GAS Journal of Engineering and Technology (GASJET)

OPEN ACCESS

Volume 2, Issue 4, 2025 Journal Homepage: https://gaspublishers.com/gasjet-home/ ISSN: 3048-5800
Email: editor@gaspublishers.com

Cardiovascular Disease (CVD) Prediction with Machine Learning (ML) K-Neighbor (KNN), Decision Tree (DT), and Random Forest (RF) Classifiers

Sahabdeen Aysha Asra[1], Dr. K.H. Ramanayaka[2], R.D.Madushan Prabod Kumara[3], W.H.Thurushi Thathsarani[4]

[1,2,3,4] Department of Information Technology, Faculty of Humanities and Social Sciences, University of Ruhuna

Received: 09.06.2025 | **Accepted:** 11.07.2025 | **Published:** 18.07.2025

***Corresponding Author:** R.D.Madushan Prabod Kumara

DOI: *10.5281/zenodo.16107733*

Abstract **Original Research Article**

Cardiovascular disease (CVD) has recently exceeded all other reasons of death universal in both every nation. Early detection of cardiac circumstances and ongoing therapeutic supervision by specialists can lower death rate. It's not always possible to properly observe patients daily, and a doctor cannot deliberate with a patient for a whole day since it needs more expertise, intellect, and time. Manual methods are complex to subject when used to diagnose CVD. In this sense, Machine Learning (ML) procedures are trustworthy and effective sources to identify and classify individuals with CVDs. According to the suggested study, we used ML algorithms to recognize and forecast human CVD, and we used the CVD dataset to evaluate the act of those algorithms using various metrics, including classification accuracy, precision, F1-score and recall. In this research, we developed and researched models for CVD prediction using the patient's various heart attributes as well as CVD detection using ML techniques such as K-Neighbors Classifier (KNN), Decision Tree Classifier (DT), and Random Forest Classifier (RF) on the dataset made openly available in Kaggle website. The experiment result shows RF classifier has the highest accuracy among other classifiers. Almost it predicts 87 % of highest accuracy.

Keywords: Machine Learning (ML), K-Nearest Neighbors Classifier (KNN), Decision Tree Classifier (DT), Random Forest Classifier (RF), Cardiovascular disease (CVD).

Citation: Asra, S. A., Ramanayaka, K.H, Kumara, R. D. M. P., & Thathsarani, W. H. T. (2025). Cardiovascular disease (CVD) prediction with machine learning (ML) K-neighbor (KNN), decision tree (DT), and random forest (RF) classifiers. *GAS Journal of Engineering and Technology (GASJET), 2*(4), [1-9].

INTRODUCTION

Cardiovascular disease (CVD) is the systemic disorder, which is the major cause of death throughout the world over. It is important to make an early detection since CVD has a tendency to be silent in terms of its progression. Researchers have attempted to discover some major risk variables and enhance prediction with the use of Machine Learning (ML) methodologies. ML models can more predictably determine the occurrence of CVD with the use of standard and well-structured data sets. Data pre-processing and feature selection are also necessary to improve the value of predictions.

This paper has suggested a ML CVD prediction model containing K-Neighbor Classifier (KNN), Decision Tree Classifier (DT), and Random Forest Classifier (RF). Furthermore, the ML classifier performance is evaluated with the help of numerous metrics of performance assessment, such as accuracy, F1-score, precision, and recall. The suggested approach has been discussed on the Kaggle dataset. The accuracy of the ML classifiers suggested has also been compared with other classifiers. In this study, we proposed a ML classifier for cardiovascular illness prediction that includes KNN, DT, and RF. In addition, the performance of the ML classifier is measured using a variety of performance assessment criteria, including accuracy, F1-score, precision, and recall. On the Kaggle dataset, the suggested approach has been examined. The suggested ML classifiers' accuracy has also been evaluated in comparison to other classifiers.

Research Questions

a) What are the primary challenges encountered in the management of heart health, particularly in routine patient monitoring and consistent data collection?

b) What methods currently exist for the prediction of cardiovascular diseases, and how effective are they?
c) How can machine learning contribute to overcoming the challenges of heart health management and improve cardiovascular disease prediction and monitoring?

