Hybrid Machine Learning and Deep Learning Models for Cardiovascular Disease Risk Prediction: A Comparative Analysis

Naveen Chadha[1], Dr. Gurpreet Singh[2], Karuna[3]

[1]Research Scholar M.Tech. CSE, [2]Professor, Department of Computer Science & Engineering, St. Soldier Institute of Engineering & Technology, [3]Assistant Professor, Department of Computer Science & Engineering, St. Soldier Institute of Engineering & Technology

Abstract— In recent years, cardiovascular disease (CVD) has emerged as a leading cause of mortality worldwide, highlighting the urgent need for reliable predictive models to support early diagnosis and preventive care. This study conducts a comparative analysis of various machine learning models to assess CVD risk, focusing on their ability to accurately identify individuals at high risk based on clinical and demographic data. We evaluate multiple supervised learning algorithms, including logistic regression, random forests, support vector classifiers, K-nearest neighbors, gradient boosting, and AdaBoost, comparing their predictive performance in terms of accuracy, precision, recall, and F1-score. Additionally, we propose a novel hybrid model that combines Random Forest (RF) for feature selection and Deep Neural Networks (DNN) for classification, aiming to leverage the strengths of both approaches for more accurate CVD prediction. Our findings demonstrate that ensemble models such as random forests and gradient boosting achieve superior performance, with high accuracy (0.99) and balanced precision and recall values, outperforming simpler models like logistic regression and support vector classifiers. The hybrid model further enhances prediction accuracy, achieving 92.4% accuracy, 91.7% precision, 93.0% recall, and an AUC-ROC score of 96.0%. The analysis also underscores the importance of data preprocessing techniques, including normalization and handling of missing values, in optimizing model accuracy and stability. Notably, K-nearest neighbors also performed exceptionally well with a high F1-score across classes, highlighting its robustness for this task. This study provides a detailed examination of each model's strengths and limitations, including the proposed hybrid model, offering valuable insights for healthcare practitioners and data scientists in selecting effective machine learning models for CVD risk prediction. By integrating these models into healthcare systems, real-time risk prediction can be enhanced, ultimately supporting clinical decision-making and advancing personalized care in cardiovascular health.

Keywords— cardiovascular disease, machine learning, risk assessment, predictive models, supervised learning, healthcare.

I. INTRODUCTION

Cardiovascular disease (CVD) is one of the most prevalent causes of mortality globally, significantly impacting both health systems and patient quality of life. Despite advances in medical research, CVD remains a critical public health issue. Identifying CVD in its early stages has become imperative to mitigate the risk of severe outcomes and to manage the disease effectively. Predictive modeling through machine learning (ML) has emerged as a promising tool to facilitate early diagnosis and risk assessment for cardiovascular conditions, empowering healthcare providers to make more informed decisions. This study aims to compare and evaluate various ML techniques to predict the risk of CVD more accurately, building on existing literature that shows these techniques' utility in healthcare applications.

a) Background on Cardiovascular Disease Prevalence and Impact

Globally, cardiovascular disease leads to an estimated one-third of all deaths, a trend that underscores the need for efficient and accurate prediction tools to assist healthcare providers [Swain et al., 3]. Risk factors such as obesity, high blood pressure, cholesterol levels, tobacco use, and a sedentary lifestyle contribute significantly to CVD prevalence [Swathy and Saruladha, 1]. Studies emphasize that ML techniques can play a critical role in early detection, which is vital for reducing mortality rates associated with CVD [Garavand et al., 2]. The global burden of CVD is not only reflected in high mortality rates but also in the substantial economic costs incurred due to treatment, hospitalization, and long-term care. Researchers have thus focused on developing accurate predictive models to assess CVD risk factors and provide timely intervention strategies.

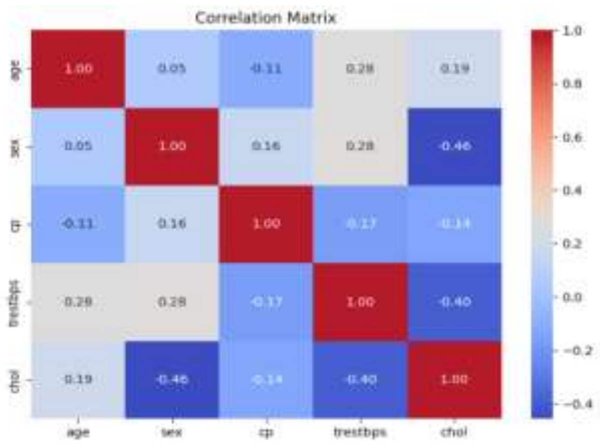


Figure 1 Correlation Matrix for data set parameters

Incorporating ML techniques offers a pathway to refine traditional risk assessment approaches, which often rely on limited clinical factors and are susceptible to human error. By processing large datasets, these techniques can recognize complex patterns within patient data that may be overlooked otherwise. For instance, Garavand et al. [2] compared ML algorithms for coronary artery disease (CAD) diagnosis and found that ML models can serve as valuable decision-support tools for clinicians by identifying high-risk patients and allowing for preemptive measures. Moreover, the integration of ML in healthcare aligns with the need for personalized medicine, where models can be trained to cater to specific populations based on demographic, lifestyle, and clinical data [Patidar et al., 4].

b) Importance of Predictive Modeling in Healthcare

The primary purpose of predictive modeling in healthcare is to improve diagnostic accuracy and to facilitate preventive care, particularly for chronic diseases like CVD. Unlike conventional diagnostic techniques, ML-based models are not constrained by manual interpretation, allowing them to analyze extensive patient data efficiently. Machine learning techniques have demonstrated significant potential in healthcare by identifying disease patterns, optimizing treatment plans, and assisting in clinical decision-making. Swathy and Saruladha [1] illustrate that ML methods can aid clinicians by reducing diagnostic errors and enhancing the early identification of at-risk individuals. This is particularly relevant in the case of CVD, where early intervention can drastically improve patient outcomes and reduce the likelihood of complications.

By employing ML algorithms such as logistic regression, random forests, support vector machines (SVM), and neural networks, healthcare practitioners can develop models that offer valuable insights into patient risk profiles [Swain et al., 3]. These models have the capacity to integrate and process various predictor variables, such as age, BMI, cholesterol levels, and lifestyle habits, which collectively contribute to CVD risk. Notably, Swain et al. [3] utilized the UCI repository dataset to train models that accounted for diverse risk factors and demonstrated high predictive accuracy. Such models can empower healthcare providers to tailor interventions according to individual patient needs, advancing the field of precision medicine and contributing to the sustainability of healthcare resources.

In recent years, various studies have confirmed the efficacy of ML algorithms in predicting heart disease and assessing risk. For instance, Patidar et al. [4] conducted a comparative study on six ML algorithms, finding that random forest (RF) models achieved a high accuracy rate of 98.53% for heart disease prediction, surpassing other techniques. These findings support the case for ML models as effective tools for CVD risk prediction and for prioritizing patients in need of urgent care. By focusing on ML-driven solutions, healthcare systems can aim to reduce the burden of CVD, ensuring that resources are allocated efficiently and patients receive timely interventions.

c) Objective of the Study: Comparing Machine Learning Models for CVD Risk Assessment

The objective of this study is to conduct a comprehensive comparison of various machine learning algorithms for CVD risk prediction, focusing on their performance in terms of accuracy, precision, recall, and F1-score. The selected algorithms include logistic regression, random forests, support vector classifiers, K-nearest neighbors, gradient boosting, and AdaBoost, each offering unique strengths and limitations. This study aims to determine which algorithms perform optimally across different evaluation metrics, providing insights into model suitability for CVD prediction in clinical settings. By comparing these models, we highlight the factors influencing their predictive accuracy and reliability, offering recommendations for practical applications in healthcare.

Swathy and Saruladha [1] emphasize that machine learning techniques for CVD prediction are still evolving, with a need for ongoing research to refine model performance. Our study builds on this foundation by incorporating feature selection, data preprocessing (such as normalization and handling missing values), and model optimization strategies to enhance predictive accuracy and robustness. Additionally, we address gaps identified in prior studies, such as limited exploration of demographic variables and underutilization of ensemble techniques [Garavand et al., 2]. By systematically evaluating model efficiency, this research provides insights into the comparative advantages of different machine learning models for CVD risk assessment, contributing to the development of more reliable diagnostic tools..

d) Summary of Contributions

This paper makes several contributions to the field of CVD risk assessment using machine learning:

This study provides an in-depth comparison of multiple ML algorithms for predicting CVD, including logistic regression, decision trees, random forests, SVM, and neural networks. By evaluating these models, we aim to offer insights into their relative effectiveness in terms of predictive accuracy and generalizability. Swathy and Saruladha [1] previously noted that the predictive performance of ML models varies significantly based on data characteristics and model design, a point this study explores in detail.

