that is well oxygenated and keeps the necessary hormones to ensure optimum blood pressure levels. It is also prone to any form of disruption to its normal operation and the resultant occurrence of heart related diseases, also referred to as CVD [1][2][3]. Numerous disorders influencing the heart and blood vessels are collectively known as cardiovascular disease (CVD). These include vascular anomalies, arrhythmias, peripheral arterial disease, coronary artery disease (CAD), rheumatic heart disease (CHD), and cardiomyopathies [4][5][6].

Heart disease can be effectively treated and prevented by means of early diagnosis and treatment, before it results in severe complications and mortality. Nonetheless, the availability of diagnostic facilities, trained medical practitioners, and medical infrastructure is problematic, and therefore, it is difficult to have early and accurate diagnosis, especially in developing countries. To address these issues, machine learning (ML) and the enhanced use of technology in the medical field of a decision-support system is becoming an appealing solution [7][8][9]. Health organizations face increasing pressure to deliver affordable and high-quality healthcare services, where precise diagnosis and proper treatment planning are essential to prevent misdiagnoses and improve patient outcomes. Early CVD diagnosis is not only cost effective in healthcare provision but also it helps to reduce the number of deaths vastly.

Despite the popularity of diagnostic procedures (such as angiography) with medical professionals, it is limited by the invasive nature, time-intensive, and expensive processes, making it less accessible due to the low-resource conditions. The emergence of ML and DL has transformed healthcare as it can automatically predict and identify diseases based on the data. These technologies enhance the ACC of medical diagnostics, support disease classification, and provide clinicians with actionable insights for early intervention [10][11]. The increasing availability of large-scale healthcare data has further strengthened the role of ML and DL models in improving the precision of disease identification, risk assessment, and prevention strategies [12][13][14].

A. Motivation and Contribution

The motivation for this research is the critical and pressing need for reliable and affordable diagnostic tools on a global scale. The major killer on a global scale is still heart disease and stroke. Although angiography and electrocardiogram (ECG) interpretation are efficient traditional diagnostic procedures, they are often unavailable in healthcare systems with limited resources due to the unique knowledge and technology they require. Advanced computer technology and the proliferation of medical data have made ML and DL potent instruments for automating early disease identification and aiding physicians in making prompt judgments. This study is therefore motivated by the potential of ML-driven approaches to transform cardiac diagnosis into a more proactive, data-driven, and cost-effective process that enhances healthcare

accessibility and quality. The area of healthcare applications benefits greatly from this study's numerous important findings:

- Developed a robust ANN-based predictive model achieving ACC, PRE, REC, and F1 for heart disease prediction.
- Enhanced model performance by the implementation of a streamlined pre-processing pipeline that includes management of missing values, feature selection using a Chi-Square test, and normalization.
- Conducted a comprehensive comparative analysis demonstrating the superiority of the proposed ANN model over MLP, DNN, and Random Forest models.
- Validated the model's clinical applicability through ROC analysis and loss/ACC curve evaluation, confirming high generalization, stability, and reliability for healthcare applications.

B. Justification and Novelty

This study proposes a model for cardiac disease prediction using ANNs that effectively uses state-of-the-art techniques for pre-processing, feature selection, and normalization. Unlike previous approaches, this model demonstrates near-perfect performance across all evaluation metrics (ACC, PRE, REC, F1), ensuring balanced prediction while minimizing false positives and negatives. The novelty lies in the comprehensive methodology that combines rigorous data pre-processing, optimized ANN architecture, and extensive validation to achieve superior generalization on the cardiovascular disease dataset. Its ability to outperform traditional models such as MLP, DNN, and RF establishes its potential for reliable clinical decision-making in early and precise heart disease detection.

C. Structure of the Paper

This paper is organized in the following way Section II provides a summary of current studies that have examined the use of ML to the prediction of heart disease. Section III describes the methods in depth, including data preparation and model creation. Section IV presents the findings and discusses them in relation to the model's performance analysis. Section V concludes with a review of the work and some recommendations for further research.

II. Literature Review

Extensive research studies on predictive modelling for heart disease have been thoroughly reviewed and analyzed to guide and strengthen the foundation of this work.

Khedkar et al. (2025) investigates the use of a heart disease dataset to test both traditional and novel methods for classifying cardiac conditions. Additionally, a KNN classifier is utilized to complement the prototype-based results by considering similarity-based classification. Results from a comparison show that few-shot-inspired methods, when used with baseline models, provide useful information on the resilience and performance of classifications on healthcare datasets. The 98.5% ACC achieved by the suggested strategy is quite remarkable. These results highlight the possibility of improving categorization ACC in the diagnosis of cardiac disease by combining prototype-based approaches with conventional models [15].

