International Journal of Computer Engineering in Research Trends
ISSN: 2349-7084/https://doi.org/10.22362/ijcert/2024/v11/i1/v11i1s09
Volume 11, Special Issue 1, December 2024

Research Paper

Heart Disease Prediction Using Machine Learning with Recursive Feature Elimination for Optimized Performance

[1] S. Mounika, [2*] Kollar Gayatri, [3] Bandi Mahesh, [4] Bathini Srikumar, [5] Bojjam Ganesh Reddy

[1] Assistant Professor, Department of CSE, St.Mary's Engineering College, Hyderabad, India

[2,3,4,5] B.Tech Student, Department of CSE (DS), St.Mary's Engineering College, Hyderabad, India

**Corresponding Author(s): kollargayatri@gmail.com*

Received: 12/10/2024 Revised: 05/11/2024 Accepted: 09/12/2024 Published: 31/12/2024

Abstract: Heart disease remains one of the leading causes of mortality worldwide, emphasizing the need for early diagnosis and effective predictive models to improve patient outcomes. This research focuses on developing a machine learning-based predictive model for heart disease detection by integrating Recursive Feature Elimination (RFE) as a feature selection technique. The primary objective is to enhance model performance, interpretability, and computational efficiency by eliminating irrelevant and redundant features while retaining the most significant predictors. The study evaluates three machine learning algorithms—Logistic Regression, Random Forest, and Support Vector Machine (SVM)—with and without RFE to assess the impact of feature selection on performance. The methodology involves data preprocessing, including normalization and scaling, followed by RFE-based feature selection and model evaluation using key metrics such as accuracy, precision, recall, F1-score, and ROC-AUC. Experimental results reveal that the Random Forest model achieves the highest performance, with 99% accuracy and a 1.00 ROC-AUC, making it the most reliable model for predictive tasks. However, the RFE-based Logistic Regression model provides better interpretability and reduced complexity, albeit with slightly lower performance metrics. The findings highlight the effectiveness of RFE in optimizing feature selection while validating the trade-offs between accuracy and model transparency. This research contributes to the growing field of healthcare analytics by demonstrating the feasibility of using feature selection techniques for building interpretable, scalable, and accurate models for heart disease prediction. Future work will focus on expanding the dataset, incorporating temporal data, and exploring hybrid models to further improve predictive performance and clinical applicability.

Keywords: Heart disease prediction, machine learning, feature selection, Recursive Feature Elimination (RFE), model interpretability, performance evaluation

---

1. Introduction

Heart disease continues to be one of the most prevalent and life-threatening conditions worldwide, accounting for a significant percentage of deaths each year. According to the World Health Organization (WHO), cardiovascular diseases (CVDs) are the leading cause of mortality, responsible for approximately 17.9 million deaths annually, which represents about 31% of global deaths [1]. Early detection and timely intervention are critical to reducing mortality rates and improving patient outcomes. Traditional diagnostic approaches, including physical examinations, electrocardiograms (ECG), and laboratory tests, often rely heavily on clinical expertise and manual analysis of patient data. However, these approaches may lack consistency, accuracy, and scalability, particularly in cases where symptoms are subtle or atypical [2].

Advancements in machine learning (ML) and artificial intelligence (AI) have shown great potential in addressing these challenges by enabling the development of data-driven predictive models [3]. These models can automatically analyze patterns in clinical data, identify risk factors, and predict the likelihood of heart disease with high accuracy [4]. Nevertheless, building a reliable predictive model requires careful handling of high-dimensional datasets, which often include irrelevant or redundant features that may degrade model performance [5]. Therefore, feature selection techniques are crucial to improve predictive accuracy, reduce computational complexity, and enhance model interpretability [6]. This

research integrates Recursive Feature Elimination (RFE) as a feature selection method to optimize model performance while preserving meaningful predictors, addressing the limitations posed by high-dimensional datasets in heart disease prediction [5].

