IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.471 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 14, Issue 12, December 2025

DOI: 10.17148/IJARCCE.2025.141236

The Impact of Outlier Management on Machine Learning Algorithms and Deep Learning Algorithms Performance for Heart Disease Prediction

Er. Harjasdeep Singh[1], Udey Partap Singh[2], Sushil[3], Yudhveer[4]

Assistant Professor, Department of CSE, MIMIT Malout[1]

Student, Department of CSE, MIMIT Malout[2]

Student, Department of CSE, MIMIT Malout[3]

Student, Department of CSE, MIMIT Malout[4]

Abstract: Heart disease is one of the most common and serious health problems in the world today. Predicting it early can help save lives by allowing people to get the right treatment on time. In this study, we compare how well **machine learning** and **deep learning** models can predict heart disease using patient data such as age, blood pressure, cholesterol level, and other health factors. Several popular machine learning algorithms like **Logistic Regression, Decision Tree, Random Forest,** and **SVM** are tested, along with deep learning models such as **ANN, CNN,** and **RNN**. Each model's performance is measured using accuracy, precision, recall, F1-score, and ROC-AUC. Our findings show that deep learning models generally perform better in terms of accuracy and can capture complex patterns in the data more effectively. However, traditional machine learning models are easier to understand and require less computational power. Overall, this comparison helps highlight the strengths and limitations of both approaches and can guide future work in building better heart disease prediction systems.

Keywords: Heart disease prediction, Machine learning, Deep Learning, SVM, Logistic regression, Decision tree, Random Forest, ANN, CNN, RNN, Outliers, comparative analysis.

I. INTRODUCTION

Cardiovascular diseases (CVDs), particularly heart disease, remain the leading cause of mortality worldwide, responsible for millions of deaths annually across all continents. According to global health statistics, nearly 17.9 million deaths occur each year due to CVDs, accounting for 31% of all deaths globally [1]

The burden of these diseases is not limited to developed countries; developing regions face even higher risks due to lifestyle changes, limited access to healthcare, and delayed diagnoses. Early detection and prediction of heart disease is therefore critical to reducing mortality rates and enabling preventive healthcare interventions. Traditionally, heart disease diagnosis relies on clinical evaluations, imaging techniques, and invasive tests such as angiography. However, these approaches are often costly, time-consuming, and subject to human error. Additionally, symptoms of heart disease can vary widely across patients—sometimes appearing atypical depending on age, gender, and commodities. This complexity has driven researchers to explore advanced computational methods such as Machine Learning (ML) and Deep Learning (DL) for more accurate, automated, and scalable prediction systems Such a system can be seen as a classification task as the goal is to make a prediction (i.e., diagnosis) on a new case based on the available records and features (of previously known cases). Such classification tasks are considered to be one of the most challenging tasks in medical informatics [2]. Machine learning (ML) is a set of mathematical algorithms, a subset of the wider field of artificial intelligence (AI) algorithms, which offer the potential to provide innovative decision support solutions to many problems in a big data environment therefore offering advances beyond the rule based engines that proliferate in many fields of science. Explainable AI is therefore a hot topic and significant in all heavily regulated industries. The US DARPA has invested significantly in research in this field while other research teams have begun to define the concept of explainable AI with respect to several problem domains [3]

A basic comparison between classic ML models (SVM) and DL models (long short term memory [LSTM]/gated recurrent unit [GRU]-RNN) was investigated. However, human hand-crafted features were used in that study, which cannot give

International Journal of Advanced Research in Computer and Communication Engineering
Impact Factor 8.471 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 14, Issue 12, December 2025
DOI: 10.17148/IJARCCE.2025.141236

the audience a view of using state-of-the-art techniques that can learn high-level features automatically from the heart sound via DL. In fact, we can see some recently published literature giving some encouraging results showing the trend on learning heart sound features in an unsupervised learning paradigm. However, a comprehensive study on the state-of-art representation learning paradigms on heart sound classification task is lacking. To this end, we introduce this work includes transfer learning, sequence-to-sequence learning and end-to-end learning approaches for the heart sound classification. For the reason that the relationship of the models usually is not clear, we utilize the Shapely values to evaluate the global features contributions. To the best of our knowledge, this is the first time to present the comprehensive investigation on heart sound classification task.[5]

Malavika G. et al. investigated the use of ML algorithms to predict heart disease. The heart disease data set from the UCI repository was used for this study. They used various ML algorithms, including LR, KNN, SVM, NB, DT, and RF, to predict heart disease, and their performances were compared. The results showed that RF (91.80%) had the highest accuracy in predicting heart disease, followed by NB (88.52%) and SVM (88.52%). The authors concluded that ML algorithms could be a useful tool in predicting heart disease and could potentially help doctors diagnose and treat patients more accurately . Sahoo G. K. et al. compared the performance of LR, KNN, SVM, NB, DT, RF, and XG Boost Machine Learning models for predicting heart disease. The Cleveland heart disease data set from the UCI ML repository was used to train the models. Comparing the results of the tested ML algorithms, the RF algorithm performed the best, with a classification accuracy of 90.16%. The exploration of various ML techniques for predicting coronary artery disease is addressed in . The study used a data set of 462 medical instances, and nine features from the South African heart disease data set. It consists of 302 healthy records and 160 records with coronary heart disease. In this study, the k-means algorithm, along with the synthetic minority oversampling technique, were used to solve the problem of imbalanced data. A comparative analysis of four different ML techniques, such as LR, SVM, KNN, and artificial neural network (ANN), can accurately predict coronary artery disease events from clinical data. The results showed that SVM had the highest accuracy performance (78.1%) .In Ahmad G. N. et al.'s study, Cleveland, Hungarian, Switzerland, Stat log, and Long Beach VA datasets were combined to obtain a larger data set compared to existing heart disease datasets. They compared the performances of LR, KNN, SVM, Nu-Support Vector Classifier (Nu-SVC), DT, RF, NB, ANN, algorithms for heart disease classification. In this study, the authors claimed that the best classification accuracy of 100% was achieved with the RF algorithm .The main objective of this study is to use the meta-heuristic method, such as the Jellyfish algorithm, to select the optimum features from the heart disease data set and use it in the Machine Learning method to classify the healthy and non-healthy heart disease data. Some of the features do not have more efficiency in the classification of heart disease. The Jellyfish has some advantages, such as the high speed of convergency, and high accuracy to find the features. For this reason, this algorithm has been selected[6].

Coronary Heart disease is the major factor for the people's deaths throughout the world, and it is necessary to detect and predict the disease in the earlier stages because time plays a vital role to save the coronary patient, which we call a "Golden hour" is very much necessary we need to get some more advanced technologies to predict the heart disease so that we can save the patients as early as possible. Therefore we can reduce the death rate treatment at the right time. From this paper, we can conclude that we have used most of the machine learning and deep learning ensemble algorithms so that it can predict heart disease at an early stage so that the patient's life can be saved[8,9,10].