2. The Primary Objectives of this Project's Development are:

a) To identify important risk factors for CVD illness based on medical data.
b) To evaluate feature selection techniques and comprehend their operation.
c) To create a ML model to predict the likelihood of acquiring CVD in the future using KNN, DT and RF.
d) To find the highest accuracy among other classifiers.

RECENT STUDIES

The most important organ, the heart, transfers blood to all other areas of the human body. The mind and other organs will cease working if the heart fails to function correctly, which causes the individual to pass away in just a matter of minutes. Thus, proper cardiac function is essential. Heart conditions are becoming one of the main killers on a global scale. Because of this, many scholars from across the world started concentrating on exploiting the massive datasets to predict heart-related illnesses[1].

Complex training methods like ensemble learning (i.e., EHBM-DNN), meta-logistic regression and optimization algorithms, like Particle Swarm Optimization (PSO) and Firefly Algorithm have attained levels of accuracy up to 95.3% [2][3][4]. Techniques for feature selection such as Fast Correlation-Based Filter (FCBF) and hybrid Support Vector Machine-Genetic Algorithm (SVM-GA) were used on top of performance to improve performance [5]. One advantage of ensemble models is that they are accurate and reliable and therefore they are very useful in underdeveloped regions [6][7]. The smart healthcare systems are realizable through integration with the Internet of Medical Things (IoMT) that utilizes the stored data on the cloud accessible to both the patient and the professional [8]. The key is feature selection, which can enhance CVD forecasting, and the image analysis using ML (e.g., pooling area curves) that can be used to detect a vascular problem such as the blocked artery and plaque [9] [10].

Several ML algorithms which have displayed potential in working with dataset such as the Z-Alizadeh Sani CAD dataset which has 54 features include DT, Logistic Regression (LR), SVM, RF, and XGBoost. The feature selection was performed with the help of Pearson correlation; 8 attributes were found key ones. The types of models used were LR and SVM, and they delivered the best results characterized by an Area Under the Curve (AUC) of 0.98 [11].

Electrocardiogram (ECG) has a good potential in the detection of CVD because it is non-invasive and low cost, however due to inter- and intra-individual variability there is a challenge [12]. The prediction has been enhanced by the use of classifiers, like CART, AdaBoost (AB), Multinomial Naive Bayes (MNB), Linear Discriminant Analysis (LDA), and Extra Trees (ET) on the data after K- fold cross-validation and hyperparameter optimization [13]. An intelligent diagnosis system Coronary Artery Disease Diagnosis (C-CADZ) is based on Fixed Analysis of Mixed Data (FAMD) features extractor and Synthetic Minority Over-sampling Technique (SMOTE) balancing algorithm during the dataset balancing. Using 96 features by analyzing the models, there was increased accuracy and LR achieved accuracy of 90% with sensitivity of 92.18% and specificity of 81.34%[14].

METHODS AND MATERIALS

1. Proposed system

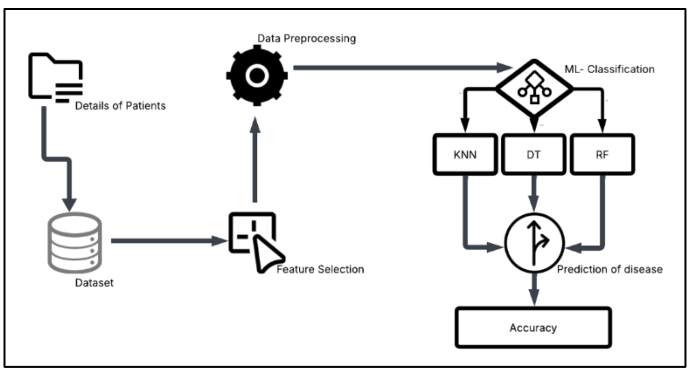

Figure 1: System Architecture

According to the Figure 1, an extensive collection of data on patients will be prepared, and then the process of feature selection, which will allow determining the most promising characteristics predictive of CVD, will be carried out. Upon a choice of the pertinent data, the information shall be cleaned up, standardized and transformed into a form of analysis. After that, various classification algorithms (such as KNN, DT, RF, and so on) will be utilized with this preprocessed data to precisely determine cardiac conditions. Last but not least, the performance of each one of the classifiers will be measured with a variety of metrics, these being: (1) the overall accuracy, (2) the precision, (3) the recall, and (4) the F1-score, to decide which of the two methods provides the most trustworthy predictions.