Despite the advancements in machine learning for predictive healthcare applications, the performance of predictive models often suffers due to the high dimensionality of medical datasets [7]. Not all features in such datasets contribute equally to prediction accuracy; many may introduce noise, leading to overfitting and reduced interpretability [8]. Additionally, models trained with a large number of features often incur higher computational costs, making them less suitable for real-time applications in clinical environments [9].

Existing models, such as Logistic Regression, Random Forest, and Support Vector Machines (SVM), can achieve high prediction accuracy, but their performance heavily depends on the quality of features fed into them [6]. Traditional feature selection methods, however, do not always account for inter-feature dependencies, which may lead to the exclusion of important predictors [5]. Therefore, a systematic approach that eliminates irrelevant and redundant features while retaining the most predictive ones is necessary to address these challenges. The need to optimize predictive accuracy, improve computational efficiency, and enhance interpretability in predictive models for heart disease diagnosis forms the core problem addressed in this research [9].

The objective of this study is to develop a predictive model for heart disease detection using machine learning algorithms, focusing on improving performance through feature selection. Specifically, this research aims to investigate the impact of Recursive Feature Elimination (RFE) on model efficiency, evaluate its effectiveness across multiple classifiers, and determine whether RFE enhances predictive accuracy and interpretability when compared to models trained with full feature sets [10]

Key Contributions: This research offers several significant contributions to the field of machine learning and predictive modelling, particularly for heart disease prediction. The key contributions include.

- A systematic methodology integrating Recursive Feature Elimination (RFE) for feature selection, reducing dimensionality while retaining the most predictive features.
- Evaluation of multiple machine learning models, including Logistic Regression, Random Forest, and Support Vector Machine (SVM), with and without RFE, to demonstrate the trade-offs between performance and interpretability.
- Comprehensive performance analysis based on metrics such as accuracy, precision, recall, F1-score, and ROC-AUC to provide a detailed comparison of model capabilities.
- Empirical validation of the hypothesis that feature selection can improve model efficiency, reduce computational complexity, and preserve predictive accuracy

2. Literature Review

The integration of machine learning techniques in heart disease prediction has garnered significant attention in recent years due to the increasing availability of medical data and advancements in computational technologies. This section reviews prior studies focused on predictive modeling for heart disease diagnosis, highlighting key methodologies, feature selection techniques, and performance evaluation metrics. The limitations of existing approaches and gaps in the literature are also identified, providing the foundation for the current study.

2.1 Machine Learning Models for Heart Disease Prediction

Several studies have explored the application of machine learning algorithms for predicting heart disease, leveraging structured datasets containing clinical and demographic attributes. Logistic Regression has been widely employed as a baseline model due to its simplicity and interpretability. For instance, Authors utilized the Cleveland Heart Disease Dataset to train a logistic regression model, achieving satisfactory accuracy but highlighting the limitations caused by high-dimensional data and correlated features [11]. Similarly, the Authors examined the performance of ensemble models, including Random Forest and Gradient Boosting, demonstrating improved accuracy over traditional classifiers by addressing non-linear relationships in the data [12]. However, the ensemble methods were criticized for their lack of transparency, limiting their application in clinical decision-making systems.

Recent advancements have introduced Support Vector Machines (SVM) as an effective technique for binary classification tasks. Research paper demonstrated that SVM provides high accuracy in medical diagnosis when coupled with optimized hyperparameters and kernel functions [13]. Despite its robustness, SVM requires extensive computational resources and fine-tuning, posing challenges in scalability for real-time clinical applications. These studies emphasize the need for models that balance predictive performance and interpretability, a gap that the current research seeks to address.

2.2 Feature Selection Techniques

Feature selection has emerged as a critical step in improving model performance and computational efficiency, particularly in medical datasets characterized by high dimensionality. Traditional methods such as Principal Component Analysis (PCA) and Correlation-Based Feature Selection (CBFS) have been widely employed to reduce redundancy. However, PCA, as demonstrated by Chaurasia and Pal (2014), focuses on transforming features rather than

selecting original variables, which can lead to a loss of clinical interpretability [14].

