Spectrum of Engineering Sciences
ISSN (e) 3007-3138 (p) 3007-312X Volume 3, Issue 3, 2025

APPROACHES TO PREDICT CARDIOVASCULAR ISSUE USING MACHINE LEARNING METHOD

Waddiat U Zahra[1], Fahim Uz Zaman[2], Gohar Mumtaz[3], Muqaddas Salahuddin*[4], Muhammad Zohaib Khan[5], Syed Akmal Sultan[6], Sammia Hira[7], Fakhra Parveen[8]

[1]*Faculty of Computer Science and Information Technology, Superior University, Lahore, 54000, Pakistan*
[2]*Department: Digital Technologies, Newcastle College University Centre (NCUC) Rye Hill Campus, Scotswood Rd, Newcastle upon Tyne NE4 7SA, United Kingdom*
[3,4,7,8]*Faculty of Computer Science and Information Technology, Superior University, Lahore, 54000, Pakistan*
[5]*Department of Information Technology, Shaheed Mohtarma Benazir Bhutto Institute of Trauma, Karachi, Pakistan*
[6]*Department of Orthopedic Surgery, Shaheed Mohtarma Benazir Bhutto Institute of Trauma, Karachi, Pakistan*

[1]alizyali427@gmail.com, [2]faheemuzzaman@hotmail.com, [3]gohar.m@superior.edu.pk, [*4]muqaddassalahuddin60@gmail.com, [5]zohaib_khan2017@yahoo.com, [6]syedakmalsultan@gmail.com, [7]hirac53oort@gmail.com, [8]parveenfakhra13@gmail.com

DOI: https://doi.org/10.5281/zenodo.15212872

Keywords
Cardiovascular Diseases (CVD), Quadratic Discriminant Analysis (QDA), Support Vector Machine (SVM), Multi-layer Perceptron (MLP), and Stochastic Gradient Descent (SGD).

Article History
Received on 07 March 2025
Accepted on 07 Apil 2025
Published on 17 April 2025

Corresponding Author: *

Abstract
Cardiovascular illnesses remain a major health concern, requiring efficient detection techniques. Despite valuable research, gaps remain in predictive models, particularly due to imbalanced datasets, leading to biased predictions. This study employs machine learning to detect cardiac issues, including myocardial infarction, addressing dataset imbalance. It evaluates Fuzzy C-Means Clustering, Quadratic Discriminant Analysis (QDA), Support Vector Machine (SVM), Multi-layer Perceptron (MLP), and Stochastic Gradient Descent (SGD). The findings offer insights into enhancing myocardial infarction prediction and improving cardiovascular disorder diagnosis.

INTRODUCTION

The cardiovascular system maintains physiological balance by circulating oxygenated blood. Cardiovascular diseases (CVDs), including congenital abnormalities, peripheral arterial disease, arrhythmias, cerebrovascular diseases, coronary artery disease (CAD), and cardiomyopathies, disrupt this function. Coronary heart disease (CHD) is the most common, accounting for 64% of cases globally. CVDs significantly impact both men and women, making CAD a leading cause of death [1,2].

The World Health Organization (WHO) reports 17.9 million annual deaths from CVDs and predicts 24.5 million by 2030 due to risk factors like smoking, obesity, high cholesterol, diabetes, and high blood pressure [3,4]. Timely medical intervention is crucial, as early diagnosis can save lives.

Machine learning (ML) has transformed healthcare with its ability to process data and identify complex patterns. However, challenges like imbalanced datasets hinder the accuracy of predictive models [5,6].

The following are the primary goals of this study:

- Analyze and compare machine learning algorithms, including Fuzzy C-Means Clustering, Multi-layer Perceptron (MLP), Support Vector Machine (SVM), Quadratic Discriminant Analysis (QDA), and Stochastic Gradient Descent (SGD), for CVD prediction.
- Improve algorithm efficiency and apply pre-processing techniques to enhance prediction accuracy.
- Provide medical professionals with a tool for early detection and personalized treatment.

This study evaluates supervised and unsupervised ML models Fuzzy C-Means Clustering, SVM, MLP, QDA, and SGD used for classification, regression, and clustering. The results demonstrate improved precision, recall, accuracy, f1-score, sensitivity, and specificity. The paper is structured as follows: Section 2 reviews literature, Sections 3 and 4 present the proposed approach and results, and Section 5 concludes the study.

1. Literature Review

Healthcare organizations seek innovative methods to manage rising heart disease-related deaths, improving service quality and timely healthcare delivery. The Internet of Things (IoT) has emerged as a key approach, enabling cardiac health monitoring. Safa and Pandian [7] proposed a stress evaluation method using physiological data, where K-Nearest Neighbor (KNN) outperformed SVM and Decision Tree (DT).

Heart disease prediction relies on techniques like KNN, DT, genetic algorithms, and Naïve Bayes (NB) [8]. Mohan et al. [9] introduced a hybrid approach achieving 88.4% accuracy. Kaur et al. [10] explored data mining techniques, highlighting SVM's superior accuracy. Using CAD technology, [11] evaluated Multi-Layer Perceptron for cardiac disorder prediction. Animesh Hazra et al. [12] compared C4.5, K-Means, DT, SVM, and NB to enhance prediction accuracy. DT algorithms were also used in [13] with SVM for better classification. Wong et al. [14] conducted a meta-analysis on ML applications for CVD prediction. AI-enhanced ECG analysis for at-risk patients was examined in [15]. Linda et al. [16] developed a clinical decision support system for exercise recommendations in heart patients.