a. Dataset

The dataset, with 270 total findings, has no missing values. There should be feasible to establish if a patient has CVD based on the presented attribute data. The cost matrix displays the financial impact of improperly categorizing the true as well as anticipated values. It demonstrates that there is no expense when absence is predicted when that is the real value, no cost when presence is predicted when that is the true value, as well as a cost of 5 when absence is predicted when that is the genuine value. Age, sex, type of chest pain, resting blood pressure, fasting blood sugar levels, serum cholesterol, maximum heart rate, resting electrocardiographic results, exercise-induced angina, ST depression brought on by exercise, slope of the peak exercise, number of major vessels, and target ranging from 0 to 2, where 0 is the absence of cardiovascular illness, are some of the attributes. The data collection is created as a data frame using the panda's package in Python and is in the CSV (Comma Separated Value) format.

b. Feature Selection

The process of feature selection can contribute to the enhancement of the forecasting effectiveness with the help of correlation matrix to determine the influencing factors of the patient such as the sex, the type of chest pain, the cholesterol level, etc.

c. Pre-Processing of data

Before the actual training process, some data pre-processing is carried out such as column renaming, variable encoding, missing value measurement and scaling attributes to enhance accuracy and stability of the models.

d. ML Classifiers

The system runs machine learning algorithms such as KNN, RF, and DT to anticipate the risk of possible CVDs, in the future, which uses input features.

i) KNN Classifier

KNN is an easy and fast way of categorizing new data following how it matches the data that was already stored, via the similarity to the nearby information.

ii) DT Classifier

The DT classifier divides data according to important characteristics and produces the simulation of human decision-making into a tree-based logic.

iii) RF Classifier

RF is an algorithm based around DT structures that mix subset-trained DTs to increase precision and prevent overfitting by averaging decisions.

e) Disease Prediction

We provide an overview of the system for predicting CVD using several classifier methodologies. The methods use KNN, RF and DT classifiers.

f) Accuracy Measures

After assessing the precision of each of the three ML approaches, a forecast model is developed. Consequently, the objective is to use a range of assessment criteria, including the confusion matrix, accuracy, ROC curves.

2) Proposed Method Concepts

Table 1 : Method Concepts

Tool/Library	Description
Python	An object-oriented language that uses dynamic types and garbage collection; it also has many paradigms such as Object-Oriented Programming (OOP), functional, and procedural programming.
Jupyter Notebook	An interactive tool, students and professionals use to write executable live code along with outputs of equations, plots, and videos.
NumPy	Library of large structured arrays and matrices of many dimensions, extending high level operations of mathematics on multi-dimensional data.
Pandas	It is a library to manipulate and analyze data that gives data structures and functions to manipulate structured data such as tables and time series.

RESULTS AND DISCUSSION

i. Reading the Dataset

The "read_csv()" method from the Pandas library may read a dataset from a big text file. The Pandas library's "read_csv()" method is an effective resource for reading and analyzing huge datasets kept in text files. The following figure shows the results of dataset reading in our ML system.

```
df = pd.read_csv('E:\MSC\Second Year - Second Semester\ML\Assignment 2\Predicting-Heart-Disease-master\dataset_heart.csv')

df.info()

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 270 entries, 0 to 269
Data columns (total 14 columns):
 #   Column                               Non-Null Count  Dtype
---  ------                               --------------  -----
 0   age                                  270 non-null    int64
 1   sex                                  270 non-null    int64
 2   chest pain type                      270 non-null    int64
 3   resting blood pressure               270 non-null    int64
 4   serum cholestoral                    270 non-null    int64
 5   fasting blood sugar                  270 non-null    int64
 6   resting electrocardiographic results 270 non-null    int64
 7   max heart rate                       270 non-null    int64
 8   exercise induced angina              270 non-null    int64
 9   oldpeak                              270 non-null    float64
 10  ST segment                           270 non-null    int64
 11  major vessels                        270 non-null    int64
 12  thal                                 270 non-null    int64
 13  heart disease                        270 non-null    int64
dtypes: float64(1), int64(13)
memory usage: 29.7 KB
```

