Journal of Information Systems Engineering and Management
2025, 10(47s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

Optimizing Cardiovascular Disease Prediction with a Hybrid Gradient Descent Adaptive Algorithm and Random Forest Classifier

B. Kalaivani[1], A. Ranichitra[2]
[1]*Research Scholar, Department of Computer Science Sri S.Ramasamy Naidu Memorial College (Affiliated to Madurai Kamaraj University) Sattur-626 203, Tamilnadu, India*
kalaivanisrnm2008@gmail.com
ORCID: 0009-0000-1802-4586
[2]*Assistant Professor, Department of Computer Science Sri S.Ramasamy Naidu Memorial College Sattur-626 203, Tamilnadu, India*
ranichitra117@gmail.com ORCID: 0000-0001-6071-0635

ARTICLE INFO

Received: 29 Dec 2024

Revised: 12 Feb 2025

Accepted: 27 Feb 2025

ABSTRACT

Cardiovascular diseases (CVDs) constitute a major global health burden, thereby emphasizing the critical need for the development of robust and accurate predictive models to ensure early detection and prompt clinical intervention. This study introduces a novel Hybrid Gradient Descent Adaptive Algorithm and Random Forest (Hybrid GD-AA-RF) model, which combines the strengths of AdaGrad and AdamW optimizers for effective feature selection and integrates them with a Random Forest (RF) classifier to enhance predictive accuracy. The proposed model employs Z-score normalization during preprocessing to standardize features and ensure consistency. Feature selection leverages Gaussian-based differential entropy for information gain, while the hybrid optimization technique combines AdaGrad's adaptive learning rates with AdamW's weight decay regularization to prioritize critical but infrequent features, minimizing overfitting and improving generalization.

The RF classifier dynamically adjusts Tuning parameters, like the number of estimators and maximum tree depth, optimizing its performance on high-dimensional medical datasets. The Hybrid GD-AA-RF model was evaluated on a combined heart disease dataset with 12 attributes and 1190 records. Comparative analysis with cutting-edge models demonstrated superior performance in the considered metrics, owing to the balanced attribute selection and categorization capabilities. The hybrid optimization approach avoids the complexity of metaheuristic algorithms while ensuring efficient computation and enhanced interpretability.

This model's robust generalization across diverse data distributions and populations highlights its scalability for real-world healthcare applications. Its ability to prioritize impactful predictors can assist clinicians in identifying critical risk factors, enabling early diagnosis and improved patient outcomes. Marks a significant advancement in machine learning-based diagnostics, providing a dependable and precise tool for predicting cardiovascular disease.

Keywords: AdamW, AdaGrad, CVD, Gradient Descent Optimization, Random forest.

INTRODUCTION

Cardiovascular disease (CVD) represents a critical planetary public health challenge, responsible for an estimated 17 million deaths annually and placing immense strain on healthcare systems worldwide [1]. CVD accounts for 31% of all global deaths, with strokes and heart attacks being the primary contributors. The multifaceted nature of cardiovascular disease risk factors including hypertension, diabetes, hyperlipidaemia and lifestyle-related influences

Journal of Information Systems Engineering and Management
2025, 10(47s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

such as chronic stress, tobacco use and inadequate nutrition poses significant challenges to early diagnosis and effective treatment. Timely and precise identification is imperative, as it not only enhances clinical outcomes but also reduces the economic burden associated with the management of cardiovascular conditions [2].

In recent years, machine learning, notably deep neural networks (DNNs), has shown considerable promise in medical applications, including the detection of CVD. DNNs are capable of identifying intricate patterns in large datasets, making them valuable tools in domains like computer vision, NLP, and healthcare diagnostics. However, optimizing DNNs for effective CVD prediction presents challenges related to model architecture, parameter tuning, data preprocessing, and selecting appropriate optimization techniques [3, 4]. Advanced machine learning methods are increasingly used to investigate patient data, providing insights that can lead to timely interventions [5, 6].

Heart disease broadly encompasses any condition that impairs the normal functioning of the heart. As part of achieving the goals of this work, it is essential to examine various factors, including the underlying causes, the challenges encountered by both patients and healthcare providers, and the available treatment options [7]. A crucial component of machine learning is optimization, which directly impacts model performance by adjusting parameters to minimize prediction errors [8]. Standard optimization approaches like gradient descent, back propagation and adaptive moment estimation (Adam) are used to enhance model accuracy and convergence speed [9-13]. This work focuses on developing an Optimizing Cardiovascular Disease Prediction with a Hybrid Gradient Descent Adaptive Algorithm and Random Forest Classifier model, specifically aimed at enhancing the diagnostic precision of cardiovascular conditions.

Optimization involves finding an acceptable solution among multiple possibilities, which is essential when dealing with high-dimensional healthcare data. Traditional optimization techniques may struggle with such data, but Hybrid optimization approaches, which combine multiple optimizers, have shown promising results. For example, AdaGrad's feature-specific learning rates enhance performance on sparse data, while AdamW's momentum-based adaptations improve performance on noisy datasets [11, 14]. Healthcare organizations increasingly rely on computer-aided diagnosis (CAD) systems that utilize machine learning techiques to interpret patient records and uncover critical diagnostic patterns [15]. These systems hold the potential to substantially elevate the accuracy and operational efficiency of cardiovascular disease detection, thereby ensuring timely and precise medical intervention [16].

Random Forest (RF) classifiers have proven effective in predictive healthcare, especially when combined with advanced optimization techniques such as AdaGrad and AdamW. These Hybrid approaches overcome the limitations of traditional models, including slow convergence and poor prediction accuracy, by improving feature selection, model stability, and the accuracy of heart disease diagnoses. AdaGrad's parameter-specific learning rates and Adam's momentum-based adjustments enhance performance on sparse and noisy datasets, making these optimizers particularly useful for heart disease prediction [17, 18].

This work proposes a new advanced system for heart disease diagnosis, incorporating Hybrid optimization techniques to enhance prediction accuracy. The system utilizes various factors such as age, sex, chest pain type, resting blood pressure etc., to make predictions.

The primary contributions of this work are outlined as follows:

1. Adaptive Feature Selection: The proposed Hybrid GD-AA-RF model improves heart disease prediction by combining Adam's momentum with AdaGrad's adaptive learning rates, leading to better feature selection.

2. Improved Classification: By integrating Random Forest with adaptive optimizers, the model enhances classification performance, resulting in improved accuracy, precision, recall, and F1 scores.

3. Comprehensive Evaluation: The model undergoes a thorough evaluation based on accuracy, precision, recall, and F1-score, confirming its effectiveness in early-stage heart disease prediction.

Furthermore, the integration of AdamW and AdaGrad enhances the model's efficiency and resilience. AdamW's weight decay helps control overfitting, making it well-suited for non-stationary data, while AdaGrad's adaptive learning rates prove effective for sparse data. This Hybrid approach promotes adaptive learning for rare but crucial features and regularization, leading to improved model stability and reduced overfitting [19].

Journal of Information Systems Engineering and Management
2025, 10(47s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

OBJECTIVES

As cardiovascular diseases continue to be a leading cause of death worldwide, the urgent need for advanced diagnostic tools has never been more apparent. In light of the escalating mortality rates attributed to these conditions, developing a reliable and efficient model for early heart disease detection becomes a critical priority. One significant challenge faced by healthcare institutions is the availability of accurate diagnostic tools at affordable prices. The main concerns with existing models are their accuracy, usefulness, and reliability. This study seeks to determine the most efficient machine-learning models for diagnosing heart disease with improved accuracy, sensitivity, and precision.

The Hybrid GD-AA-RF model offers several benefits:

1. Enhanced Robustness: By incorporating adaptive learning rates and weight decay, the Hybrid approach strengthens the model, addressing overfitting and improving stability.

2. Comprehensive Appraisal: The model undergoes an evaluation utilizing a diverse array of metrics, offering an in-depth analysis of its performance.

3. Clear Implementation: The structured process supports future scalability, optimization, and debugging.

This paper is organized as follows: Section II reviews the current literature on cardiovascular prediction models. Section III describes the Hybrid GD-AA-RF methodology. Section IV discusses performance metrics and provides a comparative evaluation, and Section V wraps up the study by highlighting the model's predictive capabilities and suggesting directions for future research.

