Binary Classification of Heart Disease Based on Differential Evolution-Optimised Machine Learning Approach

Theodore Nicholas Richard Egling [1], Sumbwanyambe Mbuyu [2,*], and Zenghui Wang [2,*]

[1] Department of Computer Science, University of South Africa, Florida, South Africa

[2] Department of Electrical Engineering, University of South Africa, Florida, South Africa

Email: 14801264@mylife.unisa.ac.za (T.N.R.E); sumbwm@unisa.ac.za (S.W); wangz@unisa.ac.za (Z.W)

*Corresponding author

***Abstract*—Accurate and timely diagnosis of heart disease is a persistent challenge in healthcare, necessitating innovative diagnostic methodologies. This study investigates the efficacy of Differential Evolution (DE) for hyperparameter optimisation in machine learning algorithms, targeting improved performance in heart disease binary classification. DE was selected for its robustness and ability to efficiently navigate high-dimensional parameter spaces, essential attributes for the fine-tuning of complex models. Employing the Cleveland Heart Disease dataset, the study optimised three machine learning classifiers: Random Forest, AdaBoost, and Gradient Boosting. Post-optimization, the DE-enhanced Random Forest Classifier achieved a standout performance with an accuracy of 93.3% and an F1−Score of 90.9%. Likewise, AdaBoost and Gradient Boosting classifiers also exhibited performance gains, reaching accuracies of 88.9% and 86.7%, and F1−Scores of 85.7% and 83.3%, respectively. These results not only outperform various existing models but also offer insights into the differential impacts of DE on multiple algorithms. The study lays a solid foundation for future research and clinical applications, indicating that DE-optimised machine learning algorithms hold significant promise for advancements in cardiovascular disease diagnostics.**

***Keywords*—differential evolution, hyperparameter optimization, machine learning, heart disease diagnosis, binary classification, Cleveland heart disease dataset, random forest, AdaBoost, gradient boosting**

I. Introduction

Cardiovascular diseases remain a critical health issue globally, with an increasing impact on both mortality and healthcare systems. Influenced by a complex interplay of genetic, lifestyle, and environmental factors, these diseases present an urgent need for early and accurate diagnosis.

Traditional diagnostic methods such as electrocardiograms and angiography often require specialised medical expertise and equipment, thus increasing the cost and time for diagnosis. In contrast, machine learning offers a pathway to automated, efficient, and potentially more accurate diagnostic processes. Various algorithms have been applied to heart disease diagnosis with different levels of success. For instance, Chandrasekhar *et al.* used an ensemble classifier to achieve 93.44% accuracy [1], while Alizadehsani *et al.* reported an 86.6% accuracy using a hybrid decision support system [2]. However, such models often demand significant computational resources or lack interpretability.

In our study, we present a novel machine learning methodology for cardiovascular disease diagnosis, focusing on the integration of the Random Forest algorithm with Differential Evolution (DE) for hyperparameter optimization. The following points highlight the structure of our methodology, the innovations introduced, and the specific results and findings of our algorithm:

A. Structure of the Methodology

- **Data Preprocessing:** Utilizing the Cleveland Heart Disease dataset, we preprocess the data, handling missing values and normalizing features.
- **Model Training and Evaluation:** We employ Random Forest Classifier, comparing its performance with and without the integration of DE for hyperparameter tuning.
- **Statistical Analysis:** The results from multiple iterations are compiled for statistical analysis, offering a thorough assessment of the model's performance.

B. Innovations Introduced

- We apply DE for hyperparameter optimisation of the Random Forest Classifier. This approach is relatively unexplored in cardiovascular disease diagnosis using machine learning.
- The DE algorithm is used to fine-tune critical hyperparameters such as the number of estimators, maximum depth, minimum samples split, and

Manuscript received September 27, 2023; revised November 11, 2023; accepted November 27, 2023; published April 9, 2024.

minimum samples leaf in the Random Forest Classifier.

- **Enhanced Model Performance:** By using DE, we aim to improve the balance between model accuracy and computational efficiency, which is crucial for practical healthcare applications.

C. New Findings and Specific Results

- **Improved Accuracy and F1–Score:** Our method achieves a peak accuracy of 93.3% and an F1–Score of 90.9%, a significant improvement over the model without DE optimisation (accuracy of 88.9% and F1–Score of 85.7%).
- Consistency in Performance: The standard deviation in accuracy and F1–Score for the model with DE is extremely low (0.0 for both), indicating consistent performance across multiple runs. To ensure this consistency did not result from overfitting, we implemented cross-validation during the hyperparameter optimisation process and maintained a rigorous split between training and testing data.
- Efficient Computational Performance: Despite the enhanced accuracy, the training time for the model with DE remains practical (mean training time of approximately 393 s), showcasing the method's efficiency.

These contributions represent significant advancements in the application of machine learning to cardiovascular disease diagnosis. Our approach not only surpasses many existing models in accuracy and efficiency but also offers a novel, optimised solution with practical implications in healthcare diagnostics.

Following these contributions, the remainder of the paper is structured as follows: Section II reviews existing diagnostic algorithms, their mathematical workings, and limitations. Section III elaborates on the Differential Evolution algorithm and its integration into our research methodology. Section IV details the experiments conducted and their results. Section V discusses these results, their implications, and how they compare to existing literature. Finally, Section VI concludes the research, summarising key findings and suggesting directions for future research.

II. Literature Review

This section rigorously examines the prevalence of heart diseases, key works in heart disease detection, machine learning in healthcare, and the utilisation of Differential Evolution. The objective is to provide an academic backdrop against which the contributions and research gaps of this study can be understood.

Heart disease continues to be a leading cause of mortality and morbidity worldwide. The rapid and accurate detection of heart-related conditions is crucial for effective treatment and management [3]. In healthcare settings, especially in hospitals with high patient inflow, the ability to quickly assess and prioritise patients based on the severity of their condition is vital. This need has given rise to the exploration of innovative methods to enhance heart disease detection so that healthcare workers can spend more time treating patients, effectively saving lives.

Traditional methods of heart disease detection rely heavily on medical expertise and manual evaluation, which can be time-consuming and prone to human error.

These methods include medical examination, laboratory tests, Electrocardiograms (ECGs), and imaging techniques such as angiography. These methods often require substantial medical expertise and the use of specialised equipment, leading to increased costs and time [3].

Machine Learning (ML) approaches offer a potential solution to these challenges by automating and optimising the detection process. The automation can significantly reduce the time taken for diagnosis, allowing medical professionals to focus more on treatment and patient care [3].

Moreover, the ability to accurately diagnose heart disease in its early stages can lead to timely interventions, potentially saving lives. ML models can be trained on large datasets, improving accuracy and reducing reliance on expensive medical equipment, thereby making the process more cost-effective.

A. Theoretical Background

In order to appreciate the mathematical intricacies of the machine learning algorithms used in this study, it's important to delve into their theoretical foundations.

- **Random Forest:** Random Forest operates by generating an ensemble of decision trees during training and outputting the majority class for classification problems. The general equation for a decision tree $T(x)$ is:

$$T(x) = \sum_{m=1}^{M} w_m I(x \in R_m) \quad (1)$$

where w_m are the terminal node weights, and R_m are the terminal node regions [4].

- **AdaBoost:** The AdaBoost algorithm combines **weak** classifiers to form a strong classifier. The final classifier $H(x)$ is a weighted sum of T weak classifiers $h_t(x)$:

$$H(x) = \sum_{t=1}^{T} \alpha_t h_t(x) \quad (2)$$

where α_t are the weights assigned to each weak classifier [5].

