International Journal of Innovation Studies

A HYBRID MACHINE LEARNING AND FEATURE OPTIMIZATION APPROACH FOR CARDIAC RISK PREDICTION

P. Kavitha
Ph.D Research Scholar, Department of Computer Science, Nallamuthu Gounder Mahalingam College, Pollachi, Tamil Nadu - 642001, India.

Dr R. Manicka Chezian
Associate Professor, Department of Computer Science, Nallamuthu Gounder Mahalingam College, Pollachi, Tamil Nadu - 642001, India. chezianr@gmail.com

Abstract
Cardiovascular disease (CVD) remains one of the major causes of mortality globally, emphasizing the importance of early prediction and effective risk detection. Recent breakthroughs in Machine Learning (ML) techniques have been increasingly applied to medical datasets to assist healthcare practitioners in clinical decision-making. Several studies have developed predictive architectures utilizing classification algorithms, feature selection techniques, hybrid learning models, and comparative analyses of multiple algorithms to identify bottlenecks and improve prediction capability. Despite these advancements, prior research often depends on relatively small or spatially specific datasets, lacks effective feature optimization, exhibits increased computational complexity, and provides reduced interpretability for clinicians. These limitations emphasize the need for a proficient integrated framework that productively reduces feature redundancy, enhances model generalization, and improves prediction accuracy. The proposed framework aims to develop an interpretable and accurate CVD prediction system by combining feature optimization techniques with ML classification models. The developed architecture focuses on identifying the most prominent clinical attributes, reducing redundant and correlated variables, and utilizing optimized features to train robust predictive models for early CVD detection. Primarily, data preprocessing and normalization are performed on the dataset. Dimensionality is subsequently reduced by applying Principal Component Analysis (PCA), followed by the Improved Dragonfly Optimization Algorithm (IDFOA) to select the most relevant features. The optimized features are then utilized to train deep learning-based predictive models and Gradient Boosting Machine (GBM) classifiers for disease prediction. Experimental analysis illustrates that the integrated optimization approach significantly enhances CVD prediction performance compared with baseline models. The combination of PCA and IDFOA effectively reduces redundant features while improving sensitivity and the Gini coefficient. Random Forest (RF) achieves the best predictive performance. Overall, the designed framework provides a proficient and reliable method for CVD prediction and supports intelligent clinical decision-support systems for early CVD risk assessment and improved healthcare outcomes.

1. Introduction

Cardiovascular Disease (CVD) is syndrome influencing the blood vessels and heart, comprising coronary heart diseases, stroke, and heart failure caused by smoking, metabolic risk factors, cholesterol and high blood pressure. CVD is one of the major reasons for mortality globally and represents a important public health dispute. As per global health reports, millions of deaths every year are ascribed to cardiovascular conditions namely coronary artery disease, heart failure, and stroke. Premature detection of individual at risk is necessary for decreasing the mortality rates and enhancing the long-term patient outcomes. The increasing availability of electronic health observation and clinical records of patients has created opportunity for implementing computational techniques to assist early disease prediction.

Over the past several years, Machine Learning (ML) has developed as a formidable technique computing medical data realizing underlying structure associated with disease progression. Several ML classifiers comprising Decision Trees (DT), Random Forest (RF), Support vector Machine (SVM), XG Boost, Gradient Boosting and K-nearest neighbors (K-NN) extensively deployed for CVD prediction. These models can evaluate multiple clinical features concurrently and support clinicians in recognizing high risk patients more efficiently. Moreover, ML-based prediction systems exhibit promising capability in boosting diagnostic accuracy and facilitate data-driven healthcare decision making. Even with all developments, clinical datasets often represent various difficulties that restrict the performance of models. Clinical records typically consist correlated variables, redundant features, and noisy details that may negatively impact model learning. Additionally, all medical attributes will not contribute equivalently to disease prediction. Top address all these limitations, feature optimization techniques namely dimensionality reduction and feature selection have earned attention in clinical data analysis. Integrating these techniques with ML models can efficiently increase prediction accuracy along with reducing model complexity. This approach not only enhances predictive performance but additionally support to the advancement of intelligent medical decision support systems which can efficiently provide assist healthcare practitioners in early risk detection and disease prevention.

Recent works has widely explored the utilization of ML techniques for enhancing CVD prediction and early assessment. Several studies focused on combined feature selection, ensemble learning and optimization techniques to optimize forecasting accuracy and assist medical problem solving. (Cao et al., 2025) proposed an improved ML framework with Particle Swarm Optimization (PSO). Even though their approach improved model performance lacks in recording complex real-world clinical patterns and shows increased evaluation cost. Likewise, (Hossian et al., 2024) employed a framework utilizing multiple classification algorithms. Even after identifying major cardiovascular factors through comparative evaluation framework is limited by small dataset, external validation and feature optimization. Later (Khan et al., 2024) Introduced a adapted Artificial Bee Colony combined with K-NN classifier to enhance prediction accuracy along with reduce training time but model exhibits higher computational cost. Then (Tarek et al., 2025) deployed a nature -inspired snake optimization algorithm for automated feature selection along with ML. Their proposal addressed high dimensional feature difficulties. Reátegui et al. (2025) and Bouqentar et al. (2024) designed model which ensemble learning and comparative analysis of ML algorithms combined with data balancing techniques. However, their works relied on spatial specific datasets and lacking advanced feature optimization mechanisms. These difficulties Highlight the requirement for a proficient integrated framework that combines dimensionality reduction and advanced

ISSN:2096-2487 | E-ISSN:2589-2975

metaheuristic feature optimization techniques with ML classifiers to enhance prediction accuracy and reliability.