Figure 2: Reading the dataset

	age	sex	chest pain type	resting blood pressure	serum cholestoral	fasting blood sugar	resting electrocardiographic results	max heart rate	exercise induced angina	oldpeak	ST segment	majo vessel
count	270.000000	270.000000	270.000000	270.000000	270.000000	270.000000	270.000000	270.000000	270.000000	270.00000	270.000000	270.00000
mean	54.433333	0.677778	3.174074	131.344444	249.659259	0.148148	1.022222	149.677778	0.329630	1.05000	1.585185	0.67037
std	9.109067	0.468195	0.950090	17.861608	51.686237	0.355906	0.997891	23.165717	0.470952	1.14521	0.614390	0.94389
min	29.000000	0.000000	1.000000	94.000000	126.000000	0.000000	0.000000	71.000000	0.000000	0.00000	1.000000	0.00000
25%	48.000000	0.000000	3.000000	120.000000	213.000000	0.000000	0.000000	133.000000	0.000000	0.00000	1.000000	0.00000
50%	55.000000	1.000000	3.000000	130.000000	245.000000	0.000000	2.000000	153.500000	0.000000	0.80000	2.000000	0.00000
75%	61.000000	1.000000	4.000000	140.000000	280.000000	0.000000	2.000000	166.000000	1.000000	1.60000	2.000000	1.00000
max	77.000000	1.000000	4.000000	200.000000	564.000000	1.000000	2.000000	202.000000	1.000000	6.20000	3.000000	3.00000

Figure 3: Summary of Numerical Features

ii. Summary of Numerical Features

"df.describe()" used to get summary of numerical features. It is directly display from the dataset.

iii. Data cleaning

Data cleaning is locating and fixing flaws in the dataset, such as addressing outliers, duplicates, or missing or inconsistent data. Making sure the ML method is trained on correct and trustworthy data is crucial.

```
df.info()

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 270 entries, 0 to 269
Data columns (total 14 columns):
 #   Column                               Non-Null Count  Dtype
---  ------                               --------------  -----
 0   age                                  270 non-null    int64
 1   sex                                  270 non-null    int64
 2   chest pain type                      270 non-null    int64
 3   resting blood pressure               270 non-null    int64
 4   serum cholestoral                    270 non-null    int64
 5   fasting blood sugar                  270 non-null    int64
 6   resting electrocardiographic results 270 non-null    int64
 7   max heart rate                       270 non-null    int64
 8   exercise induced angina              270 non-null    int64
 9   oldpeak                              270 non-null    float64
 10  ST segment                           270 non-null    int64
 11  major vessels                        270 non-null    int64
 12  thal                                 270 non-null    int64
 13  heart disease                        270 non-null    int64
dtypes: float64(1), int64(13)
memory usage: 29.7 KB
```

Figure 4: Data Cleaning

iv. Exploratory Analysis

Data scientists utilize Exploratory Data Analysis (EDA), which frequently makes use of data visualization techniques, to examine and study data sets and summarize their key properties.

a. Visualization of Categorical Features

The best visual representations for categorical variables are frequency tables, pie charts, and bar charts. The frequency table, pie chart, and bar graph that follow provide information on the quantity. The chart showing the frequency counts for each group.

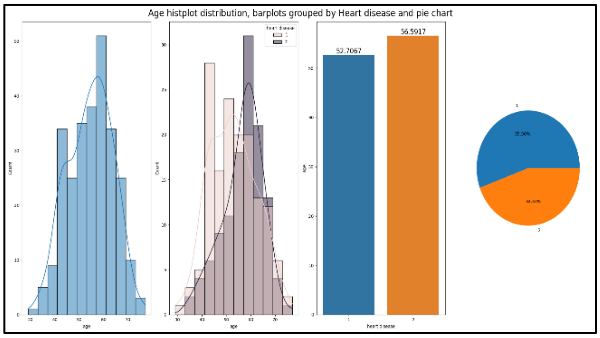

Figure 5: Distribution of Cases with Cardiovascular disease according to Age

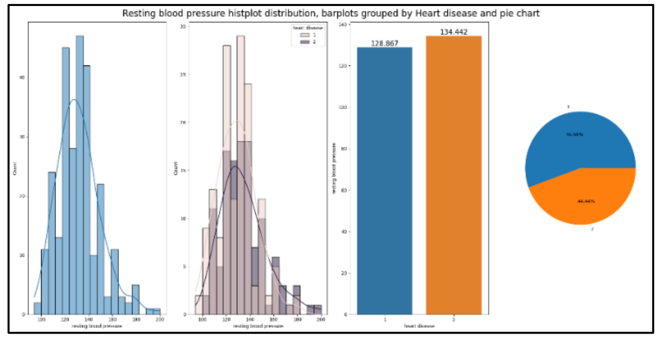

Figure 6: Distribution of Cases with Cardiovascular disease based on Resting Blood Pressure

