Published by :
https://www.ijert.org/
An International Peer-Reviewed Journal

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 06 , June - 2026

Heart Disease Prediction Using Machine Learning: Comparative Evaluation of SVM, Random Forest and ANN

Dr. Pranamika Kakati[1], Mr. Rocky Ahmed[2], Moushumi Parbin[3], Nabanita Dutta[4]

[1]Department of Computer Science, Gauhati University, Assam.

[2] Department of Computer Science, KKHSOU, Guwahati, Assam.

[3] Department of Computer Science, Gauhati University,Assam.

[4] Department of Computer Science, Gauhati University,Assam.

Abstract-Heart disease remains one of the leading causes of mortality worldwide, making early and accurate diagnosis essential for effective clinical intervention and improved patient outcomes. Machine learning techniques have emerged as promising tools for assisting healthcare professionals in predicting heart disease by analyzing patient clinical data. This study presents a comparative evaluation of three supervised machine learning algorithms: Support Vector Machine (SVM), Random Forest (RF), and Artificial Neural Network (ANN), for heart disease prediction. The Cleveland Heart Disease dataset, comprising 1025 patient records with 13 clinical attributes and one target attribute, was utilized for model development and evaluation. Data preprocessing involved exploratory data analysis, outlier detection and removal, data preparation, and train-test splitting to improve model performance. Hyperparameter tuning was performed to optimize each classifier before training. Experimental results indicate that the Random Forest classifier achieved the highest prediction accuracy of 96%, outperforming ANN and SVM, which achieved accuracies of 93.17% and 91.54%, respectively. The comparative analysis demonstrates that Random Forest provides superior predictive performance for the given dataset and highlights the potential of machine learning-based decision support systems in facilitating the early detection of heart disease.

Keywords-HeartDisease Prediction, Machine Learning, Support Vector Machine (SVM), Random Forest, Artificial Neural Network (ANN), Classification.

I. INTRODUCTION

Heart disease, also known as cardiovascular disease (CVD), is one of the leading causes of mortality worldwide. Early and accurate diagnosis is essential for timely treatment and improved patient outcomes. Conventional diagnostic methods rely on clinical examination, laboratory investigations, and medical imaging, which can be time-consuming and often require expert interpretation.

With the increasing availability of healthcare data, machine learning (ML) has emerged as an effective approach for predicting heart disease by identifying hidden patterns in clinical data. Several supervised learning algorithms, including Support Vector Machine (SVM), Random Forest (RF), and Artificial Neural Network (ANN), have been widely applied for disease prediction. However, their performance varies depending on the dataset, preprocessing techniques, and model parameters.

This study presents a comparative evaluation of SVM, Random Forest, and ANN using the Cleveland Heart Disease dataset. The dataset was preprocessed through exploratory data analysis, outlier removal, and train-test splitting before model development. The developed models were evaluated based on prediction accuracy to identify the most effective classifier for heart disease prediction. The experimental results demonstrate that the Random Forest classifier achieved the highest prediction accuracy, highlighting its effectiveness as a decision-support tool for the early detection of heart disease.

The objective of this study is to develop and compare the performance of Support Vector Machine (SVM), Random Forest (RF), and Artificial Neural Network (ANN) models for heart disease prediction using the Cleveland Heart Disease dataset, and to identify the most effective classifier for accurate and early diagnosis.

II. LITERATUREREVIEW

Heart disease prediction has become an important research area due to the increasing global mortality rate caused by cardiovascular diseases. Machine learning techniques have

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 06 , June - 2026

been widely adopted to develop automated systems for early diagnosis using clinical datasets.

Detrano et al. [6] introduced a widely used benchmark dataset for coronary artery disease prediction and proposed a probabilistic approach based on clinical attributes. Their work became a foundation for many subsequent machine learning studies in this domain.
Faieq and Mijwil [1] investigated heart disease prediction using Support Vector Machine (SVM) and Artificial Neural Network (ANN). Their results showed that both models can achieve strong classification performance depending on dataset characteristics and feature selection.

Anbuselvan [2] applied multiple machine learning algorithms such as Random Forest, SVM, and Decision Tree for heart diseaseprediction. The study concluded that Random Forest achieved the highest accuracy among all tested models.