The designed Framework focuses on implementing a hybrid ML framework for accurate CVD prediction utilizing Optimized medical features. The model proposes an integrated Principal component analysis (PCA) to reduce feature magnitude and banish multicollinearity, succeeded by the Improved Dragonfly Optimization Algorithm (IDFOA) to recognize the most related subset of predictive features. The Optimizes attributes are utilized to train ML classifiers. The architecture introduces a integrated feature optimization technique that integrates dimensionality reduction and metaheuristic feature selection. Additionally, the framework performs comparative analysis between baseline model and optimized models to illustrate the effectiveness of deployed model in assisting reliable CVD prediction and clinical decision support.

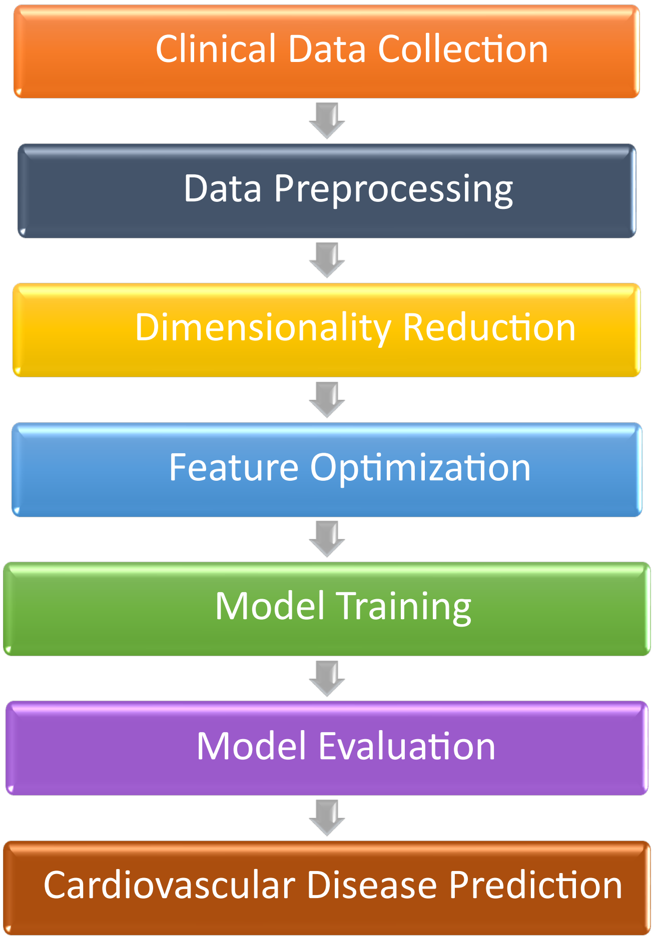

Figure 1. Framework for Cardiovascular Disease Prediction Using Hybrid Feature Optimization

2. Related works

Cardiovascular disease is a major prevalent condition. To enhance predictive accuracy, (Cao et al., 2025) formulates Machine Learning diagnostic system that ensembles feature selection, modified particle swarm optimization, and a light gradient boosting machine. The variant of PSO is employed to extract the best configuration to fabricate the MFS-DLPSO-XGBoost framework. The data analysis reveals that superior predictive power and clinical decision support system and collaborative filtering and cardioprotective.

Cardiovascular disorders (CVDs) primary cause of mortality. Lower- and middle-income countries (LMICs), like Bangladesh, are subject to various mode of CVDs, congestive cardiac failure. Primary cause of mortality in Bangladesh has currently been replaced by systemic infection and cardiac parasitosis. Hossain et al. (2024) extracted factors that are decisive with CVD and forecasting CVD exposure. It is mandatory that the ensemble learning be incorporated within a predictive model for cardiovascular risk. This study may redefine healthcare delivery by deploying new instrumentation to evaluate a client's CVD projection. Deep learning streamlines a process in cardiovascular risk assessment. An automated classifier was implemented for the data-driven heart disease classification. Khan et al. (2024) suggested a system relate on Modified Artificial Bee Colony (M-ABC) and k-Nearest Neighbours (KNN) for optimal feature subset selection. Improved Artificial Bee Colony (IABC). This study seeks to identify the feature subset selection. Hence, within the learning phase, only the specifications that render feature selection.

Machine learning has evolved as an early warning system, utilizing statistical modelling that optimize conventional approaches. Reátegui et al. (2025) modelled a cohort of 709 Ecuadorian individuals with sociodemographic and clinical characteristics to assess cardiovascular probability. The model evaluation, like ensemble learning, tree-based ensemble methods, and stacking, was collated, while rebalancing the dataset SMOTE and a composite ROS–SMOTE framework. This analysis validates the efficacy of machine learning algorithms to enhance screening accuracy and cardiovascular care, reinforcing the significance of biometrics and blood pressure indices.

Wan et al. (2025) analyses three key pillars: (1) performance evaluation in subclinical detection (2) implementation strategies, and (3) compliance audit and moral consequences. It provides a comprehensive review of machine learning based heart disease diagnostics and addresses governance and ethical implications inherent in its deployment. Conclusively, this study identifies directions for further research in this dynamic landscape.

Noroozi et al. (2023) evaluate the impact of model refinement for heart disease classification. The statistical feature screening with the high dimensional specifications designated surpassed existing techniques regarding algorithm specificity. Hence, metaheuristic algorithms enhanced calibration from classifier performance metrics.