ML algorithms have shown effectiveness in disease prediction. Kohali et al. [17] used Adaboost for breast cancer (98.57% accuracy), SVM for diabetes (85.71%), and Logistic Regression (LR) for heart disease (87.1%). This study evaluates supervised and unsupervised ML techniques for CVD prediction and proposes a combined approach for improved precision, discussed in the next section.

2. Proposed Methodology

This section outlines the development of the cardiovascular disease (CVD) prediction system. The steps involved are:

1. Retrieving the dataset from Google Drive.
2. Applying pre-processing: data cleansing, feature importance analysis, and data standardization.
3. Using Fuzzy C-Means clustering as an unsupervised method to improve precision, recall, and address dataset imbalance.
4. Splitting the dataset into training (75%) and testing (25%) subsets.
5. Training and testing the final CVD prediction system using machine learning models: Stochastic Gradient Descent (SGD), Support Vector Machine (SVM), Multi-layer Perceptron (MLP), Quadratic Discriminant Analysis (QDA), and Fuzzy C-Means Clustering.

Figure 1 provides a block diagram of the proposed method for the CVD.

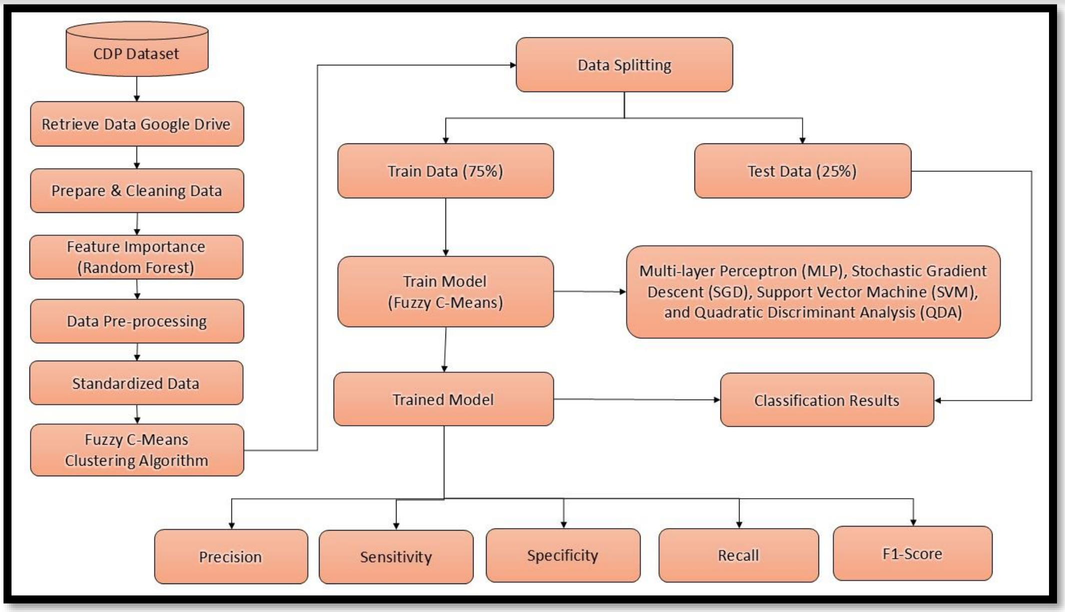

Figure 1: Block Diagram of the Proposed Method for CVD

The flowchart shows data extraction, cleaning, standardization, and normalization. Processed data (75% training, 25% testing) is used for clustering and classification. Data preprocessing and classification play distinct roles in the proposed method.

2.1. Pre-processing

Data is processed using fuzzy C-Means clustering to retrieve relevant information after gathering the necessary preprocessing data.

Data Collection

Data is collected from Kaggle's Cardiovascular Disease Prediction (CDP) dataset, which contains 70,000 instances described by 13 variables [18] and two target classes (cardiovascular: 34,979; non-cardiovascular: 35,021), providing essential traits for research analysis and cardiovascular prediction (Table 1).

Table 1: Displays the Original Cardiovascular Disease Prediction Dataset

Parameters of the Dataset	Characteristics of CDP
Age	Patient age
Height	Patient height
Weight	Patient weight
Gender	Patient Sex: female; male
Systolic blood pressure	Have you ever been fatigued, worn out, or exhausted?
Diastolic blood pressure	Have you frequently experienced ineffective coughing?
Cholesterol	Have you ever felt suffocated or out of breath?
Glucose	Have you ever felt a sore, itchy, or painful throat?
Smoking	Have you had frequent bone or body pain?
Alcohol Intake	Have you ever felt your face or nose feel full?
Physical Activity	Have you ever had fluids flow down your mouth or drip?
Cardiovascular Disease	Patient Severity; None, Mild or Moderate.

Figure 2 shows cardiovascular disease prevalence by gender (1: females, 2: males), with females exhibiting higher incidence than males. Figure 3 displays distribution by class (0: non-cardiovascular, 1: cardiovascular), where non-cardiovascular cases are more frequent.