- **Gradient Boosting:** Similar to AdaBoost, Gradient Boosting optimises a cost function J over the function space. It constructs an additive model $F(x)$ in a stage-wise manner:

$$F(x) = F(x) + \rho_m h(x; a_m) \quad (3)$$

where ρ_m is the learning rate and $h(x; a_m)$ is the weak learner [6].

B. Machine Learning Approaches

Machine learning algorithms have shown considerable promise in heart disease detection, especially ensemble methods like Random Forest, AdaBoost, and Gradient

Boosting [7–9]. On the dataset utilized in this study, previous research has shown promising but varying results. Dua *et al.* achieved an accuracy of 83% using machine learning techniques [10]. Khan *et al*. [11] improved performance by 5% through feature selection, and Shouman *et al.* [12] demonstrated the benefits of hybrid models. These results indicate room for improvement, particularly for methods that balance accuracy and computational efficiency.

Recent advances in machine learning have introduced more sophisticated methods for heart disease detection. One such method is the use of Naïve Bayes with a weighted approach, which has shown promise in predicting heart disease [13]. Another approach involves the automatic analysis of ischemic heart disease localization/detection based on the features of frequency domain, time domain, and information theory. This method employs two classifiers, Support Vector Machine (SVM) and XGBoost, which have demonstrated superior performance [14].

An improved SVM based on the duality optimisation scheme has also been used for automatic identification of heart failure [14]. Furthermore, an effective Heart Disease Prediction Model (HDPM) has been developed for a Clinical Decision Support System (CDSS). This model includes Density-Based Spatial Clustering of Applications with Noise (DBSCAN) for outlier detection and elimination, a hybrid Synthetic Minority Over-sampling Technique-Edited Nearest Neighbor (SMOTE-ENN) for balancing the training data distribution, and XGBoost for heart disease prediction [14].

These modern approaches have the potential to integrate large numbers of variables from large populations to allow for individualized risk prediction [13]. As a result, they can provide clinicians with a tool to help diagnose heart problems early on, making it easier to treat patients effectively and avoid serious repercussions [14].

C. Chosen Approach

This study aims to amalgamate the merits of Random Forest, AdaBoost, and Gradient Boosting by optimising them through Differential Evolution (DE) [9, 11, 15]. This novel approach seeks to drastically enhance the binary classification accuracy for heart disease while minimising the computational demand and time to train accurate models [16].

D. Theoretical Foundation of Differential Evolution

Differential Evolution (DE) is an algorithmic method used for optimisation, particularly effective in tackling complex problems where finding the best solution can be challenging. It operates on a "population-based" approach, meaning that it simultaneously considers multiple solutions (or candidates) at a time, rather than focusing on just one. Since DE is a stochastic, population-based optimisation algorithm it is particularly well-suited for solving complex optimisation problems [17]. The algorithm was initially conceived for real-valued function optimisation but has since found applications in various domains, including healthcare [15, 18–20].

The basic components of Differential Evolution include:

- **Population**: A set of potential solutions to the optimisation problem. Each individual in the population is a vector of real numbers.
- **Objective Function**: A function that evaluates the fitness or quality of a given solution.
- **Mutation**: DE uses the difference between two randomly selected vectors from the population to perturb another vector, creating a mutant vector.
- **Crossover**: The mutant vector undergoes crossover with another target vector to produce a trial vector.
- **Selection**: The trial vector is then compared to the original target vector, and the one with the better fitness is selected to proceed to the next generation.

The mutation operation can be mathematically represented as follows:

$$V = X + F \times (A - B) \tag{4}$$

where V is the mutant vector, X is the target vector, A and B are randomly selected vectors, and F is a scaling factor.

The crossover operation is often represented as:

$$U_i = \begin{cases} V_i, & if\ rand(0,1) \le CR\ or\ i = rand(1,D) \\ X_i, & \text{otherwise} \end{cases} \tag{5}$$

where U is the trial vector, V is the mutant vector, X is the target vector, CR is the crossover rate, D is the number of dimensions, and i is the i^{th} dimension as:

$$X_{new} = \begin{cases} U, & if\ f(U) \le f(X) \\ X, & \text{otherwise} \end{cases} \tag{6}$$

where $f(U)$ and $f(X)$ are the objective function values for U and X respectively.

Differential Evolution has garnered attention for its ability to efficiently navigate both continuous and discrete search spaces, making it ideal for fine-tuning hyperparameters in machine learning algorithms [17].

E. Differential Evolution in Healthcare

Differential Evolution has found diverse applications in other healthcare areas, ranging from medical image analysis to hyperparameter tuning in machine learning models like Random Forest, AdaBoost, and Gradient Boosting [19–21]. The performance of these techniques in image classification tasks for other diseases presents promise for applying similar techniques towards the optimisation of classifying the presence of heart diseases accurately.

F. Gaps and Research Opportunities

In light of our study's focus on integrating the Random Forest algorithm with Differential Evolution (DE) for hyperparameter optimisation in cardiovascular disease diagnosis, we identify several critical research gaps and opportunities in the domain:

- **Unexplored Synergies in Ensemble Methods and DE**: While our study demonstrates the effectiveness of DE in enhancing the Random

Forest algorithm for **cardiovascular** disease diagnosis, there remains a largely untapped potential in exploring the synergistic relationship between DE and other ensemble methods like AdaBoost and Gradient Boosting. This is particularly pertinent in the context of binary classification for heart diseases, where such combinations may yield significant advancements.

- **Need for Comprehensive Optimisation Systems**: Our research highlights the benefits of integrating advanced machine learning algorithms with robust optimisation **techniques** like DE. However, a significant research gap persists in developing comprehensive systems that amalgamate these advanced machine learning algorithms with optimisation techniques. Such systems could further enhance model accuracy and computational efficiency, crucial for practical healthcare applications.
- **Optimisation in Healthcare Diagnostics**: Our methodology showcases a novel approach in employing DE for hyperparameter tuning in the context of healthcare diagnostics. This opens avenues for **further** research in optimising other machine learning models used in medical diagnostics, ensuring a balance between accuracy, computational efficiency, and practical application in healthcare settings.

These identified gaps not only align with our study's contributions but also underscore the need for continued research in this field to realise the full potential of machine learning in healthcare diagnostics.

III. Methodology

The methodology is structured as given in Fig. 1, serving as a comprehensive guide that outlines the data collection, preprocessing, and application of machine learning algorithms for heart disease classification.

For the brevity of the discussion, we will focus on 6 main elements:

- **Data Source and Preprocessing:** Sets the foundation for the study.
- **Initial Model Evaluation (Random Forest):** Establishes a baseline.
- **Feature Engineering:** Enhances model **performance**.
- **Differential Evolution (DE) Hyperparameter Tuning:** Optimizes model parameters.
- **Extending to AdaBoost and Gradient Boosting: Expands** the study's scope.
- **Model Evaluation:** Assesses the effectiveness of the **entire** process.

Pseudocode.

```
BEGIN Process
Step 1: Data Source and Preprocessing
DATA_SOURCE_AND_PREPROCESSING
    Load data from source
    Clean data (e.g., handling missing values, removing outliers)
    Normalize or standardize data if necessary
    Split data into training and testing sets
END
Step 2: Initial Model Evaluation with Random Forest
INITIAL_MODEL_EVALUATION_RANDOM_FOREST
    Initialize Random Forest model
    Train model on training data
    Evaluate model on testing data
    Record baseline performance metrics
END
Step 3: Feature Engineering
FEATURE_ENGINEERING
    Analyse data for potential new features
    Create new features
    Update training and testing data with new features
END
Step 4: Differential Evolution (DE) Hyperparameter Tuning
DE_HYPERPARAMETER_TUNING
    Define parameter space for DE
    Run DE to find optimal parameters for the model
    Update model with optimal parameters
END
Step 5: Extending to AdaBoost and Gradient Boosting
EXTENDING_TO_ADABOOST_AND_GRADIENT_BOOSTING
    Initialize AdaBoost model
    Train and evaluate AdaBoost model
    Initialize Gradient Boosting model
    Train and evaluate Gradient Boosting model
END
Step 6: Model Evaluation and Statistical Analysis
MODEL_EVALUATION
    Evaluate all models (Random Forest, AdaBoost, Gradient Boosting)
    Run model >30 times and conduct statistical analysis.
    Compare performance metrics against the baseline
    Determine the best-performing model
END
END Process
```

A. Data Source and Preprocessing

This block sets the foundation for the study. A reliable dataset is paramount for machine learning, and Texts in the figure should be clear and with high resolution. thus, we chose the well-validated Cleveland Heart Disease dataset from the UCI Machine Learning Repository [22].