This survey aims to explore, summarize, and critically analyze the most recent and state of the art research papers in order to find research gaps for future studies. This research has been conducted systematically, to help readers gain the knowledge of previous researches conducted in the domain of heart failure and risk detection. The limitations and future work provided in Table 2 can assist researchers in fulfilling the need for future research and gap in research. Whereas, the discussion section can help them direct[11].

This article summarizes 64 major studies that use DL, ETDL, and integrated DL methods to predict HDs. Through rigorous analysis of these studies, CNN emerges as a prominent DL technique, while hybrid DL-based ETDL becomes the leading ETDL method, and hybrid DL is the most commonly used integrated DL method. Moreover, there has been an increasing amount of research conducted in recent years using a combination of DL and other technologies. In addition, we discuss various existing datasets, among which the UCI Cleveland heart disease data set is used by 62% of researchers and is the most widely used data set . It is noteworthy that Python is the preferred programming language for implementing these technologies. The main research comes mainly from journals and is published by well-known publishers such as IEEE, Springer, Elsevier, etc. Although some progress has been made in this field, researchers still face numerous challenges, one of the most significant of which is the lack of larger and more diverse datasets. This severely limits the use of DL technology to improve the accuracy and reliability of HDP[12].

International Journal of Advanced Research in Computer and Communication Engineering
Impact Factor 8.471 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 14, Issue 12, December 2025
DOI: 10.17148/IJARCCE.2025.141236

Cardiovascular disease, commonly referred to as heart disease, is a major concern in the medical field. According to recent estimates from the World Health Organization (WHO), over 20.5 million people are dying from cardiovascular disease, accounting for 31.5% of all global deaths. It is also projected that by 2030, the annual death toll will rise to 24.2 million. This paper critically reviews and summarizes the research works from 2014 to 2024 on predicting heart disease risk using machine learning and deep learning algorithms. It also highlights the feature selection techniques applied in previous studies and the key features identified to enhance heart disease risk prediction. Additionally, this work explores the hyperparameter tuning methods used in state-of-the-art studies to boost the performance of machine learning models. The findings indicate that SVM and RF techniques are the most commonly used, offering better accuracy in heart disease prediction. Models based on NB, KNN, and ANN also performed well in most cases. However, the accuracy of these models varies depending on factors such as the tool/software used, dataset size, feature set, number of instances, data preprocessing, feature selection methods, and the choice of classifier. There remains significant research potential in addressing issues like missing data, outliers, and overfitting. Current works often lack hyperparameter optimization, which is crucial for enhancing machine learning performance but is complex and time-consuming. It is recommended to explore new feature selection methods for heart disease datasets and perform hyperparameter tuning to improve prediction accuracy heart disease risk [13].

In[14,15,16],This work offers a thorough assessment of machine learning models for predicting cardiac disease, utilizing a consolidated dataset from many sources. We employed fifteen distinct models, including XG Boost, Random For est, Bagged Trees, and conventional classifiers such as Logistic Regression and Naive Bayes, by picking seven - essential features. XG Boost and Bagged Trees attained the highest accuracy of 93%, closely followed by Random Forest and KNN at 91%. To guarantee model robustness, we utilized k-fold cross-validation (K=10, K=5), where in Random Forest showed enhanced stability (94% for K=10, 92% for K=5), while XG Boost displayed a marginal decrease. The ROC-AUC scores showed that the models were even more reliable. Random Forest and Bagged Trees got 95%, XG Boost got 94%, and GBM got 92%. Furthermore, tests such as precision, recall,and F1-score proved that these models worked, with Bagged Trees and Random Forest showing a strong balance between sensitivity and specificity.

In [18,19] ,The effectiveness of machine learning (ML) algorithms for predicting coronary heart disease (CHD) is examined in this study. CHD prediction was performed using both test and training datasets, with 30% of the data allocated for model evaluation and 70% for training. Using the Mutual Information feature selection technique significantly enhanced the models' prediction accuracy by reducing irrelevant input variables. One way that the class imbalance in the NHANES data set was fixed was by using the synthetic minority oversampling technique (SMOTE). This helped the classification models work better. By accurately predicting the likelihood of CHD, this study has the potential to aid doctors and hospitals in initiating timely treatment, ultimately reducing the risk of fatalities associated with misdiagnosis or delayed intervention. The study used MATLAB to run the machine learning techniques. The results show that the proposed model, which combines feature selection and SMOTE for class imbalance, was more accurate than traditional machine learning techniques and the convolutional neural network (CNN) model. Cardiovascular diseases claim thousands of lives each year, with many of these deaths attributed to misdiagnosis or underdiagnosis of heart conditions. This study uses selected features and well-known ML methods to predict CHD from real-world datasets. The results demonstrate that the proposed model significantly improves the accuracy of CHD prediction, offering more reliable forecasts than other traditional approaches. Despite the promising results and potential uses of the proposed ML-based PSO-ANN for coronary heart disease prediction, there are several limitations to consider. The performance and reliability of ML models depend on the quality and availability of testing and training datasets, and we employed CHD and private datasets in our study, which may face limitations in availability, representativeness, and data quality.

SMOTE generates synthetic minority class samples to overcome class imbalance, but its effectiveness depends on the dataset and situation, and class imbalance can cause models to perform well for the majority group but poorly for the minority class, which is often the main concern. To address this issue, it's essential to discuss and compare different approaches for handling class imbalances along with SMOTE, such as ensemble methods or algorithm selection. Future research will encompass the examination of diverse optimizers for artificial neural networks (ANN) alongside implementing advanced deep learning methodologies. Models like recurrent neural networks (RNN), long short-term memory (LSTM), and other deep architectures will help us better predict coronary heart disease and make the diagnosis more reliable. This investigation may facilitate more thorough data analysis and enhanced model efficacy, hence enabling sophisticated applications in clinical environments.

II. LITERATURE REVIEW

The current study selected twenty-four articles concerning machine learning algorithms for predicting and diagnosing heart diseases. The vast majority of scholars applied the DT classification algorithm. Since the DT is the most

DOI: 10.17148/IJARCCE.2025.141236

compressive approach among all algorithms of machine learning, it reflects essential features in the data set. In heart disease, some parameters affect the patient, such as blood pressure, blood sugar, age, sex, genetics, and other factors. By observing the decision tree, physicians can determine which parameter affects the outcome. In addition, they can identify which population group is most impacted.The key benefit of applying Decision Trees is the ability to examine information in a class-specific manner. A root-to-leaf traversal implies a special class division based on full knowledge that the decision-maker should consider. Such an algorithm is usually beneficial because it informs everyone about the data's general nature in terms of variables. It supports quantitative and qualitative data and works well with massive datasets with low dimensionality. Moreover, it efficiently handles heavily skewed data without data transformation and is robust to outliers [1].

The methods used were SVM, multi-layer perceptron neural networks and decision trees. We highlight the importance of applying ten fold cross validation to compare the uncertainty of the trained models generated from two datasets with different characteristics. One data set, from the UCI repository, consists of timing data extracted from the PQRST features in ECG wave forms. This has a relatively large number of features per record but a small number (approximately 400) of records and is also substantially imbalanced between the classification categories. The second data set is from the Kaggle repository and has a small number of features per record but nearly seventy thousand records. This data set records features such as the presence or absence of hypertension, cholesterol levels etc. Both datasets record age and gender among their features[3].