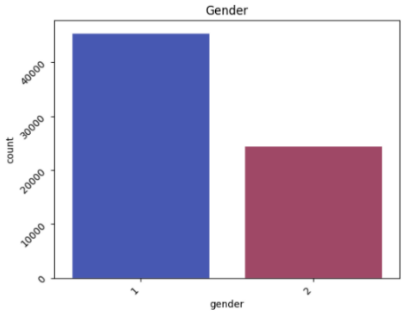

Figure 2: Quantity of Facts Male and Female Cardiovascular Disease Patients

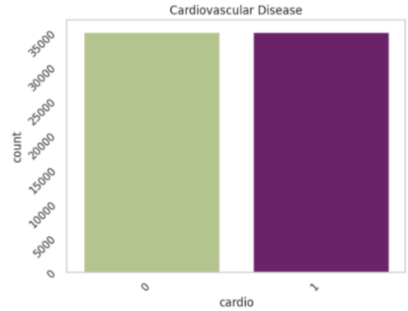

Figure 3: Quantity of Cardiovascular Disease Patients

FEATURE SELECTION: Attributes are rated by relevance lower-scoring features are eliminated to reduce dimensionality, while higher-scoring ones influence predictions. RFE with XGBoost, SVM, and RF iteratively removes less significant features to enhance efficiency and forecast accuracy [20]. The CDP model's component importance is shown in two tables, with Figure 4 graphically representing these scores and Table 2 detailing variable values.

Table 2: Cardiovascular Disease Prediction Dataset with Feature Importance (CDP)

Sr. No.	Feature	Feature Importance Score
1	age	0.308162
2	ap_hi	0.172671
3	weight	0.172423
4	height	0.157008
5	ap_lo	0.090245
6	cholesterol	0.038594
7	gluc	0.016222
8	gender	0.015581
9	active	0.013170
10	smoke	0.008399
11	alco	0.007525

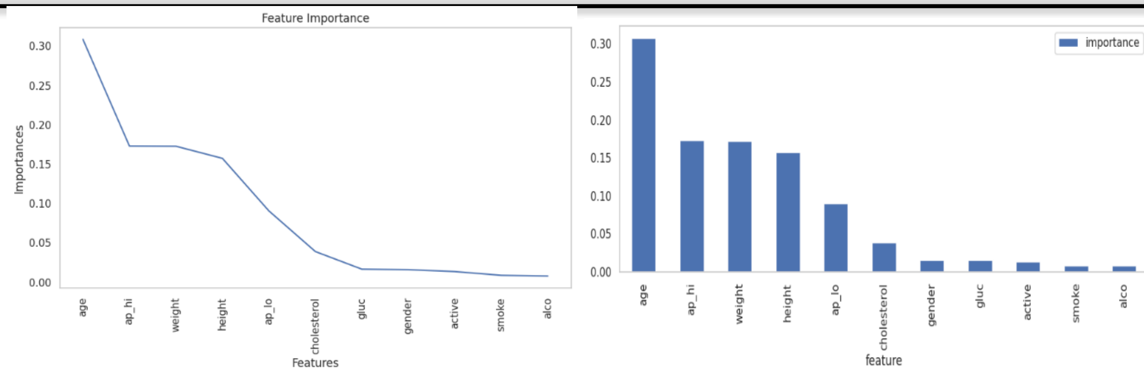

Figure 4: Feature Importance Dataset of Cardiovascular Disease Prediction (CDP)

This study investigates key approaches for cardiovascular disease prediction: pre-processing, induction, simulation, data standardization, post-processing, complexity measurements, crucial elements, and system efficiency. Data collection starts the process, followed by normalization and standardization that produce an improved dataset (Table 3). Complex data is shown in Figure 5, where the X value is a black circle and the Y value is a pale pink circle, without fuzzy C-Means clustering.

Table 3: Analyzed dataset used to forecast the cardiovascular diagnosis is displayed.

```
11-Dimension
[[ 1.62363599 0.71370255 -0.60810188 ... -0.2364393   0.61189903
  -0.98114275]
 [-0.77230005 -0.6592671   0.69038451 ... -0.2364393   0.4158916
   0.99475516]
 [-0.65671828 0.31040163 -0.57347554 ... -0.2364393 -1.95735553
   1.10581882]
 ...
 [ 0.60508663 -0.82955287 0.8840627 ...  3.43335127 -1.95735553
   0.66857836]
 [-0.72027928 -0.06903202 -0.12145167 ... -0.2364393 -1.95735553
   1.04474243]
 [-0.6917805   0.43681076 -0.04898113 ... -0.2364393   0.50882299
  -0.98114275]]
```

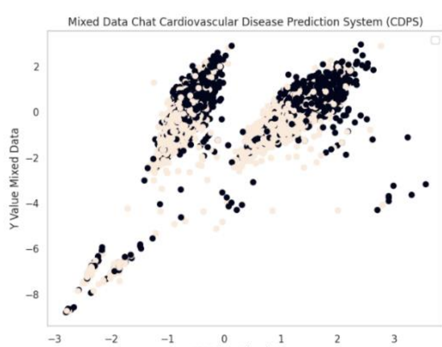

Figure 5: Cardiovascular Disease Prediction using Mixed Data Chat (CDP)

FUZZY C-MEANS CLUSTERING METHOD. Clustering is a widely used unsupervised learning technique that groups data into clusters. Fuzzy C-Means (FCM) assigns membership values to each data point based on its distance from a cluster centroid, allowing points to belong to multiple clusters [19,20]. Preprocessing creates a mixed data representation, refining large datasets by removing redundancies. FCM identifies three clusters, assigning probabilities to determine membership. This forms an agreement matrix showing relationships between clusters and samples. Applied to the multi-class Cardiovascular Disease Prediction (CDP) dataset, FCM structures unorganized data, as illustrated in Figures 6 and 7 through cluster diagrams and line graphs.