Varshini et al. (2025) introduce the model at an early stage of prediction in a timely intervention remains a challenge. Attempts have been made for the following models: XGBoost, RF, Gradient Boosting, LR, SVM, CNN, and Bayesian Networks. One-hot encoding, ANOVA F-test-based feature selection, and SMOTE used for advanced pre-processing for higher performance. In ensemble methods, XGBoost was the best method with ACC of about 97.32% while the second highest ACC rate was given by Random Forest with about 95.89% [16].

Biju, Mallya and P (2025) introduces a state-of-the-art bagging ensemble model that leverages Quantum Support Vector Classifiers (QSVC) to assess the risk of heart disease. Testing involved rigorous experimentation on the Cleveland data set for which QSVC has demonstrated better performance compared with benchmark classifiers and other quantum-inspired models as well, achieving an excellent ACC of 90.16%. In addition to its impressive performance metrics, the model incorporates SHAP (Shapley Additive explanations) analysis, which significantly enhances its interpretability [17].

Akther et al. (2024) seeks to discover an algorithm that consistently performs well across many datasets and incorporates it into an all-encompassing, user-centric method for predicting cardiac disease. Principal component analysis is one method in feature engineering that can aid with prediction. They also considered a feature selection method known as Lasso, although principal component analysis (PCA) produced the best results with the provided datasets. Using PCA, the XGBoost algorithm obtains a remarkable 97% AUC rate in disease prediction on the other dataset. It also produces a similarly amazing ACC rate and an F1 of over 99% [18].

El-Sofany (2024) used SMOTE, or the Synthetic Minority Oversampling Technique, to right the data distribution's wrong. With ACC rates of 97.57%, sensitivity of 96.61%, specificity of 90.48%, PRE of 95.00%, F1 of 92.68%, and AUC of 98%, the results showed that the XGBoost classifier worked ideally when fed the combined datasets and SF-2 feature subset. Because the proposed method was both simple and inexpensive, many were optimistic about its potential to help the healthcare industry predict the onset of early-stage heart disease [19].

Maragatharajan, Naveen and Nirmala (2024) Introduces cutting-edge research to predict heart diseases. This application designed for accurate and personalized assessment of cardiovascular health. This study considered two well-known machine learning models namely Extreme learning machine algorithms (ELM) and Crow search algorithm (CSA). The collected dataset consists of 10000 samples. They have attained 92.5% ACC with the data set. Healthcare providers and individuals aiming for proactive cardiac care benefit greatly from this tool because of the implementation of machine learning models, which guarantee scalability, adaptability, and precision [20].

Mahalakshmi and Rajakumari (2023) aim to prioritize the identification and avoidance of cardiovascular diseases in their early stages. The primary objective is to develop and discover a trustworthy approach for the evaluation of coronary artery disease (CAD) and its prognosis. The idea plans to use the Cleveland dataset and other well-established databases on cardiovascular disorders to do this. With an impressive 100% ACC rate, the Random Forest algorithm has proven to be the most effective of all the strategies tested [21].

additionally the amazing progress made by ML and ensemble models, there are still many unanswered questions regarding the prediction of cardiovascular illness. Most existing studies focus on achieving high classification ACC using well-known algorithms such as XGBoost, Random Forest, CNN, and QSVC, often on standard datasets like Cleveland or synthetic combinations. While these approaches demonstrate promising results, there is limited exploration of integrating prototype-based, few-shot, or quantum-inspired methods with advanced feature selection and balancing techniques to improve generalization across diverse, real-world datasets. Additionally, interpretability and timely prediction for early-stage heart disease remain underexplored, particularly in models combining deep learning with explainable AI methods like SHAP. There is also a scarcity of research addressing model scalability, robustness across heterogeneous datasets, and cost-effective real-time deployment in practical healthcare settings. These gaps highlight (Table I) the need for hybrid, interpretable, and adaptive frameworks capable of delivering accurate, early, and reliable heart disease predictions