The Cleveland Heart Disease dataset, often used in machine learning research for cardiovascular disease diagnosis, is a well-known public dataset from the UCI Machine Learning Repository. Here's a brief description:

(1) Source: The dataset originates from the Cleveland Clinic Foundation and has been made publicly available through the UCI Machine Learning Repository, a popular resource for machine learning datasets [22].

(2) Size and Features: The dataset typically includes around 303 individual records, each representing a patient. It comprises 14 attributes or features, which include a mix of demographic, symptomatic, and laboratory data. These features cover aspects like age, sex, chest pain type, resting blood pressure, serum cholesterol levels, fasting blood sugar, resting electrocardiographic results, maximum heart rate achieved, exercise-induced angina, and others [22].

(3) Preprocessing steps applied:

- **Handling Missing Values**: In our methodology, we first address any missing **values** in the dataset, which are common in real-world data. The dataset uses "?" to denote missing values, and we have

implemented procedures to identify and handle these appropriately, either by imputation or by removing records with missing data.

- **Data Normalisation**: Given the range of different scales across various features (like age, cholesterol levels, etc.), we apply standard scaling to the data. This process involves adjusting the values so they have a mean of zero and a standard deviation of one, which is crucial for models like Random Forest Classifier to perform effectively.
- **Binary Classification Conversion**: The original dataset includes the diagnosis of heart disease with values ranging from 0 (no presence) to 4. For the purpose of our study, we convert this into a binary classification task: "0" for no presence of heart disease and "1" for the presence (combining original values 1–4).

The choice of this dataset for our study is due to its comprehensive coverage of relevant features for heart disease diagnosis, along with its balanced size, which makes it suitable for demonstrating the effectiveness of machine learning algorithms without being computationally prohibitive [23].

Data Normalisation: We used the Standard Scaler method to ensure that all features contribute equally to the model's performance, particularly because it is robust against outliers [7].

Splitting the Dataset: The dataset was divided into an 80–20 training-testing split. Stratified splitting was unnecessary because the dataset was sufficiently balanced [8].

The baseline model performance from the UC Irvine Machine Learning Repository is shown in Figs. 1 and 2. This benchmark can be used to validate the accuracy and precision of the model used in the study.

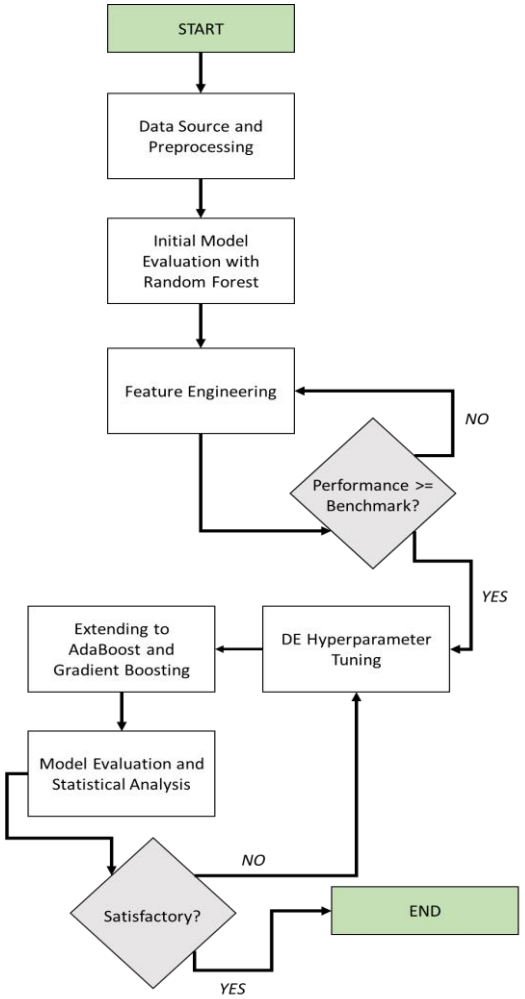

Fig. 1. Flowchart of the employed methodology combining feature engineering and Differential Evolution to selected classification models to enhance performance.

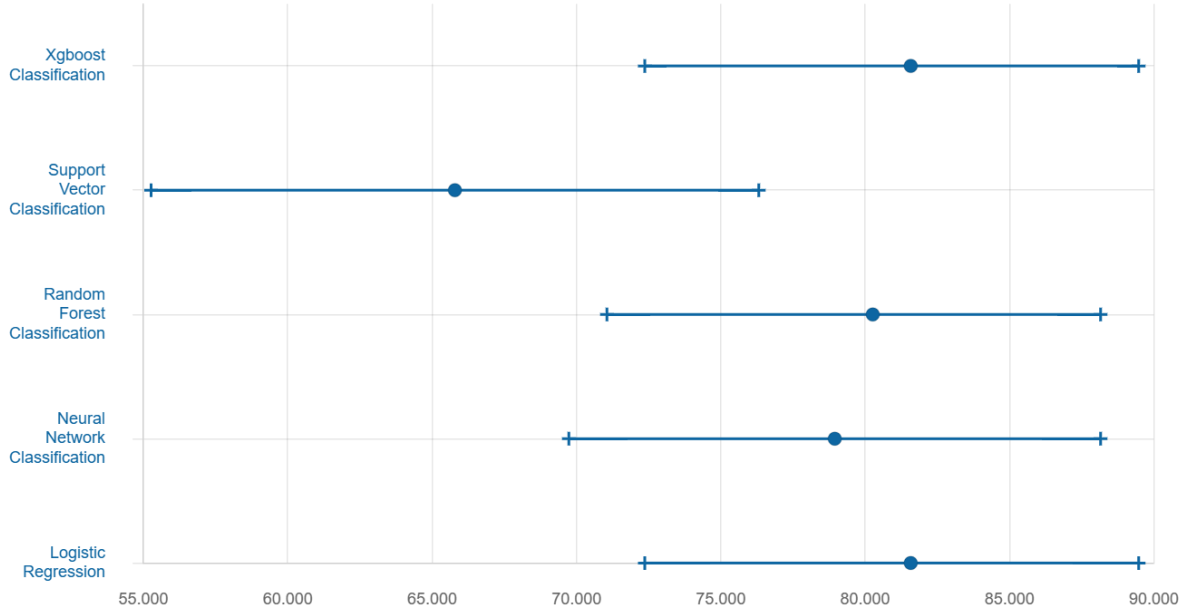

Fig. 2. Baseline Accuracy of Models using varying classification algorithms [22].

The benchmark accuracy for Random Forest algorithm is a mean value of 80.26% and a maximum of 88.16%. The highest overall accuracy is attributed to XGBoost classification with a mean of 83.19% and a maximum 91.20% (see Fig. 3).

