Journal of Engineering and Technology for Industrial Applications

ISSN ONLINE: 2447-0228

ITEGAM-JETIA

Manaus, v.12 n.57, p. 115-123. January/February, 2026.

DOI: https://doi.org/10.5935/jetia. v12i57.2771

RESEARCH ARTICLE OPEN ACCESS

CARDIOVASCULAR RISK PREDICTION USING AI

Mehek Mohd Meraj Qureshi*[1], Hariram Chavan[2]

[1]B.E. degree in Information Technology from M.H. Saboo Siddik College of Engineering at Mumbai University, Mumbai. I am currently pursuing my postgraduate degree in Artificial Intelligence (AI) from K.J Somaiya Institute of Technology, Mumbai..
[2]A Professor in the Department of Information Technology at K.J Somaiya Institute of Technology, Mumbai. He holds a B.E (Computer Engineering) and M.E (Computer Science and Engineering) from WCE Sangli, Shivaji University. He holds a PhD in Computer Engineering from VJTI, Mumbai University.

[1]https://orcid.org/0009-0001-6505-5403, [2]https://orcid.org/0000-0003-1876-6971

Email: *mehek.q@somaiya.edu, hariram@somaiya.edu

ARTICLE INFO

Article History
Received: October 2, 2025
Revised: November 20, 2025
Accepted: January 1, 2026
Published: January 31, 2026

Keywords:
Machine Learning
Explainable AI
Predictive Systems
Healthcare
Cardiovascular Disease.

ABSTRACT

A silent epidemic, cardiovascular diseases (CVDs) continue to be the primary reason for loss of life around the globe. This emphasizes the critical need for early-stage risk prediction models for CVDs. This implementation aims to use an existing model for CVD risk prediction, which will be AI-driven and reliable. The main objective is to use machine learning techniques on medical datasets, it also focuses on transparency of the model along with user experience to ensure its implementation in the healthcare sector. The model utilized key clinical features that were selected using distinct feature selection methods, namely analysis of variance, mutual information, and chi-square. The feature groups selected were CRF-1, CRF-2, CRF-3, CRF-4, and CRF-5, respectively. The dataset was preprocessed with SMOTE (Synthetic Minority Oversampling Technique), which helps to resolve class imbalance. All machine learning algorithms were evaluated using Stratified K-Fold Cross Validation and Randomized Search Cross Validation. The final model's interpretability was enhanced by integrating Explainable AI (XAI) techniques, especially Shapley Additive Explanations (SHAP) and Local Interpretable Model-agnostic Explanations (LIME). Among the algorithms tested, the Multi-Layer Perceptron (MLP) Classifier with CRF-1 achieved an AUC of 0.98 with an accuracy of 96.72%. This makes the prediction more interpretable for the healthcare professionals and patients.

I. INTRODUCTION

The implementation focuses on the cardiovascular system present in the body. The cardiovascular system has the responsibility of circulating oxygenated blood in the body. If this cardiovascular system malfunctions, it can lead to various cardiovascular diseases (CVDs). These malfunctions arise from factors like blocked arteries or impaired heart function and thus pose significant health risks across the globe. This highlights the condition's prevalence on the rise, and it is therefore crucial to implement effective strategies to manage and reduce its impact. The rapid integration of computer science into healthcare has shown significant advancements in biomedical devices. These biomedical devices are responsible for generating vast amounts of complex medical data. The traditional manual analysis of these data is not only time-consuming but also inefficient.

Machine Learning (ML) models are capable of processing these large medical datasets and extracting meaningful insights, thus helping healthcare professionals in the decision-making process. Significant advancements have been made in recent years in machine learning to develop predictive models for cardiovascular disease by leveraging both private and publicly available datasets. Some research has successfully addressed data imbalance using techniques like SMOTE. [1] Similarly, other studies have highlighted how the integration of ML is helpful in identifying the early symptoms of cardiovascular risks. The evaluation of various algorithms has demonstrated ML's capacity to enhance predictive accuracy and assist healthcare professionals. [2]

Building on these foundations, the Cardiovascular Risk Prediction (CVRP) model aims to develop an ML model that analyzes various patient attributes and generates predictions to assist doctors and physicians as a pre-report. To ensure the reliability of our model, we assess the accuracy of multiple algorithms and fine-tune them for comparative analysis.

II. THEORETICAL REFERENCE

II.1 LITERATURE SURVEY

Numerous researchers have applied diverse ML techniques to predict CVDs. These researchers leverage the utilization of well-known datasets such as the Cleveland, UCI, and Framingham Heart Disease datasets. As detailed in Table 1, an author Hosam F. El Sofany et al. demonstrated the effectiveness of combining multiple classifiers like Decision Tree, SVM, Naïve Bayes, XGBoost, and AdaBoost for enhanced prediction. Another study in 2023, Mana Saleh Al Reshan et al. proposed a distinct method by introducing a hybrid deep learning-based architecture that used preprocessing and feature learning techniques to analyze these same public datasets.

Table 1: Related work.