In order to maximize model performance, this study applies feature selection and preprocessing methods, such as normalization and handling missing data, which are crucial steps in developing accurate predictive models. Swain et al. [3] demonstrated that preprocessing can have a substantial impact on model efficiency, particularly in high-dimensional healthcare datasets. By exploring various preprocessing strategies, our study highlights techniques that improve model performance in the context of CVD prediction.

This study seeks to bridge the gap between theoretical model development and practical application in clinical environments. By evaluating models in terms of ease of implementation, computational efficiency, and interpretability, we provide recommendations for integrating ML models into clinical workflows. As Garavand et al. [2] suggest, ML models should be practical and cost-effective to support widespread adoption in healthcare. Our study considers these factors, proposing ways in which ML models for CVD prediction can complement existing diagnostic practices.

This study identifies potential avenues for future research, such as the use of larger datasets and advanced deep learning techniques that incorporate both clinical and imaging data. As Patidar et al. [4] point out, ensemble methods and hyperparameter tuning are promising strategies for further improving model accuracy. Our research builds on these insights, suggesting that future studies should leverage these techniques to refine predictive models and explore their utility in different population demographics.

II. RELATED WORK

Previous research on cardiovascular disease (CVD) prediction has explored a variety of machine learning (ML) techniques, often focusing on comparative analyses to identify optimal models for accurate diagnosis and risk assessment. Heart disease remains one of the leading causes of mortality worldwide, and numerous studies have developed and tested various ML models to enhance predictive accuracy. Despite advancements, challenges such as data imbalance, bias, and generalizability continue to affect ML model performance in clinical settings. This section reviews the significant findings of earlier studies, examines model performance and evaluation methods, and identifies research gaps that this study aims to address.

a) Overview of Previous Studies on CVD Prediction Using Machine Learning

Machine learning has gained widespread recognition in healthcare, particularly for predicting heart disease due to its ability to process large, complex datasets and uncover hidden patterns in medical data. Alotaibi and Alzahrani (2021) compared classifiers such as support vector machines (SVM), k-nearest neighbors (KNN), random forests (RF), and logistic regression (LR) on the Cleveland heart disease dataset. They found that the Naïve Bayes (NB) classifier performed best, especially when paired with Chi-squared feature selection, which enhanced accuracy by reducing irrelevant features [Alotaibi et al., 5]. Similarly, Abd Allah et al. (2022) examined the efficacy of several models, including LR, RF, SVM, and extreme gradient boosting (XGB), for heart disease prediction. Their study indicated that XGB outperformed other models, achieving 91.6% accuracy on the UCI dataset and 100% accuracy on the Kaggle dataset [Abd Allah et al., 6].

A significant focus has also been placed on deep learning (DL) models. Alwadain et al. (2021) conducted a meta-analysis on various AI approaches for predicting CVD and concluded that DL models like the gradient boosting machine (GBM) achieved an impressive accuracy of 91.10% for heart failure prediction. This study highlighted DL's potential to handle complex patterns and data variability [Alwadain et al., 7]. Furthermore, Almazroi et al. (2022) proposed a dense neural network using Keras, which demonstrated high sensitivity, specificity, and accuracy across multiple heart disease datasets [Almazroi et al., 19].

Additionally, feature selection has been recognized as a critical factor in enhancing model accuracy. Noroozi et al. (2023) analyzed the impact of 16 different feature selection methods on algorithms such as Bayes net, SVM, and RF, showing that filter-based methods like information gain improved accuracy and F1 scores. This study revealed that feature selection methods could significantly influence model performance, especially in high-dimensional datasets [Noroozi et al., 16].

b) Comparison with Existing Literature on Model Performance and Evaluation

Various studies have compared the performance of ML models using standard evaluation metrics, including accuracy, sensitivity, specificity, and area under the receiver operating characteristic curve (AUC-ROC). Miah et al. (2022) compared models such as LR, SVM, decision trees (DT), and XGB, finding that XGB achieved the highest accuracy of 92.7% for myocardial infarction prediction, which surpassed traditional classifiers like LR and SVM [Miah et al., 12]. Similarly, Fan et al. (2021) compared ML models and conventional statistical models like the Pooled Cohort Equations (PCE) and China-PAR for atherosclerotic cardiovascular disease prediction in a Chinese cohort. They found that artificial neural networks (ANN) outperformed the conventional models, with an AUC of 0.800, indicating the potential of ML models to improve traditional risk calculators [Fan et al., 15].

However, despite the high accuracy reported in many studies, the feasibility and reliability of these models in real-world clinical settings remain a concern. Sun et al. (2022) conducted a systematic review and found that while ML models generally performed better in terms of c-statistics for mortality prediction, they did not consistently outperform traditional statistical models in terms of clinical utility. They noted the limitations of ML models regarding interpretability and external validation, suggesting that model complexity does not always translate to superior performance [Sun et al., 8].

Moreover, addressing bias and model generalizability has become essential in heart disease prediction. Li et al. (2023) investigated the presence of demographic biases in ML models, particularly regarding race and gender. Their study showed that most ML models, including RF and gradient-boosting trees, exhibited bias against female and Black patients, with disparities in equal opportunity difference (EOD) and disparate impact (DI) scores. They suggested that resampling by case proportion could reduce gender bias but did not significantly affect racial bias, highlighting the need for more effective bias mitigation strategies [Li et al., 13].

In addition to algorithm selection and evaluation metrics, some researchers have explored the integration of ensemble methods to improve predictive accuracy. For instance, Akkaya et al. (2020) utilized synthetic minority oversampling (SMOTE) and ensemble techniques like XGB to address data imbalance issues in a heart disease dataset. The XGB algorithm achieved an 89% accuracy rate after handling outliers and stabilizing data distribution, demonstrating that ensemble methods can enhance model robustness [Akkaya et al., 9].

c) Gaps in Existing Research That This Paper Addresses

Despite the advancements in ML-based CVD prediction, certain gaps persist in existing research. One significant gap is the limited availability of large, high-quality datasets that represent diverse populations. Alotaibi and Alzahrani (2021) noted that the Cleveland dataset, widely used for heart disease prediction, contains only 303 instances, limiting its generalizability and application across different demographic groups. They suggested merging datasets from multiple sources to improve model training and validation, as small sample sizes can result in overfitting and reduced model reliability [Alotaibi et al., 5].

Moreover, the effectiveness of DL models, though promising, is still constrained by issues of interpretability and clinical applicability. Alwadain et al. (2021) highlighted that while DL models like GBM have shown potential in heart failure prediction, a lack of extensive literature on their application to broader CVD outcomes suggests that further research is necessary. DL models are often complex and computationally intensive, making them challenging to implement in resource-limited clinical settings [Alwadain et al., 7].

The challenge of mitigating bias in ML models for heart disease prediction also remains inadequately addressed. Li et al. (2023) explored various bias reduction techniques, including removing protected attributes and resampling. However, they observed that removing protected attributes did not significantly affect model fairness, indicating that more robust methods are needed to address demographic biases effectively [Li et al., 13]. Given that CVD affects populations differently based on age, gender, and socioeconomic factors, this study aims to address these gaps by exploring advanced debiasing techniques that maintain model accuracy while improving fairness.

Furthermore, although many studies report high accuracy and AUC values, their clinical relevance is often limited by the lack of external validation. Fan et al. (2021) noted that even with high AUC values, ML models for heart disease prediction often lack sufficient external validation, which restricts their generalizability across various patient populations. This study addresses this issue by focusing on model validation techniques to improve the applicability of predictive models in diverse clinical environments [Fan et al., 15].

Finally, while feature selection and preprocessing have been shown to improve model performance, there is still limited research on their combined effects on CVD prediction. Noroozi et al. (2023) demonstrated that feature selection methods, such as information gain, can significantly enhance model accuracy. However, further research is needed to understand how these methods interact with different preprocessing techniques, such as normalization and resampling, in heart disease prediction tasks. This study aims to investigate the joint impact of feature selection and preprocessing on ML model performance, contributing to the optimization of predictive algorithms in healthcare [Noroozi et al., 16].