TABLE I. OVERVIEW OF RECENT STUDIES ON HEART DISEASE PREDICTION USING MACHINE LEARNING

Author	Proposed Work	Dataset	Key Findings	Challenges/recommendation
Khedkar et al. (2025)	Introduced a few-shot-inspired prototype-based classifier with KNN integration	Heart Disease Dataset	Achieved 98.5% ACC, showing robustness by combining prototype and traditional classifiers	Encourages blending few-shot learning with baseline models to enhance classification ACC
Varshini et al. (2025)	Compared ML, DL, and Bayesian models with advanced preprocessing (ANOVA F-test, SMOTE)	UCI Cleveland & Kaggle Heart Disease Dataset	XGBoost achieved 97.32% ACC; RF was second-best with 95.89%	Recommends early-stage prediction methods and emphasizes preprocessing for better performance
Biju, Mallya and P (2025)	Proposed a Quantum SVM-based bagging ensemble model with SHAP-based explainability	Cleveland Heart Disease Dataset	Achieved 90.16% ACC; enhanced interpretability via SHAP analysis	Promotes quantum machine learning integration and model transparency
Akther et al. (2024)	Explored 11 algorithms across 4 Kaggle datasets with PCA and Lasso feature selection	Multiple Kaggle Heart Disease Datasets	XGBoost with PCA reached 99% ACC and F1, AUC of 97%	Suggests PCA as optimal feature reduction method; promotes algorithm consistency
El-Sofany (2024)	Applied SMOTE and SHAP for XGBoost model explainability and mobile integration	Combined Heart Disease Datasets	Achieved 97.57% ACC, 98% AUC; mobile app for user-side prediction	Recommends low-cost, fast AI tools for practical and interpretable healthcare use
Maragatharajan, Naveen and Nirmala (2024)	Developed a personalized assessment tool using ELM and CSA	Heart Disease Dataset (10,000 samples)	Attained 92.5% ACC; scalable for public and clinical use	Promotes integration of scalable ML for proactive personal heart health monitoring
Mahalakshmi and Rajakumari (2023)	Focused on early detection and prevention using RF algorithm	Cleveland Heart Disease Dataset	Reported 100% ACC using Random Forest	Encourages highly accurate models for early-stage coronary artery disease detection
Khedkar et al. (2025)	Introduced a few-shot-inspired prototype-based classifier with KNN integration	Heart Disease Dataset	Achieved 98.5% ACC, showing robustness by combining prototype and traditional classifiers	Encourages blending few-shot learning with baseline models to enhance classification ACC

III. RESEARCH METHODOLOGY

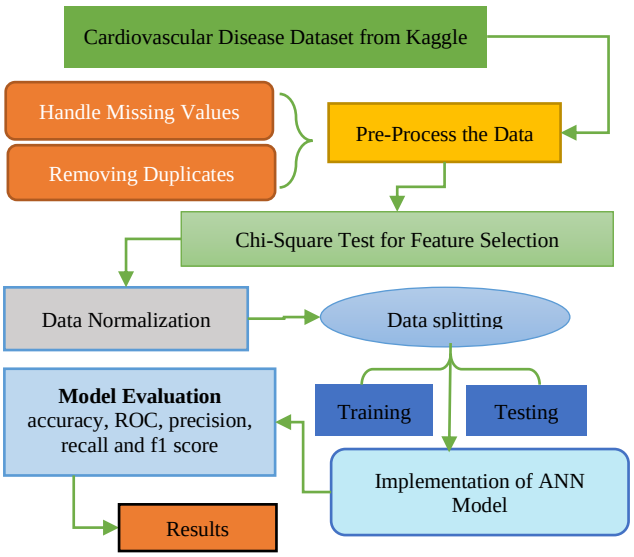

Fig. 1. Proposed Flowchart for Heart Disease Prediction

The cardiovascular disease dataset is obtained from Kaggle as part of the approach for heart disease prediction. Then, there are pre-processing processes to deal with missing information and eliminate duplicates. To ensure consistent scaling, the data is normalized before feature selection using the Chi-Square test. A classification model based on ANNs can be used by splitting the dataset into training and testing groups. While assessing the model's efficacy, several metrics are utilized, including ACC, PRE, REC, F1, and ROC. The results are then assessed to determine the model's predictive potential. In Figure 1 shows the whole process workflow.

This section lays out, in detail, the proposed flowchart for cardiac disease prediction modelling.

A. Data Collection

The Kaggle dataset on cardiovascular disease is utilized in this research. Information regarding the health status of over 300,000 adults is included in this dataset. Twelve numerical features and five categorical features make up its seventeen traits. Data visualizations such as bar plots and heatmaps were used to examine attack distribution, feature correlations etc., are given below:

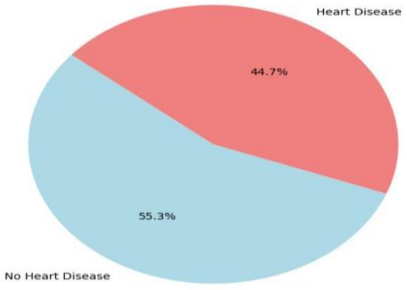

Fig. 2. Data Visualization in Heart Disease in Heart Dataset

Figure 2 shows the pie chart illustrates the distribution of heart disease cases within the dataset. It shows that 55.3% of the individuals do not have heart disease, while 44.7% are affected by it. This relatively balanced distribution indicates a moderate class imbalance, which is important to consider when developing predictive models, as it may impact the ACC and sensitivity of the classification.