Year	Author	Dataset Used	Technique used
2024	Hosam F. El-Sofany et al.[1]	Cleveland HD dataset, private HD dataset (from Egyptian specialized hospitals)	SVM, NB, DT, XGBoost, KNN, RF, Voting, AdaBoost
2023	Mana Saleh Al Reshan et al.[6]	Cleveland HD dataset	DL, Hybrid architecture, feature learning, data preprocessing
2021	Liu et al.[7]	UCI Heart Disease	RF, KNN, SVM, NB
2021	Aqsa Rahim et al [8]	Framingham, Cleveland HD and Cleveland	LR, KNN, SMOTE
2020	Hussein et al. [9]	Cleveland HD	DT, KNN
2020	Akbar et al. [10]	Cleveland HD	RF, SVM, Naïve Bayes

Source: Authors, (2026).

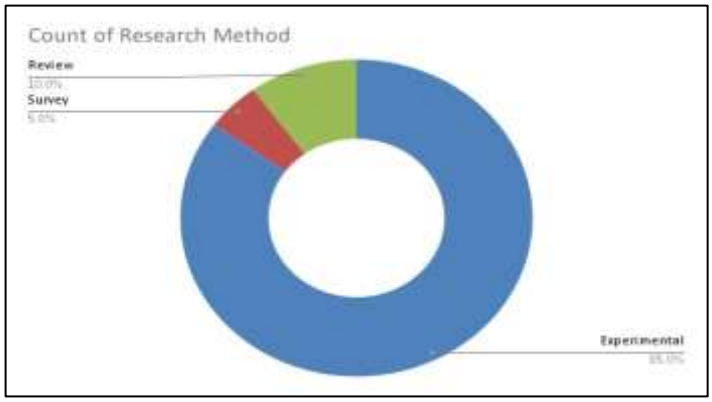

Figure 1: Distribution of different Research Methods in Research Papers
Source: Authors, (2026).

In the Figure 1, the distribution of different research methods used by researchers in 20 papers is shown. A total of 85% of the research is experimental, as shown by the largest slice. For review-based studies, it is a smaller portion of 10%. The smallest slice represents that a survey-based approach was used, accounting for 5%. Figure 2 provides a summary of ML algorithms that are most frequently used in the research papers. The size of each bar corresponds to the total number of research papers that have used a particular algorithm. This helps to identify the most prevalent algorithms in this specific domain. It also indicates which techniques are currently favored by the research community.

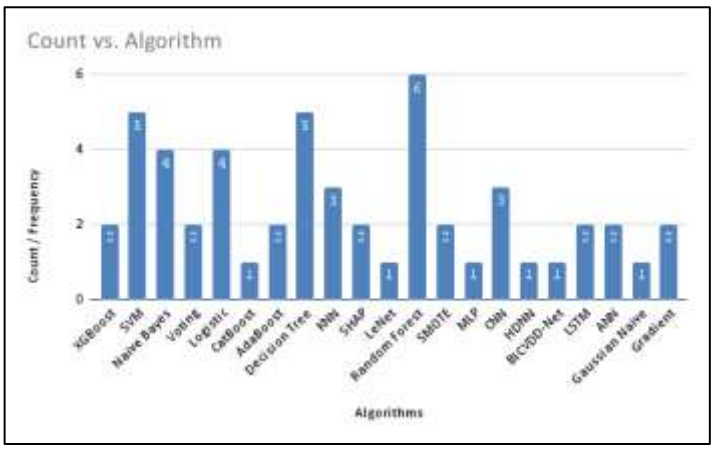

Figure 2: Frequency of Algorithms Used in Research Papers.
Source: Authors, (2026).

II.1.1 Working Principles of Classifiers

i) XGBoost / LightGBM: XGBoost/LightGBM train trees sequentially, correcting previous errors (boosting). LightGBM specifically optimizes by leaf-wise tree growth and uses Gradient-based One-Side Sampling (GOSS) for efficiency.
General form of prediction (As shown in equation 1)

$$\hat{y}_i = \sum_{k=1}^{K} f_k\,(xi), \qquad f_k \in F \tag{1}$$

where f_k are decision trees and K is the number of trees.
Objective function (As shown in equation 2):

$$\text{Obj} = \sum_{i=1}^{n} l\,(y_i, \hat{y}_i) + \sum_{k=1}^{K} \Omega\,(f_k) \tag{2}$$

where l is the loss and Ω is the regularization term, which helps to avoid overfitting. ii) MLP (Multi-Layer Perceptron): It is characterized by feedforward structure and multiple layers.

Forward pass equation (Refer to equation 3):

$$h^{(l)} = \sigma\,(W^{(l)} h^{(l-1)} + b^{(l)} \tag{3}$$

where $h^{(l)}$ is the output of layer l, W is weight matrix, b is bias, and σ activation (ReLU, sigmoid, etc.).
Final prediction (As shown in equation 4):

$$\hat{y} = \sigma(W^{(L)} h^{(L-1)} + b^{(L)} \tag{4}$$

Training uses backpropagation with gradient descent (represented by equation 5):

$$W \leftarrow W - \eta \frac{\partial L}{\partial W} \tag{5}$$

where L is the loss function and η is the learning rate.

iii) Random Forest: It builds multiple independent decision trees on different bootstrapped samples of the dataset. It trains all trees independently in parallel (bagging). It is represented in equation 6.