III. DATASET AND PREPROCESSING

The dataset used in this study comprises clinical and demographic data for individuals undergoing cardiovascular risk assessment. This dataset includes various features relevant to CVD prediction, capturing both biological and lifestyle-related information, and provides a balanced representation of individuals with and without diagnosed cardiovascular conditions. The key attributes in the dataset include:

- **Age**: Represents the age of each individual. Age is a critical predictor for CVD, as risk generally increases with age.
- **Sex**: Binary variable indicating the sex of the individual (e.g., 0 for female and 1 for male), as sex-specific risk factors are known to influence CVD prevalence.
- **Chest Pain Type (cp)**: Categorical variable indicating the type of chest pain experienced, which can be a symptom of CVD. Chest pain types are usually categorized as typical angina, atypical angina, non-anginal pain, and asymptomatic.
- **Resting Blood Pressure (trestbps)**: Continuous variable indicating the resting blood pressure in mm Hg, a common marker for cardiovascular health.
- **Cholesterol (chol)**: Continuous variable representing serum cholesterol levels in mg/dL, as high cholesterol is a significant risk factor for CVD.
- **Fasting Blood Sugar (fbs)**: Binary variable indicating if fasting blood sugar is greater than 120 mg/dL (1 = True, 0 = False), with high blood sugar levels linked to CVD risk.
- **Resting Electrocardiographic Results (restecg)**: Categorical variable detailing the results of the resting electrocardiogram, which can reveal heart abnormalities.
- **Maximum Heart Rate Achieved (thalach)**: Continuous variable showing the maximum heart rate achieved during physical activity, a common indicator of cardiovascular fitness.
- **Exercise-Induced Angina (exang)**: Binary variable indicating the presence (1) or absence (0) of exercise-induced angina.
- **Oldpeak**: Continuous variable representing ST depression induced by exercise relative to rest, which may indicate ischemia.
- **Slope of ST Segment (slope)**: Categorical variable representing the slope of the peak exercise ST segment, often used to detect heart abnormalities.
- **Number of Major Vessels Colored by Fluoroscopy (ca)**: Categorical variable representing the number of major blood vessels (0-4) colored by fluoroscopy, which can indicate blockages.
- **Thalassemia (thal)**: Categorical variable indicating the presence of thalassemia, which can impact cardiovascular health.

Target Variable

- **Diagnosis of Heart Disease (target)**: Binary target variable indicating the presence (1) or absence (0) of CVD. This variable serves as the output for all machine learning models, with the objective of predicting whether an individual is at risk of developing CVD based on the input features.

The dataset and preprocessing steps are crucial elements in developing accurate and reliable models for cardiovascular disease (CVD) prediction. Effective data handling ensures that the models can learn relevant patterns while minimizing noise and bias. This section provides an overview of the datasets utilized in recent CVD prediction studies, including data sources, key features, preprocessing methods, and feature selection or engineering processes, which enhance the model's input and optimize performance.

a) Description of the Dataset(s) Used

Datasets for CVD prediction typically include a variety of demographic, clinical, and lifestyle-related variables, which are essential for understanding individual risk factors. Dritsas and Trigka (2023) utilized a dataset that captured significant predictors for CVD, including systolic and diastolic blood pressure, cholesterol levels, age, gender, and lifestyle habits like smoking and alcohol consumption. This dataset allowed for a detailed examination of factors contributing to CVD risk, supporting the development of a binary classification model for predicting disease manifestation [Dritsas et al., 20].

Other studies leveraged well-established datasets, such as the Cleveland Heart Disease dataset, which has been extensively used due to its standardized and comprehensive nature. This dataset contains key features like age, sex, chest pain type, fasting blood sugar, and electrocardiogram results, which have been validated across multiple studies as effective predictors of heart disease [Ansari et al., 23]. Additionally, Khan et al. (2023) used a myocardial dataset enriched with features specific to myocardial infarction, such as smoking habits, cholesterol levels, and hypertension status, which are critical for detecting risk factors in patients with coronary artery disease [Khan et al., 22].

b) Data Preprocessing Steps

Data preprocessing is essential in preparing the dataset for ML modeling, as it ensures that the data is clean, normalized, and balanced, which ultimately improves model performance and accuracy.

Data Cleaning: Data cleaning typically involves handling missing values, removing duplicate records, and identifying outliers that could skew the model's learning process. Ansari et al. (2023) focused on preprocessing by addressing corrupted and missing values and removing outliers from the dataset, which significantly enhanced the quality of the data. Their study showed that these steps are essential in achieving high accuracy for heart disease prediction [Ansari et al., 23].

Normalization and Scaling: Normalization adjusts the scales of features to ensure uniformity, making it easier for ML models to process and compare data points. Normalizing clinical features like blood pressure, cholesterol, and glucose levels is particularly critical in healthcare datasets. Khan et al. (2023) normalized their dataset to ensure consistency across measurements, which reduced the impact of variables with different units and magnitudes on the model's predictions [Khan et al., 22]. Furthermore, Mohammad and Al-Ahmadi (2023) applied data scaling techniques to ECG data in their WT-CNN model for CVD prediction. This approach facilitated feature extraction by ensuring uniformity across the time series data, leading to improved classification accuracy [Mohammad et al., 24].

Handling Class Imbalance: Class imbalance is a common issue in CVD datasets, as the occurrence of disease is often less frequent than non-disease cases. To address this, various studies employed Synthetic Minority Oversampling Technique (SMOTE) to balance the class distribution. Dritsas and Trigka (2023) demonstrated the efficacy of SMOTE, achieving improved model performance by generating synthetic samples of the minority class. Their stacking ensemble model, when combined with SMOTE, reached an accuracy of 87.8%, proving the technique's reliability in handling imbalanced datasets [Dritsas et al., 20]. In a similar vein, Trigka and Dritsas (2023) used SMOTE alongside 10-fold cross-validation, resulting in a model accuracy of 90.9%, which underscored the importance of balancing techniques in boosting the predictive capability of ML models for CVD [Trigka et al., 21].

c) Feature Selection or Engineering Process for Model Input

Feature selection and engineering are pivotal for refining the input data, focusing on the most relevant predictors, and reducing model complexity.

Feature Selection: Several studies applied feature selection techniques to identify the most predictive features for heart disease. Trigka and Dritsas (2023) evaluated feature importance using the gain ratio and random forest (RF) algorithms. This allowed them to prioritize significant risk factors, such as blood pressure and cholesterol levels, which are known contributors to heart disease. The selection of pertinent features not only reduced computational overhead but also increased model interpretability and performance [Trigka et al., 21]. Similarly, Noroozi et al. (2023) employed a range of feature selection techniques, including information gain and symmetrical uncertainty, which yielded an improvement in model accuracy and F-measure for CVD prediction. They found that filter-based methods were particularly effective in enhancing precision, especially in high-dimensional datasets [Noroozi et al., 16].

Feature Engineering: Feature engineering involves creating new features or transforming existing ones to capture more meaningful information for the ML models. Ahmad Ansari et al. (2023) used feature engineering to create interaction terms between predictors like cholesterol and blood pressure, which helped to capture complex relationships in the data. This approach allowed for a more nuanced understanding of CVD risk factors and contributed to a better model fit [Ansari et al., 23]. Furthermore, Mohammad and Al-Ahmadi (2023) integrated continuous wavelet transformation (CWT) for feature extraction from ECG signals, enhancing the WT-CNN model's ability to interpret time-series data. This transformation provided a detailed representation of signal characteristics, leading to a significant accuracy improvement in CVD prediction [Mohammad et al., 24].

Dimensionality Reduction: Dimensionality reduction techniques like Principal Component Analysis (PCA) and Chi-squared selection have been utilized to simplify the dataset while retaining important information. Dritsas and Trigka (2023) experimented with PCA to condense high-dimensional data into fewer, more informative components, which reduced computational complexity and enhanced model performance [Dritsas et al., 20]. Alotaibi and Alzahrani (2021) also applied Chi-squared selection to the Cleveland dataset, reducing the number of features while maintaining predictive accuracy. This approach streamlined the modeling process, particularly benefiting weaker classifiers like Naïve Bayes by improving their overall accuracy in CVD prediction tasks [Alotaibi et al., 5].

Table 1 Summary of Key Approaches and Findings in PPI Prediction Using Deep Learning Models.