Table 4: Fuzzy C-Means Clustering Centroid Value

array ([[-3.00822552e-08, 3.08821124e-07, -9.16188718e-08, -2.57805183e-07, -1.03929298e-07, -3.10428289e-07, -1.47675000e-07, -9.90066481e-08, -8.01518498e-08, 3.35048900e-07, -9.37375561e-07], [3.00820519e-08, -3.08821099e-07, 9.16188517e-08, 2.57805300e-07, 1.03929274e-07, 3.10428243e-07, 1.47674902e-07, 9.90064923e-08, 8.01517281e-08, -3.35048769e-07, 9.37375640e-07]])

Table 5: C-Means Fuzzy Two Clusters Cardiovascular Disease Prediction (CDP) dataset with preprocessing

[[1.62363599 0.71370255 -0.60810188 ... -0.2364393 0.61189903 -0.98114275 0] [-0.77230005 -0.6592671 0.69038451 ... -0.2364393 0.4158916 0.99475516 1] [-0.65671828 0.31040163 -0.57347554 ... -0.2364393 -1.95735553 1.10581882 1] ... [0.60508663 -0.82955287 0.8840627 ... 3.43335127 -1.95735553 0.66857836 1] [-0.72027928 -0.06903202 -0.12145167 ... -0.2364393 -1.95735553 1.04474243 0] [-0.6917805 0.43681076 -0.04898113 ... -0.2364393 0.50882299 -0.98114275 0]]

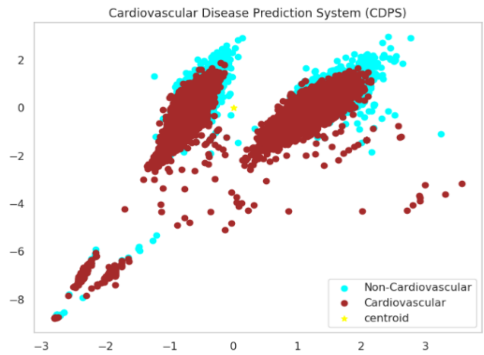

Figure 6: Fuzzy C-Means Two Clusters (CDP)

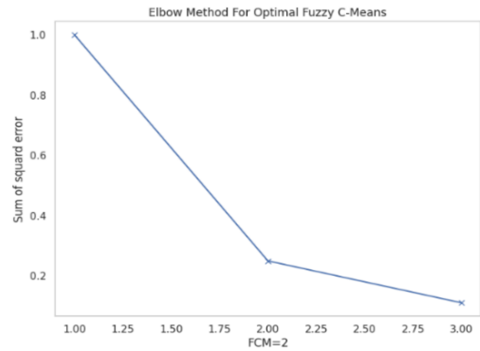

Figure 7: Fuzzy C-Means Sum of Squared Error Line Chart

Clustering feature detection is evaluated using various clustering methods, F-measure, precision, and recall. Standardized benchmarks classify results as legitimate, suspicious, or unlawful, ensuring accurate system assessment and performance evaluation.

3.2 CLASSIFICATION

Classification assigns data points to categories using supervised learning, testing various algorithms to determine the best predictor for a dataset.

MULTI-LAYER PERCEPTRON (MLP) ALGORITHM. Multi-Layer Perceptron (MLP) is an optimization method with input, hidden, and output layers, used for complex data patterns in supervised learning. It maps data into multidimensional space, enabling precise predictions **[21, 22].**

This process maps data into multidimensional space, enabling the model to match genuine labels and predict future datasets. Figure 8's confusion matrix evaluates MLP's accuracy.

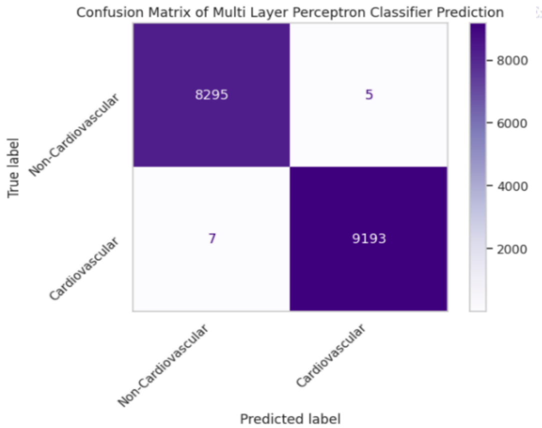

Figure 8: MLP Confusion Matrix Algorithm

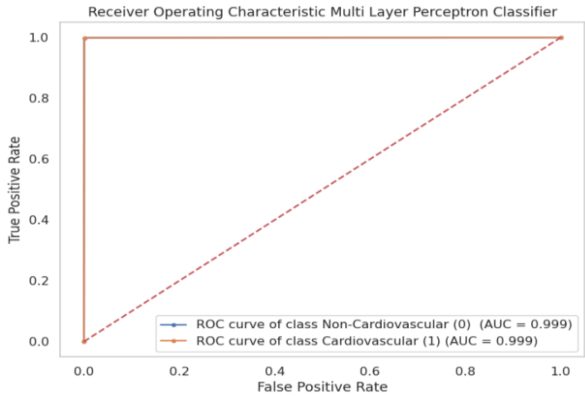

Figure 9: Shows the MLP Receiver Operating Characteristic (ROC) Curve

The confusion matrix validated the MLP model for detection and prediction tasks. Figure 9 shows results from applying MLP to a synthetic dataset. The ROC curve further evaluates model accuracy, providing insights into prediction trends and improving estimation.