$$\hat{y} = \frac{1}{T} \sum_{t=1}^{T} h_t(x) \tag{6}$$

where T = number of trees, $h_t(x)$ is the prediction from tree t and $\hat{y}$ is final aggregated prediction. iv) AdaBoost: It combines weak learners into a strong classifier. After each round, it increases the weights of misclassified samples, and then the next learner is allowed to focus more on those "hard" cases. The final prediction is a result of the weighted vote of all the weak learners.
Mathematical Representation:

a. Assign initial equal weights:

$$w_i = \frac{1}{N},\ i = 1,2,\ldots,N \tag{7}$$

b. Train weak learner $h_m(x)$ Compute weighted error.

$$\epsilon_m = \frac{\sum_{i=1}^{N} w_i I(y_i \neq h_m(x_i)}{\sum_{i=1}^{N} w_i} \tag{8}$$

c. Compute learner weight (importance):

$$\alpha_m = \frac{1}{2} \ln\left(\frac{1-\epsilon_m}{\epsilon_m}\right) \tag{9}$$

d. Update sample weights:

$$w_i \leftarrow w_i \cdot e^{\alpha_m I(y_i \neq h_m(x_i))} \tag{10}$$

e. Final strong classifier:

$$H(x) = sign\left(\sum_{m=1}^{M} \alpha_m h_m(x)\right) \tag{11}$$

II.1.2 CVRP System Architecture

The CVRP model is designed to analyze the risk of having CVD using various ML classification algorithms. The system architecture comprises three main layers: Frontend, Backend, and Database (As shown in Figure 3). The database used is available on the Kaggle platform. It comprises of 13 key clinical features that help to analyze the potential risks of having CVD. These attributes include a patient's age, gender, chest pain types, maximum heart rate reached, cholesterol, and more. The frontend is designed to provide a user-friendly interface that can be used to input health data by healthcare professionals and patients. This was made possible by using Gradio, it allows user to enter health data and receive an instant output about their cardiovascular risk level. The backend is responsible for the core logic. It includes data preprocessing using SMOTE, the evaluation of the ML algorithms, and the selection of the MLP Classifier model based on performance. The backend also integrates XAI techniques, which helps maintain the transparency of the model. A downloadable PDF report is generated for a summary of results.

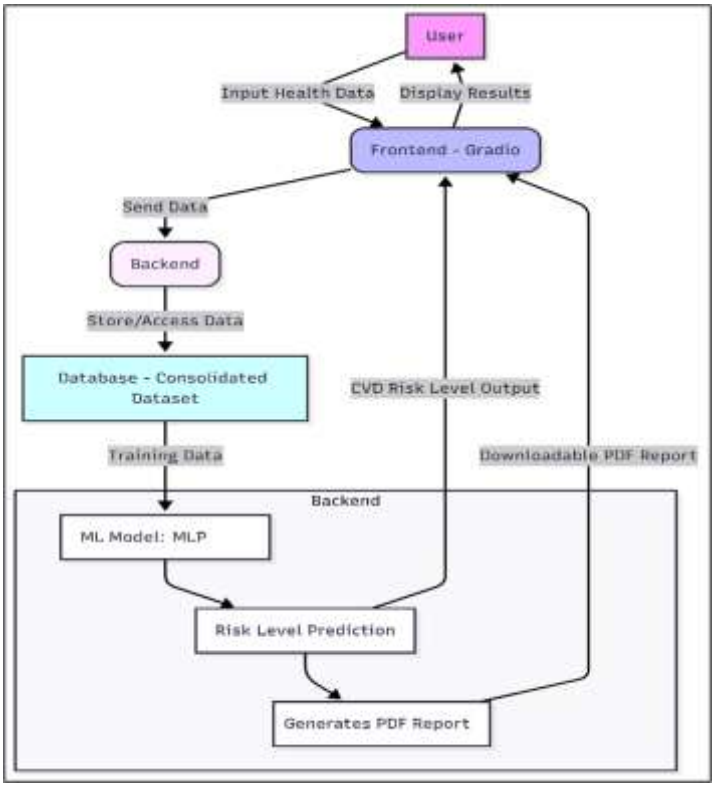

Figure 3: The CVRP System Architecture.
Source: Authors, (2026).

III. MATERIALS AND METHODS

The approach used for the CVRP model is explained in this section. The first step starts with gathering and preprocessing the dataset to prepare it for model training.

III.1 DATA PREPROCESSING

The dataset used is publicly available, it consists of 13 features, an entire summary of these features and their descriptions is mentioned in Table 2. For building any model, the most crucial task is data preprocessing. As we are dealing with medical data, data preprocessing was done with the help of tools like NumPy and pandas to ensure that we have a clean and ready dataset for training. All the missing values were handled carefully, and duplicated records were eliminated. This helps to prevent any bias or errors in the training process. The processed data was then split into two portions: 80% for training, 20% for testing.

Table 2: Features in the Dataset.