RELATED WORKS

The prediction of heart disease has become a focal point in recent research, driven by its profound impact on public health. In response, researchers have explored a diverse range of machine learning techniques, including advanced optimization algorithms, to significantly improve the accuracy and efficiency of predictive models, thereby enhancing early detection and intervention capabilities.

El-Shafiey et al. (2022) proposed an integrated approach combining Genetic Algorithm (GA) and Particle Swarm Optimization (PSO) for fine-tuning the parameters of the RF model, aiming to improve its accuracy in predicting heart disease. Their model, GAPSO-RF, utilizes multivariate analysis for feature identification, improving prediction accuracy. Despite its computational efficiency, the model faces challenges such as temporal complexity and high computational costs in smaller datasets [20].

Abdollahi and Nouri-Moghaddam (2022) developed an ensemble method for diagnosing heart disease, focusing on feature selection to enhance classification accuracy while reducing computational time. Their stacked GA model, applied to the UCI heart disease dataset, demonstrated shorter computational times and better generalization but also raised concerns about overfitting and computational overhead [21].

Doppala (2023) introduced a Hybrid GA-RBF model for coronary disease prediction, focusing on feature selection for better prediction accuracy. Although this model improved prediction effectiveness, it still struggled with high computational costs and long processing times [22].

Reyad et al. (2023) modified the Adam optimization algorithm to dynamically adjust the learning rate, which improved convergence speed and model accuracy for deep neural networks. However, tuning the additional parameters in this approach could be computationally intensive [23].

Dogo et al. (2018) explored RMSProp and Adam in the domain of Convolutional Neural Networks (CNNs) and found that these optimization techniques significantly improved model performance, but the high computational costs of running CNNs with large datasets remained a limitation [24].

Torthi et al. (2024) introduced the BAPSO-RF model, which integrates the Bat Algorithm (BA) and PSO with Random Forest to optimize parameter selection for heart disease prediction. The model demonstrated remarkable performance achieving 98.71%, accuracy, 98.67% precision, 98.23% recall, and 98.45%. Despite its superior

Journal of Information Systems Engineering and Management
2025, 10(47s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

performance, the model faces challenges related to computational expense and the need for further feature selection to reduce training time [25].

Radosavljević et al. (2015) developed a Hybrid PSO and Gravitational Search Algorithm (GSA) for power flow optimization, highlighting the benefits of Hybrid models in multi-variable environments but pointing out their temporal complexity, which can lead to performance bottlenecks in real-time applications [26].

Maddikunta et al. (2020) addressed resource management in IoT networks by combining energy-efficient communication protocols with machine learning, improving resource allocation. However, their Hybrid optimization approach faced challenges with adaptability when network topology changed frequently [27].

Wang et al. (2018) proposed a Hybrid strategy combining Adam and Stochastic Gradient Descent (SGD), improving model consistency. However, the approach's scalability issues, particularly with larger datasets were noted [28].

2.1. Research Gap

Limitation 1: Small Datasets and Lack of Combined Data: Previous studies [7] on cardiovascular disease (CVD) prediction has been limited by small datasets and the absence of combined datasets, which negatively affect model accuracy and generalization. This investigation overcomes these limitations by leveraging a more extensive, combined dataset of 1190 instances, improving model performance. However, subsequent research should prioritize the utilization of even larger and more heterogeneous datasets, while investigating the integration of data from various sources to further improve the precision and resilience of cardiovascular disease prediction models.

Limitation 2: Basic Machine Learning Models: Many existing models rely on basic machine learning algorithms, limiting their predictive power. In this research, more advanced Hybrid gradient descent techniques are used, leading to better performance. Future studies should continue to explore and integrate more sophisticated machine learning models to improve diagnostic precision and model reliability.

Limitation 3: High Computational Costs and Slow Processing Times: Despite advancements with Hybrid models like GAPSO-RF and BAPSO-RF, issues such as high computing costs, slow processing times, and scalability limitations remain [22-28].

To address these challenges, the proposed Hybrid GD-AA-RF model, which combines AdaGrad and AdamW Hybrid optimizers with Random Forest, offers a solution by improving classification performance, speeding up convergence, and enhancing feature selection. This model also uses adaptive learning rates and weight decay, improving stability, reducing overfitting, and increasing predictive accuracy.

METHODS

This work focuses on data preprocessing using the Z-score method, a novel feature selection method (GDE), Hybrid GDO optimization method, and classification approaches to construct a structured machine learning pipeline for predicting cardiovascular health outcomes. This approach ensures the prediction model's accuracy and interpretability. The system architecture aims to support decision-making using a binary classification approach to determine the presence of disease, with human health data as input. Unlike standard expert systems based on basic if-then-else logic, this approach leverages advanced machine learning algorithms for a more intelligent and adaptable system. Figure 1 illustrates the refined cardiovascular disease prediction architecture using Hybrid Gradient Descent Optimization-based machine learning techniques.

Initially, data is collected from a public dataset sourced from the UCI ML Repository, in the form of a .csv file containing 12 features and 1,190 instances. Data preprocessing includes feature engineering, addressing missing values, eliminating redundancy, and mitigating outliers using the Z-score method with a threshold of 3. Feature selection is performed using a novel Gaussian-based information gain method (GDE), selecting 9 out of 12 features. The selected features undergo optimization and are processed through binary classification algorithms for both training and testing. Gradient Descent Optimization (GDO) is employed for model training. Several models, including Hybrid GD-AA-RF, LR, NB, SVM, KNN, and DT, are assessed and compared based on accuracy, precision, recall, and F1 score. This intelligent system adopts a comprehensive approach to accurately identify cardiovascular disease.

Journal of Information Systems Engineering and Management
2025, 10(47s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

3.1 MATHEMATICAL FORMULATION FOR THE PROPOSED TECHNIQUE

The proposed model's operation for predicting heart disease is represented through a mathematical framework. The framework has numerous mathematical steps that are followed:

Step 1: Data Acquisition

The publicly accessible heart disease information, comprising 1190 records and 12 features, was used to assess the suggested technique [29]. 20 percent of the dataset was designated for evaluation, while the remaining 80 percent was utilized in the model training phase. The objective is to use the given medical information to categorize whether a patient has a cardiac problem. This data is frequently used to assess different data processing methods and serve as a benchmark. Table1 provides an in-depth overview of the dataset [25]. Additionally, it was confirmed that the dataset was balanced, with 53% of cases representing patients with cardiac disease (628 cases) and 47% representing normal individuals (561 cases).