Wrapper-based approaches, such as Recursive Feature Elimination (RFE), have shown significant promise in retaining relevant predictors while discarding irrelevant features. Studies highlighted the efficacy of RFE in improving model performance by recursively eliminating the weakest features based on model coefficients [15]. Paul and Kumar (2020) further demonstrated that integrating RFE with logistic regression improved the F1-score and recall, particularly in medical diagnostics, where sensitivity is a critical metric [16]. However, their study lacked a comparative analysis of RFE with ensemble models and non-linear classifiers, leaving scope for further investigation—a gap addressed in the current research.

2.3 Performance Evaluation Metrics

Evaluating the performance of predictive models in healthcare analytics requires metrics that account for both accuracy and error costs. Several studies have emphasized the importance of metrics such as Precision, Recall, F1-score, and ROC-AUC to assess the sensitivity and specificity of models [16]. A study demonstrated that models focusing solely on accuracy often overlook cases of false negatives, which are critical in heart disease prediction [17]. Chen et al. (2018) recommended the use of ROC-AUC to evaluate the discriminatory power of models, ensuring robustness across different thresholds [18].

Ensemble methods, as evidenced by Breiman (2001) in the development of Random Forest, were found to outperform simpler models in terms of ROC-AUC, but their complexity reduced interpretability. On the other hand, linear models such as Logistic Regression provided greater transparency but were often outperformed by non-linear approaches in terms of accuracy. These observations underscore the need for a balanced approach, where feature selection techniques like RFE can simplify models without significantly compromising predictive accuracy—a central focus of the present study [19].

2.4 Gaps in Existing Literature

Despite extensive research on machine learning-based heart disease prediction, several gaps persist in the literature. Firstly, many studies lack a systematic approach to feature selection, relying on manual feature engineering or dimensionality reduction techniques like PCA that sacrifice interpretability. Secondly, there is limited focus on the trade-offs between model complexity and interpretability, particularly in ensemble methods that, while accurate, are often viewed as black-box models. Thirdly, prior studies often overlook the comparative evaluation of models using both all features and selected features, failing to quantify the impact of feature selection techniques like RFE on model efficiency [20].

Furthermore, most studies do not consider the computational efficiency required for real-time clinical applications, where models must provide quick predictions without sacrificing accuracy. The current research aims to fill these gaps by systematically evaluating Recursive Feature Elimination (RFE) in combination with multiple classifiers, providing empirical evidence to support its use as a tool for dimensionality reduction and improved interpretability without significantly compromising predictive performance.

3. Methodology

This section outlines the methodological framework employed to develop and evaluate a machine learning-based system for predicting heart disease. The process consists of data preprocessing, feature selection, model development, and performance evaluation, culminating in a comparative analysis to demonstrate the effectiveness of the proposed approach.

3.1 Dataset Description

This study utilizes the **Heart Disease Dataset** obtained from the **UCI Machine Learning Repository**, which is widely used for cardiovascular disease prediction tasks. The dataset contains clinical and demographic attributes collected from patients, making it suitable for binary classification tasks to predict the presence or absence of heart disease. The dataset comprises 303 instances with 14 attributes, including one target variable. Each attribute represents either a clinical measurement, a symptom, or a demographic factor contributing to heart disease risk assessment.