Sr. No	Attribute	Variable Name	Attribute Description	Values of Attributes
1	Age	age	Age of the person	Age in number
2	Gender	gen	The gender of the patient (0 represents female and 1 represents male)	F = 0, M = 1
3	Chest_Pain_Types	cpt	The type and severity of chest pain felt by the patient (ASY, ATA, NAP, TA)	0,1,2,3
4	Resting_BP	rbp	Denotes the patient's resting blood pressure (mm Hg)	Values between 94 and 200
5	Cholesterol	chol	Reflects the chol level measured (mg/dL)	Values between 126 and 564
6	Oldpeak	op	During physical exertion, the maximum heart rate	Values between 0 to 9.0
7	Resting_ECG	recg	Shows the patient's resting electrocardiogram. (LVH, Normal, ST)	0,1,2
8	Fasting_BS	fbs	Range of fasting blood sugar (greater than 120 mg/dL (1 = true, 0 = false))	0,1
9	Exercise_Angina	ang	Exercise-induced chest pain. (1=yes, 0=no)	0,1
10	Slope	st	ST-segment slope during peak exercise (Upsloping=2, Flat=1, and Downsloping=0), which helps assess the patient's cardiac response to stress.	0,1,2
11	Max_HR	thalach	shows the maximum heartbeat of the patient	71 to 202
12	Fluoroscopy	ca	Fluoroscopy colored main vessels count	0 to 3
13	Thallium	ths	Thallium stress (negative=0, positive=1 and inconclusive=2)	0 to 2
14	Target		This is the target or output variable in the dataset, representing binary classification. A value of "0" = lower likelihood of heart disease, while "1" = a higher risk. This classification is determined based on patterns across the other 13 input attributes.	0,1

Source: Authors, (2026).

III.2 CLASS IMBALANCE IN THE DATASET

The data was successfully preprocessed, and now we had to tackle the issue of data imbalance. Medical datasets usually have this common problem where one outcome occurs far less frequently than the other. To resolve this problem, we applied SMOTE to the training data. SMOTE balanced the dataset by creating synthetic examples for the less frequent class. This ensures that the model would not become biased towards the more common outcome (majority class). This step was crucial for building a more reliable CVRP model that provides fair and accurate predictions across all classes, especially for those at higher risk.

III.3 THE CVRP MODEL IMPLEMENTATION

To find the most accurate and consistent model for our problem, we evaluated five different machine learning algorithms, including RF, MLP, XGBoost, LightGBM, and AdaBoost. After extensive testing, the MLP Classifier was chosen according to the performance metrics. (As shown in Table 4) The final CVRP model used the MLP Classifier due to its several advantages (as shown in Table 5 and Figure 6). It learns non-linear mappings via backpropagation. It is useful for capturing interactions from features that may not be explicitly modelled in tree-based methods. It also allowed us to compare traditional machine learning with deep learning approaches in this medical domain.

III. 3.1. Model Selection and Hyperparameter Tuning

Model selection is a crucial step that helps in selecting the best model architecture or algorithm from a set of models. On the other hand, hyperparameter tuning consists of optimizing the settings of the chosen model in order to improve its performance. Feature selection methods was used to identify distinct feature that has an impact on the model's output. These methods include analysis of variance, mutual information, and chi-square. As shown in Table 3, the feature groups selected were named as Cardiac Risk Factors (CRF), a total of five groups, namely CRF-1, CRF-2, CRF-3, CRF-4, and CRF-5, were used.

Table 3: Selected Feature groups (CRF).

Selected Group of Features	Attributes
CRF-1	age, gen, chol, cpt, thalach, op, ang, ca, st, ths
CRF-2	chol, op, fbs, cpt, thalach, ang, ca, st, ths
CRF-3	gen, chol, op, recg, fbs, cpt, thalach, ang, ca, st, ths
CRF-4	op, cp, thalach, ths, ang, chol, ca, st, fbs, recg, gen, rbp
CRF-5	age, gen, chol, rbp, op, recg, fbs, cpt, thalach, ang, ca, st, ths

Source: Authors, (2026).

We employed Stratified K-Fold Cross Validation in combination with Randomized Search CV. Stratified K-Fold Cross Validation was useful because it ensured that each data fold maintained the same proportion of target classes, which is essential for medical data. Randomized Search CV efficiently found the optimal hyperparameters for each algorithm, which saved the time of manual tuning.

III.4 INTEGRATING EXPLAINABLE AI (XAI) IN CVRP MODEL

Our CVRP model goes beyond a simple yes/no prediction. It extends the output to three meaningful risk levels: low, moderate, and high. This makes the results more practical in the healthcare domain. To ensure the model is transparent and trustworthy, we integrated XAI tools. We have used SHAP values to generate a plot (Fig.4) that clearly explains the impact of each feature on a prediction. We also used LIME to generate visuals (Fig.5), which help users understand exactly how a specific prediction was made.