STOCHASTIC GRADIENT DESCENT (SGD) ALGORITHM. optimizes objective functions, especially in linear models like SVM and Logistic Regression. It updates parameters after analyzing each data point, making it efficient for large datasets [23]. After training, it predicts category labels, aiming for high accuracy.

Figure 10 presents the confusion matrix for SGD predictions.

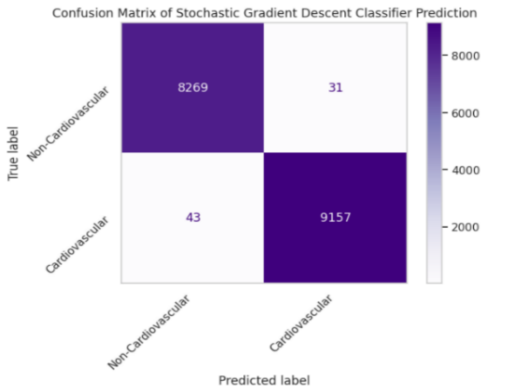

Figure 10: Stochastic Gradient Descent (SGD) Confusion Matrix Algorithm

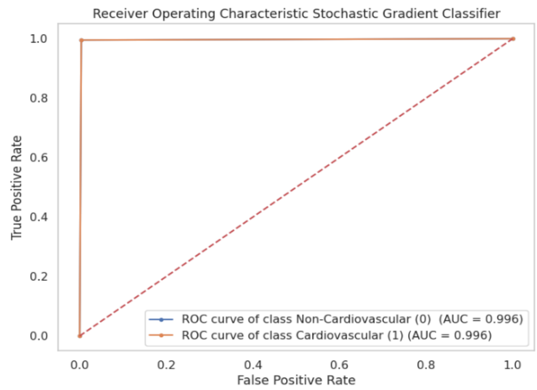

Figure 11: Shows SGD Receiver Operating Characteristic (ROC) Curve

The Receiver Operating Characteristic (ROC) curve evaluates a classifier's predictive ability as the threshold changes. Often used with cost-benefit analysis, it helps balance decision impacts. Figure 11 illustrates the ROC curve for the Stochastic Gradient Descent (SGD) model, highlighting its predictive power and accuracy optimization.

SUPPORT VECTOR MACHINE ALGORITHM. SVM is a supervised learning method used for classification and regression, ensuring maximum separation between classes in a high-dimensional space. It employs mathematical techniques like kernel functions (sigmoid, RBF, polynomial) to determine optimal decision boundaries using support vectors **[24, 25]**. Once trained, SVM predicts labels for new and existing data, aiming for high accuracy. Figure 12 presents confusion matrices demonstrating its predictive performance.

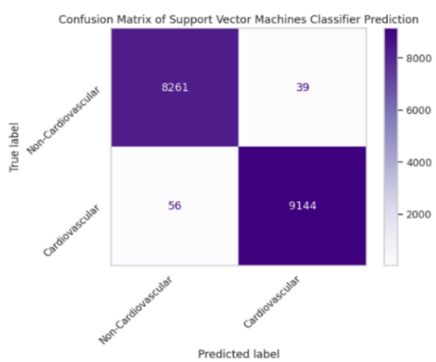

Figure 12: SVM Confusion Matrix Algorithm

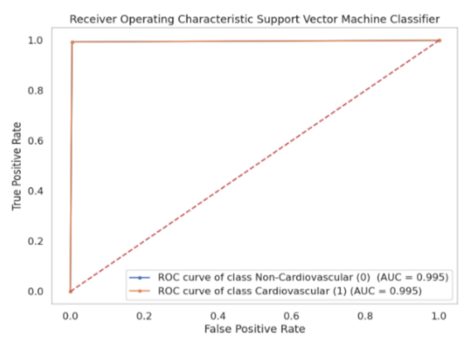

Figure 13: Shows SVM Receiver Operating Characteristic (ROC) Curve

The confusion matrix played a key role in accurate labeling for detection and prediction, making SVM a suitable choice. Figure 13 presents result from applying SVM to a synthetic dataset. The ROC curve further evaluates performance, with user feedback refining accuracy, improving overall model precision.

Quadratic Discriminant Analysis (QDA) ALGORITHM. Quadratic Discriminant Analysis (QDA) assumes Gaussian class distributions, creating a quadratic decision boundary using Bayes' rule. Unlike Linear Discriminant Analysis (LDA), QDA excels with unequal class distributions but can be computationally intensive with many predictors and prone to overfitting with small sample sizes [26,27]. It effectively determines class probabilities for accurate classification. Figure 14 presents the resulting confusion matrix, highlighting QDA's ability to classify data even with imbalanced distributions or limited training data [28,29].