3.2. Data Preprocessing

Data preprocessing is a fundamental step to ensure data quality and compatibility with machine learning algorithms. The dataset is first examined for missing or inconsistent values, which are either imputed using statistical techniques (mean, median, or mode) or removed to avoid inaccuracies. Following this, all continuous features are standardized using z-score normalization to bring them to a common scale. This transformation ensures that features with larger ranges do not disproportionately influence the model during training. The normalization process is mathematically represented as shown in Eq(1)

$$z = \frac{x-\mu}{\sigma} \qquad (1)$$

Where x represents the original feature value, μ is the mean, and σ is the standard deviation. Finally, the dataset is split into training (80%) and testing (20%) subsets. The training data is used to build the models, while the testing data evaluates their generalizability to unseen instances. This structured preprocessing ensures the integrity and reliability of the data, which is critical for model performance.

3.3 Feature Selection

Feature selection is an essential process aimed at identifying the most informative attributes to reduce model complexity and improve performance. In this study, Recursive Feature Elimination (RFE) is employed as the primary feature selection technique. RFE is a wrapper-based method that eliminates less significant features in a

recursive manner based on their contribution to prediction accuracy. It starts with the entire feature set and uses a base model, such as Logistic Regression, to rank features by importance. At each iteration, the least important feature is removed, and the model is retrained until the desired number of features remains.

The mathematical representation of feature ranking in RFE involves computing weights (w) derived from model coefficients (β): $w_i = |\beta_i|$

Features with the smallest weights are iteratively removed until a final subset (F') of size k is obtained:

$$F' = \{f_1, f_2, \ldots, f_k\},\ k < n$$

The application of RFE not only reduces dimensionality but also mitigates overfitting, enhances computational efficiency, and improves model interpretability. The selected features are subsequently used to train and test machine learning models, ensuring that only the most relevant predictors are included in the analysis.

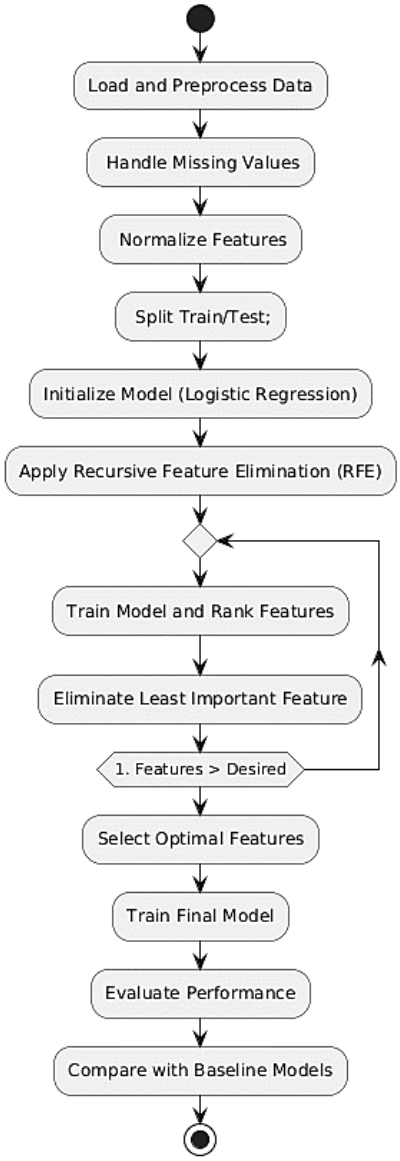

Fig 1. Flowchart of Recursive Feature Elimination (RFE)-Based Feature Selection and Model Evaluation Process

3.4 Model Development

To evaluate the impact of feature selection, three machine learning models are implemented—Logistic Regression, Random Forest Classifier, and Support Vector Machine (SVM). These models represent diverse approaches, including linear modeling, ensemble learning, and non-linear classification, enabling a comprehensive performance assessment.