This code will create a chart where the y-axis is the number of persons and the x-axis is age. Figure 5 shows that those with cardiovascular illness are represented by the colour orange, while those without the condition are shown by the colour blue. Histograms, grouped histograms, grouped barplots, pie charts are display in the figure.

Conferring to the Figure 6 histograms, grouped histograms, grouped barplots, pie charts were generated to identify how resting blood pressure connected. In this approximately 100 to 175 was mentioned in the figure.

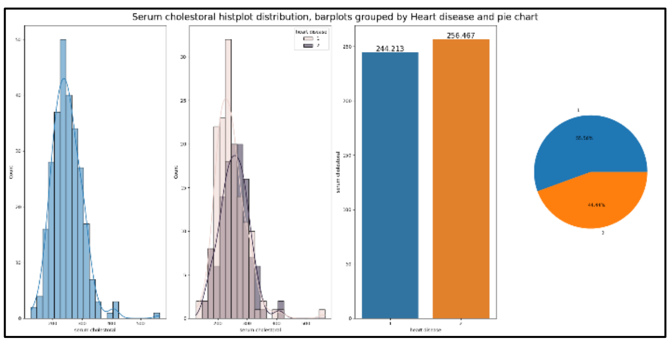

Figure 7: Distribution of Cases with Cardiovascular disease based on previous exposure to Serum Cholesterol

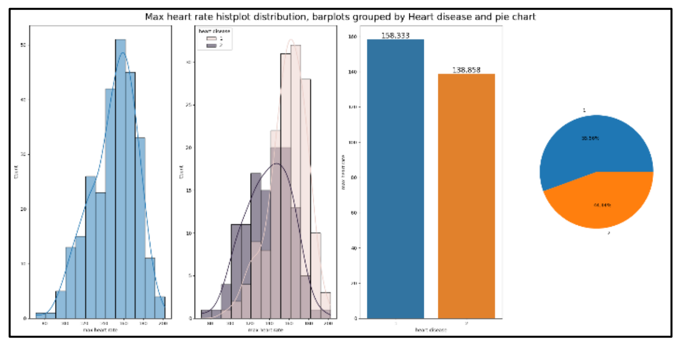

Figure 8: Distribution of Cases with Cardiovascular disease according to Maximum Heart Rate

Conferring to the Figure 7 histograms, grouped histograms, grouped barplots, pie charts were generated to identify distribution of cases with CVD based on previous exposure to serum cholesterol. In this approximately 120 to 450 was stated in the figure.

According to the Figure 8 histograms, grouped histograms, grouped barplots, pie charts were generated to identify how maximum heart rate connected with CVD. In this approximately 100 to 200 bpm was mentioned in the figure.