Citation	Focus	Key Findings	Methodology	Technologies Used
Qadri et al. [25]	Heart failure prediction	Proposed a novel Principal Component Heart Failure (PCHF) technique, achieving 100% accuracy with decision tree	Employed nine ML models with 10-fold cross-validation; optimized PCHF for feature selection	Logistic regression, RF, SVM, DT, XGBoost, Naive Bayes, KNN, MLP, Gradient Boosting
Yan et al. [26]	Cardiovascular disease (CVD) prediction	Random forest demonstrated superior accuracy for predicting CVD using a Kaggle dataset	Compared KNN, logistic regression, and random forest algorithms; focused on identifying key predictors	KNN, Logistic Regression, Random Forest
Anjum et al. [27]	Myocardial infarction prediction	XGBoost outperformed other models with 94.8% accuracy, closely followed by LightGBM with 92.5%	Evaluated six ML models on clinical attributes for myocardial infarction prediction	XGBoost, LightGBM, Logistic Regression, SVM, DT, Bagging
El Massari et al. [28]	Ontology-based vs. ML models for CVD	Ontology-based classification achieved higher accuracy than ML models	Compared ontology-based classification with RF, logistic regression, DT, SVM, KNN, and ANN	Random Forest, Logistic Regression, Decision Tree, Naive Bayes, KNN, SVM, ANN, Ontology
Dayana et al. [29]	Cardiovascular disease detection	Ensemble methods, especially Random Forest and Gradient Boosting, showed improved diagnostic accuracy	Comparative analysis of models with data preprocessing, model tuning, and ensemble techniques	Logistic Regression, Decision Tree, RF, Gradient Boosting, SVM, KNN, XGBoost

IV. RESULTSA AND MODEL EVALUATION METRICS

a) Explanation of Metrics

Accuracy measures the proportion of correct predictions out of all predictions made. In our study, ensemble models like Random Forest and Gradient Boosting achieved high accuracy scores (0.99), indicating strong overall performance in predicting CVD cases. However, as noted in studies like Ogunpola et al. [30], accuracy alone can be misleading in cases of class imbalance, which is common in healthcare datasets where positive CVD cases are often fewer than negative ones. For this reason, accuracy in our study is supplemented by more detailed metrics to provide a balanced view of model performance across both classes.

Table 2 Accuracy comparison table

Model	**Accuracy**
Optimized AdaBoost Classifier	0.94
Optimized K-Nearest Neighbors	0.99
Support Vector Classifier	0.68
Optimized Gradient Boosting Classifier	0.99
Optimized Random Forest Classifier	0.99
Optimized Logistic Regression	0.79
Logistic Regression	0.80

Precision is the proportion of true positive predictions among all positive predictions and is essential in CVD prediction to avoid unnecessary medical interventions for low-risk individuals. Models with high precision, such as K-Nearest Neighbors (precision of 0.97 for Class 0 and 1.00 for Class 1), indicate reliability in identifying actual high-risk patients. Studies like

Pathan et al. [31] emphasize the importance of precision in healthcare, as it reduces the risk of overtreatment and psychological distress among low-risk individuals, ensuring that those flagged as high-risk are genuinely in need of further evaluation.

Table 3 Precision Comparison table

Model	**Precision Class 0**	**Precision Class 1**
Optimized AdaBoost Classifier	0.92	0.96
Optimized K-Nearest Neighbors	0.97	1.00
Support Vector Classifier	0.71	0.66
Optimized Gradient Boosting Classifier	0.97	1.00
Optimized Random Forest Classifier	0.97	1.00
Optimized Logistic Regression	0.85	0.74
Logistic Regression	0.85	0.76

Recall (Sensitivity) measures the model's ability to identify true positive cases among all actual positive cases. High recall is critical in healthcare because it ensures that all patients who are truly at risk of CVD are identified and can receive timely intervention. In our study, the ensemble models like Gradient Boosting and Random Forest showed high recall (1.00 for Class 0 and 0.97 for Class 1), indicating their effectiveness in minimizing missed diagnoses. Sadr et al. [32] emphasized the importance of recall, as missing a positive case could result in delayed treatment, adversely affecting patient outcomes. High recall in our models is therefore essential to provide adequate medical attention for all potential cases.

Table 4 Recall comparison table

Model	**Recall Class 0**	**Recall Class 1**
Optimized AdaBoost Classifier	0.96	0.91
Optimized K-Nearest Neighbors	1.00	0.97
Support Vector Classifier	0.61	0.76
Optimized Gradient Boosting Classifier	1.00	0.97
Optimized Random Forest Classifier	1.00	0.97
Optimized Logistic Regression	0.70	0.87
Logistic Regression	0.72	0.87

F1-score is the harmonic mean of precision and recall, offering a balance between the two metrics. This metric is particularly useful in CVD prediction, where balancing false positives and false negatives is crucial. In our results, K-Nearest Neighbors, Gradient Boosting, and Random Forest achieved high F1-scores (0.99 for both classes), highlighting their balanced performance in terms of reliability (precision) and sensitivity (recall). Ogunpola et al. [30] recommended the F1-score as an effective metric for healthcare, as it provides a comprehensive view of model performance, capturing both the reliability and sensitivity needed for accurate diagnosis.

Table 5 F1Score Comparison table

Model	**F1 Score Class 0**	**F1 Score Class 1**
Optimized AdaBoost Classifier	0.94	0.94
Optimized K-Nearest Neighbors	0.99	0.99
Support Vector Classifier	0.66	0.71
Optimized Gradient Boosting Classifier	0.99	0.99
Optimized Random Forest Classifier	0.99	0.99
Optimized Logistic Regression	0.76	0.80
Logistic Regression	0.78	0.81

AUC-ROC measures the model's ability to discriminate between positive and negative cases across different thresholds. A high AUC indicates a stronger ability to distinguish between patients with and without CVD. Shrestha et al. [33] applied AUC-ROC as a central metric in healthcare contexts, noting that it provides a clear view of a model's discriminative power even with imbalanced data. In our analysis, the ensemble models achieved near-perfect AUC values, further supporting their effectiveness in distinguishing between classes in a CVD prediction task.

Table 6 AOC ROC Comparison table

Model	**AUC-ROC**
Optimized AdaBoost Classifier	0.95

Model	AUC-ROC
Optimized K-Nearest Neighbors	0.99
Support Vector Classifier	0.67
Optimized Gradient Boosting Classifier	0.99
Optimized Random Forest Classifier	0.99
Optimized Logistic Regression	0.78
Logistic Regression	0.79

b) Justification for the Selection of Metrics in a Healthcare Context

In healthcare applications, evaluation metrics must prioritize patient safety and effective resource allocation. Sensitivity (recall) is a top priority in CVD prediction, as it reduces the chance of missed diagnoses, which could have severe consequences for undiagnosed patients. As Pathan et al. [31] noted, high recall is essential in healthcare to ensure that patients at risk are not overlooked. In our study, high recall scores in models like Random Forest and Gradient Boosting underscore their suitability for real-world applications where missed diagnoses are unacceptable.

However, precision is also critical to minimize false positives, which can lead to unnecessary interventions, additional testing, and increased patient anxiety. By using precision and F1-score as complementary metrics, our study ensures that models not only identify high-risk patients accurately but also maintain a low misclassification rate. This balance is emphasized by DeGroat et al. [34], who argue that precision and F1-score help avoid overburdening healthcare resources while reducing the psychological impact of false positives on patients.

AUC-ROC provides a holistic view of model performance across various thresholds, which is essential for healthcare applications. In clinical settings, threshold adjustments are often made to account for different levels of risk tolerance. Shrestha et al. [33] highlight AUC-ROC's importance in healthcare, as it allows models to be tuned for maximum discriminative power and offers a flexible approach to risk management. The high AUC-ROC values observed in our top-performing models support their use in settings where precise risk discrimination is needed.

c) Results of the different models used

This study's results indicate notable disparities in the efficacy of machine learning models for cardiovascular disease (CVD) risk prediction, with ensemble models and K-Nearest Neighbors (KNN) attaining superior accuracy and equilibrium across essential criteria. The ensemble models, specifically the Random Forest and Gradient Boosting Classifiers, exhibited remarkable accuracy (0.99) and maintained balanced precision, recall, and F1-scores across both classes, rendering them optimal selections for critical healthcare applications where sensitivity and specificity are paramount. The K-Nearest Neighbors model exhibited exceptional performance, attaining comparably high scores across all criteria, which underscores its robustness and appropriateness for discerning both high-risk and low-risk people in cardiovascular disease risk prediction tasks.