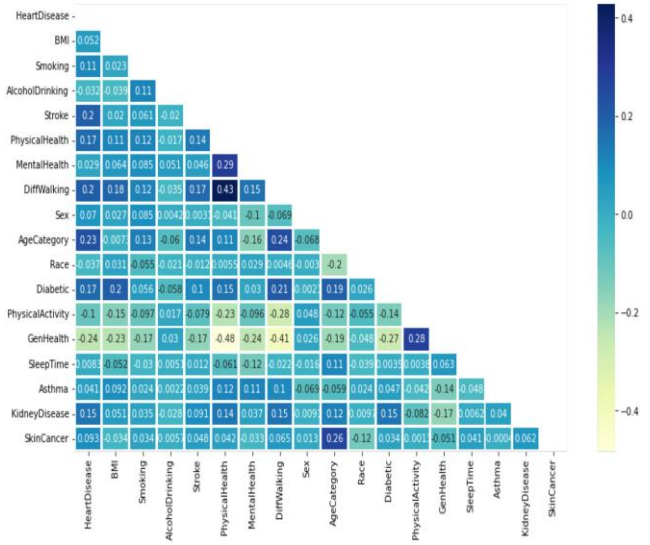

Fig. 3. Features Heatmap of the Dataset

Figure 3: The correlation heatmap shows the connections between different health variables and cardiovascular disease graphically. Darker shades indicate stronger correlations, either positive or negative. The matrix shows that heart disease has a modest positive correlation with factors such as stroke, physical health, and age category, while features like sex, race, and skin cancer exhibit weaker or negligible correlations. Notably, physical activity and general health also show moderate correlations with heart disease, suggesting their potential relevance in predictive modelling. Overall, the heatmap provides valuable insights for feature selection in building effective heart disease prediction models.

B. Data Pre-Processing

The preparation of data started by gathering the Kaggle data on cardiovascular diseases, which was later merged and cleaned to maintain consistency. Relevant characteristics were obtained and the data was pre-treated to cater to missing values, eliminate duplicates, and identify outliers. Other steps involved data transformation and normalization. The step-by-step pre-processing procedures are as follows:

- **Handle missing value:** Removing the offending rows or columns, replacing them with a constant, or employing more sophisticated imputation methods such as the mean/median/mode imputation method, regression imputation, or KNN imputation are all viable options for filling in the missing values.
- **Removing Duplicates:** Duplicate rows may be detected and eliminated in data pre-processing with multiple methods including checking to find identical rows or by using distinct identifiers.

C. Chi-Square Test for Feature Selection

Machine learning's feature selection method narrows down a bigger pool of input variables to the most important ones, thereby improving model performance and interpretability. Feature selection alleviates overfitting, lowers computing cost, and enhances model ACC and generalization by decreasing the amount of features. The chi-square test shows promise as a feature selection statistic for categorical data. For the purpose of identifying the most relevant characteristics to the target variable, it evaluates the independence of categorical features from the target variable. A popular method for examining relationships between sets of category data is the chi-square test.

$$X_i^2 = \sum \frac{(O-P)^2}{P} \quad (1)$$

Here O is seen value and P is expected value. In Equation (1), X_i^2 is chi squared is denoted.

D. Data Normalization

Normalization of records occurred under the min-max technique to ensure that the records are limited within a range of 0 through 1. The objective was to make the most efficient use of the classifiers and minimize the effect of outliers. The following mathematical formula was used to conduct the procedure of normalization Equation (2):

$$X' = \frac{X - X_{min}}{X_{max} - X_{min}} \quad (2)$$

X_{min} represents the minimum value, X_{max} stands for the maximum value, X' stands for the feature's normalized value, and X is the beginning value of the feature.

E. Data Splitting

The model's efficacy was assessed by splitting the entire dataset into a training set and a testing set. In particular, the model's parameters and training were done using 70% of the data, whereas testing and performance evaluation were done with 30%.