Technology integration, like Machine Learning (ML), has materialized as a primary catalyst to enhanced accuracy and early diagnosis. Hence, (Alwakid et al., 2025) optimized dimensionality reduction and machine learning in cardiac classification. It utilizes methodologies like fisher's score, and backward feature eliminations a substance of cardiovascular biomarkers into a optimized subset. This study employs high-fidelity datasets, fair causal feature selection, and robust evaluation frameworks to validate algorithmic equity in the AI-driven diagnostics.

Cardiovascular disease optimization relies on accurate detection of cardiac pathologies utilizing validated diagnostic modalities and high-fidelity discrimination between binary outcome. Bilal et al. (2025) enhancing diagnostic precision by aggregating model stacking: Random Forest (RF), K-Nearest Neighbor (KNN), and Support Vector Machine (SVM). Assessments implement NHMP annotated cardiac imaging dataset at Nishter Hospital in Multan, Punjab, Pakistan. Feature selection techniques, such as ANOVA, Boruta with XGBoost, Boruta with Random Forest, Chi Square, and LASSO, are incorporate refined to reduce variance, essentially precision oriented. Two model variants are evaluated: one implementing unanimous voting (VH) and the other weighted ensemble method (VS).

The primary objective of (Tarek et al., 2025) was to introduce an analytical approach to cardiovascular disease (CVD) data called CVD-SO, which utilizes snake optimization (SO). Five deep learning techniques are implemented in categorizing clinical data standardization. By integrating deep learning and the unified framework, we can formulate a CVD high-fidelity diagnostic model. The disease prevalence, CVD-mitigation of morbidity and mortality, and mitigating CVD by enabling preventive measures.

Bouqentar et al. (2024) are mitigating the emerging global health crisis driven by cardiovascular diseases (CVDs), the leading cause of mortality, consistent with the World Health Organization (WHO). With the evolution of Machine Learning (ML) and Deep Learning (DL) methodologies incorporated into Artificial Intelligence (AI), these methodologies are essential for predictive maintenance of CVDs. The enhanced ML platform provides a clinical workflow automation system for cardiovascular risk stratification, with translational potential. Ischemic heart disease is one of the primary mortality factors, reinforcing imperative diagnostic optimization. Deokar et al. (2025) introduce an ML-driven architecture for interpretable machine learning with high performance. The ensemble method yielded 98.20% with an 80:10:10 holdout method. These analyses validate the efficacy of clinical workflow automation, early warning systems and shift left CVD-consequences. The incorporation of advanced AI into clinical processes is a critical determinant of patient outcomes and clinical quality improvement.

Table 1. Comparative Study of Cardiovascular Prediction Systems

Author Name	Study Focus	Technology/Model	Key Contributions	Limitations
Cao et al., 2025	Development of an enhanced Machine learning Framework for Cardiovascular disease prediction	XG Boost, Pearson correlation analysis, Particle Swarm Optimization (PSO),	Introduces hybrid prediction framework, Multistage feature Selection, Improved hyperparameter optimization	Lacking to capture complex real world clinical pattern and generalizability, Limited clinical explainability, increased computational cost
Hossain et al., 2024	Identifying significant risk factors and predicting cardiovascular disease	Crosstab analysis, chi-Square Test, Logistic Regression, Naïve bayes, Decision tree, AdaBoost Classifier,	Developed a CVD prediction Framework, identifies important risk factors, comparative analysis of	Limited data size, lacking feature optimization, model diversity and absence of external validation

		Random Forest, Bagging Tree,	multiple algorithms,	
Khan et al., 2024	Focus on improving heart disease prediction accuracy through optimal feature selection and machine learning classification	Modified Artificial Bee Colony (M ABC) algorithm, K-nearest neighbors (KNN),	Introduces a combined approach using M-ABC for attribute selection and KNN for classification, reduced computational complexity and training time	Computational cost is high for large datasets, lacking comparison with ensemble algorithms, limited model generalizability
Reátegui et al., 2025	Predicting cardiovascular disease risk using artificial intelligence and machine learning techniques	SMOTE, Hybrid ROS-SMOTE, Decision Tree, Random Forest, Gradient Boosting, XG Boost, LIGTHGBM, AdaBoost, Bagging,	Identifies important predictors of cardiovascular disease, comparative analysis of multiple ensemble models,	Limited datasets size which limit model robustness and generalization, regional dataset bias affect applicability
Wan et al., 2025	Focus on machine learning applications for cardiovascular disease diagnosis	Logistic Regression, Support vector Machine, K nearest neighbors, Artificial Neural Networks	Summarizes recent advancement in ML techniques for CVD prediction, outlines practical employment in healthcare environments	Lacking to propose model, conclusion rely completely on previous studies, different studies analyzed in review cause performance comparison challenges
Noroozi et al., 2023	Investigates the impact of different feature selection techniques on the performance of ML algorithms for	Heuristic optimization strategies, Bayes Net, Naïve Bayes, Multivariate Linear Model, Support Vector Machine,	Evaluates 16 feature selection techniques, improved sensitivity and specificity using wraper method, balanced	Better accuracy is not achieved, limited model generalizability caused by smaller dataset, lacking integration of