v. Visualization of Numerical Features

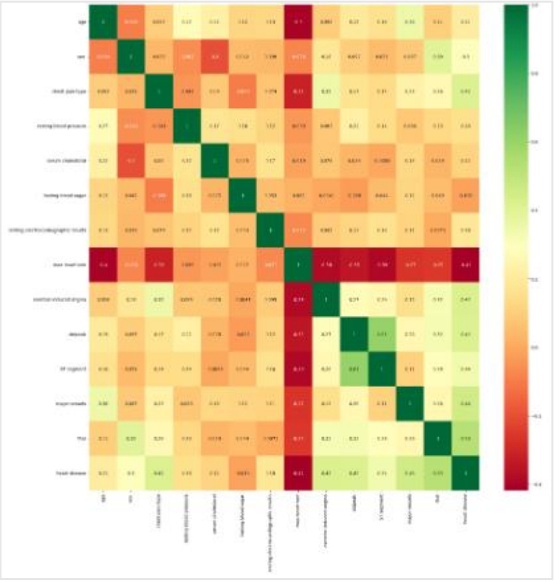

Figure 9: Heat Map

The link between each characteristic is visualized in the Figure 9 using Pearson's correlation analysis. This aids in finding the feature in the data repository that is closely connected to the class feature. Figure 9 displays the Pearson's correlation matrix for each characteristic. Some of the traits have strong correlations, as seen in Figure 9. For instance, the association between age and total resting blood pressure is 0.27. Age and cholesterol have a strong relationship, with a correlation value of 0.22. Additionally, there is a strong 0.25 association between the number of main vessels and age. The maximal heart rate attained and slope have a strong association of 0.4. In contrast, there is a negative link between age and characteristics like resting electrocardiogram and exercise-induced angina.
The nicest thing about this kind of plot is how easy it is to design it and how much information Split Dataset for Training and Testing.

vi. Data pre-processing

b. Encoding

An essential pre-processing step in ML is data encoding. It describes the process of transforming textual or category data into numerical representation so that it may be utilized as input for processing algorithms. Since most ML algorithms don't function with text or categorical information, encoding is necessary.

c. Standardization

By standardizing the data, the delivery will be changed to have a mean of 0 as well as a normal deviation of 1.

vii. Modeling

Determine which classification model has the maximum accuracy by training them all on the training set. We evaluated the precision of KNN, RF, and DT.

viii. Model evaluation

To evaluate the model, we used KNN, RF and DT. Also we choose to get model evaluation are accuracy, recall, f1-score and precision. We can infer from a comparison of the 3 models that RF produces the best accuracy with about 87% accuracy. The average recall, accuracy, F1-score and precision which is shown in the Table 2, is 0.88, or 88%. During test 1 recall, accuracy, F1-score and precision were all at 100%. This represents the highest number of percentages. In the meantime, DT improved recall, accuracy, F1-score and precision in test 2 by 76%. Therefore, the average for F1 score, recall, accuracy, and precision was 88%.

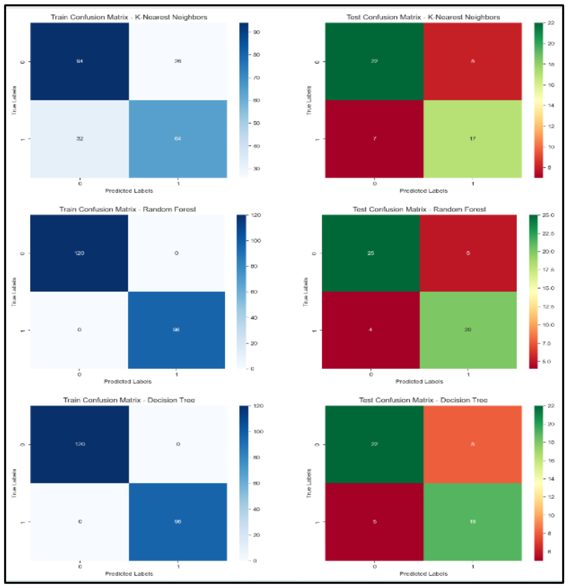

Figure 10: Confusion metrics of train and test data

The confusion matrix shows the amount of TP (64), TN (94), FP (26), and FN (32) predictions in train dataset. In this case, TP (17), TN (22), FP (8), and FN (7) predictions were identified in test model. This is confusion metrics of KNN classifier. In the train dataset, there were 96 TP predictions, 120 TN predictions, 0 FP predictions, and 0 FN predictions. In this instance, the test model detected the predictions TP (20), TN (25), FP (6), and FN (4). This is the RF algorithm's confusion measure. This is the DT algorithm's confusion measure. In the train dataset, there were 96 TP predictions, 120 TN predictions, 0 FP predictions, and 0 FN predictions. In this instance, the test model detected the predictions TP (19), TN (22), FP (8), and FN (5).