In this study, as of now, we have compared five machine learning algorithms such as Decision Tree Classifier with Grid search, DT, RF, LR, and KNN. Precision, Recall, F1-score, confusion matrix, ROC curve, and accuracy are the different evaluation metrics calculated here.

Here, the Decision Tree classifier got more accuracy, and it is 92%. In this study, we developed a novel ensemble model which is a voting method that combines four different algorithms such as Naïve Bayes, Deep learning, Generalized Linear Model, and Random Forest. For this model, we got an accuracy of 85.44%. Figure 5 depicts the same[4].

In this study, we segmented the audio recordings in HSS into 10-s-based clips, which means an accurate prediction of heart status is needed from a shorter duration (around 30 s in HSS) of the audio recording. In addition, we added a binary classification task (normal/abnormal detection) as a subtask in this work. Both the classic ML and the cuttingedge DL methods were investigated and compared by using our open-source toolkits, which can be easily reproduced. In this benchmark study, the best result for the 3-class classification task was 48.8% of UAR (chance level: 33.3%). The best result for the binary classification task was 58.7% of UAR (chance level: 50.0%) [5].

The Jellyfish algorithm is a swarm-based meta heuristic algorithm that can be used with ML methods to optimize hyperparameters. The optimum features obtained from the dataset were used in the training and testing stages of four different ML algorithms (ANN, DT, AdaBoost,*Diagnostics* **2023**, *13*, 2392 16 of 17 and SVM). Then, the performances of the obtained models were compared. The results show that the accuracy rates of all ML models improved after the dataset was subjected to feature selection with the Jellyfish algorithm. The highest classification accuracy (98.47%) was obtained with the SVM model trained using the dataset optimized with the Jellyfish algorithm. The Sensitivity, Specificity, Accuracy, and AUC for SVM without using the Jellyfish algorithm were obtained at 98.21%, 97.96%, 98.09%, and 90.21%, respectively. However, by using the Jellyfish algorithm, these values have been obtained as 98.56%, 98.37%, 98.47%, and 94.48%, respectively[6].

On the test data set, the fine-tuned XG Boost model upheld its exceptional performance by achieving a precision score of 99.14% and 90.00%, signifying its adeptness in accurately categorizing positive cases. Moreover, the recall score, at 98.29% and 84.38%, holds particular significance. The F1 Score exhibits resilience at 98.71% and 87.10%. The model's overall accuracy on the test data hovers at 98.50% and 86.89%. These remarkable outcomes underscore the XG Boost model's aptness for heart disease classification[7,8,9].

This review encompasses a wide range of ML applications for predicting heart disease, organized into five main themes: detection and diagnostics, ML models and algorithms, feature engineering and optimization, new technologies in healthcare, and cross-disease AI applications. The findings indicate that while DL models, especially hybrid CNN-LSTM architectures, tend to surpass traditional methods, the success of any model heavily relies on high-quality data, effective feature engineering, and clinical interpretability [10,11].

This article summarizes 64 major studies that use DL, ETDL, and integrated DL methods to predict HDs. Through rigorous analysis of these studies, CNN emerges as a prominent DL technique, while hybrid DL-based ETDL becomes

International Journal of Advanced Research in Computer and Communication Engineering
Impact Factor 8.471 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 14, Issue 12, December 2025
DOI: 10.17148/IJARCCE.2025.141236

the leading ETDL method, and hybrid DL is the most commonly used integrated DL method. Moreover, there has been an increasing amount of research conducted in recent years using a combination of DL and other technologies. In addition, we discuss various existing datasets, among which the UCI's Cleveland heart disease dataset is used by 62% of researchers and is the most widely used dataset[12].

Cardiovascular disease, commonly referred to as heart disease, is a major concern in the medical field. According to recent estimates from the World Health Organization (WHO), over 20.5 million people are dying from cardiovascular disease, accounting for 31.5% of all global deaths. It is also projected that by 2030, the annual death toll will rise to 24.2 million. This paper critically reviews and summarizes the research works from 2014 to 2024 on predicting heart disease risk using machine learning and deep learning algorithms.The findings indicate that SVM and RF techniques are the most commonly used, offering better accuracy in heart disease prediction. Models based on NB, KNN, and ANN also performed well in most cases[13,14,15].

This paper undertook an exhaustive investigation of ML approaches to accurately predict heart disease, a challenge in binary classification that utilized tabular datasets. By conducting an extensive examination, this study scrutinized a variety of preprocessing methods that sought to enhance the performance of the models by optimizing the data. Additionally, it utilized a range of ML models to determine their effectiveness in detecting heart disease. Evaluation metrics, which are fundamental for validating the performance of models, were meticulously implemented in order to guarantee a strong assessment of the predictive capability of each model. The results of our study emphasize the considerable capacity of ML to improve the prediction of heart disease[16,17].

The effectiveness of machine learning (ML) algorithms for predicting coronary heart disease (CHD) is examined in this study. CHD prediction was performed using both test and training datasets, with 30% of the data allocated for model evaluation and 70% for training. Using the Mutual Information feature selection technique significantly enhanced the models' prediction accuracy by reducing irrelevant input variables. One way that the class imbalance in the NHANES dataset was fixed was by using the synthetic minority oversampling technique (SMOTE). This helped the classification models work better[19].

III. METHODOLOGY

This section of the paper presents the mathematical details of the machine learning kernels that we have used to produce the results that we report in the next section. First we present the salient details for each of the classification algorithms used focusing in particular on the definition of the tunable hyper-parameters which are available for each.

This section discusses the structure of the Model. Data collection is the first step. There are different datasets available online or we can collect real-time datasets from hospitals. In this proposed methodology we have used Kaggle dataset which contains 303 records of patients with 14 attributes including the result in table 1. Then data preprocessing is needed. Since this was an online dataset, there were no missed values or redundant values. Split the dataset into two as Training set and Testing set. Choose the best correlated feature using correlation matrix. Train the model with the training set. Now the model has been trained with the labels and tested the model with Testing set. Performance of the model will be evaluated. Different evaluation metrics used here are Precision, Recall, F1-score, Accuracy, Confusion matrix and ROC curve. Then compared the accuracy of different models.The framework of the proposed model. These are briefly discussed in the following subsections.

Data preprocessing

Data preparation is the first step in creating a ML model, signaling the beginning of the process. Real-world data is frequently insufficient, inconsistent, inaccurate, and lacks essential attribute values. Data preparation involves cleaning, organizing, and formatting raw data to be suitable for ML models. In this study, we have employed a variety of data preparation approaches and MI Score of the features which are presented in respectively.

Categorical data encoding

Category data encoding is the process of converting categorical variables into numerical variables to make them usable in ML models. Categorical data comprises variables grouped into different categories, such colors, locations, or types of items. Given that the majority of ML models rely on mathematical equations, it is essential to transform categorical data into numerical data to prevent any issues. We have converted the category data in the datasets into numerical values. We utilized the LabelEncoder() function from the sklearn package for this task.