	heart disease prediction	LogitBoost, J48 Decision Tree, Random Forest	optimization of feature subsets	hybrid optimization framework
Alwakid et al., 2025	Focused on improving Cardio vascular disease prediction using advanced AI based feature and ML models	Chi-Square, Principal component analysis, Random Forest, XG Boost, Decision Tree, Logistic Regression,	Introduces combined multiple feature selection methods and dimensionality reduction techniques, addresses transparency and bias mitigation	Potentially indicates overfitting for small datasets, data originated from single source may not generalize well, No clear validation
Bilal et al., 2025	Studies on improving cardiac disease diagnosis using an ensemble ML Framework integrated with feature selection techniques	Analysis of variance, Boruta with XG Boost, Boruta with Random Forest, Chi-square Test, LASSO, K-nearest neighbors, support vector machine	Hybrid ensemble model combining RF, KNN, SVM classifiers, identifies most informative attributes and reduce irrelevant features,	Limited generalizability of model, indicates overfitting if the dataset is limited, increased computational cost and model complexity
Tarek et al., 2025	Focused on cardiovascular disease prediction through automated feature selection using nature inspired optimization algorithm along with ML classifiers	Snake optimization, nature-inspired metaheuristic algorithm, Machine Learning Algorithms	Supports automated screening and decision support, effective addresses issues such as high dimensionality, redundant attributes and model complexity	Limited dataset leads to overfitting, Lacking dataset diversity information, reduced interpretability for clinical practitioners
Bouqentar et al., 2024	Focus on developing a machine learning based	Decision Tree, Random Forest, Support vector machine, Logistic	Performs systematic comparison of multiple ML	Dependency of two benchmark datasets restrict the real-world

	system for early prediction and diagnosis of cardiovascular disease	Regression, Adaptive Boosting, K nearest neighbors, hyper tuning techniques	algorithms, early disease detection using optimized classification, provides decision support capability	clinical variability, Lacking integration of advanced feature selection
Deokar et al., 2025	Developed a ML-based framework for accurate and reliable classification of cardiovascular disease (CVD)	Multiple feature selection techniques, Novel feature ranking algorithm, Multiple ML algorithms, ensemble-based ML model	New feature ranking method is introduced to better identifies predictors for CVD, reduces high dimensional feature space	Limited dataset information regarding size and diversity, Black box system makes clinical interpretation more challenging, lacking real world validation

3. Methodology

The designed framework applied hybrid feature optimization and machine learning design for CVD forecasting. The process succeeded by a systematic workflow that comprises data preprocessing, metaheuristic feature selection, dimensionality deduction and classification modelling. Firstly, the medical dataset is prepared via preprocessing methods to make sure data consistency. Eventually, attribute optimization methods are employed to point the most related features which are then utilized to fit ML models for potential CVD prediction

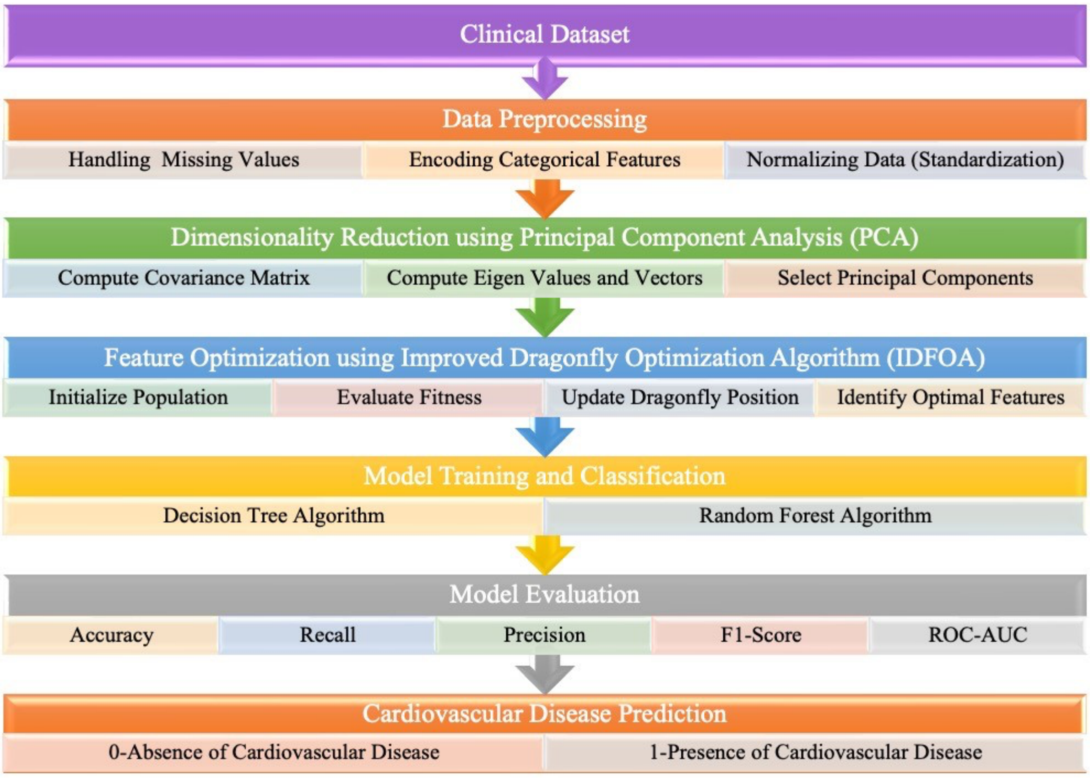

Figure 2. Workflow for PCA–IDFOA Based Cardiovascular Disease Prediction

3.1 Data Description

The dataset employed in this methodology is the Cleveland heart disease dataset (Kaggle Version). This dataset consists of clinical details collected from patients undergoing CVD analysis and contains various clinical features that are commonly utilized to recognize effective heart reliable conditions. The dataset comprises of 13-14 clinical attributes presenting physiological measurements, demographic information and diagnostic test outcomes. The focused variable is binary in nature which indicates presence or absence of CVD where '0' denotes the deficiency of CVD and '1' represents the presence of CVD. The dataset consists of significant medical features and independent variables. The availability of several physiological and diagnostic variables makes this dataset suitable for deploying ML techniques to identify major predictors and develop significant CVD predictive models.