F. Proposed Artificial Neural Network (ANN) Model

The model, which is defined by several layers with multiple nodes/neurons per layer, mimics the organization of biological brain neurons; this approach for supervised deep learning achieves this goal. Each neuron in the model receives input from all of the neurons in the layer below it and sends out signals to all of the neurons in the layer above it. To train a model's neurons, first feed them inputs, multiply those inputs by the weights assigned to each neuron, and then add a bias. This total is then fed into a function like Sigmoid or Lu. Neurones in the subsequent layer get the result of the applied function as input. Neurons at the last layer provide the ultimate forecast. Output of each neuron shows in Equation (3). With hidden layer neurons {15,7,3,1}, the greatest ACC was attained.

$$f(b + \sum xiwi) \quad (3)$$

- x_i are the inputs from the previous layer neurons
- w_i are the weights for that neuron
- b is the bias term.

G. Evaluation Metrics

PRE, ACC, REC, and F1 were used to assess the classification algorithms' performance. The ACC of the models in identifying positive and negative scenarios was assessed by calculating TN, FN, TP, and FP. By minimizing misclassifications and maximizing the detection of real positives, these metrics give a thorough evaluation of each algorithm's performance. The calculation of F1, REC, ACC, and PRE is shown in the following matrix:

Accuracy: A measure of how well the trained model predicted outcomes relative to the whole dataset (input samples). The Equation (4) is-

$$Accuracy = \frac{TP+TN}{TP+Fp+TN+FN} \quad (4)$$

Precision: PRE measures how well a model predicts positive occurrences relative to all positive occurrences. PRE indicates. How good the classifier is in predicting the positive classes is expressed as Equation (5)-

$$Precision = \frac{TP}{TP+FP} \quad (5)$$

Recall: This statistic counts how many times the accuracy of the positive event predictions was higher than the number of times it should have been lower. The formula for it in mathematics is Equation (6)-

$$Recall = \frac{TP}{TP+FN} \quad (6)$$

F1 score: REC and PRE are both balanced out by this combination, which is like a harmonic mean of the two. Its range is [0, 1]. Mathematically, it is expressed as Equation (7)-

$$F1 - score = 2 \times \frac{Precision \times Recall}{Precision + Recall} \quad (7)$$

Receiver Operating Characteristic Curve (ROC): ROC is a graphical data set that displays the ratio of true positives to false positives at different decision-cut-off points. The REC or sensitivity is also called the TPR, and the FPR is equivalent to 1-specificity.

IV. Results and Discussion

This section presents the experimental setup and evaluation of the suggested model's performance. The experiments were run on MATLAB 2019b on a windows 10 Pro with 8GB RAM and indelicate i5-1135G7. Table II shows that the ACC, PRE, REC, and F1 metrics were used for model evaluation after training on the Kaggle dataset for cardiovascular illness. The developed Artificial Neural Network (ANN) was able to deliver excellent results and produce 98.75% ACC, 99% PRE, 99% REC, and 99% F1. These measures reveal that the model is capable of proper classification of cases, false positives and negatives are reduced, and that the model represents a balanced result, which indicates its strength and clinical applicability toward the prediction of heart disease.

TABLE II. Experiment Results of Proposed Models for Heart Disease Prediction of Healthcare Applications on Cardiovascular Disease Data

Performance Matrix	Artificial Neural Network (ANN)
Accuracy	98.75
Precision	99
Recall	99
F1-score	99

Figure 4 displays the ANN model's training and validation ACC after 100 epochs. The two curves exhibit an upward pattern, which implies there is constant improvement in the learning process of the model. The training ACC starts at approximately 65% and steadily grows reaching approximately 98% in the last epochs. Equally, validation ACC is much closer to the training curve at times, and it also approaches 99, indicating that it generalizes effectively, with no serious overfitting. The fact that there exists only a small distance between the two curves over the course of the training procedure confirms that the model is robust and stable in its performance over the unseen data, and, hence, it is sound and reliable in predicting heart diseases.