Sharma et al. [3] performed a comparative study of different machine learning techniques and observed that ensemble-based and tree-based models generally outperform traditional classifiers in terms of accuracy and robustness.

Gudadhe et al. [7] developed a decision support system using SVM and ANN for heart disease classification. Their study showed that ANN provides competitive performance compared to SVM in medical diagnostic applications.

Overall, existing studies demonstrate that machine learning algorithms such as SVM, ANN, and Random Forest are widely used for heart disease prediction. However, most studies focus on limited model comparisons and often lack a unified experimental setup with consistent preprocessing and evaluation metrics. This makes it difficult to fairly compare model performance across different works. Therefore, there is a need for a systematic and fair comparative analysis of SVM, Random Forest, and ANN under identical experimental conditions, which motivates this study.

III. METHODOLOGY

The proposed methodology consists of four major stages: dataset acquisition, data preprocessing, model development, and performance evaluation. Initially, the Cleveland Heart Disease dataset was collected from the UCI Machine Learning Repository. The dataset was then preprocessed using exploratory data analysis (EDA) and outlier detection techniques to improve data quality before model training. Three supervised machine learning algorithms, namely Support Vector Machine (SVM), Random Forest (RF), and Artificial Neural Network (ANN), were implemented and evaluated for heart disease prediction. The overall workflow of the proposed methodology is illustrated in Fig. 1.

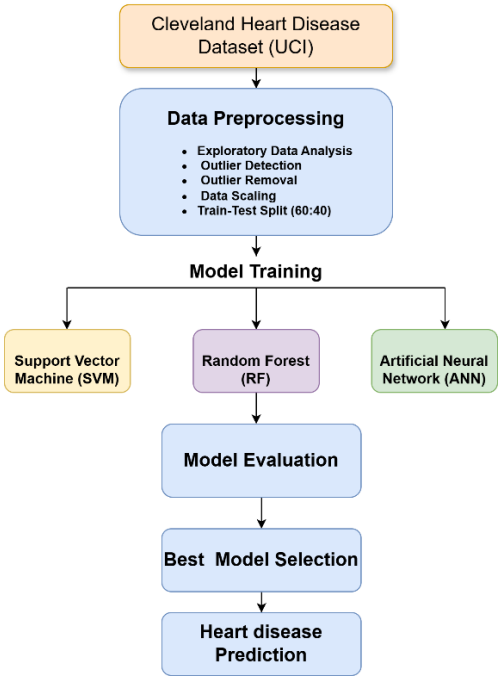

Fig. 1. Proposed methodology for heart disease prediction using SVM, Random Forest, and ANN.

A. Dataset

The Cleveland Heart Disease dataset, originally introduced by Detrano et al. [6] and obtained from the UCI Machine Learning Repository [9], was used in this study. The dataset contains 1025 patient records with 13 clinical input attributes and one target attribute indicating the presence or absence of heart disease. The input attributes include age, sex, chest pain type, resting blood pressure, serum cholesterol, fasting blood sugar, resting electrocardiographic results, maximum heart rate achieved, exercise-induced angina, oldpeak, slope of the ST segment, number of major vessels, and thal. These clinical attributes were used as input features for training and evaluating the machine learning models.

B. Data Preprocessing

Data preprocessing was performed to improve the quality of the dataset before model development. Initially, exploratory data analysis (EDA) was carried out to understand the distribution of the clinical attributes and identify potential outliers. Box plot visualization was employed for outlier detection, and the identified outliers were removed to reduce their influence on model performance. After preprocessing, the dataset was divided into training and testing subsets using a 60:40 ratio for model development and evaluation.

C. Model Development

Three supervised machine learning algorithms, namely Support Vector Machine (SVM), Random Forest (RF), and Artificial Neural Network (ANN), were developed and evaluated for heart disease classification. Different model

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 06 , June - 2026

configurations were investigated to obtain the best predictive performance.

