International Journal of Electrical Engineering & Emerging Technology, Vol. 08, No. 01, JAN 2026, pp 01-10

Machine Learning-based Analysis, Diagnosis and Prediction of Cardiovascular Diseases

[a] **Manahil Siddique,** [a] **Laraib Kehar,** [b] **Maham Mahnoor,** [a] **Ansaha Bhatti,** [a] **Sarmad Shams**
[a]Institute of Biomedical Engineering and Technology, LUMHS, Jamshoro, Pakistan
Corresponding author e-mail: manahil.siddique19@gmail.com
[**Received on: 31-07-2025 Accepted on**: 06-01-2026 **Published on:** 06-01-2026]

***Abstract*—cardiovascular diseases encompass a variety of conditions that impact the heart and blood vessels, presenting significant global health challenges. The condition of CVD patient's diagnosis, detection, prediction and analysis are difficult tasks. In the dataset, the CVD medical attributes base creates a machine learning (ML) classifier (algorithms) model, which mainly targets patients who are more likely to have or not have a heart disease. Supervised ML models are given by decision trees, random forests, neural networks, k-nearest neighbors, support vector machines and logistic regression, which help in creating a model frame and plotting the dataset CVDs for processing and performance analysis of heart patients. Effected factors, such as coronary artery disease, stroke and heart failure, are among the primary causes of illness and death worldwide. This model is a useful approach to improve the accuracy rate of prediction of the condition of the patient. These ML model, which are very comfortable, cheaper and advanced technology in the medical sector, help in detecting the accurate condition of heart failure patients for diagnosis and predicting the current situation of the patient and the accuracy rate of classifier performance is 75.41% up to 90.16%**

Index terms*—*CVD patients, ML classifier model and performance analysis.

I. INTRODUCTION

Heart failure is often known as chronic heart failure. Heart failure is a condition that develops when a patient's heart does not pump enough blood for the body's needs and Cardiovascular diseases encompass a wide range of conditions that impact the heart and blood vessels[1] [2] and they remain a leading cause of mortality worldwide.

According to the World Health Organization (WHO), CVDs comes number one as fatal illness globally; more people die annually from CVDs than from any other cause. An estimated

17.7 million people died from CVDs, representing 31% of all global deaths. Of these deaths, an estimated 7.4 million were because of coronary heart disease and 6.7 million were due to stroke. Over three-quarters of CVD deaths take place in low and middle-income countries. Out of the 17 millions premature deaths (under the age of 70) due to no communicable diseases, 82% are in underprivileged countries and CVDs cause 37% [3][4].

Traditionally, cardiovascular diseases have been diagnosed through clinical assessments, electrocardiograms (ECG), echocardiography, stress tests and blood tests. While these methods are effective to an extent, they rely heavily on the clinicians expertise and the patient's symptom presentation. Early stage heart failure is often asymptomatic or presents with vague symptoms, leading to delayed diagnosis. Furthermore, the manual interpretation of test results can sometimes result in misdiagnosis or overlooked early warning signs due to human error or variability in clinical judgment[5][6].

These limitations highlight the need for more accurate, consistent and early diagnostic approaches. This is where machine learning models have demonstrated significant promise. Due to the unpredictable and complex nature of heart failure characterized by nonspecific symptoms and minimal early manifestation machine learning algorithms can analyze vast amount of patient data to detect patterns and risk that might not be evident through traditional methods. By utilization patient records, includes information on age, sex, chest pain, blood pressure, cholesterol, fasting blood glucose, resting electrocardiographic, thalach, Exang, old peak, slope and coronary artery (Ca) thal[1]. Machine learning model can provide reliable predictions and support early detection. This not only enhance diagnostic precision but also contributes to improved patient outcomes and quality of life[7][8][9][10].

II. METHODLOGY

The patient of CVDs dataset utilized in this study was aqiure from the Public Health Dataset[11]. Which received approximately 26.5K views over the past 30 days. This dataset includes patient records related to heart disease detection. It contains a total of 76 attributes and is divided into 14 subsets. A detailed description of the dataset's columns and their respective values is provided in Table 1[12]. The experimental methodology follow is illustrated in Fig. 1[8]

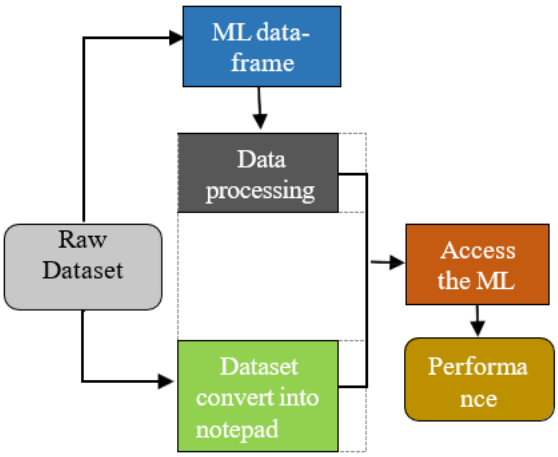

Figure 1. Illustration of Analysis, Diagnosis and Prediction of CVDs

III. RAW DATASET

The public health raw dataset of heart failure clinical patient records is in the form of an Excel sheet in which all attributes are set.