3.2 Data Preprocessing

Clinical datasets usually include inconsistencies, missing values, and variations in feature scales that may negatively impact the performance of predictive models. So, various preprocessing levels are carried out to process data for forthcoming analysis. Firstly, the dataset is inspected to identify missing values. Any missing entries are imputed using appropriate techniques to sustain data integrity. Categorical attributes are encoded into numerical presentation to assemble them suitable for ML algorithms. Additionally feature scaling is employed to normalize the dataset and confirm that all features contribute equivalently during model training. Normalization or standardization techniques are used to modify the attribute values into comparable range. These preprocessing levels surpass the consistency of the learning procedure and boost the overall execution of the predictive models.

$$Z = \frac{X - \mu}{\sigma} \tag{1}$$

Where,
Z = Normalized value after standardization,
X = Real feature value
μ = Average of feature values
σ = Standard deviation of feature

3.3 Dimensionality Reduction

Dimensionality is performed to reduce redundancy and minimize the intricacy of the clinical dataset along with preserving the most informative attributes. In clinical dataset several datasets might execute high correlation or overlapping details, which can negatively impact the performance and proficiency of ML models. Decreasing the number of attributes helps enhancing computational efficiency, reduce noise, and improve model generalization. In the proposed approach Principal Component Analysis (PCA)is deployed as a feature engineering method. PCA revamps the original jointly dependent variables into a set of non-associated principal components while preserving the maximum possible variance present in the data. BY showcasing the real attributes space into a lower feature space, PCA assist to vanish multicollinearity among attributes and retains the significant details required for prediction. The reduced feature set received from PCA serves as an optimized representation of dataset and it is parallelly used for upcoming attribute selection and ML model fitting. This process enhances the capacity of learning algorithms and provides better predictive performance.

$$C = \frac{1}{n-1} (X - \bar{X})^{T} (X - \bar{X}) \tag{2}$$

Where,
C = Covariance Matrix
X = data matrix
$\bar{X}$ = mean of dataset
n = number of samples

$$Z = X\mathcal{W} \tag{3}$$

Where,
Z = Transformed feature space
X = eigenvector matrix
$\mathcal{W}$ = original feature matrix

3.4 Metaheuristic Feature Selection

Clinical dataset often consists redundant features that may enhance the computational complexity and reduce performance. In the designed framework metaheuristic optimization approach such as Improved Dragonfly optimization Algorithm (IDFOA) is deployed to choose most informative attribute from the dataset. The IDFOA is a decentralized optimization method inspired by the adaptive and rigid swarming and exploitation abilities of the algorithm, legalizing effective search for the optimal attribute subset. Through this optimization process, the algorithm recognizes the most related clinical features that contribute to CVD prediction. The selected optimal attribute subset is later utilized as input for training the ML classification models.

$$Fitness = \alpha * \text{Accuracy} - \beta * \left(\frac{N_s}{N_t}\right) \tag{4}$$

Where,

$Accuracy$ = performance metrics
N_s = feature subset size
N_t = feature count
α and β = model parameters

3.5 Dataset Splitting

Data splitting is performed to compute the predictive performance of the ML models utilizing independent training and testing record. This method confirms that the designed approach is capable of extrapolating well to unseen data and assist to prevent overfitting. The database is binary partition namely fitting and evaluation dataset utilized for studying the model generalization is used to evaluate the framework of predictive capability. In designed model, the database is partitioned using the hold-out validation method where 70% of the model is fed into the dataset and 30% database is used for evaluation dataset. The training data is utilized for developing ML models using optimized feature set whereas, the testing set used to determine performance of the fitted models. The splitting method supports an unbiased evaluation of the classification models and confirms reliable prediction performance assessment.

3.6 Machine Learning Classification

Machine learning classification techniques are deployed to build predictive models for CVD detection using the optimized feature subset procured from the feature optimization process. These models learn patterns from the training dataset and generate predictions for unseen instances in the testing dataset. The designed models used two wisely known ML classifiers namely Decision Tree and Random Forest to perform CVD prediction. The Decision Tree algorithm develops a networked structure of decision rules in accordance with clinical features which accrediting interpretable classification outcomes that are ideal for clinical decision making. An ensemble method is a bagging that gradient boosting incorporates predictive performance and decrease high variance. The execution of designed frameworks is analyzed performance benchmarks namely precision, sensitivity, specificity and c-ststistic. By compiling the outputs of multiple trees via majority voting, (Random Forest) support more consistent and reliable prediction performance. The execution of the predicted model is evaluated with the help pf standard analyzing metric to determine the significance of the accomplished model.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{5}$$

$$Precision = \frac{TP}{TP + FP} \tag{6}$$

$$Recall = \frac{TP}{TP + FN} \tag{7}$$

$$F1 = 2 \times \frac{Precision \times Recall}{Precision + Recall} \tag{8}$$

Where,
TP = True Positive,
TN = True Negative,
FP = False Positive and
FN = False Negative