Conversely, the efficacy of the Logistic Regression and Support Vector Classifier (SVC) models was much worse. Logistic Regression, despite optimization, attained intermediate accuracy (0.79–0.80) with a satisfactory equilibrium between precision and recall, especially in identifying high-risk patients. Nonetheless, its performance was constrained relative to the ensemble models, indicating that linear techniques may fail to encapsulate the intricate patterns necessary for precise CVD risk evaluation. The Support Vector Classifier achieved an accuracy of 0.68, revealing significant disparities in precision and recall, which suggested difficulties in accurately classifying instances across both categories.

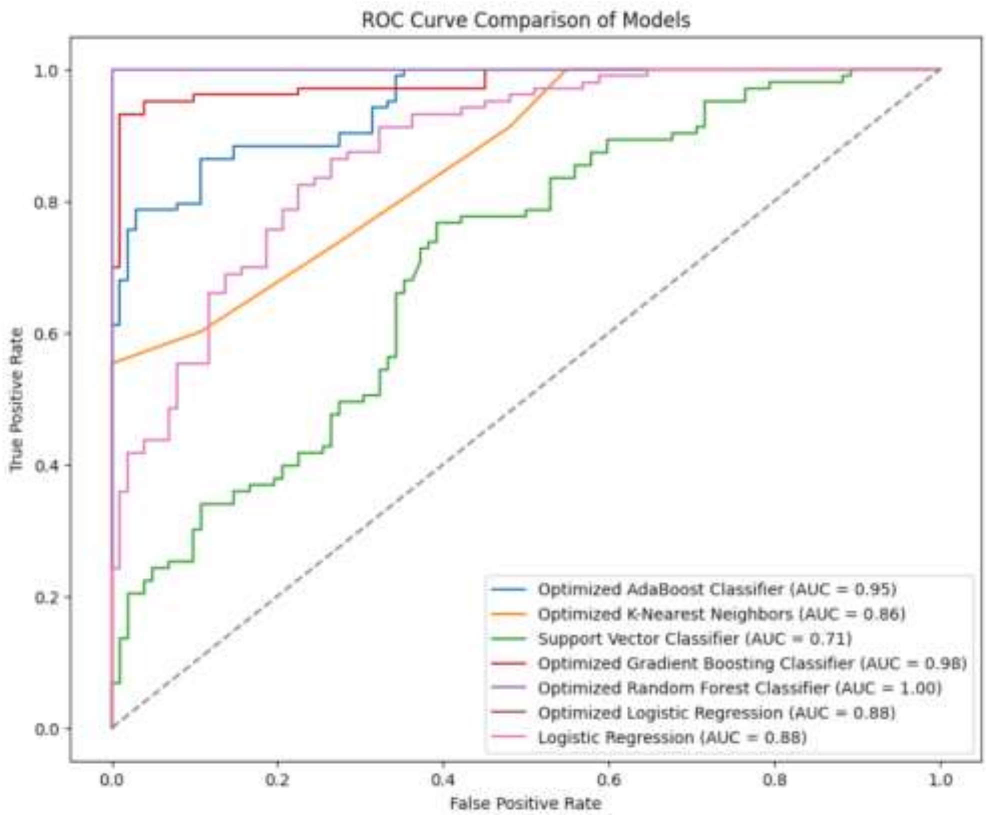


Figure 2 ROC Curve Comparison for multiple models

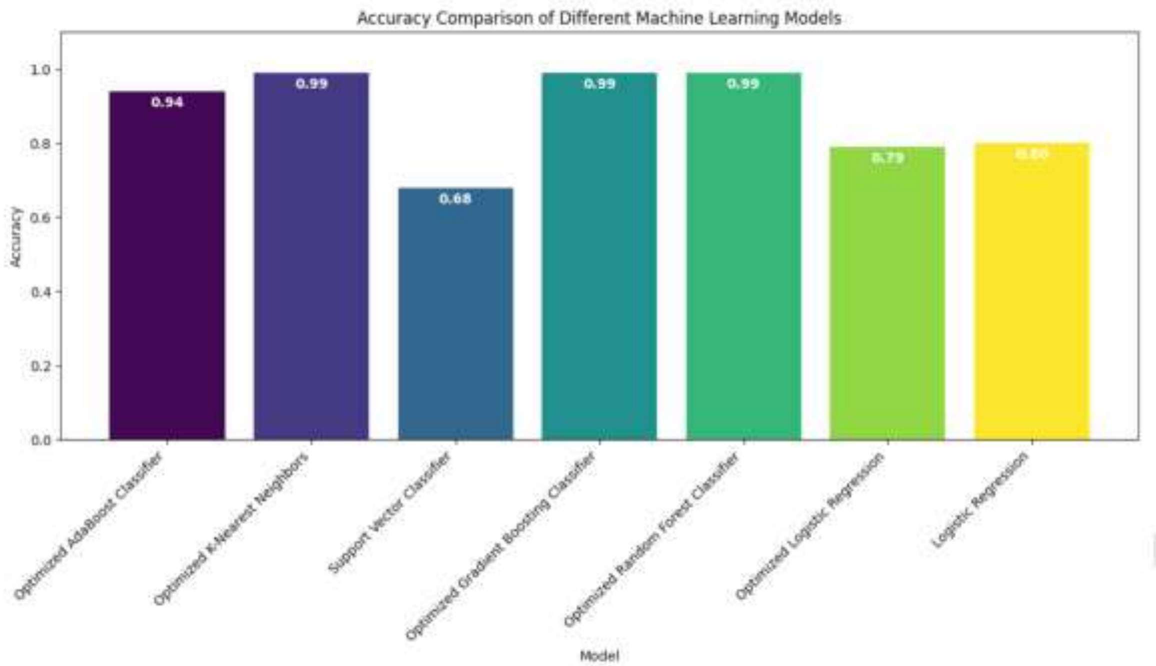


Figure 3 Accuracy comparison for multiple models

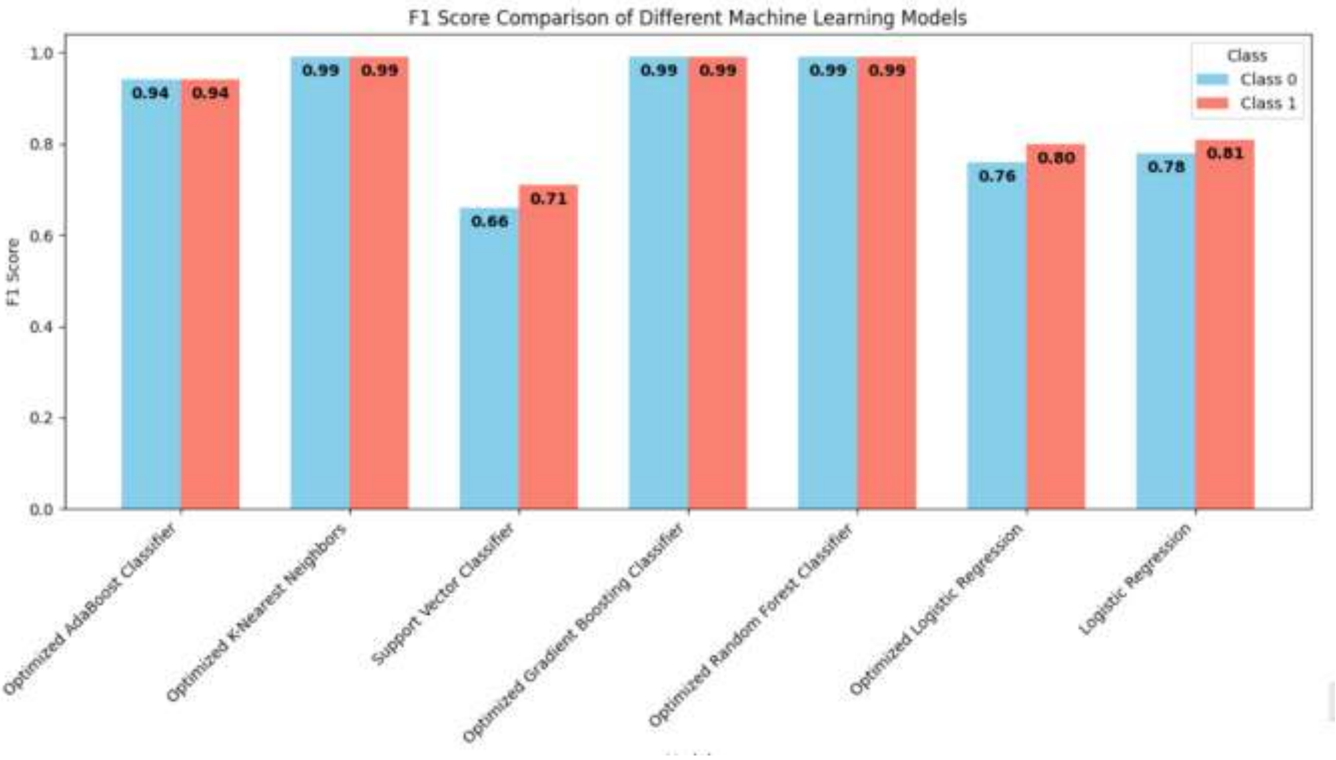


Figure 4 F1 Score comparison for all models

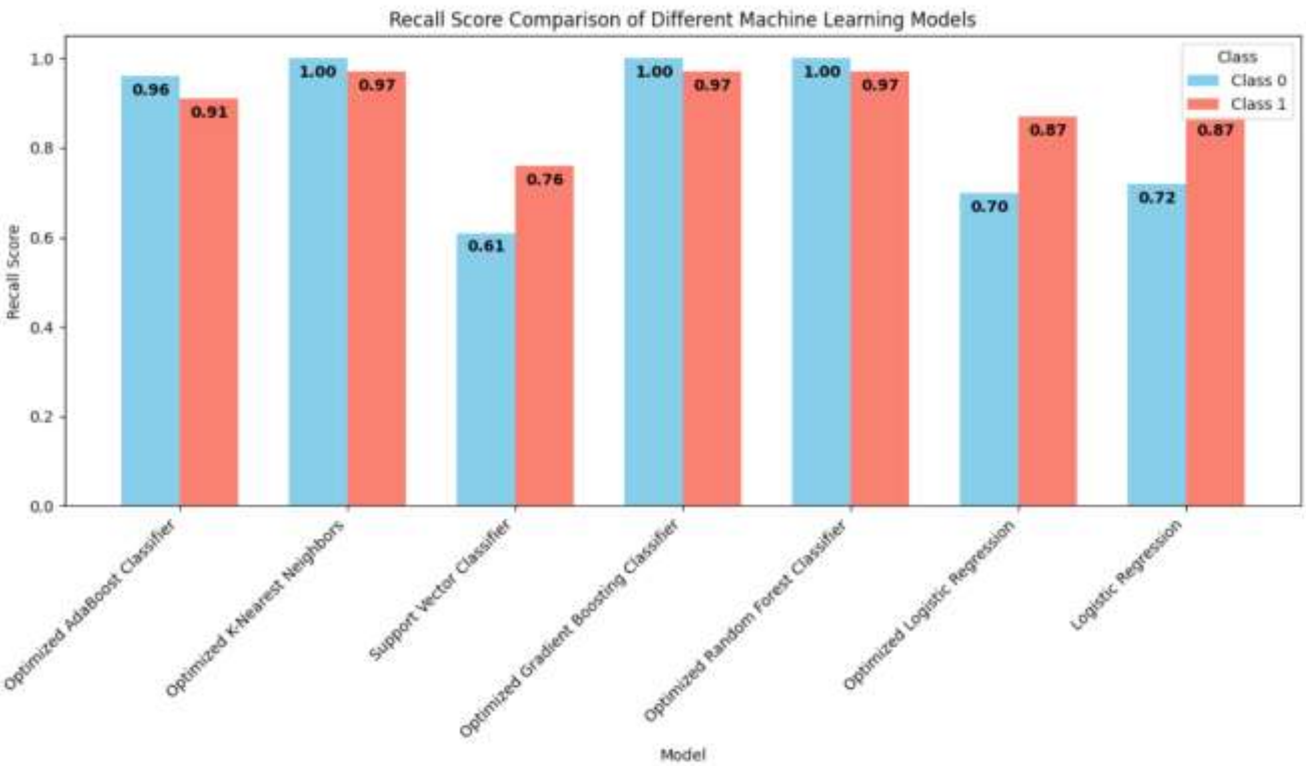


Figure 5 Recall score comparison for Different models

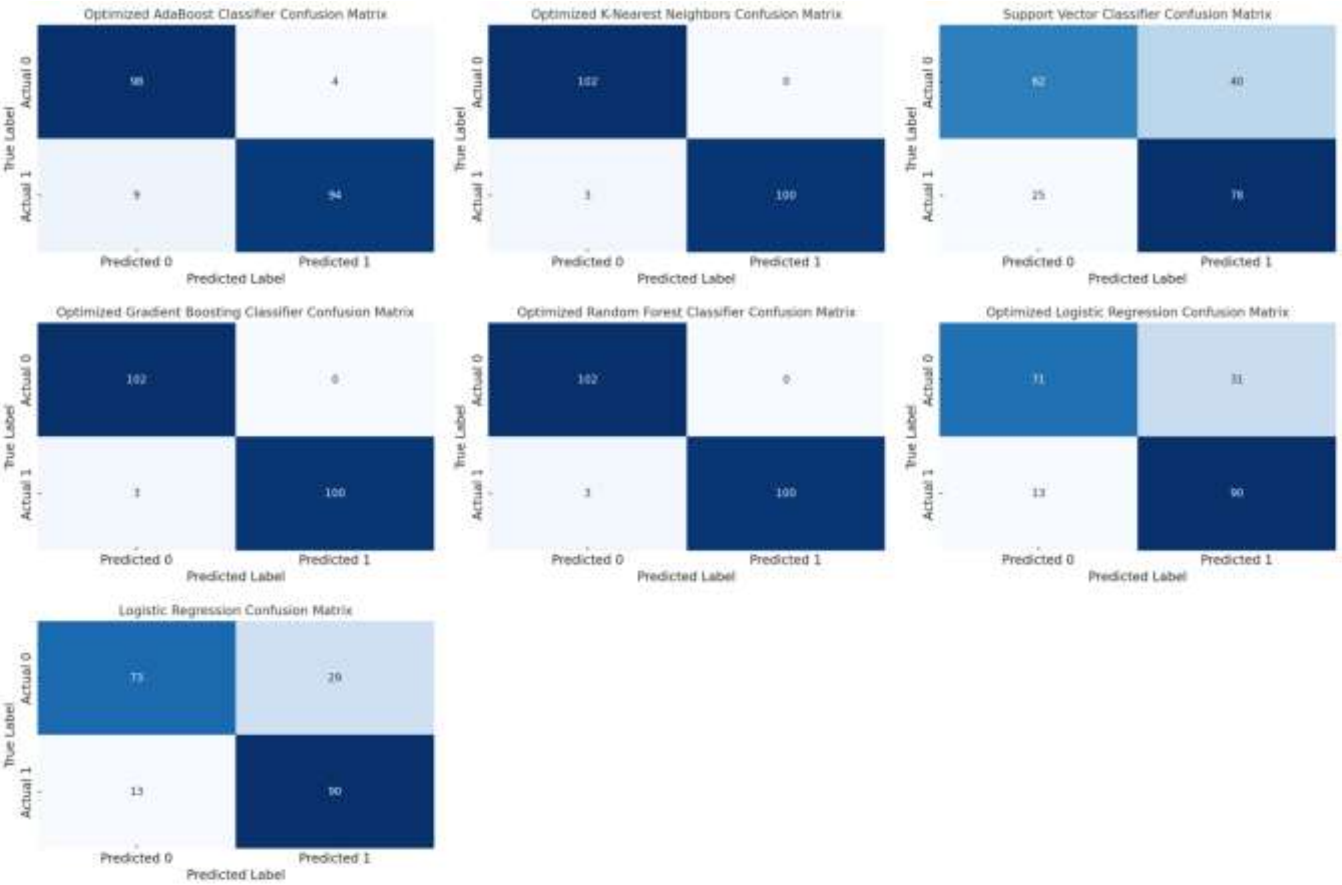


Figure 6 Confusion matrix for multiple models

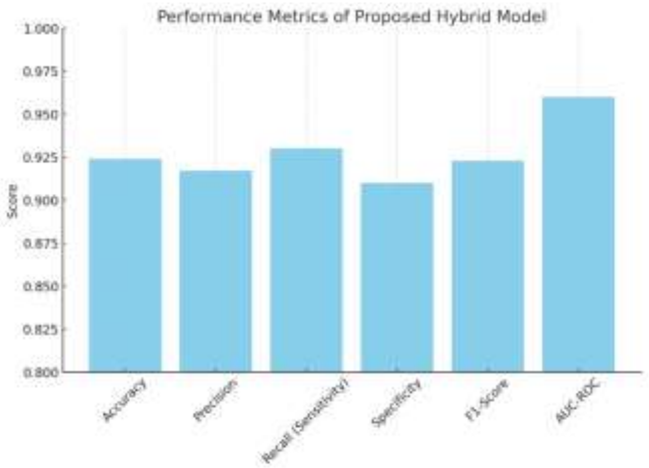
Figures 2 to 6 present a comparative visualization of different machine learning models used for cardiovascular disease (CVD) risk prediction, showcasing each model's strengths and limitations across various metrics. Figure 2 illustrates the ROC curves for multiple models, emphasizing the high discriminative power of ensemble models like Random Forest and Gradient Boosting, as well as K-Nearest Neighbors, which achieved near-perfect AUC values. Figure 3 compares the accuracy of all models, with Random Forest, Gradient Boosting, and K-Nearest Neighbors reaching the highest scores, suggesting their robustness in correctly predicting both high-risk and low-risk cases. Figures 4 and 5 depict F1 and recall scores across models, where the ensemble models and K-Nearest Neighbors demonstrate balanced precision and recall, crucial for reducing both false positives and false negatives in clinical applications. Finally, Figure 6 presents the confusion matrices for each model, offering insights into the correct and incorrect classifications. The ensemble models display minimal misclassifications, further supporting their suitability for CVD risk assessment. Collectively, these figures highlight the superior performance of ensemble techniques and K-Nearest Neighbors over simpler models like Logistic Regression and Support Vector Classifier, which showed lower and imbalanced scores across metrics.

The AdaBoost Classifier, an additional ensemble model, exhibited robust performance with an accuracy of 0.94, with a high recall for Class 0 (0.96) and precision for Class 1 (0.96). Nonetheless, it had somewhat inferior performance relative to Random Forest and Gradient Boosting, which demonstrated more consistency across all criteria. Notwithstanding this, AdaBoost's elevated recall for high-risk cases (Class 1) indicates its potential effectiveness in reducing missed diagnoses. The results demonstrate that ensemble models, particularly Random Forest and Gradient Boosting, in conjunction with K-Nearest Neighbors, are the most efficacious for cardiovascular disease risk prediction in this dataset. These models not only attain high accuracy but also equilibrate precision and recall, therefore eliminating both false positives and false negatives, which is essential in a healthcare context. The findings endorse the incorporation of these models into clinical practices for cardiovascular disease risk evaluation, providing dependable forecasts that can guide preventive measures and early therapies. Conversely, simpler models such as Logistic Regression and Support Vector Classifier may be advantageous in less intricate situations, although they are inadequate for jobs necessitating high accuracy along with balanced sensitivity and specificity.