IV. NOTEPAD

Notepad is a text editor programs commonly found on Microsoft Windows operating systems, used to view and edit plain text flies with a text file extension. The CVDs dataset, originally in Excel format, is convert into text format and the resulting content is displayed as plain text. Fig. 2 illustrates the CVDs dataset in notepad form.

```
age,sex,cp,trestbps,chol,fbs,restecg,thalach,exang,oldpeak,slope,ca,thal,target
63,1,3,145,233,1,0,150,0,2.3,0,0,1,1
37,1,2,130,250,0,1,187,0,3.5,0,0,2,1
41,0,1,130,204,0,0,172,0,1.4,2,0,2,1
56,1,1,120,236,0,1,178,0,0.8,2,0,2,1
57,0,0,120,354,0,1,163,1,0.6,2,0,2,1
57,1,0,140,192,0,1,148,0,0.4,1,0,1,1
56,0,1,140,294,0,0,153,0,1.3,1,0,2,1
44,1,1,120,263,0,1,173,0,0,2,0,3,1
52,1,2,172,199,1,1,162,0,0.5,2,0,3,1
57,1,2,150,168,0,1,174,0,1.6,2,0,2,1
54,1,0,140,239,0,1,160,0,1.2,2,0,2,1
48,0,2,130,275,0,1,139,0,0.2,2,0,2,1
49,1,1,130,266,0,1,171,0,0.6,2,0,2,1
64,1,3,110,211,0,0,144,1,1.8,1,0,2,1
58,0,3,150,283,1,0,162,0,1,2,0,2,1
50,0,2,120,219,0,1,158,0,1.6,1,0,2,1
58,0,2,120,340,0,1,172,0,0,2,0,2,1
```

Figure 2. Show the Excel-based dataset convert into a plain text (.txt) file, allowing it to be viewed and edited in Notepad

V. MACHINE LEARNING

The ML modeling is designed to process and analyze input data to generate accurate diagnostic outcomes. To build this model, a structured dataset is used, containing multiple attributes relevant to heart disease risk factors. Each column in the dataset represents a medical or demographic features associated with a patient's cardiovascular health, in Table.1 illustrate the description of the dataset's features

This feature serve as critical inputs in the training phase of model. For instance, attributes like chest pain type, resting blood pressure and blood cholesterol have long been recognized as clinical indicators for heart related conditions. Similarly, thalach (maximum heart rate achieved) and exercise induced angina provide insights into cardiovascular function under stress. The inclusion of categorical variable such as sex adds diversity to the dataset and help in building more personalized predictions[12].

The Supervised ML models included decision tree, random forest, neural networks, k-nearest neighbors, support vector machines[13]. Building a ML model has a few key steps: data preprocessing, data analysis, model training, testing and splitting the data, evolution and visualization[14]. Each step plays a major role in building an accurate and robust ML model by ensuring the data is correctly prepared, the model is effectively trained and the results are thoroughly evaluated and interpreted. This process help in achieving a reliable accuracy rate for the diagnosis, detection, prediction and analyzation by CVDs dataset[6]. The ML preprocessing phase is divided into two parts; firstly, building the ML framework; second, applying the framework for access the CVDs dataset and performance analysis[15].

Table 1. The table dataset arranges 14 subsets in Column, Description and Values

Column	Description	Values
age	Age	Last
sex	Gender	Male/Female (0/1)
cp	Chest pain(four types)	4 values(0 - 3)
trestbps	Resting Blood Pressure	Continuous
chol	Serum Cholesterol in mg/dL	mg/dL
fbs	Fast Blood sugar	>120 mg/dL
restectg	Resting Electrocardiogram	03 values(0 - 2)
thalach	Maximum Heart Rate Achieved	Continuous
exang	Exercise Induce angina	02 value(0 , 1)
oldpeak	ST depression induced by exercise relative to rest	Continuous
slope	The slop of the peak exercise ST segment	03 values (0 - 2)
ca	Number of major vessels colored by fluoroscopy	04 values(0 - 3)
thal	Thalassemia defect types (normal, fixed defect, reversible defect)	03 values (0 - 2)
target	Output show	02 values(0/1)

A. Decision Tree(DT)

It is graphically organized like a tree in which choosing the results results and making a decision tree in the form of tree which help in solving the classification problems and look like an upside to down tree that illustrated in figure.3 3 [1] decision tree algorithm starts at the root node; it is helping in presenting the CVDs dataset [2]. The feature of making decisions by

internal nodes. It is classified each attribute of CVDs dataset and represented in a group. [3] The branches

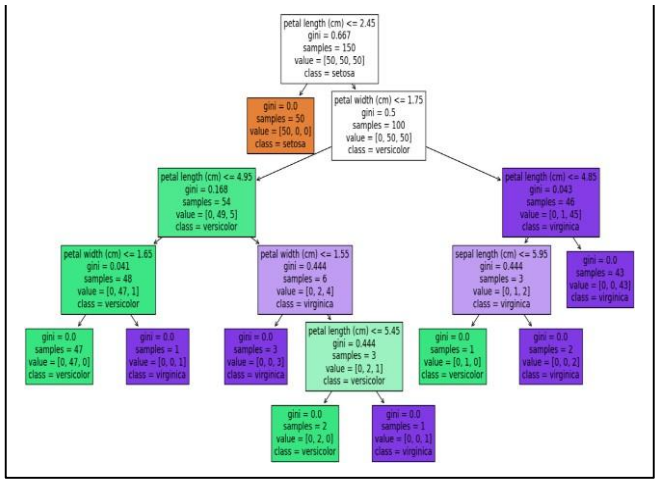

Figure 3. Graphically Organized of DT in which choosing the results and helping in solving the classification problems

help in decision making of CVDs dataset that internal nodes have. It follows the branch to the next node continuously repeating the process until it reaches the leaf nodes [4]. The leaf nodes give the final output or classification for CVDs dataset. It's feature importance shown in figure.6.