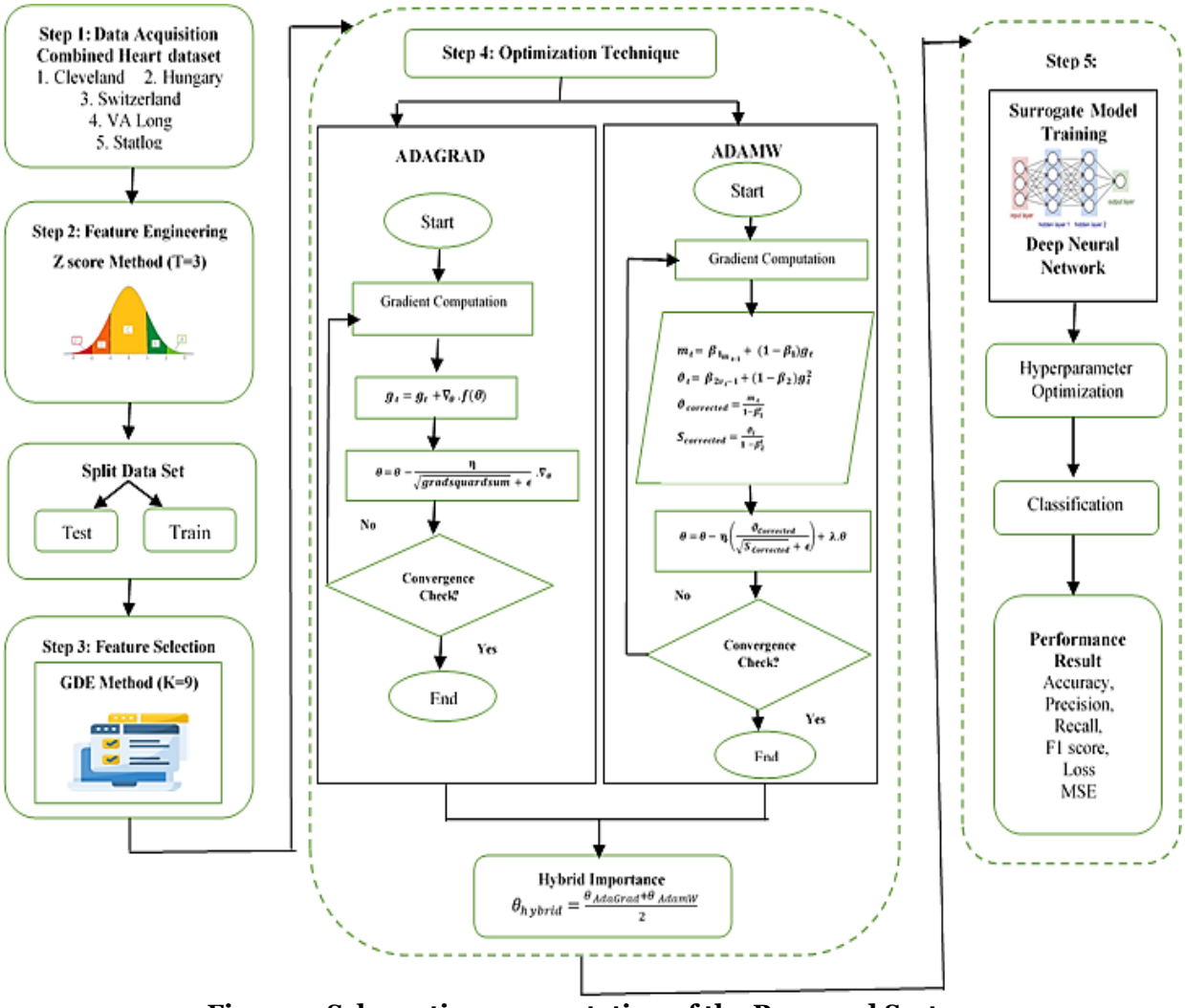


Figure 1. Schematic representation of the Proposed System

Table 1. Description of the Given Dataset [25]

S.No	Attributes	Description of Type
1	Age	Integer
2	Sex	Binary Categorical (Two Values)

Journal of Information Systems Engineering and Management
2025, 10(47s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

3	Chest Pain type (CP)	Categorical (Four Values)
4	Resting bps	Integer
5	Cholesterol	Integer
6	Fasting Blood Sugar	Binary Categorical (Two Values)
7	Resting ECG	Categorical (Three Values)
8	Max Heart Rate	Integer
9	Exercise Angina	Binary Categorical (Two Values)
10	Old Peak	Real
11	ST Slope	Categorical (Three Values)
12	Target	Binary Categorical (Two Values)

Step 2: Preprocessing

Datasets frequently include incomplete or inconsistent information, necessitating filtering and normalization prior to further analysis. In this study, the dataset is statistically standardized using Z-score method to ensure consistency [30]. The mathematical formulation of Z-score normalization is presented in Eq. (1).

$$X\ scaled = \frac{X - Mean(X)}{Std(X)} \quad \text{----- (1)}$$

This function normalizes the features in X by subtracting the mean and dividing by the standard deviation for each column. Standardization enhances model performance, particularly in gradient-based optimization algorithms, by ensuring uniform scaling across features. This transformation aids in efficient model convergence and ensures that each feature contributes equally to the predictive outcome. Following this process, the pre-processed dataset is prepared for subsequent stages such as feature selection and classification.

Step 3: Feature Identification Using Novel Method

Feature identification is crucial in optimizing machine learning models, particularly when handling complex, high-dimensional datasets, such as those found in medical applications. Information gain evaluates on how much uncertainty is reduced after the split. For continuous data, Gaussian differential entropy is used in the eq (2) and eq(3)

$$\textbf{Gaussian Based Differential Entropy(S)} = 0.5 * \sigma\sqrt{2e\pi} \text{ ---------- (2)}$$

$$\textbf{GDE for IG} = \boldsymbol{GDE}\ (\boldsymbol{S}) - \sum_{\infty}^{-\infty}[\left(\frac{|\boldsymbol{S_V}|}{|\boldsymbol{S}|}\right) * \boldsymbol{GDE}\ (\boldsymbol{S_V})] \text{ ---------- (3)}$$

Step 4: Hybrid of Gradient Descent Optimization

The model comprises three layers: the input, hidden, and output. Each layer contains a fixed number of neurons specific to its function. The implementation of Hybrid Gradient Descent (GD) –AA-RF involves several steps, including the initial selection of weights for the inputs, followed by the feed-forward process for the accumulation of inputs, weight updates during backpropagation GD, and bias calculation.

The Adaptive Gradient Algorithm, or AdaGrad, dynamically modifies the learning rate for each parameter according to its frequency of updates, which is especially beneficial for sparse or less frequent features mechanism, which accumulates squared gradients over time, prioritizes rare but essential features by assigning larger updates, thus reducing the risk of underestimating these features [31, 32]. The AdaGrad parameter update rules are expressed in eq (4-5).

Journal of Information Systems Engineering and Management
2025, 10(47s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

$$g_t = g_t + \nabla_\theta . f(\theta) \quad \text{----- (4)}$$

$$\theta = \theta - \frac{\eta}{\sqrt{gradsquardsum} + \epsilon} . \nabla_\theta \quad \text{------ (5)}$$

Here, η is the learning rate, ∇_k is the gradient at step k, ϵ prevents division by zero. The importance of pertinent characteristics in the optimization process is supported by AdaGrad, which gradually reduces the learning rates for frequently updated parameters.

AdamW (Adaptive Moment Estimation with Weight Decay) is an optimized version of the Adam algorithm that includes explicit weight decay, making it highly suitable for reducing overfitting in noisy datasets, such as medical data [33]. Unlike Adam, which indirectly includes regularization, AdamW directly incorporates a weight decay component. This separation prevents overshooting during prolonged training sessions, especially for parameters with high learning rates [34].

In Adam, each parameter's learning rate is adjusted dynamically, leveraging the benefits of both momentum and RMSProp by tracking the exponential moving averages of the gradient (m_t) and the squared gradient (ϑ_t) [35].

AdamW modifies Adam's update rules by directly adding a weight decay term, which regularizes the model without interfering with the adaptive learning rate calculations. The following methods demonstrate how to modify parameters using equations [6–10] [23]:

$$m_t = \beta_{1_{m_{t-1}}} + (1 - \beta_1)\boldsymbol{g_t} \quad \text{------ (6)}$$

$$\vartheta_t = \beta_{2v_t - 1} + (1 - \beta_2) g_t^2 \quad \text{------ (7)}$$

$$\vartheta_{corrected} = \frac{m_t}{1 - \beta_1^t} \quad \text{----- (8)}$$

$$S_{corrected} = \frac{\vartheta_t}{1 - \beta_2^t} \quad \text{----- (9)}$$

$$\theta = \theta - \eta . \left(\frac{\vartheta_{Corrected}}{\sqrt{S_{Corrected}} + \epsilon} \right) + \lambda . \theta \quad \text{----- (10)}$$

Where, λ is the weight decay (L2 regularization) ,β1 is the exponential decay rate.

The default values for β1 and β2 are 0.9 and 0.999, respectively. $\vartheta_{\text{corrected}}$ and $\vartheta_{\text{corrected}}$ are correction biases for m_t and ϑ_t respectively. This helps to minimize the risk of overfitting and enhances model generalization in high-dimensional medical datasets.

Next, outlines a Hybrid approach to feature selection by combining the unique optimization strategies of AdaGrad and AdamW in eq (11).