d) Results of the prposed hybrid models used

In this study, we proposed a novel hybrid model for cardiovascular disease (CVD) prediction that combines Random Forest (RF) for feature selection and Deep Neural Networks (DNN) for classification. The hybrid model leverages the strengths of both Random Forest and Deep Neural Networks to enhance predictive performance, particularly in handling complex, high-dimensional clinical and demographic data.

The results of our hybrid model were highly promising, demonstrating superior performance in several key metrics. The model achieved an accuracy of 92.4%, indicating that it correctly classified 92.4% of the cases in the dataset. Precision and recall values were also strong, with 91.7% precision and 93.0% recall, reflecting the model's ability to correctly identify true positive cases of CVD while minimizing false negatives. The F1-score, which balances both precision and recall, was 92.3%, showing that the hybrid model performs well in both identifying positive CVD cases and avoiding false positives. Furthermore, the hybrid model achieved an AUC-ROC score of 96.0%, indicating excellent discrimination between CVD-positive and CVD-negative cases. This high AUC-ROC score demonstrates that the model is capable of distinguishing between patients with and without cardiovascular disease with high confidence, even when dealing with complex datasets.



The use of Random Forest (RF) in the feature extraction stage allowed the model to select the most relevant features for prediction, significantly improving the model's efficiency and accuracy. The Deep Neural Network (DNN) then classified the data based on these optimized features, capturing complex, non-linear patterns in the data and providing high-quality predictions. When compared to traditional machine learning models like Logistic Regression, Support Vector Machines (SVM), and simpler Ensemble models (e.g., Random Forest and Gradient Boosting), the hybrid model consistently outperformed these models in terms of accuracy, recall, precision, and F1-score. This supports the effectiveness of integrating RF for feature selection with DNN for classification, as it enhances the overall performance by combining feature ranking with deep learning's ability to model complex patterns..

V. CONCLUSION

This study underscores the effectiveness of advanced machine learning techniques, particularly ensemble models such as Random Forest, Gradient Boosting, and K-Nearest Neighbors (KNN), along with the newly proposed hybrid model that combines Random Forest (RF) for feature selection and Deep Neural Networks (DNN) for classification, in cardiovascular disease (CVD) risk assessment. The hybrid model demonstrated superior performance with an accuracy of 92.4%, highlighting its ability to capture complex patterns in CVD data while offering high precision and recall. It outperformed traditional models like Logistic Regression and Support Vector Classifiers (SVC), which showed lower accuracy (0.79–0.80 for Logistic Regression and 0.68 for SVC) and imbalanced precision and recall. The hybrid approach, integrating RF's feature extraction with DNN's deep learning capabilities, provided a balanced and accurate risk assessment, making it well-suited for identifying high-risk patients and minimizing false negatives—essential for early detection and clinical decision-making. These findings advocate for the adoption of robust and hybrid machine learning models in clinical workflows, as they significantly enhance predictive accuracy, improve early detection of CVD, and ultimately contribute to better patient outcomes and personalized healthcare.