B. Random Forest (RF)

RF operates by utilization multiple decision tree during model training, where each individual tree contributes to the final classification and regression results. This ensemble approach significantly improve model performance and prediction accuracy. The significance of every feature in RF model is highlighted in figure 6, which emphasizes the need for high quality input data to enhance model performance. The feature is to improve the quality of the input dataset, which leads to better model performance. Additionally, Figure 7 presents the confusion matrix, which visualizes the comparison between the true values and predicted values of the model outputs.

C. Logistic Regression (LR)

A mathematical use if binary classification of two classes is LR, used to predict the probability of CVDs dataset target variable (0 or 1 / true or false) and figure.6 represents the logistic regression feature importance. LR confusion matrix with labels and their true and predicted value is illustrated in Fig. 7.

D. K-Nearest Neighbors (KNN)

The KNN algorithm is used to train the CVDs dataset, solve the classification and regression problems and store the dataset in memory. This supervised ML algorithm accuracy output is on the basis of different distance metrics like Euclidean distance, Q norm distance, Manhattan distance and the varying value of K shown in figure. 4, in figure. 7 represents the confusion matrix with labels for their true and predicted values. K-NN of Each attribute of the dataset features importance shown in the figure. 6

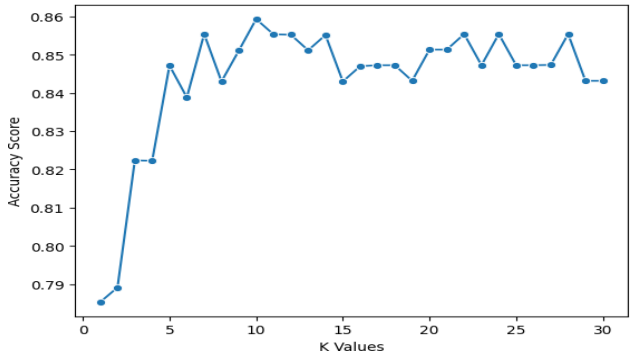

Figure 4. Shows the K-NN accuracy score and K value of dataset

E. Neural Networks (NN)

This algorithm has layers of interconnected neurons, which process the input and generate the output. The neural network imports the MLP. The architecture of NN have three layers [1] input layer have initial dataset CVDs, which have 4 neurons and fully interconnected (4×64) with hidden layer [2] Hidden layer activation function is mathematical function applied to each neuron in a NN after calculating its weight sum of input and it's represent the connection magnitudes between neurons in adjacent layers. Hidden layer 1 have 64 neurons, ReLU that fully interconnected (64×32), Hidden Layer 2 has 32 neurons, ReLU that is fully interconnected (32×3), and [3] Output Layer have 3 neurons, Softmax. Figure.5 illustrates the NN architecture, while Figure.6 highlights the importance of specific feature in classification. Figure.7 present the confusion matrix.

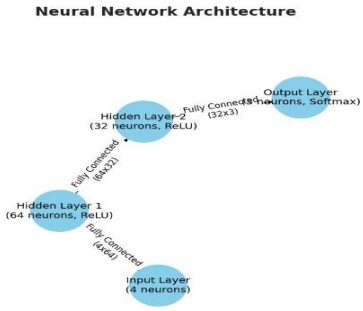

Figure 5. Neural Network architecture layers process the input and output of dataset

F. Support Vector Machines (SVM)

The SVM algorithm helps in finding a decision boundary or hyperplane that best separates different classes in a dataset. It is used for training models for classification, regression and outlier detection. For the nonlinear case, it maps the data into higher-dimensional spaces with the help of kernel functions, enabling the creation of a linear decision boundary based on the attributes. In figure.7, represent the confusion matrix with labels their true and predicted values and feature importance shown in figure.6[16].