Algorithm: Hybrid Feature Optimization-Based Cardiovascular Disease Prediction

Input: Clinical dataset $D = \{X_1, X_2, ..., X_N\}$ containing patient records with clinical attributes and target labels

Output: Optimized feature dataset D_{OPT} and predicted cardiovascular disease class $Y = \{0,1\}$

START

Step 1: Load cardiovascular disease dataset D

Step 2: Data Preprocessing

For each sample $X_i \in X$do

Handle missing or inconsistent values

Encode categorical attributes into numerical form

Normalize attributes using Z-score normalization:

$$Z = (X - \mu) / \sigma$$

End For

Step 3: Dimensionality Reduction using PCA

Compute covariance matrix C of normalized dataset

Calculate eigenvalues and eigenvectors

Select principal components with highest variance

Transform feature matrix:

$$X_{PCA} = XW$$

Step 4: Feature Selection using Improved Dragonfly Optimization Algorithm (IDFOA)

For each candidate feature subset S_i do

Train temporary classifier using selected features

Evaluate fitness function:

$$Fitness = \alpha * Accuracy - \beta * \left(\frac{N_s}{N_t}\right)$$

End For

Update dragonfly positions based on swarming behaviors

Select optimal feature subset S_{best}

Construct optimized dataset:

$$D_{OPT} = X_{PCA} S_{best}$$

Step 5: Splitting database into training sets and testing sets

Training set = 70%, Testing set = 30%

Step 6: Model Training

Train Decision Tree classifier on training dataset

Train Random Forest classifier on training dataset

Step 7: Model Evaluation

Select best classifier among them for CVD Prediction

For each test sample $x \in test\ set$ do

Predict class label y_{pred}

End For

Evaluate model execution using Accuracy, Precision, Recall, F1-score, ROC-AUC

Step 8: Prediction output Return optimized dataset D_{OPT} and predicted class Y

END

ISSN:2096-2487 | E-ISSN:2589-2975

4. Results and Discussion

The deployment of the integrated system is calculated using the optimized attribute set received via dimensionality reduction and metaheuristic attribute selection. ML classifiers namely recursive partitioning using the chosen attributes and computed on testing dataset. The model evaluation quantified through metrics namely accuracy, recall, F1-score and ROC-AUC.

Table 2. Model evaluation metrics

Class	Precision	Recall	F1-Score	Support
No CVD (0)	0.79	0.81	0.80	45
CVD (1)	0.83	0.80	0.80	46
Accuracy	-	-	0.80	91
Macro Avg	0.82	0.80	0.80	91
Weighted Avg	0.82	0.80	0.80	91

Table 3. Performance report of Random Forest Algorithm

Class	Precision	Recall	F1-Score	Support
No CVD (0)	0.87	0.89	0.88	45
CVD (1)	0.90	0.88	0.89	46
Accuracy	-	-	0.89	91
Macro Avg	0.89	0.89	0.89	91
Weighted Avg	0.89	0.89	0.89	91

Table 4. Performance analysis Between Proposed Model and Existing Work

Methods	Accuracy	Precision	Recall	F1-score
CVD-RF-PCA	0.89	0.89	0.89	0.89
CVD-PSO-XG	0.74	0.76	0.71	0.73
CVD-MABC-KNN	0.84	0.83	0.84	0.85
CVD-NB	0.83	0.85	0.80	0.86
CVD-XGB	0.84	0.83	0.82	0.84