Handling class imbalance

SMOTE has been employed as a means to rectify the class imbalance. SMOTE, which stands for Synthetic Minority Oversampling Technique, is a technique implemented in the domain of supervised learning to rectify class imbalances in

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.471 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 14, Issue 12, December 2025

DOI: 10.17148/IJARCCE.2025.141236

datasets. As opposed to producing duplicates, it generates synthetic samples from the minority class, thereby circumventing the overfitting issue associated with random oversampling. SMOTE generates a more diverse set of examples compared to a straightforward duplication of minority class samples by generating synthetic samples as a linear combination of the original samples and their neighbors. We obtained a balanced dataset with an equal number of samples in each class after implementing SMOTE. Balancing the dataset using SMOTE was essential for this study because imbalanced datasets can lead to biased model performance, where the model favors the majority class. This bias often results in poor recall for the minority class, which is particularly problematic in heart disease detection as it may lead to misclassification of patients at risk. By creating a balanced dataset, we ensured that the models had an equal opportunity to learn patterns from both classes, leading to fairer and more reliable performance metrics.

Feature correlation analysis

In which find the correlation between each dependent and independent element in

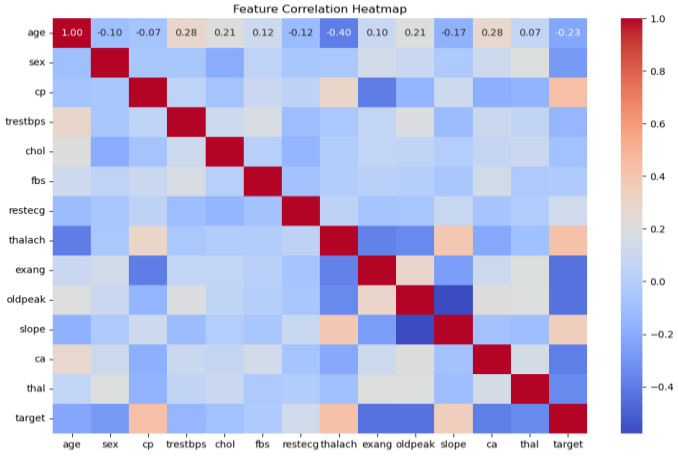

Fig 1 correlation heatmap

Feature importance analysis

Feature Importance Calculation: When constructing predictive models, it is critical to comprehend the significance of every feature in relation to the target variable. An efficacious approach to assess this level of importance is by employing the computation of Mutual Information (MI) scores. In contrast to more straightforward linear metrics, MI offers a broader measure that can encompass any type of relationship between variables, whether it be nonlinear or linear. If two variables are independent, then the score is zero. Conversely, a larger score signifies an enhanced interdependence or correlation among the variables. The MI score of a feature with respect to the target variable indicates, in the context of feature selection, the degree to which knowledge of the feature reduces uncertainty regarding the target. The feature importance graph is presented.

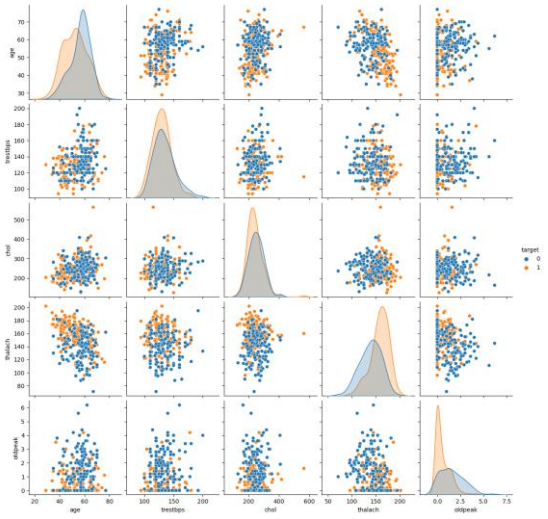

Fig 2: Pair plot

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.471 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 14, Issue 12, December 2025

DOI: 10.17148/IJARCCE.2025.141236

Feature scaling

An integral part of our research was preparing the data, with a special emphasis on feature scaling. Feature scaling is essential in ML models to standardize the range of independent variables, which aids algorithms in converging faster and achieving higher performance. Scaling the characteristics of our dataset assures that no one feature has a dominant influence because of its scale, considering the dataset's heterogeneous nature. We tested four distinct feature scaling methods in comparison to a baseline model without any feature scaling to assess their influence on the efficacy of ML models in predicting heart disease. The Standard Scaler method was used to standardize the feature distribution by removing the mean and dividing by the standard deviation resulting in features centered around zero and with a standard deviation of one. This approach is very efficient when the features are normally distributed, although it may be used with any distribution. Standard scaling helps stabilize the convergence of gradient descent algorithms by ensuring all features are on a comparable scale.

It can be represented as:

Min-Max Scaling was used to standardize the features to a certain range, usually [0, 1]. We standardized each characteristic by deleting its minimum value and dividing by the range to guarantee equal contribution to the final forecast. This method is beneficial for situations where the parameters must fall inside a certain range and is commonly employed when the distribution is non-Gaussian.

Table 1: Attributes of Dataset

S.no	Attribute Name	Description	Type
1.	age	Patient's age	int64
2.	sex	Sex of the patient	int64
3.	exng	Angina induced by Exercise (1 = yes; 0 = no)	int64
4.	caa	Count of major vessels	int64
5.	cp	Chest pain type	int64
6.	trtps	Resting blood pressure (in mm Hg)	int64
7.	chol	Cholesterol in mg/dl fetched via BMI sensor	int64
8.	fbs	FBS	int64
9.	restcg	Resting electro-cardiographic results	int64
10.	thalachh	Most heart rate attained	int64
11.	oldpeak	Exercise induced ST -segment depression relative to restfulness	Float64
12.	thall	Thalium Stress results	int64
13.	sip	The slope of the peak exercise-ST segments	int64
14.	output	0 = less chance of heart attack, 1= more chance of heart attack	int64

Machine learning algorithms

Following data preparation, various machine learning algorithms were applied to evaluate the performance of different classification techniques in predicting CHD. We worked with many models, including KNN, Support Vector Machine (SVM), LR, RF, and Naïve Bayes (NB). Each algorithm's performance was refined through hyperparameter optimization approaches.

K-nearest neighbors (KNN)

KNN is a non-parametric, supervised ML algorithm that classifies a data point based on the majority class of its k-nearest neighbors. The algorithm was tuned by optimizing the value of k, which controls the number of neighbors considered during classification. Cross-validation was employed to select the optimal k, minimizing classification error.

Support vector machine (SVM)

SVM constructs a hyperplane that best separates the data into two classes. This study used a radial basis function (RBF) kernel to allow for non-linear classification. The hyperparameters C (regularization) and γ (kernel coefficient) were optimized using a grid search to maximize accuracy.

Logistic regression (LR)

LR is a simple yet effective method for binary classification. This study applied regularized logistic regression to prevent overfitting, with $L2$ regularization used to penalize large coefficient values. The model was optimized using a stochastic gradient descent approach.