1) Support Vector Machine (SVM)

Support Vector Machine (SVM) is a supervised learning algorithm that constructs an optimal decision boundary to separate different classes. In this study, Linear, Polynomial, Radial Basis Function (RBF), and Sigmoid kernel functions were evaluated. Among these, the Polynomial kernel produced the highest prediction accuracy and was selected for comparative analysis.

2) Random Forest (RF)

Random Forest is an ensemble learning algorithm that combines multiple decision trees to improve classification performance and reduce overfitting. Each tree is trained using a randomly selected subset of the training data, and the final prediction is obtained through majority voting. The Random Forest classifier was implemented to classify patients into heart disease and non-heart disease categories.

3) Artificial Neural Network (ANN)

Artificial Neural Network (ANN) is a computational model inspired by the biological neural network of the human brain. The ANN model employed in this study consists of an input layer, hidden layer(s), and an output layer. During training, the network learns the relationship between clinical attributes and the target class by adjusting connection weights using the backpropagation algorithm. Different ANN configurations were evaluated, and the best-performing model was selected for comparison with SVM and Random Forest.

D. Performance Evaluation

The developed machine learning models were trained using the prepared training dataset and evaluated on the testing dataset. The performance of each classifier was assessed using prediction accuracy and confusion matrix analysis. Prediction accuracy was used to measure the overall classification performance, while the confusion matrix provided detailed information regarding correctly and incorrectly classified instances. The comparative analysis of these evaluation metrics was used to identify the most effective machine learning model for heart disease prediction.

IV. RESULTS AND DISCUSSION

The performance of the developed machine learning models was evaluated using the testing dataset. Support Vector Machine (SVM), Random Forest (RF), and Artificial Neural Network (ANN) were trained using the preprocessed Cleveland Heart Disease dataset. The prediction accuracy obtained from each classifier was used to compare the performance of the developed machine learning models. The experimental results demonstrate that Random Forest achieved the highest prediction accuracy among the three models, indicating its suitability for heart disease prediction.

A. Performance of Support Vector Machine (SVM)

The Support Vector Machine classifier was implemented using four different kernel functions, namely Linear, Polynomial, Radial Basis Function (RBF), and Sigmoid. The performance of each kernel was evaluated using prediction accuracy. Among the evaluated kernels, the Polynomial kernel achieved the highest prediction accuracy of 91.54% and was therefore selected for comparison with the remaining classifiers. The comparative accuracy of different kernel functions is presented in Fig.2.

	Kernels	Accuracy
0	poly	91.54 %
1	rbf	91.06 %
2	sigmoid	76.74 %
3	linear	86.01 %

Fig. 2. Performance comparison of different SVM kernel functions

B. Performance of Random Forest (RF)

The Random Forest classifier was trained using multiple decision trees generated from randomly selected subsets of the training dataset. The ensemble learning approach enabled the model to reduce overfitting and improve classification performance. Experimental results show that the Random Forest classifier achieved the highest prediction accuracy of 96.00%, outperforming both SVM and ANN

C. Performance of Artificial Neural Network (ANN)

The Artificial Neural Network model was trained using an input layer, hidden layer(s), and an output layer with the backpropagation learning algorithm. Different network configurations were evaluated to obtain the best predictive performance. The best-performing ANN model achieved a prediction accuracy of 93.17%, demonstrating competitive classification performance. The accuracy obtained by different ANN models is shown in Fig.3.

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 06 , June - 2026

	Models	Accuracy
0	Model1	92.682927
1	Model2	89.756098
2	Model3	93.170732
3	Model4	90.243902
4	Model5	91.707317
5	Model6	91.707317

Fig. 3. Performance comparison of ANN models with different configurations

D. Comparative Analysis

A comparative evaluation of the three supervised machine learning algorithms was performed using prediction accuracy as the primary evaluation metric. The prediction accuracies obtained by the developed models are presented in Table I.

Table I. Performance comparison of machine learning models

Machine Learning Model	Accuracy (%)
Support Vector Machine (SVM)	91.54
Artificial Neural Network (ANN)	93.17
Random Forest (RF)	96.00

The results presented in Table I indicate that the Random Forest classifier achieved the highest prediction accuracy of 96.00%, outperforming Artificial Neural Network (93.17%) and Support Vector Machine (91.54%). The superior performance of Random Forest can be attributed to its ensemble learning approach, which combines multiple decision trees to improve classification accuracy and reduce overfitting. Therefore, Random Forest is identified as the most effective classifier for heart disease prediction using the Cleveland Heart Disease dataset.