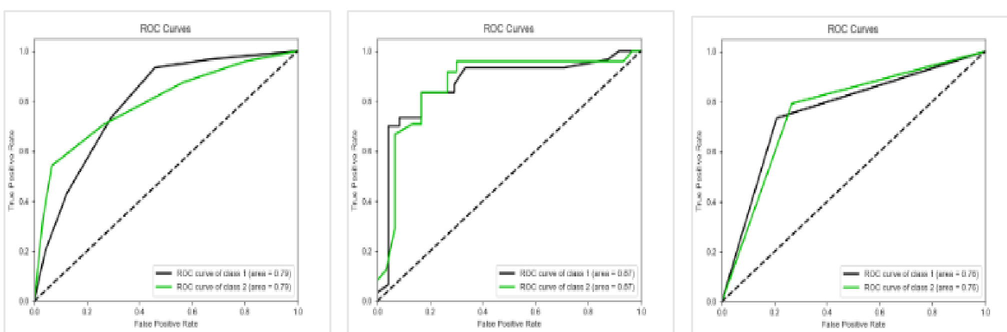

Figure 11: ROC curve of each classifier

When comparing 3 classifiers ROC curves of KNN class 1 and 2 have same value of 0.79. Then ROC curves of RF class 1 and 2 have 0.87 of highest robust value. Finally, DT classifier ROC curve of class 1 and 2 have 0.76 value.

This section discusses the consequences of the field experimentations conducted on the proposed model. These results show that KNN, RF Classifier, as well as DT give better results than other procedures, even though the bulk of research employ other program, including SVM and DT, for recognizing individuals with cardiovascular illnesses. Our algorithms perform better than those used in past research in terms of speed and accuracy. They are extremely affordable since they additionally save a lot of money.

Moreover, according to our experiment, RF and DT perform better than KNN Classifier in predicting if a patient would grow a cardiac illness. This proves that RF and DT are further effective in diagnosing a cardiac disorder. 'Table 2' shows a plot of the number of patients that are separated and forecasted by the algorithms based on the resting blood pressure, age group, chest pain, and sex. In conclusion, the additional medical characteristics that we employed from the dataset we acquired have improved our accuracy.

Table 2: Evaluation metrics of different classifiers

Classifiers	**Accuracy**	**Recall**	**Precision**	**F1-score**
KNN	0.80	0.80	0.81	0.79
DT	0.83	0.83	0.84	0.83
RF	0.87	0.87	0.88	0.87

According to the Table 2, RF has the highest accuracy of 87%. Also simultaneously F1-score and recall are 87%. Precision is increased as 88%. Next, the DT has 83% of accuracy, recall and F1-score. In addition precision improved as 84%. In last place KNN is stated. In order to that 80% accuracy and recall 81% of precision and 79% F1-score were gained. Finally the RF has the highest accuracy among other classifiers.

We found an increase in the accuracy of majority of the classifiers by comparing their performance on the normal and standardized datasets. Therefore, before using ML classifiers, standardizing the dataset is a beneficial strategy for improving accuracy. Similar to this, we have seen a notable gain in accuracy following classifier hyperparameter adjustment. As a result, algorithm tweaking is a beneficial method for raising the algorithms' accuracy. We draw the conclusion that DT and KNN classifiers exhibit generally high accuracy based on the comparison of other classifiers. However RF has the highest accuracy of 87%.

CONCLUSION

The best accuracy, 87%, is in RF. The F1-score and recall are both 87% at the same time. The improvement in precision is 88%. The DT follows with 83% accuracy, recall, and F1-score. Additionally, accuracy increased by 84%. KNN comes last and is mentioned. In order to do that, 79% F1-score, 81% precision, and 80% recall were attained. Finally, among all classifiers, the RF has the best accuracy. It is obvious that the RF has the best accuracy, recall, precision, and F1 score percentages. The KNN, last among the others, has the lowest value. This project effectively predicted CVD with 87% accuracy by resolving the feature selection issues underlying the models. Based on a study of several classifiers, we conclude that DT and KNN classifiers typically demonstrate high accuracy. The best accuracy, though, is RF with 87%.

Moreover, in order to determine the model's generalizability to fresh, untested data, it would be advantageous to evaluate the model's performance on a held-out test dataset. Future research should ultimately try to demonstrate the generalizability, interpretability, and robustness of the results as well as the clusters created by the algorithm, since these factors might enhance decision making based on the study's results and help in comprehending the results.