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.471 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 14, Issue 12, December 2025

DOI: 10.17148/IJARCCE.2025.141236

Random forest (RF)

RF is an ensemble method that constructs multiple decision trees and aggregates their results to improve accuracy. To mitigate overfitting, the number of trees in the forest (*n_estimators*) and the maximum depth of each tree was optimized. The model was further enhanced by implementing bootstrap sampling to increase its robustness.

Naïve Bayes (NB)

The Naïve Bayes classifier assumes conditional independence between features and applies Bayes' theorem to predict a class's probability. This approach is efficient for high-dimensional data and was implemented with Gaussian assumptions for continuous features.

Evaluation metrics

A variety of metrics were utilized to assess the performance of the models, with the primary ones being accuracy, recall, precision, and F1-score. The accuracy metric quantifies the proportion of correct predictions relative to the total number of predictions. The metric of accuracy quantifies the proportion of correct positive predictions relative to the overall number of positive predictions. Recall can be conceptualized as the ratio of precise positive forecasts to the sum of precise positive forecasts and erroneous negative predictions. The harmonic means of recall and precision constitutes the F1-score. The Area Under the Curve (AUC) is a metric utilized to assess the binary classification model's performance. As the AUC increases, so does the model's ability to distinguish between positive and negative instances. It is a prevalent metric utilized in ML to evaluate the performance of classification algorithms.

Confusion matrix:

A confusion matrix is a tabular format that evaluates the effectiveness of a classification model. It juxtaposes the anticipated classifications with the actual classifications, yielding insights about the model's performance efficacy. The matrix comprises four fundamental elements: True Positives (TP), True Negatives (TN), False Positives (FP), and False Negatives (FN).

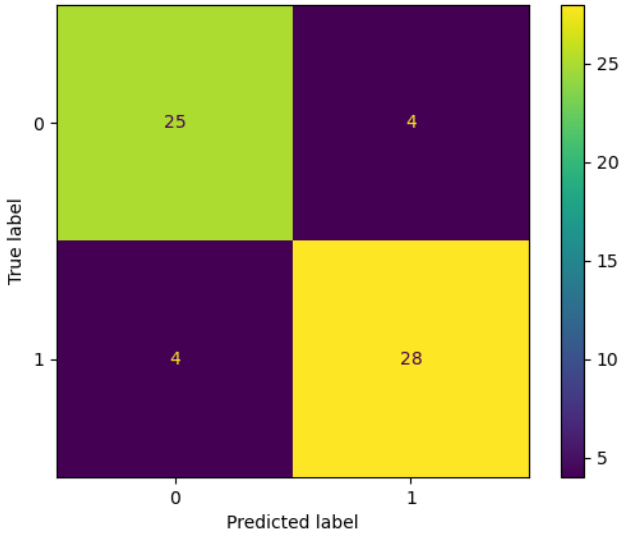

Fig 3: confusion metrix

Accuracy:

Accuracy is the fraction of accurately predicted cases (including true positives and true negatives) among all instances.

$$Accuracy = \frac{TP+TN}{TP + TN + FP + FN}$$

Precision:

Precision is also known as positivity. The predictive value is the proportion of real positive forecasts among all positive ones.

$$Precision = \frac{TP}{TP + FP}$$

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.471 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 14, Issue 12, December 2025

DOI: 10.17148/IJARCCE.2025.141236

Recall:
Recall, also known as responsiveness or true positive rate, is a metric that indicates the proportion of real positive forecasts across all positives.

$$Recall = \frac{TP}{TP+FN}$$

F1Score:
The score of F1 is the harmonic average of precision and recall, resulting in a balance of the two measures.

$$F1\ score = 2X\frac{Precision\ X\ Recall}{Precision\ X\ Recall}$$

ROC-AUC (Receiver Operating Characteristic– Area Under the curve)
The Recall (TPR) and False Positive Rate (FPR) are contrasted on the ROC curve. The classifier's performance is summarised at every level of classification by the AUC (Area Under the Curve).

$$\text{AUC} = \int_0^1 T\ P\ Rd\ (FPR)$$

Deep learning models
Deep learning models, which are subsets of artificial neural networks, have revolutionized modern artificial intelligence applications.

The **Artificial Neural Network (ANN)** is the most fundamental deep learning architecture, inspired by the structure and functioning of the human brain. It consists of interconnected processing nodes called neurons, arranged in input, hidden, and output layers. Each neuron computes a weighted sum of its inputs followed by an activation function, mathematically represented as

$$y = f\left(\sum_{i=1}^{n} w_i x_i + b\right)$$

Where xi denotes the input features, wi the corresponding weights, b the bias, and f the activation function such as sigmoid , tanh, or ReLU. ANNs are typically trained using backpropagation and gradient descent to minimize a defined loss function.

For image and spatial data, the **Convolutional Neural Network (CNN)** extends ANN architecture by introducing convolutional and pooling layers that automatically extract hierarchical features. The core operation is the convolution, defined as

$$s(i,j) = (j*k)(i.j) = \sum_m \sum_n I(i+m,j+n).k(m,n)$$

where I represents the input image, K is the kernel (filter), and S(i,j) is the output feature map. Pooling operations, such as max pooling, further reduce spatial dimensions while retaining the most relevant information. CNNs are widely used in computer vision tasks including object recognition, image classification, and medical image analysis.

In contrast, the **Recurrent Neural Network (RNN)** is designed to model sequential or temporal data, where the current output depends on both present and past inputs. The RNN maintains a hidden state ht that evolves over time as

$$h_t = f(w_{xh}x_t + w_{hh}h_{t-1} + b_n),\ y_t = g(w_{hh}h_t + b_y)$$

where x_t is the input at time step t, W_{xh}, W_{hh}, and W_{hy} are weight matrices, and f, g denote activation functions. This recurrent connection allows RNNs to retain contextual information across sequences, making them suitable for natural language processing, speech recognition, and time-series prediction. Overall, ANN, CNN, and RNN architectures

International Journal of Advanced Research in Computer and Communication Engineering
Impact Factor 8.471 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 14, Issue 12, December 2025
DOI: 10.17148/IJARCCE.2025.141236

collectively form the foundation of deep learning, enabling machines to learn complex, non-linear representations from data across diverse domains.

IV. RESULTS

To assess the performance of various ML algorithms, we examine significant research that has documented metrics, including accuracy, sensitivity, specificity, area under the curve (AUC), and computational complexity. Table 1 provides a comparative analysis of ML models for predicting heart disease. The critical insights include conventional ML models, such as LR and SVM, which demonstrate satisfactory performance but exhibit reduced accuracy when juxtaposed with DL models. Ensemble methods such as RF and KNN demonstrate enhanced performance owing to their capacity to identify intricate data patterns. DL methodologies (ANN, RNN, CNN,) yield superior accuracy but demand significant computational resources. FL methodologies ensure optimal performance while tackling privacy issues in practical implementations.

Before removing outlier:
Calculate the accuracy of the **ML models** with the outliers and compare the models and also calculate the AUC, Precision, Recall, F1-Score etc.