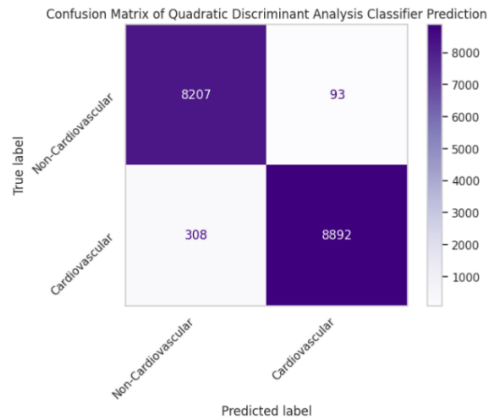

Figure 14: Quadratic Discriminant Analysis (QDA) Confusion Matrix Algorithm

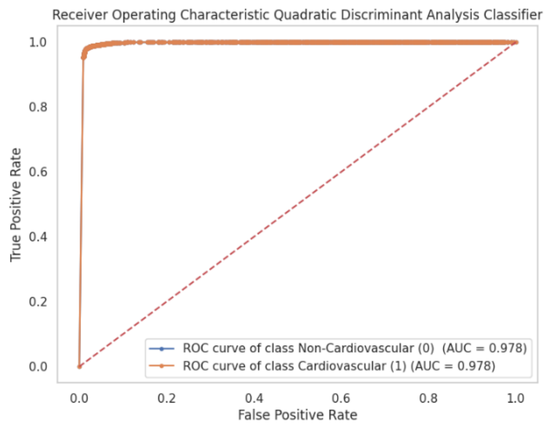

Figure 15: Shows QDA Receiver Operating Characteristic (ROC) Curve

ROC analysis evaluates a classifier's ability to adjust its discrimination threshold, aiding informed decision-making. Closely linked to cost-benefit analysis, it provides insights into model performance. Figure 15 presents the ROC curve, visually illustrating the classifier's overall effectiveness.

3. RESULTS AND DISCUSSION:

A comprehensive analysis of hybrid algorithms was conducted to assess their predictive and classification performance for cardiovascular disease. The following table presents the hybrid algorithms and their corresponding accuracies:

Table 6: Accuracy of Hybrid Models for Cardiovascular Disease (CVD)

Hybrid Algorithm	Accuracy of Algorithms
Fuzzy C-Means, Multi-layer Perceptron (MLP) **Proposed Method**	99.9314 %
Fuzzy C-Means, Stochastic Gradient Descent (SGD) **Proposed Method**	99.5771 %
Fuzzy C-Means, Support Vector Machine (SVM) **Proposed Method**	99.4571 %
Fuzzy C-Means, Quadratic Discriminant Analysis (QDA) **Proposed Method**	97.7085 %

Hybrid algorithms showed varying accuracy in predicting cardiovascular disease, with Fuzzy C-Means + MLP achieving 99.93%, SGD + Fuzzy C-Means 99.58%, Fuzzy C-Means + SVM 99.46%, and Fuzzy C-Means + QDA 97.71%, highlighting their potential for improved medical diagnostics.

Table 7: Parameter evaluations for CVD prediction using algorithm combinations

S/No.	Parameter Score	Fuzzy C-Means, Multi-layer Perceptron (MLP)	Fuzzy C-Means, Stochastic Gradient Descent (SGD)	Fuzzy C-Means, Support Vector Machine (SVM)	Fuzzy C-Means, Quadratic Discriminant Analysis (QDA)
1	Precision	0.99930661	0.99572639	0.99450991	0.97673897
2	Recall	0.99931836	0.99579557	0.99460712	0.97765845
3	F1-Score	0.99931247	0.99576051	0.99455757	0.97705055
4	Sensitivity	1.0	1.0	1.0	1.0
5	Specificity	1.0	1.0	1.0	1.0

Hybrid algorithm evaluations show excellent performance in cardiovascular disease prediction. The Fuzzy C-Means and MLP combination achieved 0.9993 precision, recall, and F1-score, with 1.0 sensitivity and specificity. Similarly, Fuzzy C-Means with SGD and SVM showed strong results, with precision, recall, and F1-scores above 0.995. The Fuzzy C-Means and QDA combination achieved 0.9767 accuracy, proving the effectiveness of hybrid models for precise cardiovascular predictions.

4. Comparative Analysis

The classifier's effectiveness was assessed by comparing various supervised ML techniques to identify the most efficient one. Previous studies on cardiovascular disease (CVD) prediction showed limited accuracy, with Naïve Bayes achieving 83%, Logistic Regression 85%, and SVM 84.5%. In contrast, hybrid models significantly improved performance. The Fuzzy C-Means and MLP combination achieved 99.93% accuracy, while Fuzzy C-Means with SGD, SVM, and QDA attained 99.58%, 99.46%, and 97.71%, respectively. These results highlight the superior accuracy of hybrid methods for CVD prediction, enhancing medical diagnostics. Detailed performance metrics are presented in Table 7.

Table 8: Performance Accuracy of CVD Models Employing Hybrid Algorithms

Hybrid Algorithm	Accuracy of Algorithms
Parthiban **et al. [30]**	Naïve Bayes 74%
Kumar Dwivedi et al. **[31]**	Naïve Bayes 83%
	Classification tree 77%
	K-NN 80%
	Logistic regression 85%
Otoom **et al. [32]**	Naïve Bayes 84.5%
	SVM 84.5%
Fuzzy C-Means, Multi-layer Perceptron (MLP) **Proposed Method**	99.9314 %
Fuzzy C-Means, Stochastic Gradient Descent (SGD) **Proposed Method**	99.5771 %
Fuzzy C-Means, Support Vector Machine (SVM) **Proposed Method**	99.4571 %
Fuzzy C-Means, Quadratic Discriminant Analysis (QDA) **Proposed Method**	97.7085 %

Figures 16 and 17 show the accuracy of different hybrid models for heart disease prediction, ranging from 97.7% to 99.9%. The Fuzzy C-Means and MLP combination achieved the highest accuracy of 99.93%, followed by Fuzzy C-Means with SGD (99.58%) and SVM (99.46%). The Fuzzy C-Means and QDA combination also performed well, reaching 97.71% accuracy. These results highlight the effectiveness of hybrid models in improving patient identification accuracy.