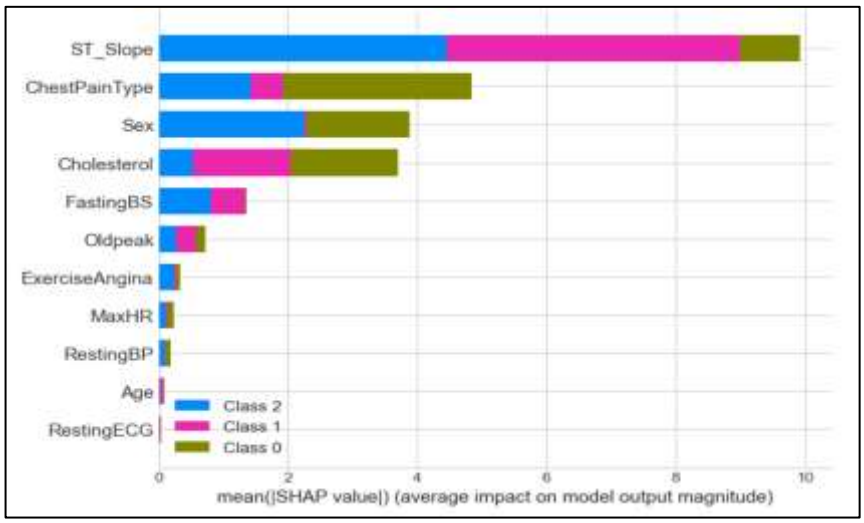

Figure 4: SHAP Bar Plot (Feature Importance).
Source: Authors, (2026).

The bar plot in Figure 5 shows the overall feature importance for the model. It helps understand that, on average, features like ST_Slope and ChestPainType are the most effective features for the predictions.

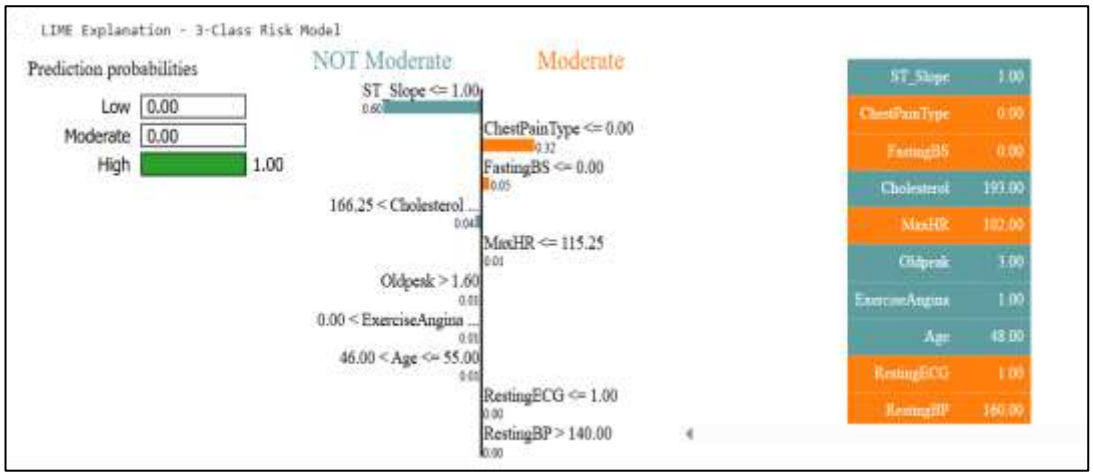

Figure 5: LIME explanation for a patient.
Source: Authors, (2026).

Figure 5: This LIME plot helps analyze which feature value positively or negatively influenced the model's result for an individual patient. In the final step, we saved the best model and built a user-friendly platform for healthcare professionals and patients. This web-based app was built using Gradio. It allows to enter health data and instantly the app will give an output about the level of cardiovascular risk. It even generates a downloadable PDF report that summarizes the results. This made the tool both intelligent and efficient to use for patients and healthcare professionals.

IV. RESULTS AND DISCUSSIONS

For the CVRP model, it is crucial to understand the model's ability to generalize across diverse patient data and detect CVDs. This can be done using performance evaluation techniques. These techniques help evaluate the accuracy and effectiveness of a model. It also helps to identify the areas for improvement (As shown in Table 4).

IV.1 PERFORMANCE EVALUATION TECHNIQUES

For evaluation of the CVRP model's performance, the following parameters were utilized:

1. **Accuracy:** It represents the percentage of correctly predicted cases of CVDs, which shows the overall effectiveness of the model. It is given as in Eq.12

$$\text{Accuracy} = \frac{True\ Positive\ (TP) + True\ Negative\ (TN)}{TP+TN+FN+FP} \tag{12}$$

2. **Precision:** This represents the proportion of true positive (TP) results out of all the positive predictions. This signifies the reliability of the model's positive outcomes. It is given as in Eq.13

$$\text{Precision} = \frac{True\ Positive\ (TP)}{TP+FP} \tag{13}$$

3. **Recall (Sensitivity):** Recall is used to identify the proportion of all actual cases of CVDs that were classified correctly as positives by the model. It is given as in Eq.14

$$\text{Recall} = \frac{True\ Positive\ (TP)}{TP+FN} \tag{14}$$

4. **F1 Score:** A consolidation of precision and recall is F1 score. It is given as in Eq.15
5.

$$\text{F1} - \text{Score} = 2\frac{Precision*Recall\ (Sensitivity)}{Precision+Recall\ (Sensitivity)} \tag{15}$$

6. **Execution Time:** It denotes the time required to process the dataset and generate predictions.