E. Graphical User Interface (GUI)

A Graphical User Interface (GUI) was developed to demonstrate the practical implementation of the proposed heart disease prediction system. Based on the comparative analysis, the Random Forest classifier was selected for deployment due to its superior performance. The GUI accepts the required clinical attributes as input and predicts whether a patient is likely to have heart disease. Sample prediction results are shown in Fig. 4 and Fig. 5.

Fig. 4. Example of a person not having heart disease

Fig. 5. Example of a person having heart disease

V. CONCLUSION AND FUTURE WORK

This study investigated the application of Support Vector Machine (SVM), Random Forest (RF), and Artificial Neural Network (ANN) for heart disease prediction using the Cleveland Heart Disease dataset. The comparative evaluation demonstrates the effectiveness of machine learning techniques in supporting the early detection of heart disease and assisting clinical decision-making. Among the evaluated models, Random Forest showed the best overall performance for the given dataset.

Future Scope: Future research may focus on evaluating the proposed approach using larger and more diverse datasets, incorporating advanced feature selection and optimization techniques, and exploring deep learning methods to further improve prediction performance and real-world applicability.

REFERENCES

[1] A. K. Faieq and M. M. Mijwil, "Prediction of Heart Diseases Utilising Support Vector Machine and Artificial Neural Network," Indonesian Journal of Electrical Engineering and Computer Science, vol. 26, no. 1, pp. 374–380, 2022.

[2] P. Anbuselvan, "Heart Disease Prediction Using Machine Learning Techniques," International Journal of Engineering Research & Technology (IJERT), vol. 9, no. 11, pp. 515–518, 2020.

[3] V. Sharma, S. Yadav, and M. Gupta, "Heart Disease Prediction Using Machine Learning Techniques," in *2020 2nd International Conference on Advances in Computing, Communication Control and Networking (ICACCCN)*, pp. 177–181, IEEE, 2020.

[4] Y. Xing, J. Wang, Z. Zhao, and Y. Gao, "Combination Data Mining Methods with New Medical Data to Predicting Outcome of Coronary Heart Disease," in *Proceedings of the 2007 International Conference

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 06 , June - 2026

on Convergence Information Technology (ICCIT 2007)*, pp. 868–872, IEEE, 2007.

[5] J. Chen, G. Xi, Y. Xing, J. Chen, and J. Wang, "Predicting Syndrome by NEI Specifications: A Comparison of Five Data Mining Algorithms in Coronary Heart Disease," in International Conference on Life System Modeling and Simulation (LSMS 2007), Lecture Notes in Computer Science, pp. 129–135, Springer, 2007.

[6] R. Detrano, A. Janosi, W. Steinbrunn, M. Pfisterer, J. J. Schmid, S. Sandhu, K. H. Guppy, S. Lee, and V. Froelicher, "International Application of a New Probability Algorithm for the Diagnosis of Coronary Artery Disease," The American Journal of Cardiology, vol. 64, no. 5, pp. 304–310, 1989.

[7] M. Gudadhe, K. Wankhade, and S. Dongre, "Decision Support System for Heart Disease Based on Support Vector Machine and Artificial Neural Network," in *2010 International Conference on Computer and Communication Technology (ICCCT)*, pp. 741–745, IEEE, 2010.

[8] M. A. M. Abushariah, A. A. M. Alqudah, O. Y. Adwan, and R. M. M. Yousef, "Automatic Heart Disease Diagnosis System Based on Artificial Neural Network (ANN) and Adaptive Neuro-Fuzzy Inference Systems (ANFIS) Approaches," Journal of Software Engineering and Applications, vol. 7, no. 12, pp. 1055–1064, 2014.

[9] D. Dua and C. Graff, "UCI Machine Learning Repository," University of California, Irvine, School of Information and Computer Sciences, 2019. [Online]. Available: https://archive.ics.uci.edu/ml