Logistic Regression (LR) Results
The recall before removing outliers scores stood at 91.7% and 89.38%, emphasizing the model's importance in correctly identifying all positive cases, particularly in scenarios where missing potential cases of heart disease is a critical concern. The F1 Score captured genuine positive cases at 88.14% and 89.52%. Regarding overall accuracy, the model achieved an accuracy score of 88.52% and the precission is 89%,
AUC-Score is 88.41%.

Decision tree result
In this we analyse the accuracy of the decision tree the recall is stood 75% and 90%. The F1 score is 81% and 83%, precision is 89% and 76%, AUC-score is 82.32%. Regarding overall accuracy, the model achieved an accuracy score of 81.96%

Random Forest result
Similarly, Evaluate the performance of the model with the no. Of random forest the recall is stood 83% and 88%. The F1 score is 86% and 84%, precision is 86% and 85%, AUC-score is 85.12%. Regarding overall accuracy, the model achieved an accuracy score of 85.24%.

K-Nearest Neighbors (KNN) Result
We commenced the analysis by employing the K-Nearest Neighbors (KNN) algorithm with varying 'k' values, representing the number of nearest neighbors considered during the predictions. Employing cross-validation, we computed scores for each 'k' value, ultimately discerning that 'k = 7' yielded the most favorable mean cross-validation score. This outcome underscores that configuring KNN with 'k = 7' exhibits significant promise. The recall is stood 75% and 62%. The F1 score is 72% and 65%, precision is 69% and 69%, F1-score is 86% and 84%, AUC-score is 68.53%. Regarding overall accuracy, the model achieved an accuracy score of 68.85%.

Support Vector Machine (SVM) Results
As the model achieved a precision score of 86% and 88%, a recall score of 86% and 88%, and an F1 Score of 86% and 88%, AUC-Score 86.85%. On the test dataset, the model exhibited an accuracy of approximately 86.88% an, affirming its consistent and accurate predictive capabilities.

Here, The comparison table of the ML models:

Model	Accuracy (%)	Precision (%)	Recall (%)	F1-score (%)	AUC-score (%)
LR	**88.52**	89	91	89	88.41
DT	81.96	89	75	81	82.32
RF	85.24	85	88	86	85.12
KNN	68.85	69	75	72	68.53
SVM	**86.88**	88	88	88	86.85

Table 2: ML model performance comparison

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.471 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 14, Issue 12, December 2025

DOI: 10.17148/IJARCCE.2025.141236

Here the comparison bar graph

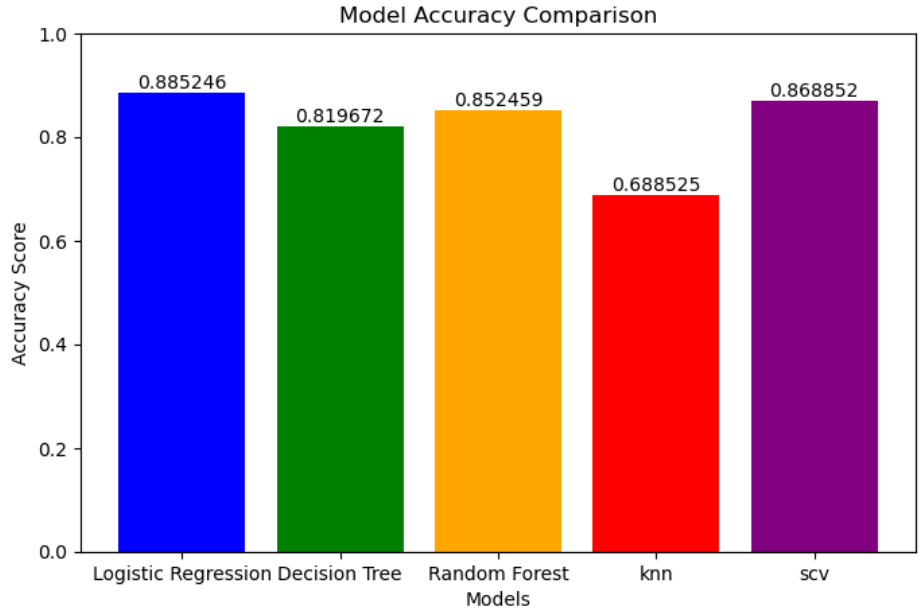

Fig 4: ML model comparison with outliers

Calculate the performance of **DL models** with the outliers and compare the models and also calculate the AUC, precision, Recall, F1-Score etc.

Artificial neural network (ANN)

In ANN we analyse the accuracy of the decision tree the recall is stood 78% and 90%. The F1 score is 84% and 83%, precision is 89% and 79%, AUC-score is 83.89%. Regarding overall accuracy, the model achieved an accuracy score of 83.60%.

Convolutional Neural Network (CNN)

In CNN we analyse the accuracy of the decision tree the recall is stood 84% and 93%. The F1 score is 89% and 93%, precision is 84% and 93%, AUC-score is 88.73%. Regarding overall accuracy, the model achieved an accuracy score of 88.52%.

Recurrent Neural Network (RNN)

In RNN we analyse the accuracy of the decision tree the recall is stood 79% and 75%. The F1 score is 77% and 77%, precision is 80% and 74%, AUC-score is 77.15%. Regarding overall accuracy, the model achieved an accuracy score of 77.04%.

Here the comparison table

Model	Accuracy (%)	Recall (%)	Precision (%)	F1score (%)	AUC score
ANN	83.60	78	89	83	83.89
CNN	**88.52**	84	93	89	88.73
RNN	77.04	75	80	77	77.15

Table 3: DL model performance comparison

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.471 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 14, Issue 12, December 2025

DOI: 10.17148/IJARCCE.2025.141236

Here the comparison bar graph

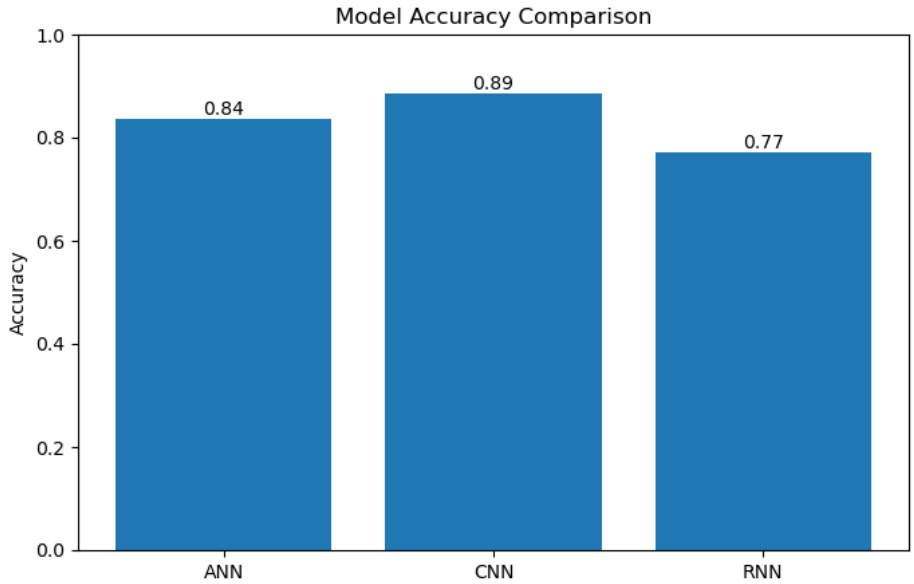

Fig 5: DL model performance with outliers

After removing outliers

Analyse the performance of ***ML and DL models*** after removing the outliers the outliers and compare the models and also calculate the AUC, precision, Recall, F1-Score etc.