REFERENCES

[1] S. J. Al'Aref *et al.*, "Machine learning of clinical variables and coronary artery calcium scoring for the prediction of obstructive coronary artery disease on coronary computed tomography angiography: Analysis from the CONFIRM registry," *Eur. Heart J.*, 2020, doi: 10.1093/eurheartj/ehz565.

[2] S. N. Pasha, D. Ramesh, S. Mohmmad, A. Harshavardhan, and Shabana, "Cardiovascular disease prediction using deep learning techniques," *IOP Conf. Ser. Mater. Sci. Eng.*, vol. 981, no. 2, 2020, doi: 10.1088/1757-899X/981/2/022006.

[3] D. C. Yadav and S. Pal, "Analysis of Heart Disease Using Parallel and Sequential Ensemble Methods With Feature Selection Techniques," *Int. J. Big Data Anal. Healthc.*, vol. 6, no. 1, pp. 40–56, 2021, doi: 10.4018/ijbdah.20210101.oa4.

[4] Savita, G. Rani, and A. Mittal, "An Optimized Machine Learning Approach for Coronary Artery Disease Detection," *J. Adv. Inf. Technol.*, 2023, doi: 10.12720/jait.14.1.66-76.

[5] Wiharto, E. Suryani, S. Setyawan, and B. P. E. Putra, "Hybrid Feature Selection Method Based on Genetic Algorithm for the Diagnosis of Coronary Heart Disease," *J. Inf. Commun. Converg. Eng.*, vol. 20, no. 1, pp. 31–40, 2022, doi: 10.6109/jicce.2022.20.1.31.

[6] Z. F. Khan and S. R. Alotaibi, "Applications of Artificial Intelligence and Big Data Analytics in m-Health: A Healthcare System Perspective," *J. Healthc. Eng.*, vol. 2020, 2020, doi: 10.1155/2020/8894694.

[7] N. Sinha, A. Joshi, and S. P. Mohanty, "iCardo 3.0: A Machine Learning Framework for Prediction of Conduction Disturbance in Heart," in *Lecture Notes in Networks and Systems*, 2024. doi: 10.1007/978-981-99-7814-4_28.

[8] O. Al Shorman, B. Al Shorman, M. Al-Khassaweneh, and F. Alkahtani, "A review of internet of medical things (IoMT) - Based remote health monitoring through wearable sensors: A case study for diabetic patients," *Indones. J. Electr. Eng. Comput. Sci.*, 2020,

doi: 10.11591/ijeecs.v20.i1.pp414-422.

[9] M. Sayadi, V. Varadarajan, F. Sadoughi, S. Chopannejad, and M. Langarizadeh, “A Machine Learning Model for Detection of Coronary Artery Disease Using Noninvasive Clinical Parameters,” *Life*, 2022, doi: 10.3390/life12111933.

[10] R. Bharti, A. Khamparia, M. Shabaz, G. Dhiman, S. Pande, and P. Singh, “Prediction of Heart Disease Using a Combination of Machine Learning and Deep Learning,” *Comput. Intell. Neurosci.*, vol. 2021, 2021, doi: 10.1155/2021/8387680.

[11] A. Gupta, R. Kumar, H. S. Arora, and B. Raman, “C-CADZ: computational intelligence system for coronary artery disease detection using Z-Alizadeh Sani dataset,” *Appl. Intell.*, 2022, doi: 10.1007/s10489-021-02467-3.

[12] G. Phadke, M. R. Rajati, and L. Phadke, “Prediction of coronary artery disease using electrocardiography: A machine learning approach,” in *Proceedings - International Conference on Machine Learning and Cybernetics*, 2020. doi: 10.1109/ICMLC51923.2020.9469585.

[13] A. Saboor, M. Usman, S. Ali, A. Samad, M. F. Abrar, and N. Ullah, “A Method for Improving Prediction of Human Heart Disease Using Machine Learning Algorithms,” *Mob. Inf. Syst.*, vol. 2022, 2022, doi: 10.1155/2022/1410169.

[14] I. C. Dipto, T. Islam, H. M. M. Rahman, and M. A. Rahman, “Comparison of Different Machine Learning Algorithms for the Prediction of Coronary Artery Disease,” *J. Data Anal. Inf. Process.*, vol. 08, no. 02, pp. 41–68, 2020, doi: 10.4236/jdaip.2020.82003.