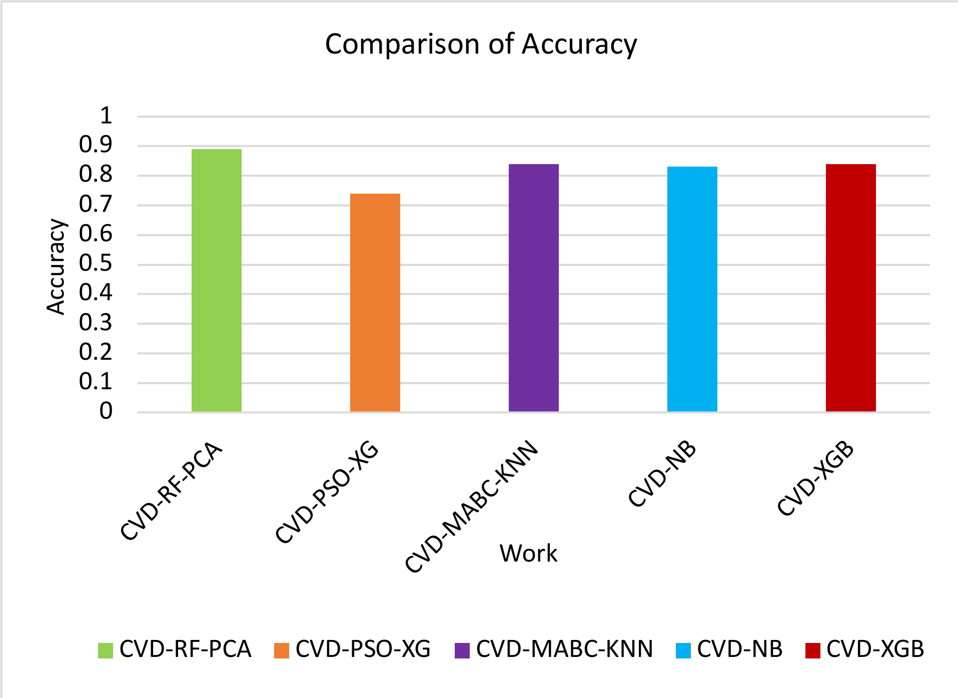

Figure 3. Comparison of Accuracy Between Proposed Work and Existing Work

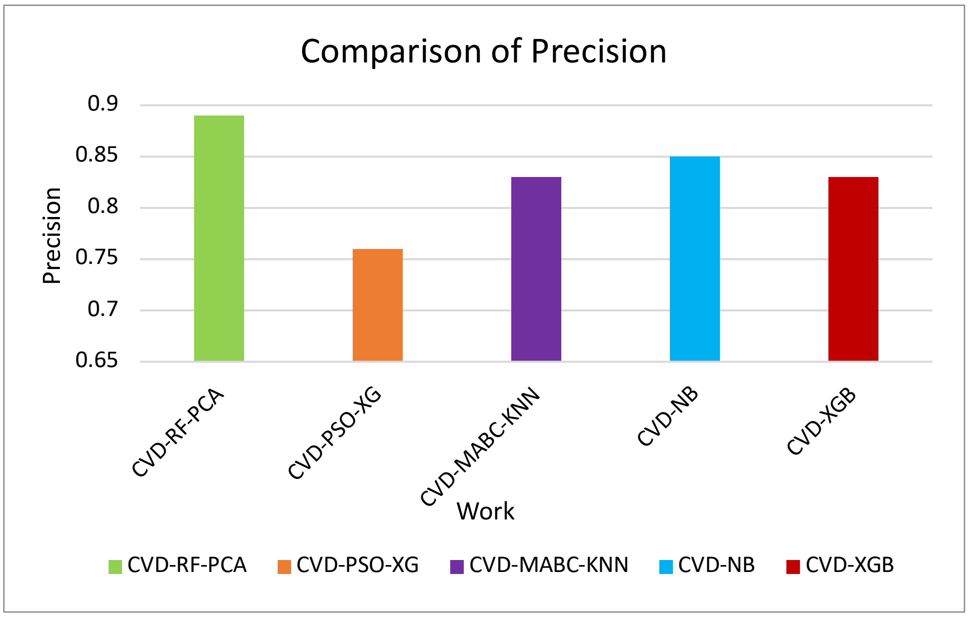

Figure 4. Comparison of Precision Between Proposed Work and Existing Work

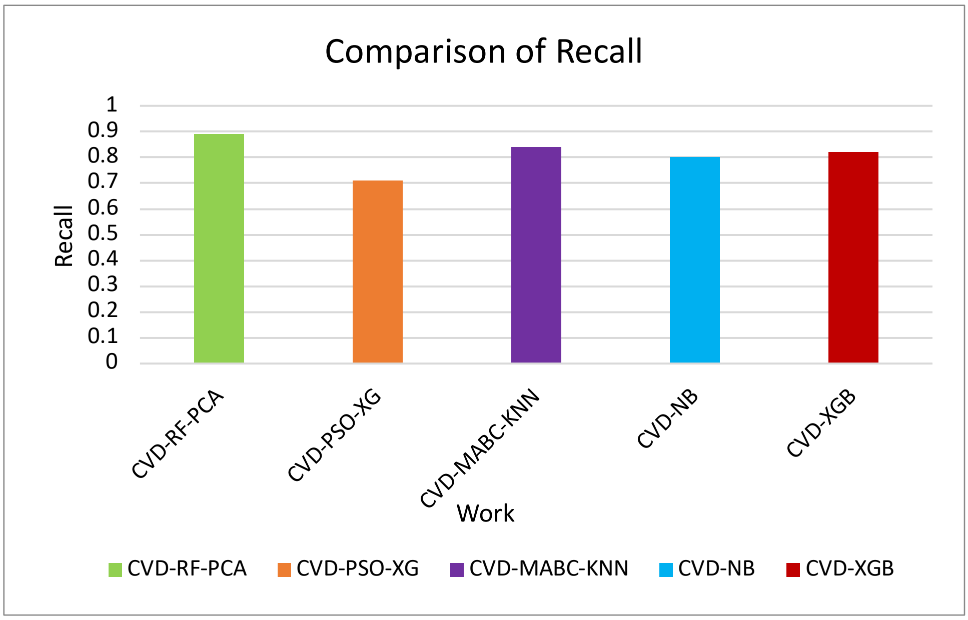

Figure 5. Comparison of Recall Between Proposed Work and Existing Work

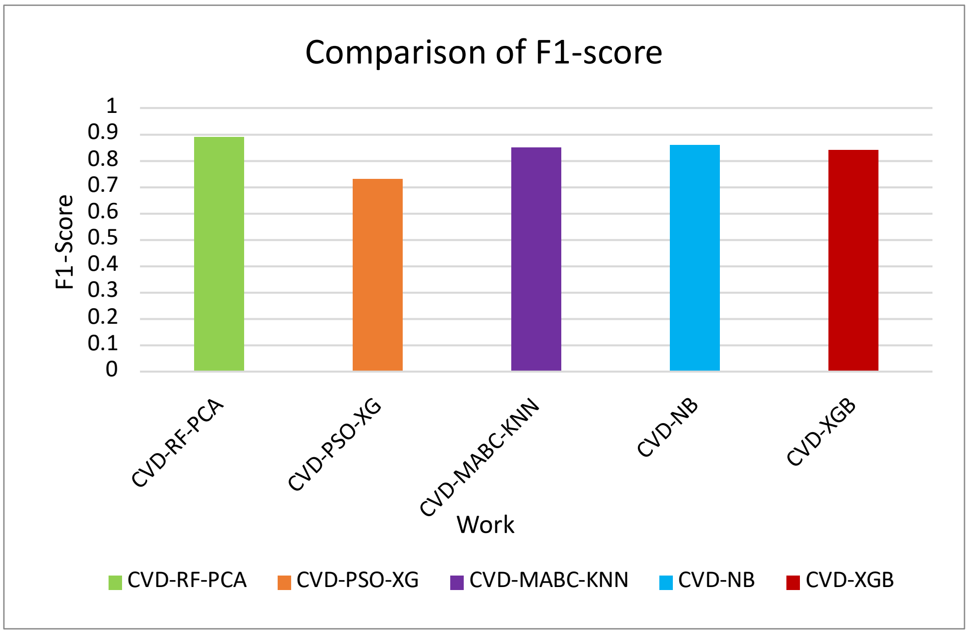

Figure 6. Comparison of F1-score Between Proposed Work and Existing Work

The experimental outcomes illustrates that the integrated feature optimization proposal enhance the prediction capability of ML models by decreasing redundant features ad improving the relevance of clinical attributes utilized for training. Among the computed classifiers, (Random Forest) model shows superior performance since its ensemble studying capability and capacity to handle complex attribute interactions. The comparative evaluation between the baseline model and optimized models shows that the combination of dimensionality reduction and metaheuristic feature selection proficiently improves classification accuracy and model consistency.

5. Conclusion

The Framework presents integrated feature optimization and ML Framework for CVD prediction. The proposed framework integrated dimensionality reduction and metaheuristic feature selection techniques to improve the accuracy of the model. By using optimized attributes, recursive binary partitioning is able to accomplish enhanced prediction accuracy and reduce model complexity. The outcomes illustrate that the integrated framework improves the capacity of the machine learning model prediction while minimizing redundant attributes. The architecture can support early CVD detection and assist the clinical decision support system. Future works may focus on validating the framework utilizing large and diverse datasets and exploring advanced techniques to further improve the performance of prediction.

6. References

Cao, K., Liu, C., Yang, S., Zhang, Y., Li, L., Jung, H., & Zhang, S. (2025). Prediction of cardiovascular disease based on multiple feature selection and improved PSO-XGBoost model. Scientific Reports, 15(1), 12406.

Hossain, S., Hasan, M. K., Faruk, M. O., Aktar, N., Hossain, R., & Hossain, K. (2024). Machine learning approach for predicting cardiovascular disease in Bangladesh: evidence from a cross-sectional study in 2023. BMC cardiovascular disorders, 24(1), 214.