Average parameters from both optimizers to form a Hybrid parameter vector:

$$\theta_{hybrid} = \frac{\theta_{AdaGrad} + \theta_{AdamW}}{2} \quad \text{----- (11)}$$

Step 5: Hyper Parameters Using Random Forest Classifier

The surrogate neural network utilizes its architecture to dynamically predict hyperparameters for the Random Forest (RF) classifier, specifically targeting the number of estimators (n_estimators) and the maximum tree depth (max_depth). This not only enhances the efficiency of RF in handling binary classification tasks for high-dimensional datasets but also ensures adaptability to different data distributions. The RF classifier's ensemble-based structure makes it highly robust, capable of managing complex data structures, reducing variance through bagging, and capturing non-linear relationships. Furthermore, its ability to prevent overfitting by selecting random subsets of features and data ensures optimal performance, even with noisy or redundant datasets. These characteristics make RF particularly well-suited for critical medical applications where interpretability and accuracy are vital.

Journal of Information Systems Engineering and Management
2025, 10(47s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

The RF classifier is trained using the most pertinent features identified through this hybrid optimization process, ensuring that the model prioritizes the most influential predictors. This methodology is assessed using various performance metrics, including accuracy, precision, recall and F1 score, offering a thorough evaluation of the model's effectiveness. The incorporation of the Hybrid GD-AA-RF model proves especially beneficial for cardiovascular disease prediction, as it adeptly manages non-linear, high-dimensional and intricate datasets commonly encountered in medical diagnostics.

The model's capability to robustly generalize across heterogeneous data distributions ensures its adaptability to different populations and datasets, making it suitable for scalable healthcare applications. In the proposed Hybrid GD-AA-RF model, AdaGrad and AdamW optimizers are seamlessly integrated with RF classification, creating a hybrid optimization strategy to enhance the feature selection and classification process. During preprocessing, Z-score standardization ensures data consistency by scaling features within a standard range, effectively managing outliers. The hybrid significance score is a key innovation, calculated by averaging independent feature importance scores derived from both AdaGrad and AdamW optimizers. AdaGrad excels at adjusting learning rates for less frequently occurring features, while AdamW's weight decay mechanism improves generalization. Together, these optimizers provide a stable learning process, enabling faster convergence and reducing the risk of overfitting or underfitting.

Additionally, the hybrid optimization strategy improves interpretability by highlighting the importance of specific features, which can assist clinicians in understanding critical predictors of cardiovascular disease. With its robust performance, the Hybrid GD-AA-RF model represents a significant advancement in machine learning-driven diagnostic systems, supporting timely and accurate decision-making for cardiovascular health outcomes. In figure 2 show the Pseudocode for Proposed (Hybrid GD-AA-RF) Model.

1. Data Preprocessing:
   a. Load dataset and separate features (X) and target (y).
   b. Scale features using Z-score normalization:
2. Define Optimization Algorithms:
   a. AdaGrad:
      i. Initialize parameters and gradient accumulator.
      ii. Update parameters with an adaptive learning rate:
   b. AdamW:
      i. Initialize moment estimates and bias correction factors.
      ii. Update parameters with weight decay:
3. Combine AdaGrad and AdamW Results:
   a. Average parameters from both optimizers to form a hybrid parameter vector:
4. Use the Random Forest Classifier on all the features.
a. Use the Random Forest classifier to train on the selected features.
   b. Generate probability predictions: rf=rf_classifier.predict_proba(X)
5. Evaluate Model: Compute model performance metrics (Accuracy, Precision, Recall, F1 Score)
6. Output Results: Display evaluation metrics to assess model effectiveness.

Figure 2: Pseudocode for Proposed (Hybrid GD-AA-RF) Model

This approach leverages the combined strengths of AdaGrad and AdamW to achieve effective feature importance, which minimizes overfitting risk and improves generalization in high-dimensional medical datasets. By integrating AdaGrad's adaptive learning rate with AdamW's weight decay regularization, the Hybrid method prioritizes important yet infrequent features while controlling model complexity, resulting in more efficient computation and higher classification accuracy. This streamlined feature selection process avoids the complexity of metaheuristic methods like the Bat Algorithm or Particle Swarm Optimization. The iterative update of parameters, balancing AdaGrad's adaptability and AdamW's regularization, ensures that key features are consistently prioritized for optimal predictive performance.

Journal of Information Systems Engineering and Management
2025, 10(47s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

RESULTS

In this context, using machine learning algorithms for accurate and efficient classification in the prediction of heart disease formation is crucial for healthcare. Given the complex nature of predicting heart disease, using suitable feature selection techniques along with optimized classifiers will help improve the performance of our models. This study focuses on exploring and implementing advanced approaches to enhance classification accuracy by applying feature selection through optimization techniques. Subsequently, several machine learning algorithms, such as LR, NB, SVM, KNN, and DT, are analyzed and benchmarked against the Random Forest (RF) algorithm, delivers enhanced performance across key evaluation metrics such as accuracy, precision, recall and F1-score in the context of heart disease prediction

4.1. Performance Metrics Analysis

The proposed Hybrid GD-AA-RF Model has been applied, and here, we present the experimental results. Several metrics such as accuracy, precision, recall, and F1-score are used to evaluate the model performance. However, in order to know more about the proposed work, it is compared with most known methods.

As demonstrated in Table 2 and Figure 3, several classifiers were evaluated using original features. LR, NB, SVM, KNN, and DT were benchmarked against the RF model. Among these, the RF model outperformed the others, achieving 94.96% accuracy, 94.66% precision, 96.12% recall, and 95.38% F1-score, establishing it as the best performer.

Table 2: Metrics Comparison across Classifiers with Original Features

Classification Models	Accuracy	Precision	Recall	F1- Score
Logistic Regression	80.67 %	81.68 %	82.95 %	82.31 %
Naïve Bayes	85.29 %	85.29 %	85.27 %	86.27 %
SVM	80.25 %	81.54 %	82.17 %	81.85 %
KNN	68.49 %	71.43 %	69.77 %	70.59 %
Decision Tree	88.24 %	89.76 %	88.37 %	89.06 %
Random Forest	**94.96** %	**94.66** %	**96.12** %	**95.38** %

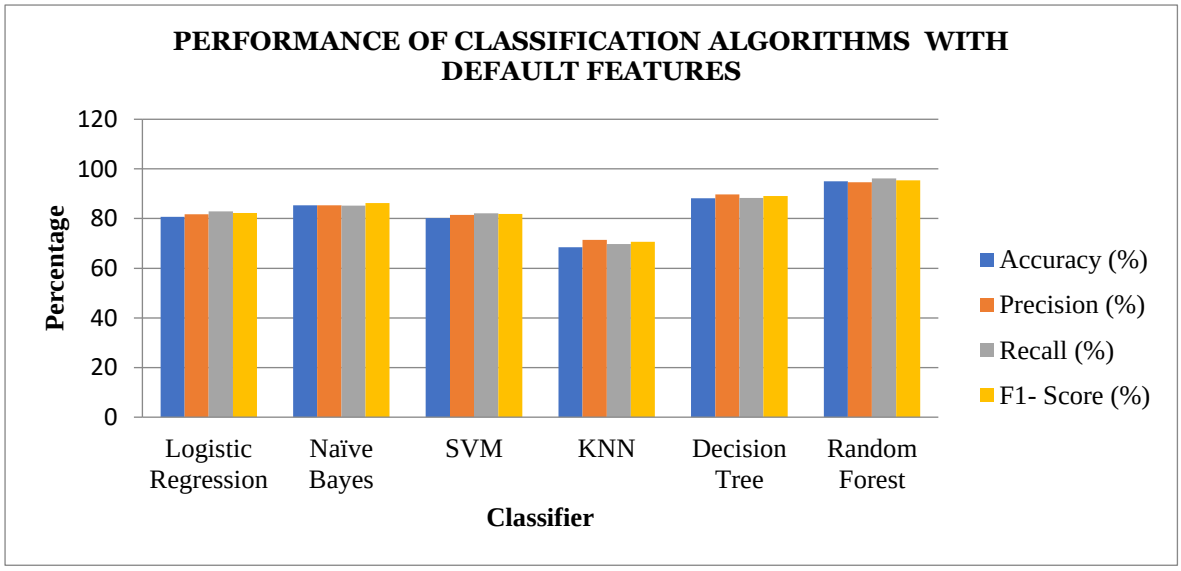


Figure 3: Performance of Classification Algorithms with default features

Journal of Information Systems Engineering and Management
2025, 10(47s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