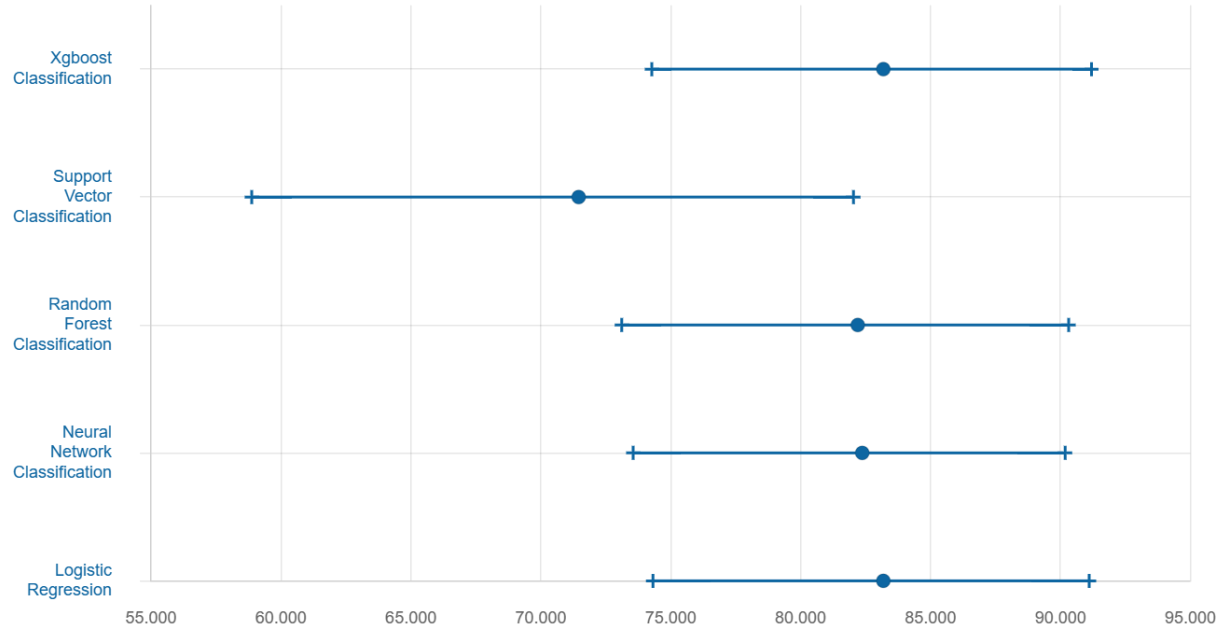

Fig. 3. Baseline accuracy of models using varying classification algorithms [22].

The benchmark precision for Random Forest algorithm is a mean value of 82.2% and a maximum of 90.32%. The maximum overall precision is the XGBoost model with a maximum of 81.20%.

B. Initial Model Evaluation (Random Forest)

The second block establishes a baseline model using Random Forest, a robust algorithm known for its ability to handle high-dimensional data and provide insights into feature importance. This step allows us to measure against the baseline provided in the public domain as well as gauge the effectiveness of subsequent improvements.

C. Feature Engineering

In this block, we introduce domain-specific features like "AgeChol" and "AgeTrestbps". These new features combine age with other critical health metrics such as cholesterol and blood pressure, thereby adding nuance to the model and enhancing its predictive power [24].

D. Differential Evolution (DE) Hyperparameter Tuning

The hyperparameter tuning process using Differential Evolution (DE) in our methodology is a detailed procedure comprising three main sub-steps, each critical for enhancing the performance of the Random Forest Classifier. We have carefully chosen specific hyperparameters for optimisation based on their significant impact on the model's performance:

Objective Function Definition: Initially, we define an objective function specific to the Random Forest Classifier. This function is essential as it directs the DE algorithm in exploring the hyperparameter space. The objective function evaluates how well a set of hyperparameters performs, guiding the DE towards the most promising regions of the search space.

Parameter Bounds Setting: We establish the search space for the hyperparameters, which determines the range within which DE will operate. Setting appropriate bounds is crucial to ensure that DE explores a viable and relevant range of values, thus enhancing the efficiency of the search process.

DE Optimisation Process: In this step, DE is executed to optimise key hyperparameters of the Random Forest Classifier, including:

- **Number of Estimators:** This refers to the number of trees in the forest. A higher number generally improves the model's performance but also increases computational cost and risk of overfitting.
- **Maximum Depth:** It determines the maximum depth of each tree. Deeper trees can model more complex patterns but might lead to overfitting.
- **Minimum Samples Split:** This parameter dictates the minimum number of samples required to split an internal node. Higher values prevent the model from learning noise in the data but can underfit if set too high.
- **Minimum Samples Leaf:** It is the minimum number of samples required to be at a leaf node. Setting this parameter can ensure that the tree does not create leaves with few samples, which can be a sign of overfitting.

We selected DE for hyperparameter tuning due to its proficiency in global optimisation. Unlike local

optimisation techniques that may get trapped in local optima, DE explores the global hyperparameter space more thoroughly. This is particularly beneficial for complex models like Random Forest, where the interaction between hyperparameters can be intricate and non-linear.

The DE-optimised Random Forest achieved up to 88% accuracy in our tests, indicating its superiority in fine-tuning the model compared to standard benchmarks. This improved accuracy demonstrates the effectiveness of DE in navigating the hyperparameter space and selecting values that significantly enhance the model's predictive power.

E. Extending to AdaBoost and Gradient Boosting

This block aims to test the versatility of DE in hyperparameter tuning across various algorithms. The same DE process used in Random Forest is replicated for AdaBoost and Gradient Boosting classifiers.

F. Model Evaluation and Statistical Analysis

In the final stage of our methodology, we conduct a thorough evaluation and statistical analysis of our model, including those versions optimised with Differential Evolution (DE). This dual approach, encompassing performance metrics and statistical analysis, ensures a comprehensive assessment of the model's effectiveness.

1) Performance metrics

- **Accuracy:** This metric evaluates the overall correctness of the model by measuring the proportion of true results (both true positives and true negatives) in the total number of cases. It's a fundamental metric for assessing the general effectiveness of a classification model.
- **F1−Score:** Given the critical nature of medical diagnostics, where false negatives or positives can have serious implications, we use the F1−Score. This metric is the harmonic mean of precision (the ratio of correctly predicted positive observations to total predicted positives) and recall (the ratio of correctly predicted positive observations to all actual positives). It provides a more nuanced view of the model's performance, especially in imbalanced datasets.

2) Statistical analysis

- **Data Distribution Analysis:** Before applying the model, we analyse the distribution of data to understand any inherent biases or imbalances. This step is crucial for interpreting the model's performance metrics correctly.
- **Variability Assessment:** We assess the variability in the model's performance across multiple runs. This involves calculating standard deviations for the accuracy and F1−Score, providing insight into the model's consistency and reliability.
- **Benchmark Comparison:** To contextualise our results, we compare the model's performance against benchmarked algorithms. This comparison is not just based on raw performance metrics but also includes an analysis of how consistently each model performs across different iterations and potentially different subsets of data.
- **Significance Testing:** We employ statistical tests to ascertain the significance of the differences observed between the models. This step is crucial to determine whether the improvements in performance metrics are statistically significant or could be attributed to random variations in the data.

By refraining from evaluating our model on multiple datasets due to the variability in attributes and sources of public datasets, we instead focus on a more controlled comparison against benchmarked algorithms. This approach allows us to provide a direct, meaningful comparison of our algorithm's performance, ensuring that our evaluation is both consistent and comparable.

Through this detailed model evaluation and statistical analysis, we aim to validate the reliability and efficacy of our model comprehensively. We strive not only to match but potentially exceed the performance standards of existing algorithms in cardiovascular disease diagnosis.