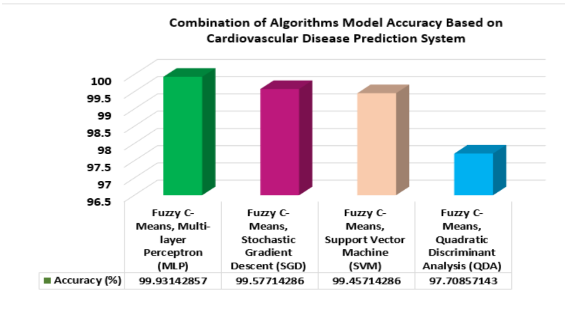

Figure 16: Algorithm Combination Model Accuracy for CDP

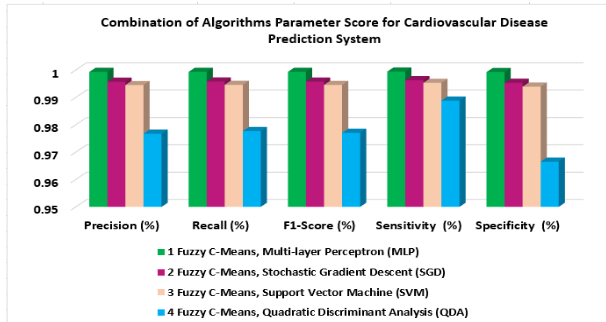

Figure 17: Combination of Parameter Score and Algorithms for CDP

The accuracy graph shows peak performance for hybrid models. If needed, accuracy can be adjusted. Currently, Score Precision, Recall, F1-Score, Sensitivity, and Specificity achieve optimal accuracy.

5. Conclusion

The study presents a technique to assist doctors in diagnosing cardiovascular diseases. Various machine learning methods, including fuzzy C-Means, MLP, SGD, SVM, and QDA, were applied to CVD datasets. The combination of fuzzy C-Means and MLP achieved the highest accuracy of 99.93%, proving its efficiency. Other combinations also showed strong results: 99.58% (Fuzzy C-Means + SGD), 99.46% (Fuzzy C-Means + SVM), and 97.71% (Fuzzy C-Means + QDA). These findings highlight the effectiveness of hybrid models for cardiac disease prediction. Future research could enhance accuracy by incorporating larger datasets and advanced ML techniques.

REFERENCES

1. World Health Organization. WHO Cardiovascular Diseases. Available online: https://www.who.int/health-topics/cardiovascular-diseases#tab=tab_1 (accessed on 19 January 2022).
2. Ramesh, A.N.; Kambhampati, C.; Monson, J.R.; Drew, P.J. Artificial intelligence in medicine. Ann. R. Coll. Surg. Engl. 2004, 86, 334. [Google Scholar] [CrossRef] [PubMed]
3. Abdellatif, A.; Mubarak, H.; Abdellatef, H.; Kanesan, J.; Abdelltif, Y.; Chow, C.-O.; Chuah, J.H.; Gheni, H.M.; Kendall, G. Computational detection and interpretation of heart disease based on conditional variational auto-encoder and stacked ensemble-learning framework. Biomed. Signal Process. Control 2024, 88, 105644. [Google Scholar] [CrossRef]
4. Tartarisco, G.; Cicceri, G.; Bruschetta, R.; Tonacci, A.; Campisi, S.; Vitabile, S.; Cerasa, A.; Distefano, S.; Pellegrino, A.; Modesti, P.A.; et al. An intelligent Medical Cyber–Physical System to support heart valve disease screening and diagnosis. Expert Syst. Appl. 2024, 238, 121772. [Google Scholar] [CrossRef]
5. Cuevas-Chávez, A.; Hernández, Y.; Ortiz-Hernandez, J.; Sánchez-Jiménez, E.; Ochoa-Ruiz, G.; Pérez, J.; González-Serna, G. A Systematic Review of Machine Learning and IoT Applied to the Prediction and Monitoring of Cardiovascular Diseases. Healthcare 2023, 11, 2240. [Google Scholar] [CrossRef] [PubMed]
6. Plati, D.K.; Tripoliti, E.E.; Bechlioulis, A.; Rammos, A.; Dimou, I.; Lakkas, L.; Watson, C.; McDonald, K.; Ledwidge, M.; Pharithi, R.; et al. A Machine Learning Approach for Chronic Heart Failure Diagnosis. Diagnostics 2021, 11, 1863. [Google Scholar] [CrossRef] [PubMed]