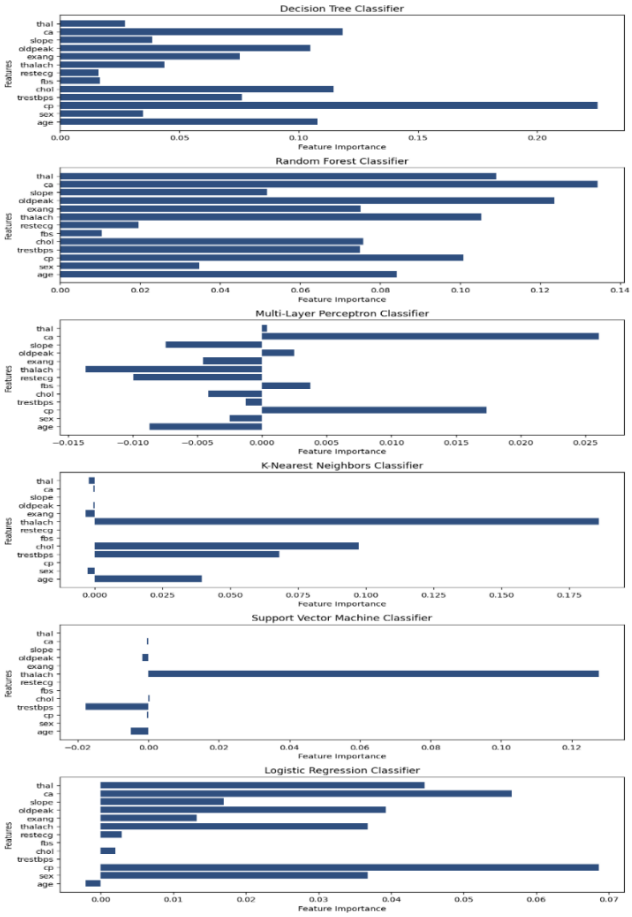

Figure 6. Feature importance of algorithms plot the input feature contributes for measurement of each classifiers prediction or accuracy of model

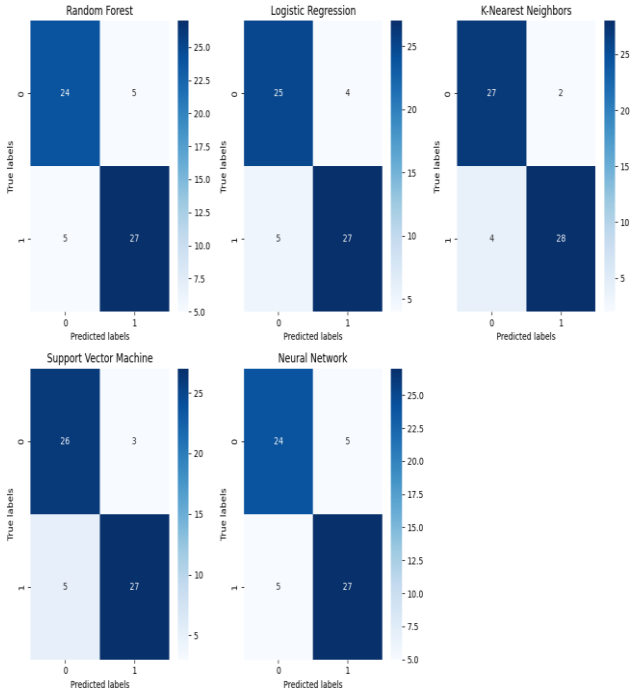

Figure 7. Confusion matrix of all classifier plot and its true and predicted values are label

VI. DATA PROCESSING

To assess algorithm performance on the CVDs dataset for early diagnosis of whether the heart failure disease has or not, this involves collecting and processing data to create a robust data frame. Table.2 shows the frame for getting information about CVDs dataset. Data preprocessing steps included [1] Table.3 describes the missing values in dataset and table.4 shows the addressing missing data [2] outlier removal Table.5 provides information about the outliers in each column of the CVDs dataset. It is important preprocessing step in model to help in performance and more robust statistical method to handle the extreme values [3] statistical measures about CVDs dataset are represented in Table.6 [4] Distribution of target variable before and after balancing in Table.7 [5] In Table.8 Converting categorical data to dummy variables and It's also known as one-hot encoding [6] Fig.8 represent the analyzing feature correlation of dataset.

Table 2. Information about dataset

S#	column	non-null Count	D type
0	age	303 non-null	int64
1	sex	303 non-null	int64
2	cp	303 non-null	int64
3	trestbps	303 non-null	int64
4	chol	303 non-null	int64
5	fbs	303 non-null	int64
6	restecg	303 non-null	int64
7	thalach	303 non-null	int64
8	exang	303 non-null	int64
9	oldpeak	303 non-null	int64
10	slop	303 non-null	int64
11	ca	303 non-null	int64
12	thal	303 non-null	int64
13	target	303 non-null	int64

Table 3. Missing value in the dataset

Attributes	missing values
Age	0
Sex	0
cp	0
trestbps	0
chol	0
fbs	0
restectg	0
thalach	0
exang	0
oldpeak	0
slope	0
ca	0
thal	0
target	0
dtype: int64	0

Table 4. Outlier Removal help in performance and more robust statistical method to handle the extreme values

Method	Count
Mean	0
Median	0
Mode	0
Forward Fill	0
Backward Fill	0
Drop rows	0

Table 5. Addressing missing data

Colum		Q1	Q2	IQR	Lower Bound	Upper Bound	Number of Outliers
age		47.5	61.0	13.5	27.25	81.25	0
sex		0.0	1.0	1.0	-1.50	2.50	0
cp		0.0	2.0	2.0	-3.00	5.00	0
trestbps		12.0	140.0	20.0	90.00	170.00	9
chol		211.0	274.5	63.5	115.75	369.75	5
fbs		0.0	0.0	0.0	0.00	0.00	45
restectg		0.0	1.0	1.0	-1.50	2.50	0
thalach		133.5	166.0	32.5	84.75	214.75	1
exang		0.0	1.0	1.0	-1/50	2.50	0
oldpeak		0.0	1.6	1.6	-2.40	4.00	5
slope		1.0	2.0	1.0	-0.50	3.50	0
ca		0.0	1.0	1.0	-1.50	2.50	25
thal		2.0	3.0	1.0	0.50	4.50	2
target		0.0	1.0	1.0	-1.50	2.50	0