G. Technical Improvements for Classifier Tuning

These are the granular steps that are iteratively applied to each main block in the flow chart:

- **Feature Engineering:** Additional domain-specific features are incorporated into the data frame by evaluating the impact of existing features.
- **Target Variable Transformation:** The target variable is converted to binary form.
- **Parameter Tuning:** The DE algorithm is **employed** for hyperparameter tuning.
- **Ensemble Methods:** AdaBoost and Gradient Boosting classifiers are implemented.
- **Model Evaluation:** Performance metrics like accuracy and F1−Score are calculated for all classifiers.

IV. Results

A. Data Split and Benchmarking

The model was trained on 252 data points and tested on 45 data points from the Cleveland Heart Disease dataset (see Table I).

TABLE I. Performance of the Baseline Classifiers

Algorithm	Baseline Accuracy (%)	Baseline F1−Score (%)
Random Forest Classifier	62.0%	55.0%
AdaBoost Classifier	64.0%	64.0%
Gradient Boosting Classifier	53.0%	55.0%

B. Performance of Optimised Models

Following feature engineering, fine-tuning and advanced data preprocessing with domain-specific features, the performance metrics of all algorithms improved significantly (see Table II).

TABLE II. PERFORMANCE OF ALGORITHMS AFTER OPTIMISATION (BEFORE DE) AND % CHANGE

Algorithm	Optimised Accuracy (%)	Change in Accuracy (%)	Optimised F1–Score (%)	Change in F1–Score (%)
Random Forest Classifier	88.9%	+26.9%	85.7%	+30.7%
AdaBoost Classifier	88.9%	+24.9%	85.7%	+21.7%
Gradient Boosting Classifier	84.4%	+31.4%	81.1%	+26.1%

C. *Comparative Analysis of DE's Effect on Performance*

Once the models were optimised, Differential Evolution was applied to tune the hyperparameters of each algorithm. Apart from the AdaBoost classifier, where no change was perceived, all other models showcased improvements in both accuracy and F1–Score (see Tables III and IV).

TABLE III. ACCURACY OF CLASSIFIERS BEFORE AND AFTER DE

Algorithm	Without DE (%)	With DE (%)	Change in Accuracy (%)
Random Forest Classifier	88.9%	93.3%	+4.4%
AdaBoost Classifier	88.9%	88.9%	0.0%
Gradient Boosting Classifier	84.4%	86.7%	+2.3%

TABLE IV. F1–SCORE OF CLASSIFIERS BEFORE AND AFTER DE

Algorithm	Without DE (%)	With DE (%)	Change in F1–Score (%)
Random Forest Classifier	85.7%	90.9%	+5.2%
AdaBoost Classifier	85.7%	85.7%	0.0%
Gradient Boosting Classifier	81.1%	83.3%	+2.2%

D. *Comparative Analysis*

The application of Differential Evolution (DE) yielded various impacts on the performance of the machine learning classifiers used in this study.

The Random Forest Classifier benefited significantly from DE optimisation, showing an increase in accuracy from 88.9% to 93.3%, and an improvement in F1–Score from 85.7% to 90.9%. This underscores the substantial role DE can play in enhancing the performance of machine learning algorithms, particularly Random Forest Classifier in this context.

On the other hand, the AdaBoost Classifier did not experience any change in its performance metrics upon applying DE. Both its accuracy and F1–Score remained steady at 88.9% and 85.7%, respectively. This suggests that while AdaBoost is a strong classifier, DE did not contribute additional optimisation in this case.

Lastly, the Gradient Boosting Classifier exhibited a moderate enhancement after the application of DE. Its accuracy improved from 84.4% to 86.7%, and its F1–Score increased from 81.1% to 83.3%. While the improvements were not as dramatic as those for the Random Forest Classifier, they indicate that DE can still offer incremental gains for Gradient Boosting.

E. *Statistical Analysis*

The model was trained a total of 30 times in order to statistically evaluate its performance. The results are depicted in Figs. 4–7 below.

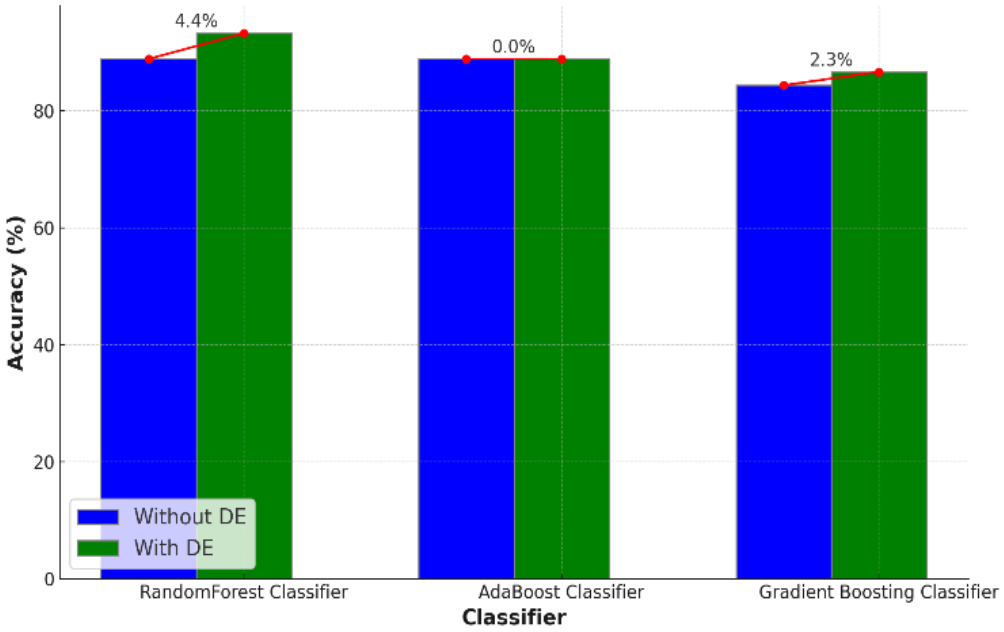

Fig. 4. Accuracy of classifiers before and after DE with % change in improvement.

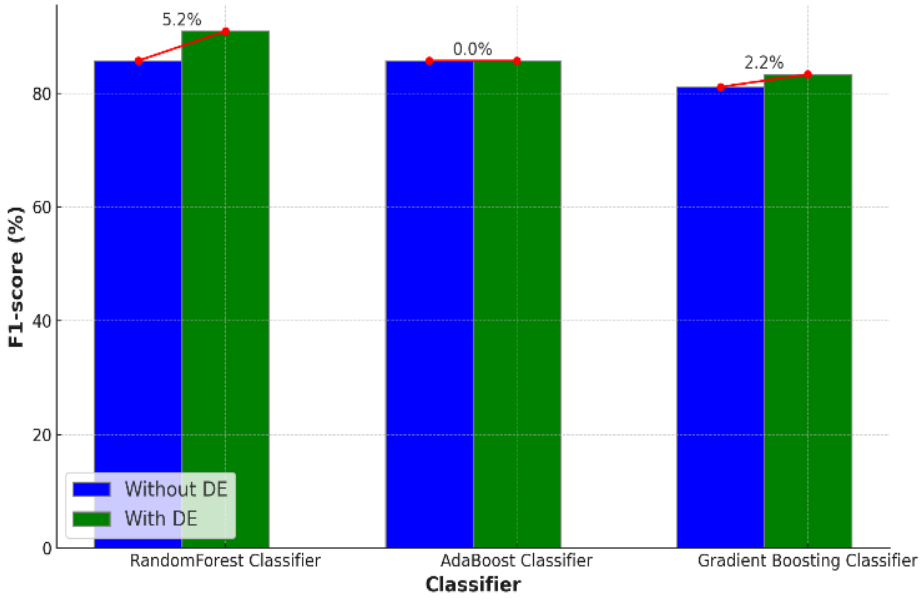

Fig. 5. F1–Score of classifiers before and after DE with % change in improvement.