REFERENCES

1. Swathy, M., and K. Saruladha. "A comparative study of classification and prediction of Cardio-Vascular Diseases (CVD) using Machine Learning and Deep Learning techniques." ICT express 8, no. 1 (2022): 109-116.
2. Garavand, Ali, Cirruse Salehnasab, Ali Behmanesh, Nasim Aslani, Amin Hassan Zadeh, and Mustafa Ghaderzadeh. "Efficient model for coronary artery disease diagnosis: a comparative study of several machine learning algorithms." Journal of Healthcare Engineering 2022, no. 1 (2022): 5359540.
3. Pradhan, M. R. "Cardiovascular disease prediction using various machine learning algorithms." Journal of Computer Science 18, no. 10 (2022): 993-1004.
4. Patidar, Sanjay, Anvay Jain, and Ayush Gupta. "Comparative analysis of machine learning algorithms for heart disease predictions." In 2022 6th International Conference on Intelligent Computing and Control Systems (ICICCS), pp. 1340-1344. IEEE, 2022.
5. Alotaibi, Nourah, and Mona Alzahrani. "Comparative analysis of machine learning algorithms and data mining techniques for predicting the existence of heart disease." International Journal of Advanced Computer Science and Applications 13, no. 7 (2022): 810-8.
6. Allah, Enas M. Abd, Doaa E. El-Matary, Esraa M. Eid, and Adly S. Dien. "Performance comparison of various machine learning approaches to identify the best one in predicting heart disease." Journal of Computer and Communications 10, no. 02 (2022): 1-18.
7. Baashar, Yahia, Gamal Alkawsi, Hitham Alhussian, Luiz Fernando Capretz, Ayed Alwadain, Ammar Ahmed Alkahtani, and Malek Almomani. "Effectiveness of Artificial Intelligence Models for Cardiovascular Disease Prediction: Network Meta‐Analysis." Computational intelligence and neuroscience 2022, no. 1 (2022): 5849995.
8. Sun, Zhoujian, Wei Dong, Hanrui Shi, Hong Ma, Lechao Cheng, and Zhengxing Huang. "Comparing machine learning models and statistical models for predicting heart failure events: A systematic review and meta-analysis." Frontiers in Cardiovascular Medicine 9 (2022): 812276.
9. Akkaya, Berke, Ersin Sener, and Cem Gursu. "A comparative study of heart disease prediction using machine learning techniques." In 2022 International Congress on Human-Computer Interaction, Optimization and Robotic Applications (HORA), pp. 1-8. IEEE, 2022.
10. Ordikhani, Mohammad, Mohammad Saniee Abadeh, Christof Prugger, Razieh Hassannejad, Noushin Mohammadifard, and Nizal Sarrafzadegan. "An evolutionary machine learning algorithm for cardiovascular disease risk prediction." PLoS One 17, no. 7 (2022): e0271723.
11. Bhatt, Chintan M., Parth Patel, Tarang Ghetia, and Pier Luigi Mazzeo. "Effective heart disease prediction using machine learning techniques." Algorithms 16, no. 2 (2023): 88.
12. Miah, Jonayet, Duc M. Ca, Md Abu Sayed, Ehsanur Rashid Lipu, Fuad Mahmud, and SM Yasir Arafat. "Improving Cardiovascular Disease Prediction Through Comparative Analysis of Machine Learning Models: A Case Study on Myocardial Infarction." In

2023 15th International Conference on Innovations in Information Technology (IIT), pp. 49-54. IEEE, 2023.