Table 6. Statistical measure

Statistical	age	sex	cp	trestbps	chol	fbs	restecg	thalach	exang	oldpeak	slope	ca	thal	target
Count	303	303	303	303	303	303	303	303	303	303	303	303	303	303
Mean	54.36	0.68	0.96	131.62	246.26	0.148	0.52	149.64	0.32	1.03	1.39	0.72	2.31	0.54
Std	9.08	0.46	1.03	17.53	51.83	0.35	0.52	22.90	0.46	1.16	0.61	1.02	0.61	0.49
Min	29.00	0.00	0.00	94.00	126.00	0.00	0.00	71.00	0.00	0.00	0.00	0.00	0.00	0.00
25%	47.50	0.00	0.00	120.00	211.00	0.00	0.00	133.50	0.00	0.00	1.00	0.00	2.00	0.00
50%	55.00	1.00	1.00	130.00	240.00	0.00	1.00	153.00	0.00	0.80	1.00	0.00	2.00	1.00
75%	61.00	1.00	2.00	140.00	274.50	0.00	1.00	166.00	1.00	1.60	2.00	1.00	3.00	1.00
max	77.00	1.00	3.00	200.00	564.00	1.00	2.00	202.00	1.00	6.20	2.00	4.00	3.00	1.00

Table 7. Distribution of target variable

Class	Before Balancing	After Balancing
1	165	165
0	138	165

Table 8. Converting categorical data to dummy variables

SN	age	sex	cp	trestbps	chol	fbs	restectg	thalach	exang	oldpeak	slope	ca	thal
298	57	0	0	140	241	0	1	123	1	0.2	1	0	3
299	45	1	3	110	264	0	1	132	0	1.2	1	0	3
300	68	1	0	144	193	1	1	141	0	3.4	1	2	3
301	57	1	0	130	131	0	1	115	1	1.2	1	1	3
302	57	0	1	130	236	0	0	174	0	0.0	1	1	2

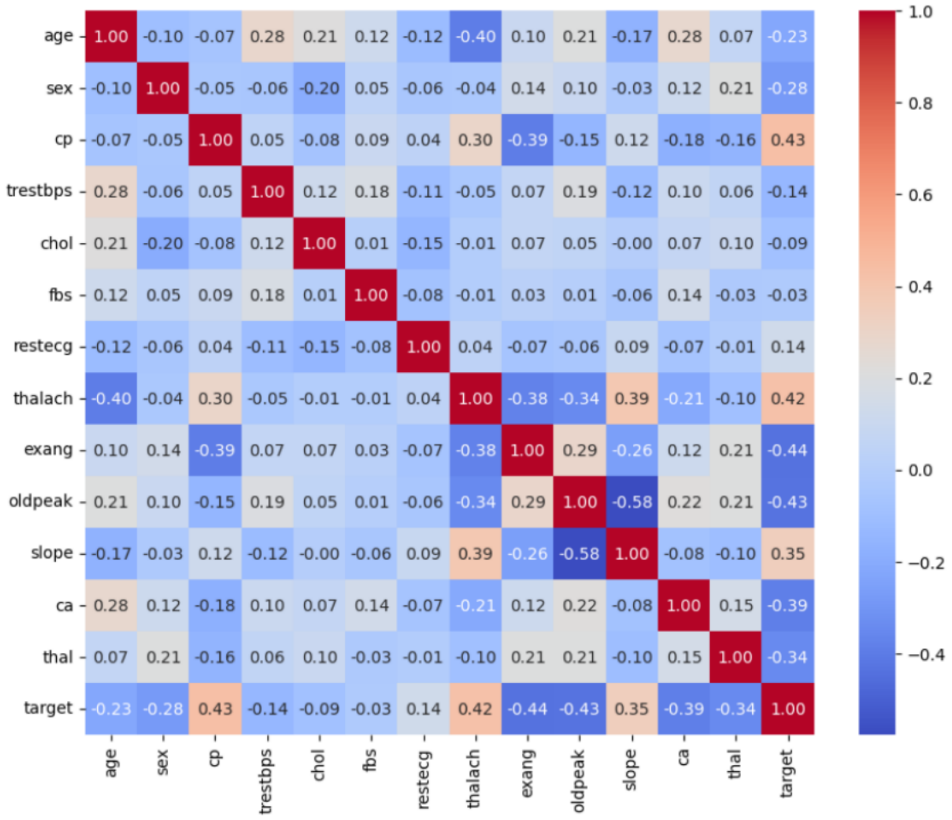

Figure 8. Feature correlation measures the relationship between all inputs of dataset and determine the how closely they vary together

A. Confusion Matrix

It's helpful for calculating attributes of dataset CVDs and calculating the actual accuracy or predictor for that all attributes. The confusion matrix is divided into two classes: predicted class and actual class[17]. The supervised ML algorithms that are used in these model calculate their [1] a confusion matrix is a practical key for classification problems where the output can belong to two or more classes. In Table.9, display the confusion matrix, which evaluates the performance of supervised ML algorithms on a test dataset with known true values. For binary classification, the matrix includes: **True Positive (TP), True Negative (TN), False Positive (FP), and False Negative (FN)** [2] Plotted confusion matrix shown in Fig.7

[3] Classification report of confusion matrix and shown their accuracy rate. The performance system of training and testing scores of algorithms shown in Fig. 9 Table.10 illustrates the prediction made by six different ML algorithms on the CVDs data[9][18].