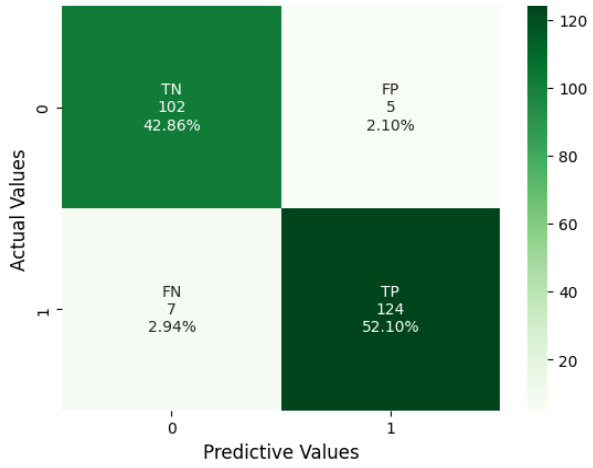


Figure 4: Confusion Matrix for Random Forest with Default Feature

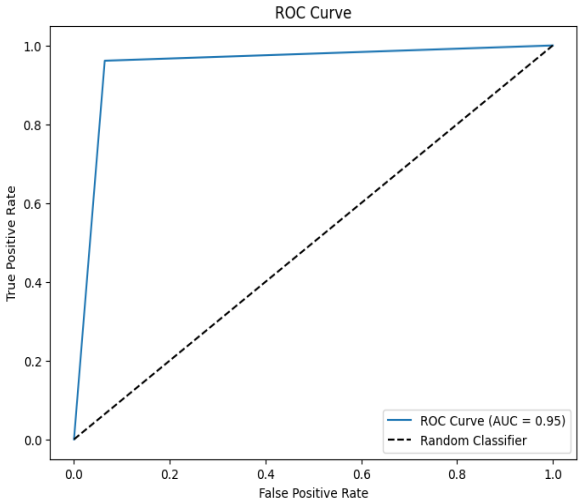


Figure 5: ROC Curve for Random Forest with Default Features

The confusion matrix of the Random Forest model with default features is presented in Figure 4; it is capable of classifying both types of heart disease and the lack of them. The high diagonal values suggest a high correct prediction rate. Figure 5 displays the ROC curve for Random Forest with Default Feature, emphasizing the balance between sensitivity and specificity, with a higher curve reflecting superior performance. This success is largely due to Random Forest's effectiveness in minimizing overfitting and improving generalization through bagging and feature randomness. These techniques enable the model to perform well on new, unobserved data, making it highly trustworthy for heart disease diagnosis. Moving forward, the Random Forest algorithm will be the focus, showcasing its ongoing effectiveness in this application.

Table 3. Simulation results produced by the proposed model with the different train-test ratio

Train/Test Ratio	Accuracy (%)	Precision (%)	Recall (%)	F1- Score (%)	MSE	LOSS
60-40	99.72	99.46	100	99.73	0.0028	0.1162
70-30	99.64	99.30	100	99.65	0.0036	0.0900
80-20	99.05	98.61	99.60	99.10	0.0095	0.0914
90-10	99.07	98.58	99.64	99.11	0.0093	0.0747

The model assessment of the Advanced Heart Disease forecasting System using a Hybrid Gradient Descent Adaptive model with various train-test ratios (60-40, 70-30, 80-20, and 90-10) revealed that the 80-20 split offers the best balance between training and evaluation is shown in Table 3. This configuration attained an outstanding accuracy of 99.05%, an F1-score of 99.10%, and kept the Mean Squared Error (MSE) and Loss values low at 0.0095 and 0.0914, respectively.

The 80-20 split is ideal because it allocates 80% of the data for training, enabling the model to learn effectively from a substantial dataset, while 20% is reserved for testing, ensuring comprehensive performance evaluation. This balance minimizes the risk of overfitting that can occur with smaller test sets, such as the 90-10 split, and shows a more robust and generalizable model's assessment ability to handle unseen data.

In comparison, the 60-40 split achieved the highest accuracy and F1-score overall, but the reduced training data limits the model's capacity to capture complex patterns in datasets with complex relationships. The 70-30 split

Journal of Information Systems Engineering and Management
2025, 10(47s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

performed competitively but showed slightly higher Loss, indicating a compromise in performance. Although the 90-10 split yielded strong accuracy and lower Loss, its small test set reduces the reliability and comprehensiveness of the evaluation process.

The 80-20 split strikes an excellent trade-off by offering sufficient training data for effective model learning while maintaining a large enough test dataset to ensure reliable, realistic, and statistically valid performance metrics. These qualities make the 80-20 split the most effective choice for achieving high accuracy, minimizing error, and ensuring the model's applicability.

Table 4. Simulation results of Model with LR, NB, SVM, K-NN and DT classifiers

Classification Models	Accuracy (%)	Precision (%)	Recall (%)	F1- Score (%)	MSE
Logistic Regression	85.29	84.78	89.31	86.99	0.1471
Naïve Bayes	84.87	85.71	87.02	86.36	0.1513
SVM	86.13	83.56	93.13	88.09	0.1387
KNN	86.97	86.76	90.08	88.39	0.1303
Decision Tree	89.08	94.12	85.50	89.60	0.1092
Proposed Model	**99.05**	**98.61**	**99.60**	**99.10**	**0.0095**

Table 4 delineates a comparative analysis of the performance metrics attained by the proposed Advanced Cardiovascular Disease Prediction System, which leverages a Hybrid Gradient Descent Optimization (GD) framework in conjunction with classifiers such as Logistic Regression (LR), Naïve Bayes (NB), Support Vector Machine (SVM), K-Nearest Neighbors (K-NN), and Decision Tree (DT). The classification component of the proposed system exhibited superior efficacy, achieving an exceptional accuracy of 99.05%, along with remarkable precision, recall, and F1-score values—underscoring its dominance over conventional approaches. The robustness of the Hybrid GD–AA–RF methodology significantly contributed to this outcome by facilitating dynamic and efficient feature selection, thereby minimizing the Mean Squared Error (MSE) to a notably low value of 0.0095.

Among traditional classifiers, the DT performed well with 89.08% of accuracy offering high precision (94.12%) and a low MSE of 0.1092. The KNN algorithm followed closely, achieving an accuracy of 86.97% with balanced metrics, including an F1-score of 88.39%. SVM demonstrated decent performance with 86.13% accuracy and a notably high recall (93.13%), but its MSE of 0.1387 indicates room for improvement.

LR and NB reached moderate results, with accuracies of 85.29% and 84.87%, respectively, and comparable precision and recall values. However, their performance falls short compared to advanced models like the Decision Tree and the Proposed Model. Furthermore, it is noted that the automatic feature selection led to improved accuracy for the proposed Hybrid GD-AA-RF model. Figure 6 presents a graphical comparison of the various machine-learning approaches.

4.2. ROC Analysis

The ROC analysis graph demonstrates the diagnostic performance of six algorithms, LR NB, SVM, K-NN, and the Proposed Model Hybrid GD-AA-RF. Each curve plots sensitivity (y-axis) against specificity (x-axis), with the AUC score indicating the algorithm's accuracy. The Hybrid GD-AA-RF model outperforms others with an AUC of 0.9905, closely approaching the ideal top-left corner, signifying excellent diagnostic ability. In contrast, model like DT show moderate performance, while K-NN demonstrate strong accuracy with AUC scores above 0.9. The graph highlights the superior predictive power of the Hybrid GD-AA-RF model for heart disease diagnosis.