7. **ROC-AUC Score:** This helps to understand the ability of the model to distinguish between the healthy patients from the at-risk patients. As shown in Figure 7.

8. **Cost:** It is the expenses associated with data storage, processing, and cloud services utilized during model implementation.

Table 4: Performance evaluation results of ML techniques using CRFs.

ML Algorithm		MLP	RF	XGBOOST	LIGHTGBM	ADABOOST
CRF-1	Accuracy %	96.72	93.44	95.89	95.08	93.44
	Sensitivity %	97.00	94.00	96.00	95.00	94.00
	F1 score %	97.00	94.00	96.00	95.00	94.00
CRF-2	Accuracy %	91.08	91.08	95.08	95.08	91.80
	Recall %	94.00	88.00	91.00	91.00	91.00
	F1 score %	92.00	92.00	95.00	95.00	92.00
CRF-3	Accuracy %	93.44	91.08	95.90	95.08	91.08
	Recall %	94.00	88.00	94.00	91.00	94.00
	F1 score %	94.00	92.00	96.00	95.00	92.00
CRF-4	Accuracy %	93.44	93.44	95.08	95.08	93.44
	Recall %	94.00	91.00	94.00	91.00	94.00
	F1 score %	94.00	94.00	95.00	95.00	94.00
CRF-5	Accuracy %	91.08	90.16	95.08	95.08	95.08
	Recall %	91.00	88.00	91.00	91.00	94.00
	F1 score %	92.00	90.00	95.00	95.00	95.00

Source: Authors, (2026).

Table 5: Performance results of MLP Classifier using CRF-1 and SMOTE.

ML Algorithm	Accuracy	Precision	Sensitivity	F1 score	AUC
MLP	96.72%	97.00%	97.00%	97.00%	0.98

Source: Authors, (2026).

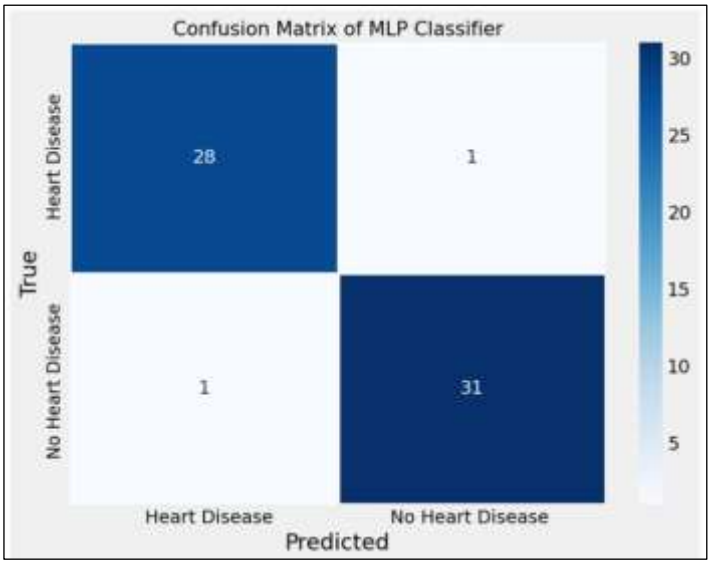

Figure 6: Confusion matrix of MLP Classifier using CRF-1 and SMOTE.
Source: Authors, (2026).

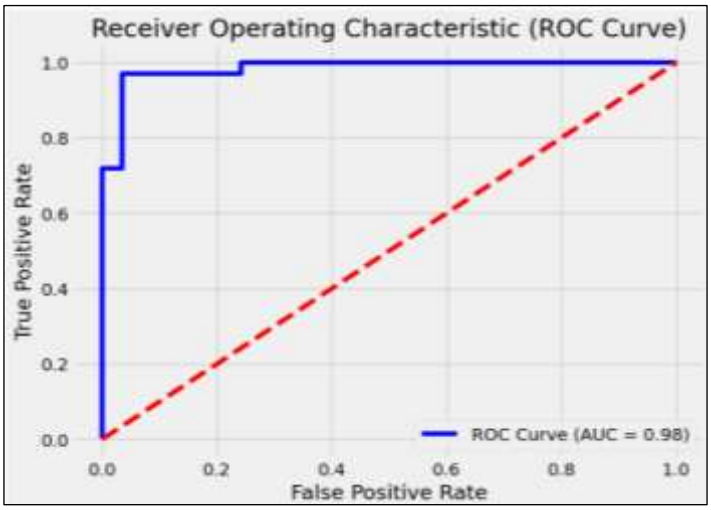

Figure 7: AUC and ROC curve of MLP Classifier using CRF-1 and SMOTE.
Source: Authors, (2026).

IV. 2. LIMITATIONS AND FUTURE SCOPE

The CVRP model's aim was to enhance the predictive accuracy of CVD prediction by integrating machine learning techniques. It will be a reliable platform that will help healthcare professionals and patients. As there were dataset limitations, the model couldn't reach a comparable accuracy with the existing solutions. Future plans include integrating datasets from different sources, including medical imaging, along with telehealth services. This will further improve the CVRP model's accuracy and effectiveness. This system can also be enhanced by analyzing and defining the stages of CVDs, which will help identify the risk of CVDs in early stages.