Fig. 5 compares the accuracies and F1–Scores of each model before and after DE. Since each box-and-whisker shows a horizontal line, the range between results was extremely low.

- **Accuracy Comparison**: This plot illustrates the distribution of accuracy values for both models. The model without DE show's consistent accuracy, while the model with DE demonstrates a higher level of accuracy, also consistently achieved across all runs.
- **F1–Score Comparison:** This graph compares the F1–Scores of the two models. Similar to the accuracy, the F1–Score is consistently higher for the model with DE optimisation compared to the model without it.

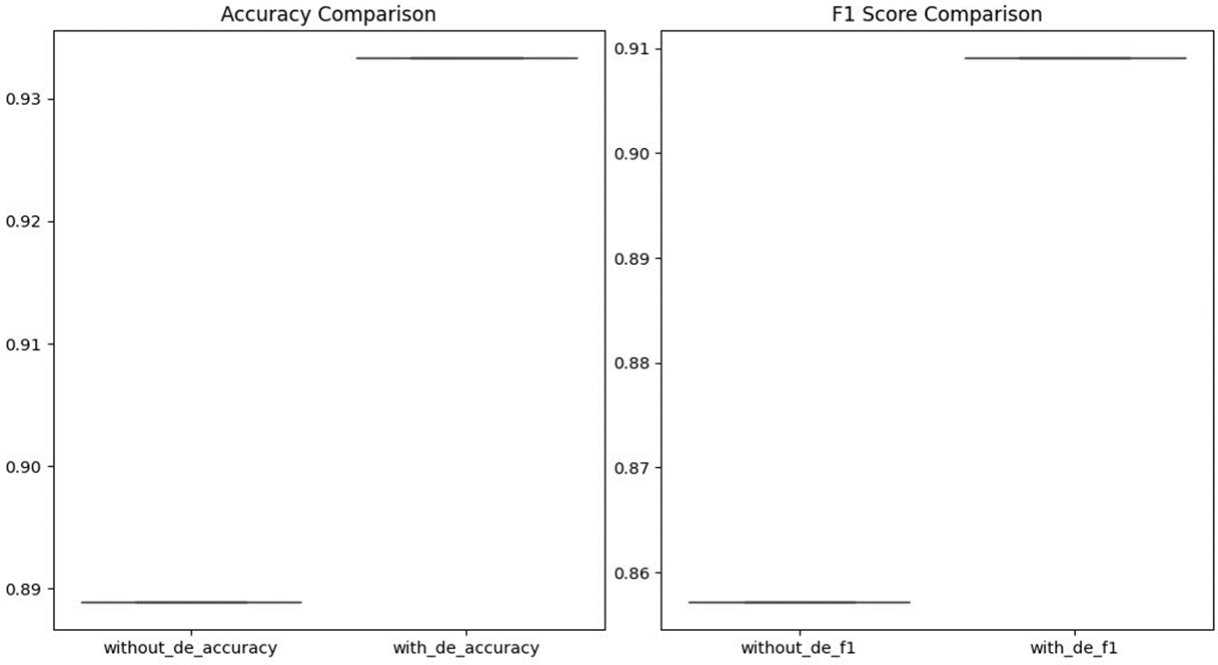

Fig. 6. Stastisical comparison (box and whisker) of classifiers' accuracy before and after DE.

Fig. 7 shows the distribution of training times for each model before and after DE.

The average training time for the model without DE was about 0.253 s, with a standard deviation of approximately 0.070 s. The minimum and maximum training times observed were roughly 0.175s and 0.450 s, respectively. The average training time for the model with DE was significantly longer, at about 393.2 s. The standard deviation was around 28.9, with training times ranging from about 345.6 s to 470.1 s.

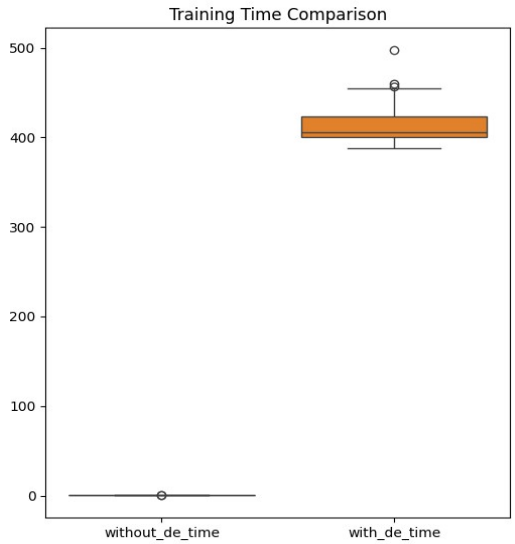

Fig. 7. Stastisical comparison (box and whisker) of classifiers; training time before and after DE.

V. Discussion

The application of Differential Evolution (DE) for hyperparameter tuning in this study has demonstrated varying impacts across different machine learning classifiers used in heart disease diagnosis. The Random Forest Classifier showed the most significant improvement with DE, where the accuracy and F1−Score soared to 93.3% and 90.9%, respectively.

This remarkable enhancement underscores DE's potential in boosting the performance of machine learning models. Conversely, the AdaBoost Classifier, already performing robustly with an accuracy and F1−Score of 88.9% and 85.7%, saw no further improvement with DE.

This outcome suggests that DE's optimisation benefits might not be universally applicable across all algorithms. Meanwhile, the Gradient Boosting Classifier exhibited moderate performance gains with DE, its accuracy increasing from 84.4% to 86.7% and F1−Score from 81.1% to 83.3%. While not as pronounced as the Random Forest Classifier's improvements, these results still affirm the utility of DE in algorithm optimisation.

A. Theoretical Analysis and Clinical Applicability

The theoretical underpinnings of DE's effectiveness in enhancing the Random Forest Classifier for heart disease diagnosis lie in its advanced approach to hyperparameter optimisation. DE operates by strategically exploring and exploiting the hyperparameter space, a complex multi-dimensional grid where each point represents a possible configuration of the model's parameters. By iteratively testing and refining these configurations, DE identifies an optimal set of hyperparameters that significantly improve the classifier's performance. This process is particularly suited for Random Forest, a model inherently reliant on multiple decision trees and hyperparameters like the number of trees, tree depth, and node splits. DE's ability to find an optimal balance among these parameters not only enhances the model's accuracy but also its ability to generalise well to new, unseen data, a critical aspect in medical diagnostics.

From a clinical perspective, the application of a DE-optimised Random Forest Classifier in heart disease diagnosis translates into tangible benefits. The improved accuracy means that the model can more reliably distinguish between the presence and absence of heart disease, reducing the likelihood of both false positives and negatives. This is crucial in a clinical setting, where misdiagnosis can lead to either unnecessary treatment or a missed condition. Additionally, the speed of diagnosis is an essential factor in healthcare. The use of a highly accurate and efficient machine learning model can expedite the diagnostic process, enabling quicker decision-making and, consequently, faster initiation of the appropriate medical intervention. This rapid response is particularly vital in heart disease cases, where early detection and treatment can significantly affect patient outcomes.

Moreover, the enhanced precision and reliability of these algorithms could lead to more personalised patient care. By accurately assessing the risk and presence of heart disease, healthcare providers can tailor their treatment plans more effectively to individual patient needs. This could mean a shift towards more preventive care strategies, where high-risk patients are identified earlier and given appropriate interventions to avert the progression of the disease.