Journal of Information Systems Engineering and Management
2025, 10(47s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

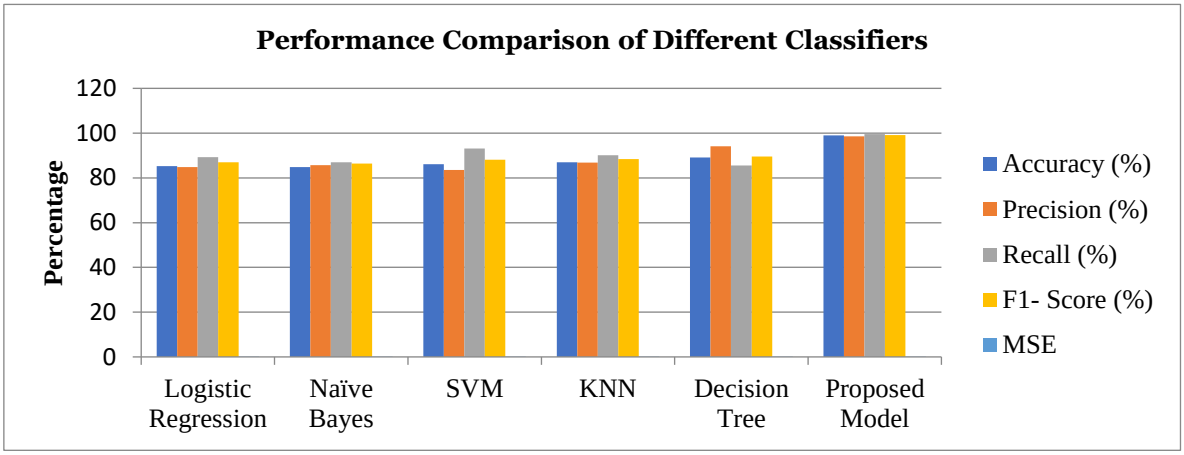


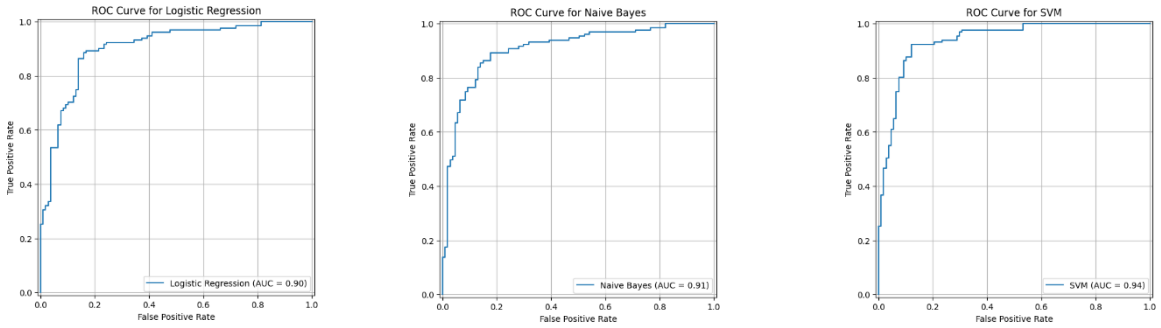
Figure. 6 Simulation results produced by LR, NB, SVM, K-NN, DT and Proposed Model.

Table 5. Values of mean square error (MSE), root mean square error (RMSE), Out-of-bag (OOB) error corresponding to no. of epochs for Proposed Model.

No. of Epoch	10	50	60	100
MSE	0.0515	0.0315	0.0095	0.0147
RMSE	0.2269	0.1775	0.0972	0.1213
OOB Error	0.1597	0.1008	0.0924	0.0977

Table 5 presents the correlation between the number of training epochs and the evaluation measures: MSE, RMSE, and OOB Error. As the number of epochs increases, MSE and RMSE initially decrease, reaching their lowest values at 60 epochs, with MSE reaching 0.0095 and RMSE at 0.0972. However, at 100 epochs, both metrics show a slight increase, suggesting overfitting, where the model starts to overlearn the training data and performs less effectively on unobserved data. Similarly, the OOB Error decreases steadily up to 60 epochs, but begins to rise after that, further supporting the observation of diminishing returns beyond 60 epochs.

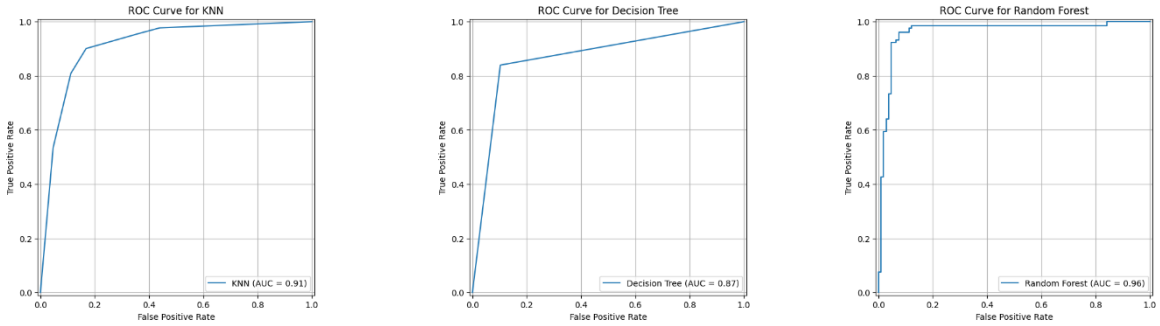
The purpose of this work is to find the optimal training duration that balances model accuracy and generalization. With the lowest values of MSE, RMSE, and OOB Error at 60 epochs, it is evident that 60 epochs achieve the best model performance. Epochs beyond 60 show signs of overfitting, highlighting the importance of avoiding excessive training. Thus, the optimal number of epochs ensures that the model remains both accurate and generalizable without overfitting.



(a) **(b)** **(c)**

Journal of Information Systems Engineering and Management
2025, 10(47s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article



(d) **(e)** **(f)**

Figure 7. Six ROC curves i.e. (a), (b), (c), (d), (e), and (f) show receiver operating characteristic (ROC) curves produced by Logistic Regression (LR), Naïve Bayes (NB), Support Vector Machine (SVM), K-Nearest Neighbor (KNN), Decision Tree(DT), and Proposed Model respectively.

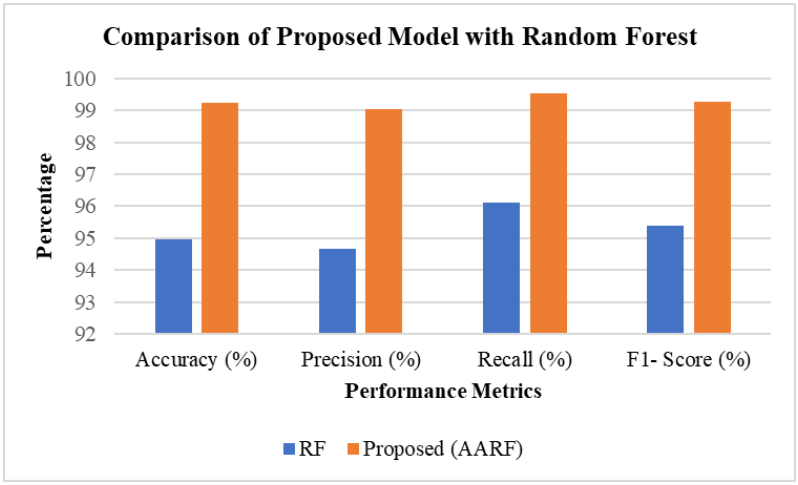


Figure 9: Comparison of Proposed Model with Random Forest

In this work, several optimization techniques were active to improve the performance of the Random Forest classifier for predicting heart disease. Bayesian Optimization (BO) with RF achieved solid performance, but its results were surpassed by other methods. Particle Swarm Optimization (PSO) combined with RF improved both accuracy and precision by effectively exploring the solution space, resulting in better model performance. The Global Differential Evolution (GDE) with LASSO method provided outstanding results by selecting the relevant features, minimizing the overfitting, and enhancing the accuracy and F1-score. Additionally, GDE combined with Genetic Algorithm (GA) showed promising outcomes by balancing exploration and exploitation in the optimization process, leading to competitive performance. Ultimately, the proposed custom optimization method, which combines these techniques, achieved the highest accuracy and overall performance, establishing it as the most effective approach for predicting heart disease.