Logistic Regression (LR) Results

The recall before removing outliers scores stood at 91% and 88%, emphasizing the model's importance in correctly identifying all positive cases, particularly in scenarios where missing potential cases of heart disease is a critical concern. The F1 Score captured genuine positive cases at 91% and 87%. Regarding overall accuracy, the model achieved an accuracy score of 89.47% and the precission is 94% and 84%,
AUC-Score is 89.76%.

Decision tree result

In this we analyse the accuracy of the decision tree the recall is stood 74% and 83%. The F1 score is 75% and 79%,precision is 88% and 86%, AUC-score is 78.06%. Regarding overall accuracy, the model achieved an accuracy score of 77.196%

Random Forest result

Similarly, Evaluate the performance of the model with the no. Of random forest the recall is stood 91% and 88%. The F1 score is 91% and 87% ,precision is 94% and 84%, AUC-score is 85.12%. Regarding overall accuracy, the model achieved an accuracy score of 89.76% .

K-Nearest Neighbors(KNN) Result

We commenced the analysis by employing the K-Nearest Neighbors (KNN) algorithm with varying 'k' values, representing the number of nearest neighbors considered during the predictions. Employing cross-validation, we computed scores for each 'k' value, ultimately discerning that 'k = 7' yielded the most favorable mean cross-validation score. This outcome underscores that configuring KNN with 'k = 7' exhibits significant promise. The recall is stood 75% and 62%. The F1 score is 72% and 65% ,precision is 69% and 69%,F1-score is 86% and 84%, AUC-score is 68.53%. Regarding overall accuracy, the model achieved an accuracy score of 68.85% .

Support Vector Machine (SVM) Results

As the model achieved a precision score of 91% and 83%, a recall score of 87% and 88%, and an F1 Score of 85% and 90%, AUC-Score 87.59%. On the test dataset, the model exhibited an accuracy of approximately 87.71% an, affirming its consistent and accurate predictive capabilities.

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.471 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 14, Issue 12, December 2025

DOI: 10.17148/IJARCCE.2025.141236

Here the model comparison table given below:

Model	Accuracy (%)	Precision (%)	Recall (%)	F1score (%)	Accord
LR	89.47	94	91	91	89.76
DT	77.19	86	83	79	78.06
RF	89.76	94	91	91	89.76
KNN	87.71	91	88	90	87.59
SVM	**91.22**	94	91	93	91.24

Table 4: ML model accuracy with removing outliers

Here the comparison barograph:

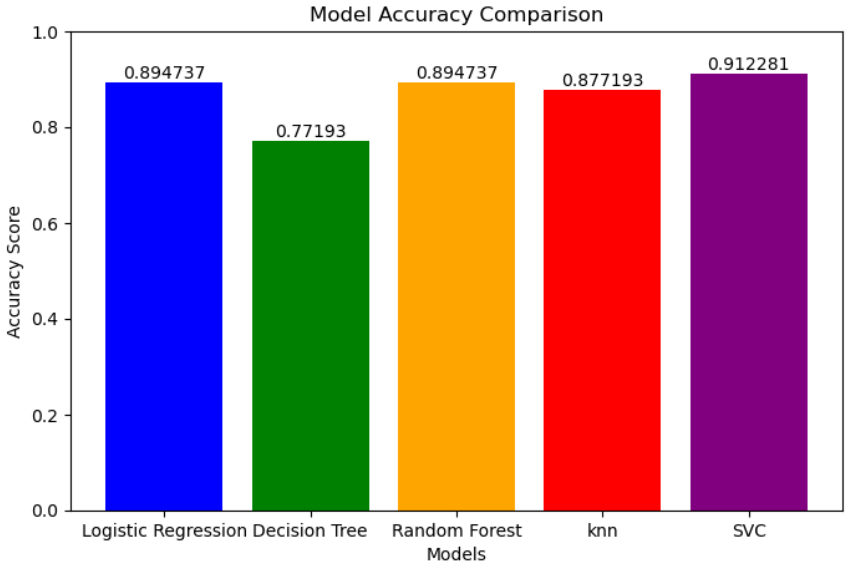

Fig 6: ML model comparison after outlier removing

Artificial neural network (ANN)

In ANN we analyse the accuracy of the decision tree the recall is stood 100% and 97%. The F1 score is 99% and 98%, precision is 100% and 96%, AUC-score is 98.52%. Regarding overall accuracy, the model achieved an accuracy score of 98.28%.

Convolutional Neural Network (CNN)

In CNN we analyse the accuracy of the decision tree the recall is stood 97% and 96%. The F1 score is 97% and 96%, precision is 97% and 96%, AUC-score is 96.35%. Regarding overall accuracy, the model achieved an accuracy score of 96.49%.

Recurrent Neural Network (RNN)

In RNN we analyse the accuracy of the decision tree the recall is stood 100% and 88%. The F1 score is 94% and 92% ,precision is 100% and 85%, AUC-score is 92.11%. Regarding overall accuracy, the model achieved an accuracy score of 92.04% .

Here the comparison table given below:

Model	Accuracy(%)	Precision(%)	Recall(%)	F1score(%)	AUCscore
ANN	**98.28**	100	100	99	98.52
CNN	96.49	97	97	97	96.35
RNN	92.04	100	100	94	94.11

Table 5: DL model comparison after removing outliers

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.471 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 14, Issue 12, December 2025

DOI: 10.17148/IJARCCE.2025.141236

Here the comparison bar graph:

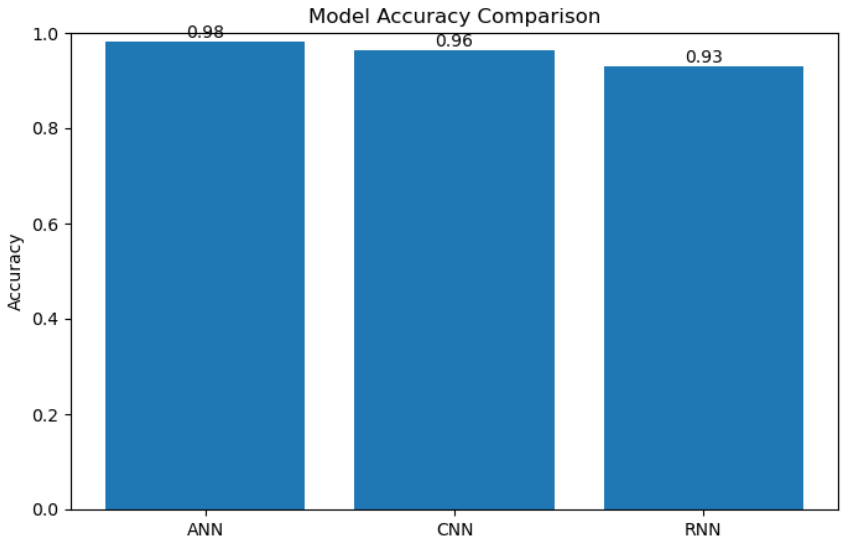

Fig 7: DL model after removing outliers

There are the comparison of Machine learning and deep learning models before and after removing the outliers