Khan, M. A., Mazhar, T., Mateen Yaqoob, M., Badruddin Khan, M., Jilani Saudagar, A. K., Ghadi, Y. Y., ... & Shahid, M. (2024). Optimal feature selection for heart disease prediction using modified Artificial Bee colony (M-ABC) and K-nearest neighbors (KNN). Scientific Reports, 14(1), 26241.

Reátegui, R., Tandazo-Malla, C., Suárez, R., & Ramírez-Cerna, L. (2025). Cardiovascular risk prediction via ensemble machine learning and oversampling methods. Scientific Reports.

Wan, S., Wan, F., & Dai, X. J. (2025). Machine learning approaches for cardiovascular disease prediction: A review. Archives of Cardiovascular Diseases.

Noroozi, Z., Orooji, A., & Erfannia, L. (2023). Analyzing the impact of feature selection methods on machine learning algorithms for heart disease prediction. Scientific reports, 13(1), 22588.

Alwakid, G., Ul Haq, F., Tariq, N., Humayun, M., Shaheen, M., & Alsadun, M. (2025). Optimized machine learning framework for cardiovascular disease diagnosis: a novel ethical perspective. BMC Cardiovascular Disorders, 25(1), 123.

Bilal, O., Hekmat, A., Shahzad, I., Raza, A., & Khan, S. U. R. (2025). Boosting Machine Learning Accuracy for Cardiac Disease Prediction: The Role of Advanced Feature Engineering and Model Optimization. The Review of Socionetwork Strategies, 19(2), 271-300.

Tarek, Z., Alhussan, A. A., Khafaga, D. S., El-Kenawy, E. S. M., & Elshewey, A. M. (2025). A snake optimization algorithm-based feature selection framework for rapid detection of cardiovascular disease in its early stages. Biomedical Signal Processing and Control, 102, 107417.

Bouqentar, M. A., Terrada, O., Hamida, S., Saleh, S., Lamrani, D., Cherradi, B., & Raihani, A. (2024). Early heart disease prediction using feature engineering and machine learning algorithms. Heliyon, 10(19).

Deokar, A. S., & Pradhan, M. A. (2025). Optimizing Cardiovascular Disease Detection Using Ranking-Based Feature Selection Machine Learning Models. Engineering, Technology & Applied Science Research, 15(5), 28172-28178.

ISSN:2096-2487 | E-ISSN:2589-2975

Ramesh, T. R., Lilhore, U. K., Poongodi, M., Simaiya, S., Kaur, A., & Hamdi, M. (2022). Predictive analysis of heart diseases with machine learning approaches. Malaysian Journal of Computer Science, 132-148.