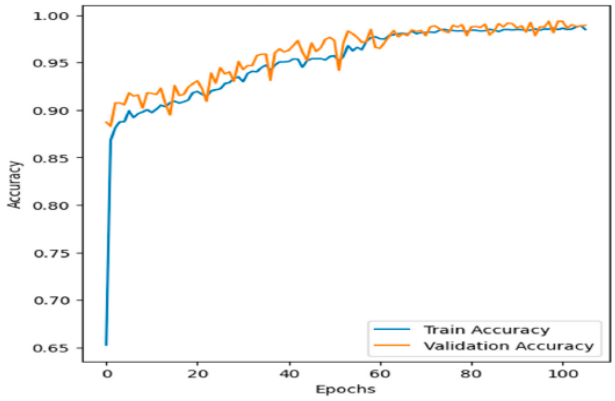

Fig. 4. Accuracy curves for the ANN Model

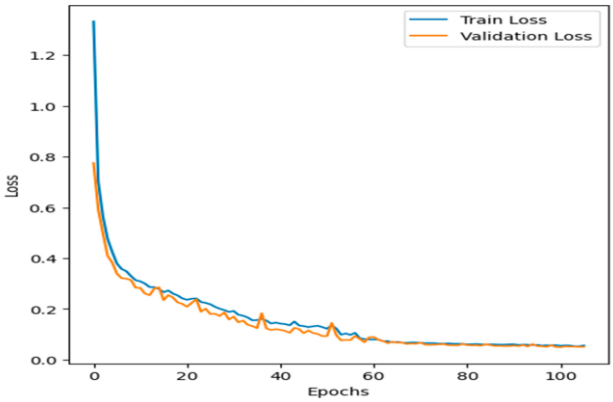

Fig. 5. Loss Curves for the ANN Model

Figure 5 displays the ANN model's loss curves for training and validation over a hundred iterations. During the initial 20 epochs, the training loss drops significantly from its high value of around 1.3 and the validation loss drops from its high value of around 0.8, suggesting that learning is taking place effectively. As training progresses, both curves continue to decline gradually and converge near zero, with minor fluctuations. The consistent reduction and close alignment of the two loss curves suggest that the model learns efficiently without overfitting, maintaining a balanced performance on both training and unseen validation data. The consistent convergence of loss values indicates that the model is stable and has good generalizability in forecasting the occurrence of heart disease.

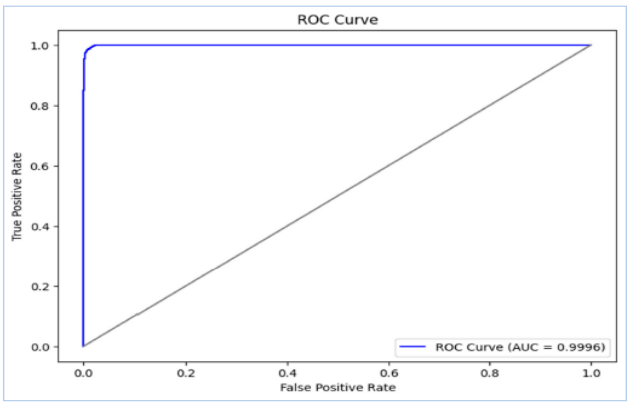

Fig. 6. ROC Curve of the ANN Model

In Figure 6, observe the ROC curve, which illustrates the ACC with which the ANN model predicted the occurrence of heart sickness. This curve compares the TPR (sensitivity) with the FPR. The blue curve is tightly shaped around the upper-left side of the graph which shows that the ACC in the distinction between positive and negative classes is high. The AUC equals 0.9996, and it implies close-to-perfect classification. This extremely high AUC value demonstrates that the model has excellent discriminatory power, effectively minimizing both false positives and false negatives, making it highly suitable for clinical decision-making in heart disease detection.

A. Comparative Analysis

Table III displays the results of a comparison of ACC with other existing models, which were used to validate the efficacy of the proposed ANN model. In this paper presents a comparative analysis of various predictive models used for heart disease prediction in healthcare applications, utilizing the cardiovascular disease dataset from Kaggle. The Multilayer Perceptron (MLP) achieved an ACC of 84.78%, with a PRE of 72%, a REC of 91%, and an F1 of 86.79%, indicating strong REC performance. The DNN demonstrated a higher ACC of 89.85%, with a PRE of 97.89%, a REC of 78.37%, and a comparatively lower F1 of 67.56%, reflecting some imbalance between PRE and REC. The RF model further improved results, attaining 90% ACC, 92% PRE, 98% REC, and a high F1 of 95%, showcasing its robustness. The proposed ANN outperformed all other models, achieving the highest ACC of 98.75%, with near-perfect PRE, REC, and F1 values of 99%, confirming its superior predictive capability and balanced performance across all evaluation metrics.

TABLE III. COMPARISON OF DIFFERENT MODELS FOR HEART DISEASE PREDICTION FOR HEALTHCARE APPLICATIONS USING THE CARDIOVASCULAR DISEASE DATASET

Models	Accuracy	Precision	Recall	F1-score
MLP[22]	84.78	72	91	86.79
DNN[23]	89.85	97.89	78.37	67.56
RF[24]	90	92	98	95
ANN	98.75	99	99	99

The proposed ANN demonstrates a clear advantage over other models by achieving the highest ACC of 98.75%, along with near-perfect PRE, REC, and F1 values of 99%. This indicates not only superior predictive capability but also a balanced performance across all evaluation metrics, minimizing both FP and FN. Compared to MLP, DNN, and Random Forest models, the ANN provides more reliable and consistent results, making it particularly effective for applications where high ACC and robust classification are critical. Its ability to outperform traditional and deep learning models highlights its suitability for complex prediction tasks requiring precise and balanced outcomes.