Table 6: Performance of Random Forest classification using Various Optimization Techniques

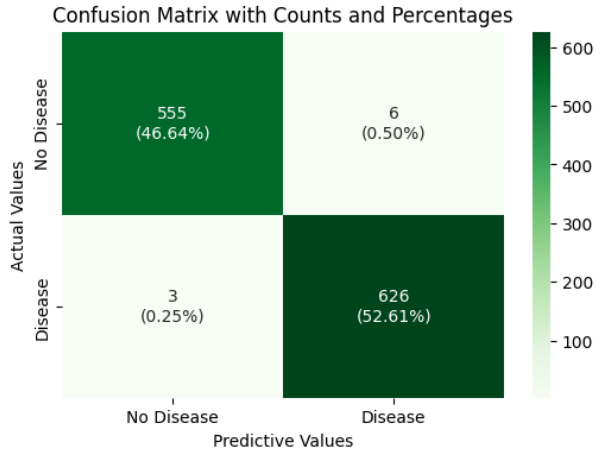
Methods	Accuracy (%)	Precision (%)	Recall (%)	F1- Score (%)
BO+RF	93.7	96.18	92.65	94.38
PSO+RF	95.38	96.95	94.78	95.85

Journal of Information Systems Engineering and Management
2025, 10(47s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

GDE_LASSO+RF	96.64	97.71	96.24	96.97
GDE_GA+RF	94.95	93.43	97.70	95.52
Proposed	**99.05**	**98.61**	**99.60**	**99.10**

The table 6 illustrate the classification performance of various optimization techniques combined with the Random Forest classifier, including Bayesian Optimization (BO)+RF, Particle Swarm Optimization (PSO)+RF, GDE_LASSO+RF ,GDE_GA+RF and the proposed Hybrid GD-AA-RF algorithm. Among these, the Hybrid GD-AA-RF method achieves the highest metrics, with 99.05% accuracy, 908.61% precision, 99.60% recall, and 99.10% F1-score, outstanding other techniques. This enhanced performance is due to Hybrid GD-AA-RF combination of AdaGrad and AdamW optimizers, where AdaGrad's adaptive learning rate addresses sparse gradients, while AdamW's momentum and adaptive learning enhance convergence and avoid local minima. Together, these elements enable Hybrid GD-AA-RF to effectively select key features and optimize the Random Forest model, resulting in improved prediction accuracy.



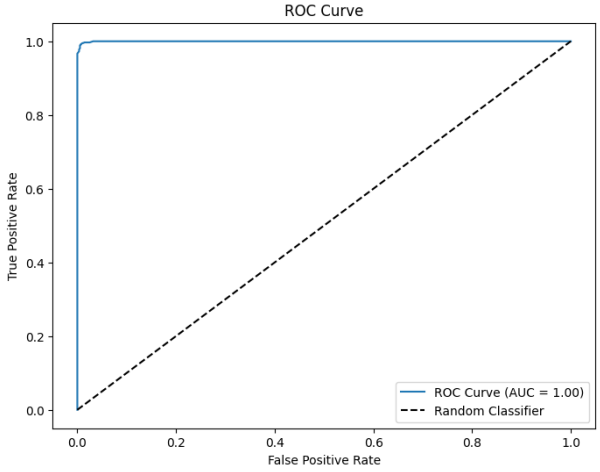


Figure 7: Confusion Matrix for Proposed Model **Figure 8: ROC Curve for Proposed Model**

The confusion matrix in Figure 7, reveals the model's precise classification of both heart disease and non-disease cases, with high values on the main diagonal, indicating minimal misclassification. The ROC curve in Figure 8 display the model's reliability, as its proximity to the upper-left corner signifies optimal trade-off, enhancing its capability to detect true cases without excessive false positives. Figures 7 and 8 validate the effectiveness of the proposed model in predicting heart disease, highlighting its high accuracy and well-balanced sensitivity and specificity. These results collectively demonstrate the model's potential to aid healthcare professionals with early diagnosis and timely intervention for heart disease.

4.3 Comparative Analysis

The fig 9 shows that the proposed Hybrid GDO- AA-RF model outperforms the standard Random Forest in all key metrics, such as accuracy, precision, recall, and F1-score, demonstrating its superior ability to accurately identify heart disease cases with minimal false positives and a high detection rate of true cases.

Also, this section compares the performance of the proposed Hybrid GD-AA-RF approach with other existing methods based on evaluation metrics like accuracy, precision, recall, and F1-score. Table 4 includes results from previous studies and benchmarks the performance of the proposed method against existing classifiers. The Hybrid GD-AA-RF classifier using trained, tested, and validated on five combined heart disease datasets, demonstrates significant improvements over other methods. The results from Table 4 show that the proposed Hybrid GD-AA-RF approach achieves an accuracy of 99.24%, precision of 99.05%, recall of 99.52%, and F1-score of 99.28%, which are

Journal of Information Systems Engineering and Management
2025, 10(47s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

superior to those reported in the literature for methods such as GAPSO-RF [20], Stacked GA [21], GA-RBF [22], and BAPSO-RF [25].

Table 4: Comparative Analysis

Author	Method	Dataset	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)
El-Shafiey., et.al [20]	GAPSO-RF	Heart Disease Dataset	95.60	97.44	92.68	94.00
Abdollahi ., et.al [21]	Stacked GA		97.57	N/A	96	N/A
Doppala ., et.al [22]	GA-RBF		85.40	95	96	95
Torthi, R., et.al [25]	BAPSO-RF		98.71	98.67	98.23	98.45
Proposed method	**Hybrid GD-AA-RF**		**99.05**	**98.61**	**99.60**	**99.10**

4.4 Discussion

This section outlines the benefits of the proposed method and addresses the shortcomings of existing approaches. Several current models present notable constraints: The model proposed in [20] is challenged by high temporal complexity, poor performance on small datasets, and elevated computational demands. The model in the [21] approach incurs substantial processing overhead and carries a heightened risk of overfitting. The technique [22] results in increased computational load and longer processing times per generation. Meanwhile, the BAPSO-RF [25] model requires considerable data resources for large-scale datasets and shows a higher error rate in multiclass classification scenarios.

In comparison, the proposed Hybrid GD-AA-RF model successfully mitigates these issues. It delivers fast convergence towards the target objective functions and efficiently transitions from exploration to exploitation in the initial stages of the optimization process. Consequently, the proposed model exhibits enhanced performance, achieving an accuracy of 99.05%, precision of 98.61%, recall of 99.60%, and an F1-score of 99.10%.

CONCLUSION

This study introduces a Hybrid AdaGrad and AdamW-based Random Forest (RF) model, termed Hybrid GD-AA-RF, to strengthen feature selection and increase the predictive accuracy for heart disease. The framework was evaluated on a combined heart disease dataset comprising 12 attributes and 1190 records, standardized using the Z-score normalization method to enhance performance. Feature selection was achieved using a novel Gaussian-based Differential Entropy (GDE) method with Information Gain, which reduced the feature, set to 9 significant predictors.

During the optimization phase, the hybrid approach combined AdaGrad and AdamW optimizers, leveraging their strengths to achieve faster convergence and balance between exploration and exploitation. AdaGrad adapted the learning rate for infrequent features, while AdamW's weight decay mechanism prevented overfitting, thereby enhancing the overall optimization technique. The RF classifier further refined the optimization by serving as the fitness function, improving the framework's classification accuracy.

The proposed Hybrid GD-AA-RF model demonstrated superior performance, achieving high metrics accuracy (99.05%), precision (98.61%), recall (99.60%) and F1-score (99.10%), which underscore its effectiveness in early and

precise heart disease prediction. The results indicate that this methodology holds significant potential as a reliable and resilient instrument for clinical diagnostic applications within the healthcare domain.

Subsequent work will involve scaling the study to more extensive datasets to further validate the performance and reliability of the Hybrid GD-AA-RF model. Additionally, incorporating image-based data such as echocardiograms or angiograms alongside tabular data will boost the model's accuracy through multimodal learning. This approach aims to create a more comprehensive diagnostic framework, effectively leveraging diverse data modalities to improve accuracy and clinical applicability in cardiovascular disease prediction.