IV. 3. D]ISCUSSIONS

The CVRP model used MLP Classifier which achieved a high accuracy of 96.72% with AUC of 0.98. The model has a user-friendly frontend, which makes it convenient to identify CVD risk by healthcare professionals and patients. The system also maintains transparency and interpretability by using XAI techniques. It makes predictions more understandable, thus making it effective in real-time usage. This system aspires to develop a high-confidence model for early risk detection and management of CVDs.

V. CONCLUSIONS

The CVRP model's aim was to enhance the overall predictive accuracy of CVD prediction. This was achieved by using a publicly available dataset and integrating machine learning techniques. CVRP will be a reliable platform that will help healthcare professionals assess patient health reports. ML algorithms are evaluated and selected depending on dataset characteristics to improve the model reliability and effectiveness in cardiovascular risk prediction. CVRP helps prioritize user experience while addressing privacy concerns and ensuring accessibility for healthcare providers. The integration of Explainable AI (XAI) introduces an approach to improving model transparency and interpretability, making cardiovascular risk predictions more reliable and understandable.

VI. AUTHOR'S CONTRIBUTION

Conceptualization: Mehek Mohd Meraj Qureshi, Hariram Chavan.
Methodology: Mehek Mohd Meraj Qureshi, Hariram Chavan.
Investigation: Mehek Mohd Meraj Qureshi, Hariram Chavan.
Discussion of results: Mehek Mohd Meraj Qureshi, Hariram Chavan.
Writing – Original Draft: Mehek Mohd Meraj Qureshi, Hariram Chavan.
Writing – Review and Editing: Mehek Mohd Meraj Qureshi, Hariram Chavan.
Resources: Mehek Mohd Meraj Qureshi, Hariram Chavan.
Supervision: Mehek Mohd Meraj Qureshi, Hariram Chavan.
Approval of the final text: Mehek Mohd Meraj Qureshi, Hariram Chavan.

VII. REFERENCES

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

[2] Z. Alom, M. A. Azim, Z. Aung, M. Khushi, J. Car, and M. A. Moni, ''Early stage detection of heart failure using machine learning techniques,'' in Proc. Int. Conf. Big Data, IoT, Mach. Learn., Cox's Bazar, Bangladesh, 2021, pp. 23–25.

[3] V. Vision Paul and J. A. I. S. Masood, "Exploring Predictive Methods for Cardiovascular Disease: A Survey of Methods and Applications," in IEEE Access, vol. 12, pp. 101497-101505, 2024, doi: 10.1109/ACCESS.2024.3430898.

[4] H. Khan, N. Javaid, T. Bashir, M. Akbar, N. Alrajeh and S. Aslam, "Heart Disease Prediction Using Novel Ensemble and Blending Based Cardiovascular Disease Detection Networks: EnsCVDD-Net and BlCVDD-Net," in IEEE Access, vol. 12, pp. 109230-109254, 2024, doi: 10.1109/ACCESS.2024.3421241.

[5] M. S. A. Reshan, S. Amin, M. A. Zeb, A. Sulaiman, H. Alshahrani and A. Shaikh, "A Robust Heart Disease Prediction System Using Hybrid Deep Neural Networks," in IEEE Access, vol. 11, pp. 121574-121591, 2023, doi: 10.1109/ACCESS.2023.3328909.

[6] Y. Liu, X. Li, and J. Ren, ''A comparative analysis of machine learning algorithms for heart disease prediction,'' Comput. Methods Programs Biomed., vol. 200, Nov. 2021, Art. no. 105965.

[7] A. Rahim, Y. Rasheed, F. Azam, M. W. Anwar, M. A. Rahim and A. W. Muzaffar, "An Integrated Machine Learning Framework for Effective Prediction of Cardiovascular Diseases," in IEEE Access, vol. 9, pp. 106575-106588, 2021, doi: 10.1109/ACCESS.2021.3098688.

[8] N. S. Hussein, A. Mustapha, and Z. A. Othman, ''Comparative study of machine learning techniques for heart disease diagnosis,'' Comput. Sci. Inf. Syst., vol. 17, no. 4, pp. 773–785, 2020.

[9] S. Akbar, R. Tariq, and A. Basharat, ''Heart disease prediction using different machine learning approaches: A critical review,'' J. Ambient Intell. Humanized Comput., vol. 11, no. 5, pp. 1973–1984, 2020.

[10] A. Singh and R. Kumar, "Heart Disease Prediction Using Machine Learning Algorithms," 2020 International Conference on Electrical and Electronics Engineering (ICE3), 2020, pp. 452-457, DOI: 10.1109/ICE348803.2020.9122958.

[11] D. Krishnani, A. Kumari, A. Dewangan, A. Singh and N. S. Naik, "Prediction of Coronary Heart Disease using Supervised Machine Learning Algorithms," TENCON 2019 - 2019 IEEE Region 10 Conference (TENCON), 2019, pp. 367-372, DOI: 10.1109/TENCON.2019.8929434

[12] M. S. Raja, M. Anurag, C. P. Reddy and N. R. Sirisala, "Machine Learning-Based Heart Disease Prediction System," 2021 International Conference on Computer Communication and Informatics (ICCCI), 2021, pp. 1-5, DOI:
10.1109/ICCCI50826.2021.9402653.