V. DISCUSSION

In the study's findings, In ML models the performance of the **SVM** is better rather than other models the model achieved a precision score of 91% and 83%, a recall score of 87% and 88%, and an F1 Score of 85% and 90%, AUC-Score 87.59%. On the test dataset, the model exhibited an accuracy of approximately 87.71% an, affirming its consistent and accurate predictive capabilities and the DL model **ANN** performed better, we analyses the accuracy of the decision tree the recall is stood 100% and 97%. The F1 score is 99% and 98%, precision is 100% and 96%, AUC-score is 98.52%. Regarding overall accuracy, the model achieved an accuracy score of 98.28%.
The study underscores the significance of algorithm selection and hyperparameter refinement for precise heart disease predictions.

VI. CONCLUSION AND FUTURE WORK

Coronary Heart disease is the major factor for the people's deaths throughout the world, and it is necessary to detect and predict the disease in the earlier stages because time plays a vital role to save the coronary patient, which we call a "Golden hour" is very much necessary we need to get some more advanced technologies to predict the heart disease so that we can save the patients as early as possible. Therefore we can reduce the death rate treatment at the right time. From this paper, we can conclude that we have used most of the machine learning and deep learning ensemble algorithms so that it can predict heart disease at an early stage so that the patient's life can be saved.

ACKNOLEGEMENTS

First and foremost, I humbly acknowledge the divine guidance and blessings of God, which have been an unwavering source of strength and inspiration throughout this journey. I extend my deepest appreciation to my Professor, **Er. Harjasdeep Singh**, for his invaluable guidance, constant support, and mentorship throughout this dissertation journey. Your expertise and guidance have played a critical role in shaping this work.

REFERENCES

[1]. Mao Yian, Jimma LB, Mihretie TB. Machine Learning algorithms for heart disease diagnosis: A systematic review. Elservier Ltd. **(2025)**.https://doi.org/10.1016/j.cpcardiol.2025.103082
[2]. Alam Zahangir Md, Rahman Saifur, Rahman Sohel. A Random Forest based predictor for medical data classification using feature ranking. Elsevier Ltd**. (2019)**.https://doi.org/10.1016/j.imu.2019.100180
[3]. Hagan Rachael, Gillan JC, Mallett Fiona. Comparison of machine learning methods for the classification of cardiovascular disease. Elservier Ltd. **(2021)**.https://doi.org/10.1016/j.imu.2021.100606

DOI: 10.17148/IJARCCE.2025.141236

[4]. Bhalla Rajni, S Sreekumari. A Comparative Study of Heart Disease Prediction using Machine Learning. ResearchGate**. (2024)**.https://www.researchgate.net/publication/381653893

[5]. Qian Kun, Bao Zhihao, Zhao Zhonghao. Learning Representations from Heart Sound: A Comparative Study on Shallow and Deep Models. Cybrog Bionic Syst**.(2024).**https://doi.org/10.34133/cbsystems.0075

[6]. Ahmad AA,zHuseyin Polat. Prediction of Heart Disease Based on Machine Learning Using Jellyfish Optimization Algorithm.Diagonotics.**(2023).**https://doi.org/10.3390/diagnostics13142392

[7]. Ogunpola, A,Saeed, F,Basurra, S, Albarrak, A.M,Qasem, S.N. Machine Learning-Based Predictive Models for Detection of Cardiovascular Diseases. Diagnostics **(2024).**https://doi.org/10.3390/diagnostics14020144

[8]. Deepika S. and Jaisankar N**.** Review On Machine Learning and Deep Learning-based Heart Disease Classification and Prediction.The Open Biomedical Engineering Journal.**(2022)**.https://openbiomedicalengineeringjournal.com

[9]. Menon SV, Kumar P, Hadi AM, Hasson SA and Lozanović J . A comprehensive review of machine learning for heart disease prediction: challenges, trends, ethical considerations, and future directions. Frontiers**.(2025)**. https://doi.org/10.3389/frai.2025.1583459

[10]. Kumar R, Garg S, Kaur R, Johar MGM, Singh S, Menon SV, Kumar P, Hadi AM, Hasson SA and Lozanović J . A comprehensive review of machine learning for heart disease prediction: challenges, trends, ethical considerations, and future directions. Frontiers. **(2025).**https://doi/10.3389/frai.2025.1583459

[11]. Khan Younas, Qamar Usman, Yousaf Nazish. Machine Learning Techniques for Heart Disease Datasets: A Survey, ResearchGate**. (2019).** https://www.researchgate.net/publication/329519531

[12]. Zhou Chunjie, Dai Pengfei, Hou Aihua Zhang Zhexing. A comprehensive review of deep learning-based models for heart disease prediction. Springer. **(2024).**https://doi.org/10.1007/s10462-024-10899-9

[13]. Reddy Vardhana VK. A Comprehensive Review on Heart Disease Risk Prediction using Machine Learning and Deep Learning Algorithms., Research Gate. **(2024).**https://www.researchgate.net/publication/385012222

[14]. Teja Darhan M, Rayalu Mokesh G. Optimizing heart disease diagnosis with advanced machine learning models: a comparison of predictive performance. BMC Cardiovascular Disorders.**(2025).**https://doi.org/10.1186/s12872-025-04627-6

[15]. Teja Darhan M, Rayalu Mokesh G. Optimizing heart disease diagnosis with advanced machine learning models: a comparison of predictive performance. BMC Cardiovascular Disorders**.(2025).**https://doi.org/10.1186/s12872-025-04627-6

[16]. Alsabhan Waleed, Alfadhly Abdullah. Effectiveness of machine learning models in diagnosis of heart disease: a comparative study. Scientific reports**. (2025).**https://doi.org/10.1038/s41598-025-09423-y

[17]. Alsabhan Waleed, Alfadhly Abdullah. Effectiveness of machine learning models in diagnosis of heart disease: a comparative study. Scientific reports**. (2025).**https://doi.org/10.1038/s41598-025-09423-y

[18]. Kumar Ankur, Dhanka Sanjay, Sharma Abhinav, Bansal Rohit. A hybrid framework for heart disease prediction using classical and quantum-inspired machine learning techniques. Scientific reports**. (2025).** https://doi.org/10.1038/s41598-025-09957-1

[19]. Rehman MU, Naseem Sahid, Rehman AU, Mahmood Tariq. Predicting coronary heart disease with advanced machine learning classifiers for improved cardiovascular risk assessment. Scientific reports. **(2025).** https://doi.org/10.1038/s41598-025-96437-1

APPENDIX

A. Dataset

Heart disease data

B. Code

Click here