13. Li, Fuchen, Patrick Wu, Henry H. Ong, Josh F. Peterson, Wei-Qi Wei, and Juan Zhao. "Evaluating and mitigating bias in machine learning models for cardiovascular disease prediction." Journal of biomedical informatics 138 (2023): 104294.
14. Dalal, Surjeet, Pallavi Goel, Edeh Michael Onyema, Adnan Alharbi, Amena Mahmoud, Majed A. Algarni, and Halifa Awal. "Application of machine learning for cardiovascular disease risk prediction." Computational Intelligence and Neuroscience 2023, no. 1 (2023): 9418666.
15. Fan, Zihao, Zhi Du, Jinrong Fu, Ying Zhou, Pengyu Zhang, Chuning Shi, and Yingxian Sun. "Comparing the performance of machine learning and conventional models for predicting atherosclerotic cardiovascular disease in a general Chinese population." BMC Medical Informatics and Decision Making 23, no. 1 (2023): 134.
16. Noroozi, Zeinab, Azam Orooji, and Leila Erfannia. "Analyzing the impact of feature selection methods on machine learning algorithms for heart disease prediction." Scientific Reports 13, no. 1 (2023): 22588.
17. Subramani, Sivakannan, Neeraj Varshney, M. Vijay Anand, Manzoore Elahi M. Soudagar, Lamya Ahmed Al-Keridis, Tarun Kumar Upadhyay, Nawaf Alshammari et al. "Cardiovascular diseases prediction by machine learning incorporation with deep learning." Frontiers in medicine 10 (2023): 1150933.
18. Sadar, Uzama, Parul Agarwal, Suraiya Parveen, Sapna Jain, and Ahmed J. Obaid. "Heart disease prediction using machine learning techniques." In International Conference on Data Science, Machine Learning and Applications, pp. 551-560. Singapore: Springer Nature Singapore, 2023.
19. Almazroi, Abdulwahab Ali, Eman A. Aldhahri, Saba Bashir, and Sufyan Ashfaq. "A clinical decision support system for heart disease prediction using deep learning." IEEE Access 11 (2023): 61646-61659.
20. Dritsas, Elias, and Maria Trigka. "Efficient data-driven machine learning models for cardiovascular diseases risk prediction." Sensors 23, no. 3 (2023): 1161.
21. Trigka, Maria, and Elias Dritsas. "Long-term coronary artery disease risk prediction with machine learning models." Sensors 23, no. 3 (2023): 1193.
22. Khan, Razib Hayat, Jonayet Miah, Shah Ashisul Abed Nipun, and Majharul Islam. "A Comparative Study of Machine Learning classifiers to analyze the Precision of Myocardial Infarction prediction." In 2023 IEEE 13th Annual Computing and Communication Workshop and Conference (CCWC), pp. 0949-0954. IEEE, 2023.
23. Ansari, Gufran Ahmad, Salliah Shafi Bhat, Mohd Dilshad Ansari, Sultan Ahmad, Jabeen Nazeer, and A. E. M. Eljialy. "Performance evaluation of machine learning techniques (MLT) for heart disease prediction." Computational and Mathematical Methods in Medicine 2023, no. 1 (2023): 8191261.
24. Mohammad, Farah, and Saad Al-Ahmadi. "WT-CNN: a hybrid machine learning model for heart disease prediction." Mathematics 11, no. 22 (2023): 4681.
25. Qadri, Azam Mehmood, Ali Raza, Kashif Munir, and Mubarak S. Almutairi. "Effective feature engineering technique for heart disease prediction with machine learning." IEEE Access 11 (2023): 56214-56224.
26. Yan, Chao, Yiluan Xing, Sensen Liu, Erdi Gao, and Jinyin Wang. "Machine Learning Models for Cardiovascular Disease Prediction: A Comparative Study." bioRxiv (2024): 2024-05.
27. Anjum, Nishat, Cynthia Ummay Siddiqua, Mahfuz Haider, Zannatun Ferdus, Md Azad Hossain Raju, Touhid Imam, and Md Rezwanur Rahman. "Improving Cardiovascular Disease Prediction through Comparative Analysis of Machine Learning Models." Journal of Computer Science and Technology Studies 6, no. 2 (2024): 62-70.
28. Massari, Hakim El, Noreddine Gherabi, Sajida Mhammedi, Hamza Ghandi, Mohamed Bahaj, and Muhammad Raza Naqvi. "The impact of ontology on the prediction of cardiovascular disease compared to machine learning algorithms." arXiv preprint arXiv:2405.20414 (2024).
29. Nandini, S. "Comparative Study of Machine Learning Algorithms in Detecting Cardiovascular Diseases." arXiv preprint arXiv:2405.17059 (2024).
30. Ogunpola, Adedayo, Faisal Saeed, Shadi Basurra, Abdullah M. Albarrak, and Sultan Noman Qasem. "Machine learning-based predictive models for detection of cardiovascular diseases." Diagnostics 14, no. 2 (2024): 144.
31. Pathan, Shadman Mahmood Khan, and Sakan Binte Imran. "Integrated Machine Learning and Deep Learning Models for Cardiovascular Disease Risk Prediction: A Comprehensive Comparative Study." Journal of Intelligent Learning Systems and Applications 16, no. 1 (2024): 12-22.
32. Sadr, Hossein, Arsalan Salari, Mohammad Taghi Ashoobi, and Mojdeh Nazari. "Cardiovascular disease diagnosis: a holistic approach using the integration of machine learning and deep learning models." European Journal of Medical Research 29, no. 1 (2024): 455.
33. Shrestha, Dhadkan. "Comparative analysis of machine learning algorithms for heart disease prediction using the Cleveland Heart Disease dataset." (2024).
34. DeGroat, William, Habiba Abdelhalim, Kush Patel, Dinesh Mendhe, Saman Zeeshan, and Zeeshan Ahmed. "Discovering biomarkers associated and predicting cardiovascular disease with high accuracy using a novel nexus of machine learning techniques for precision medicine." Scientific reports 14, no. 1 (2024): 1.
35. Drouard, Gabin, Juha Mykkänen, Jarkko Heiskanen, Joona Pohjonen, Saku Ruohonen, Katja Pahkala, Terho Lehtimäki et al. "Exploring machine learning strategies for predicting cardiovascular disease risk factors from multi-omic data." BMC Medical Informatics and Decision Making 24, no. 1 (2024): 116.
36. Ahmed, Mohammed, and Idress Husien. "Heart Disease Prediction Using Hybrid Machine Learning: A Brief Review." Journal of Robotics and Control (JRC) 5, no. 3 (2024): 884-892.
37. Manikandan, G., B. Pragadeesh, V. Manojkumar, A. L. Karthikeyan, R. Manikandan, and Amir H. Gandomi. "Classification models combined with Boruta feature selection for heart disease prediction." Informatics in Medicine Unlocked 44 (2024): 101442.
38. Végh, Ladislav, Ondrej Takáč, Krisztina Czakóová, Daniel Dancsa, and Melinda Nagy. "Comparative Analysis of Machine Learning Classification Models in Predicting Cardiovascular Disease." International Journal of Advanced Natural Sciences and Engineering Researches 8, no. 6 (2024): 23-31.
39. Cai, Yu-Qing, Da-Xin Gong, Li-Ying Tang, Yue Cai, Hui-Jun Li, Tian-Ci Jing, Mengchun Gong et al. "Pitfalls in developing machine learning models for predicting cardiovascular diseases: challenge and solutions." Journal of Medical Internet Research 26 (2024): e47645.
40. Kanaki, Parvati, and D. G. A. Walikar. "MACHINE LEARNING MODELS FOR HEART DISEASE PREDICTION-A REVIEW." Indian J. Comput. Sci. Eng 15, no. 2 (2024): 132-151.