[13] R. Katarya and P. Srinivas, "Predicting Heart Disease at Early Stages using Machine Learning: A Survey," 2020 International Conference on Electronics and Sustainable Communication Systems (ICESC), 2020, pp. 302-305, DOI:
10.1109/ICESC48915.2020.9155586.

[14] S. K. J. and G. S., "Prediction of Heart Disease Using Machine Learning Algorithms," 2019 1st International Conference on Innovations in Information and Communication Technology (ICIICT), 2019, pp. 1-5, DOI: 10.1109/ICIICT1.2019.8741465.

[15] Pabitra Kumar Bhunia, Arijit Debnath, Poulami Mondal, Monalisa D E, Kankana Ganguly, Pranati Rakshit, "Heart Disease Prediction using Machine Learning," International Journal Of Engineering Research & Technology (IJERT) NCETER – 2021 (Volume 09 – Issue 11), DOI: 10.17577/IJERTCONV9IS11071.

[16] Himanshu Sharma, M. A. Rizvi, "Prediction of Heart Disease Using Machine Learning Algorithms: A Survey," 2017 International Journal on Recent and Innovation Trends in Computing and Communication (IJRITCC), pp. 99-104, DOI: 0.17762/ijritcc.v5i8.1175.

[17] P. Kumar, S. K. Sahu, "Predicting Heart Disease with Machine Learning Techniques: A Review," 2023 IEEE ICCIS, pp. 100-104, DOI: 10.1109/ICCCIS.2023.950047.

[18] V. Sharma, S. Yadav and M. Gupta, "Heart Disease Prediction using Machine Learning Techniques," 2020 2nd International Conference on Advances in Computing, Communication Control and Networking (ICACCCN), 2020, pp. 177-181, DOI: 10.1109/ICACCCN51052.2020.9362842.

[19] K. G. Dinesh, K. Arumugaraj, K. D. Santhosh and V. Mareeswari, "Prediction of Cardiovascular Disease Using Machine Learning Algorithms," 2018 International Conference on Current Trends towards Converging Technologies (ICCTCT),
2018, pp. 1-7, DOI: 10.1109/ICCTCT.2018.8550857.

[20] S. Akbar, R. Tariq, and A. Basharat, ‘‘Heart disease prediction using different machine learning approaches: A critical review,’’ J. Ambient Intell. Humanized Comput., vol. 11, no. 5, pp. 1973–1984, 2020.

[21] M. Liu, X. Sun, Y. Liu, X. Yang, Y. Xu, and X. Sun, ‘‘Deep learningbased prediction of coronary artery disease with CT angiography,’’ Jpn. J. Radiol., vol. 38, no. 4, pp. 366–374, 2020.

[22] A. Shoukat, S. Arshad, N. Ali, and G. Murtaza, ‘‘Prediction of cardiovascular diseases using machine learning: A systematic review,’’ J. Med. Syst., vol. 44, no. 8, p. 162, Aug. 2020.

[23] N. Khandadash, E. Ababneh, and M. Al-Qudah, ‘‘Predicting the risk of coronary artery disease in women using machine learning techniques,’’ J. Med. Syst., vol. 45, p. 62, Apr. 2021.

[24] H. El-Sofany, S. A. El-Seoud, O. H. Karam, Y. M. Abd El-Latif, and I. A. T. F. Taj-Eddin, ‘‘A proposed technique using machine learning for the prediction of diabetes disease through a mobile app,’’ Int. J. Intell. Syst., vol.
2024, pp. 1–13, Jan. 2024.

[25] V. Bento, M. Kohler, P. Diaz, L. Mendoza, and M. A. Pacheco, ‘‘Improving deep learning performance by using explainable artificial intelligence (XAI) approaches,’’ Discover Artif. Intell., vol. 1, no. 1, pp. 1–11, Dec. 2021.

[26] D. Y. Omkari and K. Shaik, "An Integrated Two-Layered Voting (TLV) Framework for Coronary Artery Disease Prediction Using Machine Learning Classifiers," in IEEE Access, vol. 12, pp. 56275-56290, 2024, doi: 10.1109/ACCESS.2024.3389707.

[27] J. J. Gabriel and L. Jani Anbarasi, "Accurate Cardiovascular Disease Prediction: Leveraging Opt_hpLGBM With Dual-Tier Feature Selection," in IEEE Access, vol. 12, pp. 142427-142448, 2024, doi: 10.1109/ACCESS.2024.3470537.

[28] Available at https://www.geeksforgeeks.org/machine-learning-with-python/

[29] Available at https://www.freecodecamp.org/learn/machine-learning-with-python/

[30] Available at https://www.deeplearning.ai/courses/machine-learning-in-production/

[31] Available at https://medium.com/@darshanaweerasooriya11/introduction-to-artificial-intelligence-machine-learning-deep-learning-81b6b29d98d7

[32] World Health Organization. Cardiovascular Diseases (CVDs). Accessed: May 5, 2023. [Online]. Available: https://www.afro.who.int/ health-topics/cardiovascular-diseases.