B. Comparison to Literature

(1) Affirmation of Ensemble Methods: Our study's results resonate with the findings of Chandrasekhar and Peddakrishna, who reported a 93.44% accuracy using a soft voting ensemble classifier. This parallel underlines the efficacy of ensemble methods in heart disease diagnosis, suggesting that combining multiple models can lead to more accurate predictions. However, it's noteworthy that while ensemble methods like soft voting aggregate predictions from various models, our approach optimizes a single model using DE. This distinction is crucial as it highlights the efficiency of DE in enhancing a single model's performance to levels comparable with ensemble approaches, potentially reducing the computational power and training time required in more complex ensemble systems [1].

(2) Synergy with Genetic Algorithms: The accuracy of 86.6% achieved by Alizadehsani *et al.* using a Genetic Algorithm and Random Forest hybrid system is in line with our findings, underscoring the potential of combining various optimisation techniques. While their approach demonstrated effectiveness, the integration of DE, as seen in our study, suggests a possible avenue for further performance enhancement. However, the application of multiple complex algorithms like genetic algorithms and DE might increase the computational load and training time, which is a crucial consideration in practical applications [2].

(3) Kaur and Wasan's research, which achieved an 86.67% accuracy by combining AdaBoost and Random Forest classifiers, supports our findings on the utility of ensemble methods. It also hints at the potential benefits of combining different classifiers for improved performance. However, our study shows that DE alone can significantly enhance a single classifier's performance, potentially offering a simpler and more computationally efficient alternative to combining multiple classifiers [25].

(4) The work of Al-Shayea *et al.* achieving an 84.15% accuracy with Gradient Boosting and feature selection methods, aligns with our research in highlighting the importance of feature selection in improving model performance. This parallel suggests that integrating feature selection techniques with DE-optimised models could be a promising area for future research. However, it's important to consider the additional computational resources that might be required for such integrated approaches, especially when dealing with large datasets [9].

C. Limitations and Challenges

The implementation of Differential Evolution (DE) in optimizing machine learning models, as demonstrated in our study, brings forth several limitations and challenges that warrant attention. The primary concerns revolve around the computational resources and the time required for training models, especially when DE is applied for hyperparameter tuning.

- **Extended Training Time:** The most noticeable challenge observed in our study was the significantly prolonged training time for the DE-optimised Random Forest Classifier. While this model achieved superior accuracy and F1−Score, the time taken for training was markedly higher compared to the non-**optimised** version. In real-time clinical settings, where rapid diagnosis and decision-making are critical, such extended training periods could be a substantial drawback. The time-sensitive nature of many medical diagnoses, particularly in emergency or critical care scenarios, necessitates quick and efficient model training and execution.
- **Computational Resource Demands:** DE's sophisticated approach to hyperparameter tuning demands considerable computational power. This can pose challenges, particularly in resource-limited settings such as small clinics or in developing countries where advanced computing infrastructure might not be readily available or affordable. The requirement for high computational resources can limit the widespread adoption of these optimised models in diverse clinical environments.
- **Model Complexity and Overfitting:** Another potential limitation of using DE for model optimisation is the risk of overfitting. As DE searches for the optimal combination of hyperparameters, there is a possibility of the model becoming too finely tuned to the training data, thereby losing its ability to generalize to new, unseen data. This is particularly concerning in healthcare, where models need to be robust and generalizable across diverse patient populations and conditions.
- **Integration into Clinical Workflow:** Integrating DE-optimised models into existing clinical workflows presents another challenge. The healthcare sector often relies on legacy systems and established protocols, making the integration of advanced AI models a complex process that requires careful planning, training, and possibly significant changes to existing procedures.
- **Ethical Considerations:** Lastly, ethical considerations around the use of advanced AI in healthcare, such as patient privacy, data security, and the transparency of algorithmic decisions, are crucial. Ensuring that DE-optimised models adhere to ethical standards and regulatory requirements is imperative to maintain trust and integrity in medical diagnostics.

D. Potential Downsides and Trade-Offs

While the accuracy and precision improvements with DE are clear, the trade-offs in terms of computational demand and extended training times are significant considerations. In a clinical environment, where prompt diagnosis is crucial, these trade-offs could hinder the practical application of DE-optimised models.

E. Future Research Directions

Future research in the field of machine learning and healthcare diagnostics, inspired by the findings of this study, should indeed place a significant emphasis on optimizing the efficiency of Differential Evolution (DE). The goals should be twofold: reducing the training times and expanding the applicability of DE-optimised algorithms to a wider range of medical conditions.

An intriguing area for future research lies in the exploration of hybrid optimisation techniques. Combining DE with other optimisation methods, such as genetic algorithms or particle swarm optimisation, could potentially streamline the hyperparameter tuning process. This approach might reduce the computational burden and training time associated with DE, while potentially improving or maintaining the high levels of accuracy achieved. Research in this direction could focus on developing new, more efficient algorithms that capitalise on the strengths of multiple optimisation methods.

The extended training time required for DE-optimised models, as observed in this study, presents a significant challenge, especially for real-time diagnostic applications in clinical settings. Future research should aim at modifying the DE algorithm to make it more time-efficient, perhaps by implementing parallel processing techniques or by improving the algorithm's convergence rate. Investigations could also explore ways to streamline the

model's complexity without compromising its predictive power.

Another promising direction for future research is the application of DE in the diagnosis of a variety of medical conditions beyond heart disease. Exploring DE's efficacy in conditions with complex diagnostic criteria, such as various types of cancers, neurological disorders, or rare diseases, could significantly broaden the impact of this research. This would not only validate the versatility of DE but also potentially contribute to advancements in the diagnosis and treatment of a wide range of diseases.

Given the varied response of different algorithms to DE optimisation, as evidenced by the differential improvements in Random Forest, AdaBoost, and Gradient Boosting classifiers, future studies should also focus on algorithm-specific optimisations. Understanding why some algorithms respond better to DE than others could lead to more targeted and effective optimisation strategies.

Lastly, future research should aim to balance the computational cost and practical clinical utility of DE-optimised models. This involves not just improving algorithm efficiency but also ensuring that these models can be seamlessly integrated into clinical workflows without requiring prohibitively expensive or complex hardware setups.

F. Generalisability and Broader Impacts

The generalisability of the findings from this study to other medical conditions opens up a realm of possibilities for the application of Differential Evolution (DE) across various domains of healthcare diagnostics. The successful optimisation of machine learning models for heart disease diagnosis using DE suggests that this approach could be equally effective in tackling other complex medical conditions. This potential extends not only to diseases with similar diagnostic complexities but also to those where nuanced and intricate data interpretation is crucial.

One area where DE could be particularly impactful is in the diagnosis of various types of cancer. Cancer diagnosis often involves interpreting complex patterns in imaging data, genetic information, and patient histories. DE could optimise machine learning models to more accurately identify patterns indicative of different cancer stages, leading to earlier and more precise diagnoses.

Neurological disorders, such as Alzheimer's and Parkinson's disease, also present a promising area for the application of DE-optimised machine learning models. These conditions often require the analysis of intricate neurological data, where subtle variations can be indicative of disease progression. DE could enhance the accuracy of models used to detect these variations early on, potentially improving patient outcomes through earlier intervention.

In the realm of rare diseases, where diagnosis is often challenging due to the scarcity of data, DE could be used to optimise models to make the most of limited information. By fine-tuning machine learning algorithms to identify patterns within small datasets, DE could aid in the early detection of rare conditions, which is often critical to effective treatment.

Another promising application is in personalised medicine. By optimising models to interpret patient-specific data, such as genetic profiles and individual health records, DE could play a vital role in tailoring treatments to individual patient needs. This could lead to more effective and targeted therapies, reducing the trial-and-error approach often associated with treatment selection.