Table 9. Confusion Matrix

	Predicted Positive	Predicted Negative
Actual Positive	True Positive (TP)	False Negative (FN)
Actual Negative	False Positive (FP)	True Negative (TN)

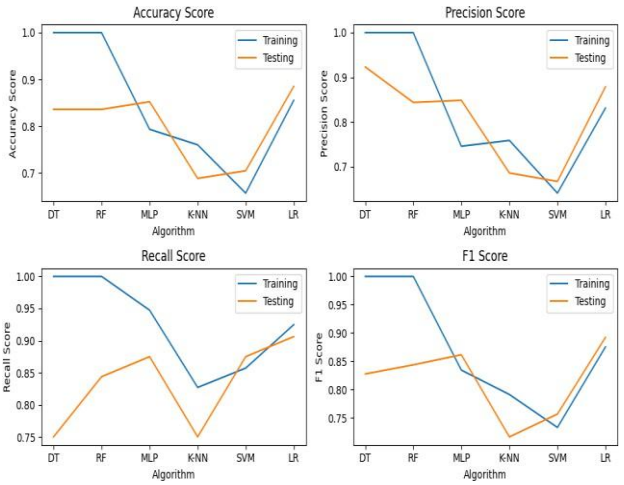

Figure 9: Performance system of training and testing scores of algorithms

Table 10. Prediction Table of CVDs dataset by Using Six ML Algorithms

S#	Algorithm	Accuracy (%)	Precision (%)	Recall (%)	F1 Score (%)
1	Training – DT Classifier	100.00	100.00	100.00	100.00
2	Training – RF Classifier	100.00	100.00	100.00	100.00
3	Training – MLP Classifier	80.99	84.02	79.40	79.83
4	Training – K-NN Classifier	76.03	76.08	75.30	75.49
5	Training – SVC	65.70	67.18	63.50	62.67
6	Training – LR	85.54	86.23	84.77	85.15
7	Testing – DT classifier	83.61	84.73	84.05	83.57
8	Testing – RF classifier	83.61	83.57	83.57	83.57
9	Testing – MLP classifier	88.52	89.01	88.25	88.41
10	Testing – K-NN classifier	68.85	68.90	68.53	68.55
11	Testing – SVC	70.49	72.81	69.61	69.09
12	Testing – LR	88.52	88.58	88.42	88.48

B. Accuracy

The rate of correct predictions from the total predictions. Accuracy provides an insight to the basic query: at what rate of time does the model make correct predictions. Measuring the accuracy of the CVD dataset using the calculation expression shown in eq.1[19][20]

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \qquad (eq.\,1)$$

C. Precision

Rate of correct positive predictions from all positive predictions. Measuring the precision of the CVD dataset using the calculation expression shown in eq.2

$$Precision = \frac{TP}{TP + TF} \qquad (eq.\,2)$$

D. Recall (Sensitivity or True Positive Rate)

Rate of correct positive predictions from all actual positives. Recall measures the completeness of the results, with high scores indicating that most relevant instances are captured, even if some irrelevant ones are also included. To measure the Recall of CVDs dataset using the calculation expression shown in eq.3

$$Recall = \frac{TP}{TP + FN} \qquad (eq.3)$$

E. F1 Score

F1 score is used to assess the performance of CVDs dataset. It represents the harmonic mean of precision and recall, providing a balanced measure. Calculation expressed in eq. 4

$$F_1 = \frac{2 * Precision * Recall}{Precision + Recall} \qquad (eq.4)$$

VII. Results and Discussions

The ML classifiers preprocessing the data-frame for CVDs dataset perform of feature importance, Correlation, training and testing and confusion matrix of algorithms which are helpful for calculating the attributes and calculating the actual accuracy or predictor for that all attributes. In Fig.10 represent the accuracy percentage and performance of all classifiers. The highest accuracy rate has K-NN classifier 90.16% and the lowest has a DT classifier 75.41%. In Fig.11 represent the precision percentage and performance of all classifier. The highest precision rate has MLP classifier 90% and the lowest has an SVC classifier 66%. Fig.12 represents the Recall percentage and performance of all classifier. The highest Recall rate have LR classifier 90% and the lowest has a DT classifier 71%. Fig.13 represents the F1 Score percentage and performance of all classifiers. The highest rate has an LR classifier of 89% and the lowest has a K-NN classifier of 71%. Fig.14 shows the net result of the CVDs dataset: how much patients are suffering (effected) from heart disease and how much are not suffering (non-effected) from heart disease.