V. CONCLUSION AND FUTURE STUDY

The provision of proper and timely forecasting of heart disease can help in reducing the mortality rate and enhancing the quality of healthcare. This work suggested an effective ANN model to predict heart disease, using a large cardiovascular dataset on the Kaggle platform and the complex pre-processing methods, such as missing values processing, duplicate elimination and user Chi-Square-based features selection, as well as the minmax normalization. The model demonstrated exceptional performance, achieving 98.75% ACC, 99% PRE, 99% REC, and 99% F1, outperforming conventional models such as MLP, DNN, and RF. Analysis of training and validation curves, along with ROC evaluation, confirmed the model's stability, strong generalization, and clinical relevance, highlighting its potential for accurate disease classification and early intervention. This model can be further developed to incorporate a wider range of diverse and large-scale datasets in order to enhance its robustness across various populations in future research. Model interpretability for clinicians can be enhanced by including explainable AI approaches like SHAP or LIME. Additionally, hybrid models combining deep learning with ensemble or quantum-inspired approaches may further optimize predictive performance. Deployment in real-time healthcare systems, including cloud- or mobile-based platforms, can make heart disease prediction more accessible, cost-effective, and practical, ultimately supporting preventive care and timely medical decision-making in clinical practice.

REFERENCES

[1] B. Şahin and G. İlgün, "Risk factors of deaths related to cardiovascular diseases in World Health Organization (WHO) member countries," *Heal. Soc. Care Community*, 2022, doi: 10.1111/hsc.13156.

[2] R. P. Mahajan, "Development of Predictive Models for Early Detection of Alzheimer's Disease Using Machine Learning," *Int. J. Curr. Eng. Technol.*, vol. 15, no. 2, pp. 115–123, 2025.

[3] A. Vellido, "Societal Issues Concerning the Application of Artificial Intelligence in Medicine," *Kidney Dis.*, vol. 5, no. 1, pp. 11–17, 2019, doi: 10.1159/000492428.

[4] M. M. Ali, B. K. Paul, K. Ahmed, F. M. Bui, J. M. W. Quinn, and M. A. Moni, "Heart disease prediction using supervised machine learning algorithms: Performance analysis and comparison," *Comput. Biol. Med.*, vol. 136, p. 104672, Sep. 2021, doi: 10.1016/j.compbiomed.2021.104672.

[5] H. Gajjar, U. Kora, and V. Gondaliya, "Telemedicine with Automated ECG Analysis," *Int. J. Eng. Res. Technol.*, vol. 2, no. 10, pp. 2253–2255, 2013.

[6] S. B. Shah, "Artificial Intelligence (AI) for Brain Tumor Detection: Automating MRI Image Analysis for Enhanced Accuracy," *Int. J. Curr. Eng. Technol.*, vol. 14, no. 06, Dec. 2024, doi: 10.14741/ijcet/v.14.5.5.

[7] X. Zhu, Z. Cheng, S. Wang, X. Chen, and G. Lu, "Coronary angiography image segmentation based on PSPNet," *Comput. Methods Programs Biomed.*, vol. 200, p. 105897, Mar. 2021, doi: 10.1016/j.cmpb.2020.105897.

[8] J. D. Huang, J. Wang, E. Ramsey, G. Leavey, T. J. A. Chico, and J. Condell, "Applying Artificial Intelligence to Wearable Sensor Data to Diagnose and Predict Cardiovascular Disease: A Review," *Sensors*, vol. 22, no. 20, 2022, doi: 10.3390/s22208002.

[9] V. K. Singh, D. Pathak, and P. Gupta, "Integrating Artificial Intelligence and Machine Learning into Healthcare ERP Systems: A Framework for Oracle Cloud and Beyond," *ESP J. Eng. Technol. Adv.*, vol. 3, no. 2, pp. 171–178, 2023, doi: 10.56472/25832646/JETA-V3I6P114.

[10] W. Chen *et al.*, "Classifying Heart-Sound Signals Based on CNN Trained on MelSpectrum and Log-MelSpectrum Features," *Bioengineering*, vol. 10, no. 6, 2023, doi: 10.3390/bioengineering10060645.