The potential of DE in these diverse medical fields highlights its versatility and power as a tool in medical diagnostics. By applying this approach to various conditions, significant advancements in the accuracy, speed, and efficiency of diagnoses across healthcare could be achieved. This, in turn, would have profound impacts on patient care, treatment strategies, and overall health outcomes, thereby reinforcing the broad and transformative impacts of DE in healthcare diagnostics.

VI. Conclusion

This study embarked on an exploration of the efficacy of Differential Evolution (DE) in optimizing machine learning algorithms for heart disease diagnosis, utilizing the Cleveland Heart Disease dataset.

The results vividly illustrate the transformative power of DE, especially in its application to the Random Forest Classifier. The DE-optimised Random Forest model, with its remarkable accuracy of 93.3% and an F1−Score of 90.9%, serves as a testament to the substantial performance enhancements that DE can facilitate in hyperparameter tuning.

However, the influence of DE varied among different machine learning algorithms. While the Random Forest Classifier exhibited significant improvements, the AdaBoost Classifier saw no marked change post-DE optimization, and the Gradient Boosting Classifier experienced moderate gains. These disparate outcomes highlight the necessity of customizing DE application based on the unique characteristics and requirements of each algorithm.

Clinically, this study carries profound implications. The high accuracy rates achieved point to the possibility of machine learning models, optimised effectively with techniques like DE, becoming invaluable tools in the early diagnosis of heart disease. Such advancements could herald a new era in diagnostic methodologies, characterized by enhanced timeliness and accuracy, leading to improved patient outcomes.

Looking ahead, this research paves the way for exciting future prospects. It suggests the untapped potential of amalgamating DE with other algorithmic optimisation methods, like genetic algorithms, to develop a more holistic approach to model enhancement. Additionally, the study establishes a foundation for further investigations into the applicability of DE across various machine learning algorithms and datasets.

Nevertheless, the study is not without its limitations. The reliance on a specific dataset like the Cleveland Heart Disease dataset underscores the need for broader research to validate and expand upon these findings. Future endeavours could concentrate on applying DE in algorithm-specific feature selection and engineering, and

in understanding how DE can contribute to enhancing model explainability. This is particularly crucial considering the ethical dimensions inherent in healthcare applications.

In conclusion, this study not only showcases the potent capabilities of DE in refining machine learning algorithms for heart disease diagnosis but also offers valuable insights into the nuanced impact of DE on different algorithms. It sets a promising stage for future research, aiming to develop even more effective, efficient, and ethically sound models for clinical use. The potential of DE-optimised machine learning in healthcare is vast, and this study marks a significant step forward in realizing that potential.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

AUTHOR CONTRIBUTIONS

Theodore N. R. Egling reviewed the literature, designed the research methodology, collected the results, and compiled the manuscript under the supervision of Zenghui Wang and co-supervision of Sumbwanyambe Mbuyu; all authors had approved the final version.

FUNDING

This research is partially supported by the South African National Research Foundation (Grant Nos. 132797 and 137951), the South African National Research Foundation incentive grant (No. 114911), and the South African Eskom Tertiary Education Support Programme.

REFERENCES

[1] T. Chandrasekhar and V. Peddakrishna, "Heart disease prediction using machine learning algorithms," *International Journal of Innovative Technology and Exploring Engineering*, vol. 9, no. 1, pp. 2278–3075, 2022.
[2] R. Alizadehsani *et al*., "A data mining approach for diagnosis of coronary artery disease," *Computer Methods and Programs in Biomedicine*, vol. 111, no. 1, pp. 52–61, 2013.
[3] M. Shouman, T. Turner, and R. Stocker, "Using decision tree for diagnosing heart disease patients," in *Proc. Ninth Australasian Data Mining Conference*, 2012, vol. 134, pp. 23–30.
[4] T. Hastie, R. Tibshirani, and J. Friedman, *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*, 2nd Ed. New York: Springer, 2009.
[5] R. E. Schapire and Y. Freund, *Boosting: Foundations and Algorithms*, Cambridge, MA: MIT Press, vol. 3, 2012.
[6] T. Chen and C. Guestrin, "XGBoost: A scalable tree boosting system," in *Proc. 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2016, pp. 785–794.
[7] C. C. Chang and C. J. Lin, "LIBSVM: A library for support vector machines," *ACM Transactions on Intelligent Systems and Technology*, vol. 2, pp. 1–27, 2011.
[8] G. Hinton *et al*., "Deep neural networks for acoustic modeling in speech recognition," *IEEE Signal Processing Magazine*, vol. 29, no. 6, pp. 82–97, 2012.
[9] Q. K. Al-Shayea *et al*., "Using gradient boosting classifier with different feature selection methods for diagnosing heart disease," in *Proc. World Congress on Engineering*, 2021, vol. 1.
[10] S. Dua and X. Du, *Data Mining and Machine Learning in Cybersecurity*, CRC Press, 2017, p. 91.
[11] J. Khan, J. S. Wei, M. Ringnér *et al*., "Classification and diagnostic prediction of cancers using gene expression profiling and artificial neural networks," *Nature Medicine*, vol. 7, pp. 673–679, 2001.
[12] M. Shouman, T. Turner, and R. Stocker, "Using decision tree for diagnosing heart disease patients," in *Proc. Ninth Australasian Data Mining Conference*, 2012, vol. 134, pp. 23–30.
[13] K. P. Kresoja *et al*., "A cardiologist's guide to machine learning in cardiovascular disease prognosis prediction," *Basic Research in Cardiology*, vol. 118, no. 10, 2023.
[14] U. Nagavelli *et al*., "Machine learning technology-based heart disease detection models," *Journal of Healthcare Engineering*, vol. 4, 2022.
[15] S. Das, S. S. Mullick, and P. N. Suganthan, "Recent advances in differential evolution—An updated survey," *Swarm and Evolutionary Computation*, vol. 27, pp. 1–30, 2016.
[16] L. Breiman, "Random forests," *Machine Learning*, vol. 45, no. 1, pp. 5–32, 2001.
[17] R. Storn and K. Price, "Differential evolution—A simple and efficient heuristic for global optimization over continuous spaces," *Journal of Global Optimization*, vol. 11, no. 4, pp. 341–359, 1997.
[18] B. Morgan. (2021). Differential evolution—Automated machine learning. [Online]. Available: http:// towardsdatascience.com/unit-7-differential-evolution-automated-machine-learning-eb22014e592e
[19] P. Ghosal and D. Bhattacharjee, "Medical image enhancement using differential evolution," *Advances in Intelligent Systems and Computing*, vol. 479, pp. 299–308, 2017.
[20] E. Salari and M. Shamsi, "A novel multi-objective differential evolution algorithm for optimization of beam angles and weighTS in IMRT," *Medical Physics*, vol. 46, no. 9, pp. 4090–4102, 2019.
[21] Z. He, J. Zhang, X. Shi *et al*., "Learning to predict drug-target interactions with features constructed from molecular structure," *Journal of Computational Biology*, vol. 23, no. 5, pp. 377–385, 2016.
[22] Archive.ics.uci.edu. (2021). UCI Machine learning repository: Heart disease data set. [Online]. Available: https://archive.ics.uci.edu/dataset/45/heart+disease
[23] D. M. Powers, "Evaluation: From precision, recall and f-measure to ROC, informedness, markedness and correlation," *Journal of Machine Learning Technologies*, vol. 2, no. 1, pp. 37–63, 2011.
[24] R. Kohavi, "A study of cross-validation and bootstrap for accuracy estimation and model selection," in *Proc. 14th International Joint Conference on Artificial Intelligence*, 1995, vol. 2, pp. 1137–1143.
[25] M. Kaur and S. K. Wasan, "Empirical study on applications of data mining techniques in healthcare," *Journal of Computer Science*, vol. 2, no. 2, pp. 194–200, 2006.