REFRENCES

[1] World Health Organization, 2020. Laboratory testing for 2019 novel coronavirus (2019-nCoV) in suspected human cases: Interim guidance, 17 January 2020. World Health Organization.
[2] G.N. Ahmad, H. Fatima, S. Ullah, A.S. Saidi, and Imdadullah, “Efficient Medical Diagnosis of Human Heart Diseases Using Machine Learning Techniques With and Without GridSearchCV”, IEEE Access, Vol. 10, pp. 80151-80173, 2022.
[3] Alzubaidi, L., Zhang, J., Humaidi, A.J., Al-Dujaili, A., Duan, Y., Al-Shamma, O., Santamaría, J., Fadhel, M.A., Al-Amidie, M. and Farhan, L., 2021. Review of deep learning: concepts, CNN architectures, challenges, applications, future directions. Journal of big Data, 8, pp.1-74.
[4] Louati, H., Bechikh, S., Louati, A., Hung, C.C. and Said, L.B., 2021. Deep convolutional neural network architecture design as a bi-level optimization problem. Neurocomputing, 439, pp.44-62.
[5] Kushwaha, P.K. and Kumaresan, M., 2021, November. Machine learning algorithm in healthcare system: A Review. In 2021 International Conference on Technological Advancements and Innovations (ICTAI) (pp. 478-481). IEEE.
[6] Jinny, S.V. and Mate, Y.V., 2021. Early prediction model for coronary heart disease using genetic algorithms, hyper-parameter optimization and machine learning techniques. Health and Technology, 11(1), pp.63-73.
[7] Nawaz, M.S., Shoaib, B. and Ashraf, M.A., 2021. Intelligent cardiovascular disease prediction empowered with gradient descent optimization. Heliyon, 7(5).
[8] Mohamed, R. and Amany M, S., 2023. A modified Adam algorithm for deep neural network optimization.
[9] Sun, S., Cao, Z., Zhu, H. and Zhao, J., 2019. A survey of optimization methods from a machine learning perspective. IEEE transactions on cybernetics, 50(8), pp.3668-3681.
[10] Sarker, I.H., 2021. Machine learning: Algorithms, real-world applications and research directions. SN computer science, 2(3), p.160.
[11] Kingma, D.P. and Ba, J.L., 2015, May. Adam: A method for stochastic optimization 3rd International Conference on Learning Representations. In ICLR 2015-Conference Track Proceedings (Vol. 1).
[12] Luo, L., Xiong, Y., Liu, Y. and Sun, X., 2019. Adaptive gradient methods with dynamic bound of learning rate. arXiv preprint arXiv:1902.09843.
[13] Ruder, S., 2016. An overview of gradient descent optimization algorithms. arXiv preprint arXiv:1609.04747.
[14] Ginsburg, B., Castonguay, P., Hrinchuk, O., Kuchaiev, O., Lavrukhin, V., Leary, R., Li, J., Nguyen, H., Zhang, Y. and Cohen, J.M., 2019. Stochastic gradient methods with layer-wise adaptive moments for training of deep networks. arXiv preprint arXiv:1905.11286.
[15] W. Li, M. Zuo, H. Zhao, Q. Xu, and D. Chen, “Prediction of coronary heart disease based on combined reinforcement multitask progressive time-series networks”, Methods, Vol. 198, pp. 96-106, 2022.
[16] Nagavelli, U., Samanta, D. and Chakraborty, P., 2022. Machine Learning Technology-Based Heart Disease Detection Models. Journal of Healthcare Engineering, 2022(1), p.7351061.
[17] Bian, K. and Priyadarshi, R., 2024. Machine learning optimization techniques: a Survey, classification, challenges, and Future Research Issues. Archives of Computational Methods in Engineering, pp.1-25.
[18] Breiman, L., 2001. Random forests. Machine learning, 45, pp.5-32.
[19] J. Zhang, “Gradient Descent based Optimization Algorithms for Deep Learning Models Training,” 2019, [Online]. Available: http://arxiv.org/abs/1903.03614.

Research Article

[20] El-Shafiey, M.G., Hagag, A., El-Dahshan, E.S.A. and Ismail, M.A., 2022. A Hybrid GA and PSO optimized approach for heart-disease prediction based on random forest. Multimedia Tools and Applications, 81(13), pp.18155-18179.
[21] Abdollahi, J. and Nouri-Moghaddam, B., 2022. A Hybrid method for heart disease diagnosis utilizing feature selection based ensemble classifier model generation. Iran Journal of Computer Science, 5(3), pp.229-246.
[22] Doppala, B.P., Bhattacharyya, D., Chakkravarthy, M. and Kim, T.H., 2023. A Hybrid machine learning approach to identify coronary diseases using feature selection mechanism on heart disease dataset. Distributed and Parallel Databases, pp.1-20.
[23] Reyad, M., Sarhan, A.M. and Arafa, M., 2023. A modified Adam algorithm for deep neural network optimization. Neural Computing and Applications, 35(23), pp.17095-17112.
[24] Dogo, E.M., Afolabi, O.J., Nwulu, N.I., Twala, B. and Aigbavboa, C.O., 2018, December. A comparative analysis of gradient descent-based optimization algorithms on convolutional neural networks. In 2018 international conference on computational techniques, electronics and mechanical systems (CTEMS) (pp. 92-99). IEEE.
[25] Torthi, R., Marapatla, A.D.K., Mande, S., Gadiraju, H.K.V. and Kanumuri, C., 2024. Heart Disease Prediction Using Random Forest Based Hybrid Optimization Algorithms. International Journal of Intelligent Engineering & Systems, 17(2).
[26] Radosavljević, J., Klimenta, D., Jevtić, M. and Arsić, N., 2015. Optimal power flow using a Hybrid optimization algorithm of particle swarm optimization and gravitational search algorithm. Electric Power Components and Systems, 43(17), pp.1958-1970.
[27] Maddikunta, P.K.R., Gadekallu, T.R., Kaluri, R., Srivastava, G., Parizi, R.M. and Khan, M.S., 2020. Green communication in IoT networks using a Hybrid optimization algorithm. Computer Communications, 159, pp.97-107.
[28] Wang, Y., Zhou, P. and Zhong, W., 2018. An optimization strategy based on Hybrid algorithm of Adam and SGD. In MATEC Web of Conferences (Vol. 232, p. 03007). EDP Sciences.
[29] Sumwiza, K., Twizere, C., Rushingabigwi, G., Bakunzibake, P. and Bamurigire, P., 2023. Enhanced cardiovascular disease prediction model using random forest algorithm. Informatics in Medicine Unlocked, 41, p.101316.
[30] Kalaivani, B. and Ranichitra, A., 2023, June. Unveiling the Impact of Outliers: An Improved Feature Engineering Technique for Heart Disease Prediction. In International Conference on IoT Based Control Networks and Intelligent Systems (pp. 469-478). Singapore: Springer Nature Singapore.
[31] Ruder, S., 2016. An overview of gradient descent optimization algorithms. arXiv preprint arXiv:1609.04747.
[32] Selvaraj, R., Satheesh, T., Suresh, V. and Yathavaraj, V., 2023. Optimized Machine Learning for CHD Detection using 3D CNN-based Segmentation, Transfer Learning and Adagrad Optimization. arXiv preprint arXiv:2305.00411.
[33] Soydaner, D., 2020. A comparison of optimization algorithms for deep learning. International Journal of Pattern Recognition and Artificial Intelligence, 34(13), p.2052013.
[34] Kingma, D.P. and Ba, J.L., 2015, May. Adam: A method for stochastic optimization 3rd International Conference on Learning Representations. In ICLR 2015-Conference Track Proceedings (Vol. 1).
[35] Sun, S., Cao, Z., Zhu, H. and Zhao, J., 2019. A survey of optimization methods from a machine learning perspective. IEEE transactions on cybernetics, 50(8), pp.3668-3681.
[36] Wang, Q., Nguyen, T.T., Huang, J.Z. and Nguyen, T.T., 2018. An efficient random forests algorithm for high dimensional data classification. Advances in Data Analysis and Classification, 12, pp.953-972.
[37] Sannigrahi, M. and Thandeeswaran, R., 2024. Predictive Analysis of Network based Attacks by Hybrid Machine Learning Algorithms utilizing Bayesian Optimization, Logistic Regression and Random Forest Algorithm. IEEE Access.