[11] M. A. Mostafiz, "Machine Learning for Early Cancer Detection and Classification: AIBased Medical Imaging Analysis in Healthcare," *Int. J. Curr. Eng. Technol.*, vol. 15, no. 03, pp. 1–10, 2025.

[12] P. Oza, P. Sharma, and S. Patel, "Machine Learning Applications for Computer-Aided Medical Diagnostics," in *Lecture Notes in Networks and Systems*, 2021, pp. 377–392. doi: 10.1007/978-981-16-0733-2_26.

[13] A. Raza, A. Mehmood, S. Ullah, M. Ahmad, G. S. Choi, and B. W. On, "Heartbeat sound signal classification using deep learning," *Sensors (Switzerland)*, vol. 19, no. 21, 2019, doi: 10.3390/s19214819.

[14] B. Sahu, J. Mishra, S. Parveen, B. Das, A. Panigrahi, and A. Pati,

"STIR-MO: An Enhanced Hybrid Cancer Predictive Model," *ICoICC 2025 - 3rd Int. Conf. Intell. Cloud Comput.*, no. July, pp. 1–6, 2025, doi: 10.1109/ICoICC64033.2025.11052206.

[15] V. Khedkar, N. Vullam, J. R. Babu, U. Bhagyalatha, S. Babu Vadde, and A. Lakshmanarao, "Hybrid Classification Approach for Heart Disease using Few Shot Inspired Machine Learning Models," in *2025 3rd International Conference on Integrated Circuits and Communication Systems (ICICACS)*, 2025, pp. 1–5. doi: 10.1109/ICICACS65178.2025.10968965.

[16] C. Varshini, V. N. Sri, N. Sravya, and M. Srinivas, "Harnessing Artificial Intelligence for Heart Disease: Comparing Machine Learning, Deep Learning, and Bayesian Network Approaches," in *2025 International Conference on Artificial Intelligence and Data Engineering (AIDE)*, 2025, pp. 436–441. doi: 10.1109/AIDE64228.2025.10987434.

[17] J. V Biju, A. P. Mallya, and K. P, "Prediction of Cardiac Disease Using Quantum Enhanced Ensemble Learning Approach," in *2025 International Conference on Innovative Trends in Information Technology (ICITIIT)*, 2025, pp. 1–6. doi: 10.1109/ICITIIT64777.2025.11041018.

[18] K. Akther, M. S. R. Kohinoor, B. S. Priya, M. J. Rahaman, M. M. Rahman, and M. Shafiullah, "Multi-Faceted Approach to Cardiovascular Risk Assessment by Utilizing Predictive Machine Learning and Clinical Data in a Unified Web Platform," *IEEE Access*, vol. 12, pp. 120454–120473, 2024, doi: 10.1109/ACCESS.2024.3436020.

[19] H. F. El-Sofany, "Predicting Heart Diseases Using Machine Learning and Different Data Classification Techniques," *IEEE Access*, vol. 12, pp. 106146–106160, 2024, doi: 10.1109/ACCESS.2024.3437181.

[20] M. Maragatharajan, P. Naveen, and E. Nirmala, "Heart Disease Classification using Crow Search Algorithm," in *2024 5th International Conference on Smart Electronics and Communication (ICOSEC)*, 2024, pp. 1432–1437. doi: 10.1109/ICOSEC61587.2024.10722048.

[21] M. Mahalakshmi and R. Rajakumari, "Delving into Heart Disease Prediction: A Comprehensive Review," in *2023 2nd International Conference on Ambient Intelligence in Health Care, ICAIHC 2023*, 2023. doi: 10.1109/ICAIHC59020.2023.10430484.

[22] A. Sharma, H. Sharma, S. Varshney, and N. Gusain, "Heart Disease Prediction: A Comparative Analysis of Machine Learning Algorithms," *Comput. J. Appl. Comput. Sci. Intell. Technol.*, vol. 3, no. 2, pp. 171–192, 2024, doi: 10.17492/computology.v3i2.2309.

[23] P. M. Goad and P. J. Deore, "Predicting and Analyzing Cardiovascular Disease through Ensemble Learning Approaches," *Int. Res. J. Multidiscip. Technovation*, vol. 6, no. 5, pp. 153–163, 2024, doi: 10.54392/irjmt24510.

[24] M. Abdullah, "Artificial intelligence-based framework for early detection of heart disease using enhanced multilayer perceptron," *Front. Artif. Intell.*, vol. 7, 2024, doi: 10.3389/frai.2024.1539588.