7. Senthil kumar mohan, chandrasegar thirumalai and Gautam Srivastva, “Effective Heart Disease Prediction Using Hybrid Machine Learning Techniques” IEEE Access 2019.
8. Amandeep Kaur and Jyoti Arora,“Heart Diseases Prediction using Data Mining Techniques: A survey” International Journal of Advanced Research in Computer Science , IJARCS 2015-2019.
9. Santhana Krishnan J and Geetha S, “Prediction of Heart Disease using Machine Learning Algorithms” ICIICT, 2019.
10. Aditi Gavhane, Gouthami Kokkula, Isha Panday, Prof. Kailash Devadkar, “Prediction of Heart Disease using Machine Learning”, Proceedings of the 2nd International conference on Electronics, Communication and Aerospace Technology (ICECA), 2018.
11. Animesh Hazra, Subrata Kumar Mandal, Amit Gupta, Arkomita Mukherjee and Asmita Mukherjee, "Heart Disease Diagnosis and Prediction Using Machine Learning and Data Mining Techniques: A Review", Advances in Computational Sciences and Technology, ISSN 0973-6107 Volume 10, Number 7 (2017) pp. 2137-2159
12. Praveen Kumar Reddy M, T Sunil Kumar Reddy, S.Balakrishnan, Syed Muzamil Basha Ravi Kumar Poluru, "Heart Disease Prediction Using Machine Learning Algorithm", International Journal of Innovative Technology and Exploring Engineering (IJITEE), ISSN: 2278-3075, Volume-8 Issue-10, August 2019
13. Muqaddas, M., Majeed, S., Hira, S., & Mumtaz, G. (2024). A Systematic Literature Review on Performance Evaluation of SQL and NoSQL Database Architectures. Journal of Compu-ting & Biomedical Informatics, 7(02). https://jcbi.org/index.php/Main/article/view/548/502
14. Krittanawong C, Virk HUH, Bangalore S, "Machine learning prediction in cardiovascular diseases: a meta-analysis". Sci Rep. 2020;10:16057. https:// doi. org/ 10. 1038/ s41598- 020- 72685-1.
15. Siontis KC, Noseworthy PA, Attia ZI, "Artificial intelligence-enhanced electrocardiography in cardiovascular disease management". Nat Rev Cardiol. 2021. https:// doi. org/ 10. 1038/ s41569- 020- 00503-2.
16. Linda PS, Yin W, Gregory PA, et al. "Development of a Novel Clinical Decision Support System for Exercise Prescription Among Patients With Multiple Cardiovascular Disease Risk Factors", Mayo Clinic Proceedings: Innovations, Quality & Outcomes,Volume 5, Issue 1,2021,Pages 193-203, ISSN 2542-4548.
17. Pahulpreet Singh Kohli and Shriya Arora, “Application of Machine Learning in Diseases Prediction”, 4th International Conference on Computing Communication and Automation (ICCCA), 2018.
18. Cardiovascular Disease Dataset, Kaggle, https://www.kaggle.com/datasets/sulianova/cardiovascular-disease-dataset
19. Feature Importance, Built-In (2011), https://builtin.com/data-science/feature-importance
20. Muhammad Zohaib Khan, Sarmad Ahmed Shaikh, Muneer Ahmed Shaikh, Kamlesh Kumar Khatri, Mahira Abdul Rauf, Ayesha Kalhoro, Muhammad Adnan, "The Performance Analysis of Machine Learning Algorithms for Credit Card Fraud Detection", iJOE – Vol. 19, No. 03, 2022, https://doi.org/10.3991/ijoe.v19i03.35331
21. Data Clustering Algorithms, Fuzzy C-Means Clustering Algorithm, https://sites.google.com/site/dataclusteringalgorithms/fuzzy-c-means-clustering-algorithm
22. Shadab Yameen Shaikh, Naseem Afzal Qureshi, Muhammad Zohaib Khan, Muhammad Ali Khan, Aisha Imroz, Muhammad Ahmed Kalwar, "Performance Analysis of Classification Algorithms for Software Defects Prediction by Mathematical Modelling & Simulations", Sindh Journal of Headways in Software Engineering, Volume 02, Issue 01, http://sjhse.smiu.edu.pk/sjhse/index.php/SJHSE/article/view/54

23. Ramzan, M. S., Nasim, F., Ahmed, H. N., Farooq, U., & Khan, H. (2025). An innovative machine learning-based end-to-end data security framework in emerging cloud computing databases and integrated paradigms: Analysis on taxonomy, challenges, and opportunities. Spectrum of Engineering Sciences, 3(2), 90–125. https://sesjournal.com/index.php/1/article/view/106
24. Multi-Layer Perceptrons (MLPs) Algorithm, Pathmind and DeepAI (2020), https://wiki.pathmind.com/multilayer-perceptron, and https://deepai.org/machine-learning-glossary-and-terms/multilayer-perceptron.
25. Parveen, F., Iqbal, S., Mumtaz, G., & Salahuddin, M. (2024). Real-time intrusion detection with deep learning: Analyzing the UNR intrusion detection dataset. Journal of Computing & Biomedical Informatics, 7(02). https://jcbi.org/index.php/Main/article/view/554
26. Afzal, M., Salahuddin, M., Hira, S., Sultan, M. F., Ahmad, S. Z., & Iqbal, M. W. (2024). A systematic literature review of understanding the human-computer interaction collaboration with user experience design. Bulletin of Business and Economics, 13(2), 723–729. https://doi.org/10.61506/01.00386
27. Stochastic Gradient Descent (SGD), Tutorials Point, https://www.tutorialspoint.com/scikit_learn/scikit_learn_stochastic_gradient_descent.htm
28. Quadratic Discriminant Analysis, Society for Elimination of Rural Poverty (SERP) (2014), https://serp.ai/quadratic-discriminant-analysis/
29. Moeez, M., Mahmood, R., Asif, H., Iqbal, M. W., Hamid, K., Ali, U., & Khan, N. (2024). Comprehensive Analysis of DevOps: Integration, Automation, Collaboration, and Continuous Delivery. Bulletin of Business and Economics (BBE), 13(1).
30. Parthiban G, Srivatsa SK. “Applying machine learning methods in diagnosing heart disease for diabetic patients”. Int J Appl Inf Syst (IJAIS). 2012;3(7):25–30.
31. Dwivedi AK. “Performance evaluation of different machine learning techniques for prediction of heart disease”. Neural Comput Appl. 2018;29(10):685–693.
32. Otoom AF, Abdallah EE, Kilani Y, Kefaye A, Ashour M. “Effective diagnosis and monitoring of heart disease”. Int J Softw Eng Appl. 2015;9(1):143–56.