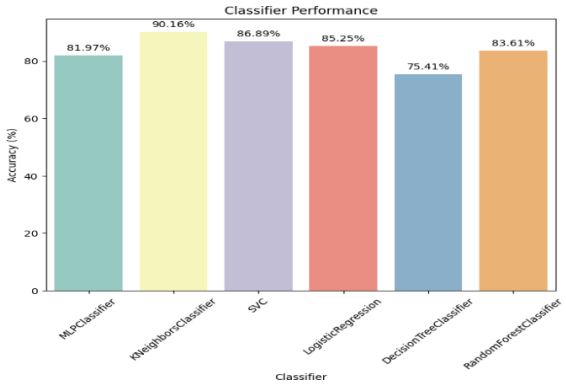

Figure 10. Accuracy of algorithms performance

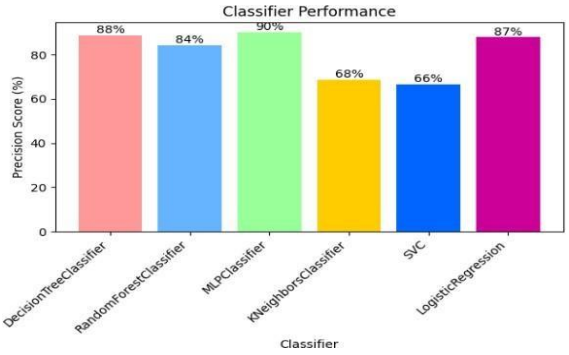

Figure 11. Precision of algorithms performance

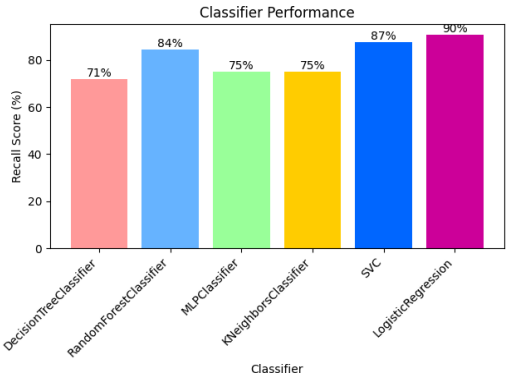

Figure 12. Recall of algorithms performance

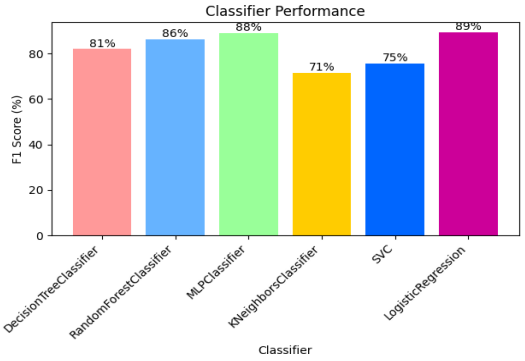

Figure 13. Score of algorithms performance

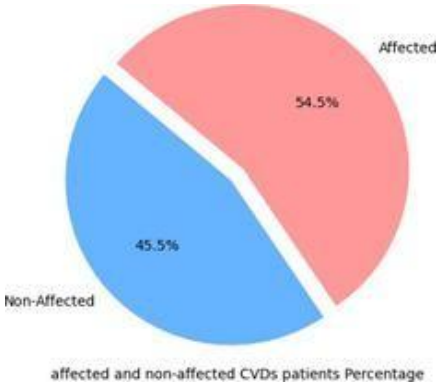

Figure 14. Net result of target of dataset how many patients are affected and how many non – affected with heart diseases

VIII. COMPARISON

The compared the supervised ML model algorithms results with highest results of pervious papers illustrate in Fig.11.

Table 11. Comparison result with previous papers

Previous Paper	Accuracy%	Precision%	Recall%	F1 Score%
Proposed model DT classifier	75.41	88	71	81
Sanni, R. R., & Guruprasad, H. S. (2021)	85.33	73	70	-
Senan, E. M., Abunadi, I., Jadhav, M. E., & Fati, S. M. (2021)	96.46	97.11	100	98.53
Proposed model RF classifier	83.61	84	84	86
Sanni, R. R., & Guruprasad, H. S. (2021)	81.66	78.26	97.29	-
Proposed model K-NN classifier	90.16	68	75	71
Sanni, R. R., & Guruprasad, H. S. (2021).	68.88	68.88	100	-
Senan, E. M., Abunadi, I., Jadhav, M. E., & Fati, S. M. (2021)	96.82	95.76	98.51	97.12
Proposed model SVC classifier	86.89	66	87	75%
Senan, E. M., Abunadi, I., Jadhav, M. E., & Fati, S. M. (2021)	92.35	95.41	96.10	95.75
Proposed model LR classifier	85.25	87	90	89%
Sanni, R. R., & Guruprasad, H. S. (2021)	81.66	78.26	97.29	-
Senan, E. M., Abunadi, I., Jadhav, M. E., & Fati, S. M. (2021)	91.05	94.52	92.39	93.44
Proposed model MLP classifier	81.97	90	75	88

IX. CONCLUSION

The ML – based approach was implemented for the diagnosis, detection, prediction and analysis of CVDs using clinical patient data. The development of ML classifiers involve a structure sequence of steps, including data preprocessing, analysis, training, testing, feature selection, correlation study and performance evaluation using confusion matrix metrics – such as accuracy, precision, recall and F1 – score. The results demonstrated that ML models, particularly supervised learning algorithms, can significantly enhance the predictive accuracy of CVDs detection, offering

accuracy rates ranging from 75.41% to 90.16%. These models provide a reliable decision support system that assists clinicians by analyzing patient histories and by identifying those at high risk. Ultimately, this approach improve early diagnosis and patient outcomes, offering a cost effective and efficient solution for both healthcare provider and patients.

X. ACKNOWLEDGMENT

We acknowledge the support and guidance of our supervisor Engr. Laraib Kehar who has always supported and encouraged us. She has guided us throughout the thesis. Our sincere gratitude to Prof. Maham Mahnoor & Head Of Department Dr. Sarmad Shams for their kind interests, valuable guidance, and encouragement.

Lastly, an acknowledgment to our family and friends for giving us moral support and to keeping our spirit high for overcoming difficulties. In addition, We are also obliged to all our respected teachers at the Institute of Biomedical Engineering and Technologies, for sharing their valuable suggestions for this thesis.

REFERENCES

[1] N. R. Colledge, B. R. Walker, and S. H. Ralston, "Davidson's Principles and Practice of Medicine 21st Edition."

[2] L. Holifield, "History and Epidemiology," *Princ. Nurs. Pract. Era COVID-19*, vol. 320, no. January, p-p 1–15, 2022, doi: 10.1007/978-3-030-94740-8_1.

[3] "Cardiovascular diseases (CVDs)." Accessed: Jan. 19, 2025. [Online]. Available: https://www.who.int/news-room/fact-sheets/detail/cardiovascular-diseases-(cvds)

[4] S. Kaptoge *et al.*, "World Health Organization cardiovascular disease risk charts: revised models to estimate risk in 21 global regions," *Lancet Glob. Heal.*, vol. 7, no. 10, pp. e1332–e1345, 2019, doi: 10.1016/S2214-109X(19)30318-3.

[5] B. D. Aderounmu, "DETECTION OF CARDIOVASCULAR DISEASES IN PATIENTS USING MACHINE By ADEROUNMU BABAJIDE December 2022," no. February, pp. 0–47, 2023.

[6] B. Mahesh, "Machine Learning Algorithms - A Review," *Int. J. Sci. Res.*, vol. 9, no. 1, pp. 381–386, 2020, doi: 10.21275/art20203995.

[7] G. Boucher, "Book Reviews: Book Reviews," *Crit. Sociol.*, vol. 37, no. 4, pp. 493–497, 2011, doi: 10.1177/0261018311403863.

[8] V. Ravi and K. Kolla, "Heart Disease Diagnosis Using Machine Learning Techniques in Python : a Comparative Study of Classification," vol. 6, no. 9, pp. 98–118, 2015.

[9] V. M. Patro and M. Ranjan Patra, "Augmenting Weighted Average with Confusion Matrix to Enhance Classification Accuracy," *Trans. Mach. Learn. Artif. Intell.*, vol. 2, no. 4, 2014, doi: 10.14738/tmlai.24.328.

[10] R. R. Sanni and H. S. Guruprasad, "Analysis of performance metrics of heart failured patients using Python and machine learning algorithms," *Glob. Transitions Proc.*, vol. 2, no. 2, pp. 233–237, 2021, doi: 10.1016/j.gltp.2021.08.028. "Cardiovascular Disease dataset." Accessed: Jan. 19, 2025. [Online]. Available: https://www.kaggle.com/datasets/sulianova/cardiovascular-disease-dataset

[11] "Computational and Mathematical Methods in Medicine - 2021 - Senan - Score and Correlation Coefficient-Based Feature.pdf."

[12] M. Pal, S. Parija, G. Panda, K. Dhama, and R. K. Mohapatra, "Risk prediction of cardiovascular disease using machine learning classifiers," Open Med., vol. 17, no. 1, pp. 1100–1113, 2022, doi: 10.1515/med-2022-0508.

[13] E. Ahmad, "Cardiovascular Diseases (CVDs) Detection using Machine Learning Algorithms," Int. J. Res. Appl. Sci. Eng. Technol., vol. 8, no. 6, pp. 2341–2346, 2020, doi: 10.22214/ijraset.2020.6376.

[14] I. Z. Sadiq, F. S. Abubakar, M. A. Saliu, B. Sanusi, A. Salihu, and A. Muhammad, "Machine learning algorithms for predictive modeling of dyslipidemia - associated cardiovascular disease risk in pregnancy: a comparison of boosting, random forest, and decision tree regression," Bull. Natl. Res. Cent., 2025, doi: 10.1186/s42269-024-01295-y.

[15] M. Palaniswami, A. Shilton, D. Ralph, and B. D. Owen, "Machine learning using support vector machines," Proc. - Int. Conf. Artif. Intell. Sci. Technol., no. May, pp. 1–8, 2000, [Online]. Available: http://people.eng.unimelb.edu.au/shiltona/publications/aisat2000.pdf

[16] M. B. Nirmala, K. R. Haarika Reddy, M. Sah, S. Shastry, and V. V. Murthy, "Cardiovascular disease prediction using machine learning algorithms," 12th Int. Conf. Adv. Comput. Control. Telecommun. Technol. ACT 2021, vol. 2021-Augus, no. June, pp. 208–214, 2021, doi: 10.17762/turcomat.v12i6.2426.

[17] J. Davis and M. Goadrich, "The Relationship between PR and ROC curves," 2019. [Online]. Available: http://portal.acm.org/citation.cfm?doid=1143844.1143874

[18] V. Ramesh, M. S. Das, and B. N. Rao, Heart Disease Detection and Prediction Using ML Algorithms in Python, vol. 1. Atlantis Press InternationalBV, 2023. doi: 10.2991/978-94-6463-252-1_38.

[19] H. Jindal, S. Agrawal, R. Khera, R. Jain, and P. Nagrath, "Heart disease prediction using machine learning algorithms," IOP Conf. Ser. Mater. Sci. Eng., vol. 1022, no. 1, pp. 0–10, 2021, doi: 10.1088/1757- 899X/1022/1/012072.