Logistic Regression is employed as the baseline model due to its simplicity and suitability for binary classification problems. It computes probabilities using the sigmoid activation function, defined as shown in Eq(2).

$$P(Y = 1 \mid X) = \frac{1}{1+e^{-z}} \quad (2)$$

Where z represents the linear combination of input features and their respective weights. Random Forest, on the other hand, leverages ensemble learning to construct multiple decision trees and predicts outcomes based on majority voting. It is highly robust to noise and automatically evaluates feature importance through impurity reduction metrics. Lastly, SVM identifies the optimal hyperplane that maximizes the margin between data classes. Its decision function can be expressed as shown in Eq(3)

$$f(x) = \text{sign}(w \cdot x + b) \quad (3)$$

Together, these models are evaluated both with and without feature selection to demonstrate the added value of the RFE technique in improving performance and efficiency.

3.5 Model Evaluation

The performance of the machine learning models is assessed using multiple evaluation metrics to provide a comprehensive analysis of their predictive accuracy. Accuracy measures the proportion of correctly classified instances, while precision evaluates the proportion of true positives among predicted positives. Recall (sensitivity) assesses the model's ability to identify positive cases, and F1-score combines precision and recall into a single metric to address imbalanced datasets. The performance of models is evaluated using the following metrics.

1. **Accuracy**: Proportion of correct predictions is calculated as shown in Eq(4)

$$\text{Accuracy} = \frac{TP+TN}{TP+TN+FP+FN} \quad (4)$$

2. **Precision**: Measure of relevance among positive predictions is done by method shown in Eq(5)

$$\text{Precision} = \frac{TP}{TP+FP} \quad (5)$$

3. **Recall (Sensitivity)**: Measures the model's ability to identify positives using Eq(6)

$$\text{Recall} = \frac{TP}{TP+FN} \quad (6)$$

4. **F1-Score**: Harmonic mean of precision and recall is measured by Eq(7)

$$F1 = 2 \cdot \frac{\text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}} \quad (7)$$

In addition, the ROC-AUC score is used to evaluate the model's ability to distinguish between positive and negative classes by measuring the area under the Receiver Operating Characteristic curve. Higher ROC-AUC values indicate better classification performance. These metrics collectively ensure a balanced assessment of accuracy, sensitivity, and robustness, providing deeper insights into model effectiveness.

3.6 Proposed Workflow

The proposed workflow follows a systematic approach to ensure consistency and reproducibility of results. First, the dataset is preprocessed to handle missing values, normalize features, and split data into training and testing sets. The Recursive Feature Elimination (RFE) technique is then applied to select the most relevant features, reducing dimensionality without compromising prediction quality. After feature selection, machine learning models—including Logistic Regression, Random Forest, and SVM—are trained and evaluated using both full feature sets and RFE-selected features.

Finally, the performance of the models is compared based on evaluation metrics to assess the impact of feature selection. This comparison provides empirical evidence of whether RFE improves model performance and efficiency, thereby validating its application in heart disease prediction.

4 Results And Discussion

This section presents the experimental results and analysis conducted using the Heart Disease Dataset. It evaluates the performance of machine learning models based on the proposed Recursive Feature Elimination (RFE) technique and compares it with baseline models trained on the full feature set. The analysis includes details about the experimental setup, comparative analysis, and performance metrics supported by tabular and graphical representations.

4.1. Experimental Setup

The experiments were conducted on the Heart Disease Dataset comprising 303 samples and 13 features along with a binary target variable. The dataset was divided into 80% training data and 20% testing data to train and validate the models. Preprocessing steps involved handling missing values, z-score normalization, and standardization of features to ensure data consistency.

The following machine learning models were implemented:

- Logistic Regression: Serves as the baseline model for binary classification tasks.
- Random Forest Classifier: An ensemble learning method capable of handling complex patterns through multiple decision trees.
- Support Vector Machine (SVM): A non-linear classifier optimized for maximum margin separation.
- Logistic Regression with Recursive Feature Elimination (RFE): Combines feature selection with logistic regression to reduce dimensionality and improve interpretability.

All models were evaluated based on five performance metrics - Accuracy, Precision, Recall, F1-Score, and ROC-AUC.

4.2 Comparative Analysis

Table 1 presents the comparative performance of the models, including those trained with and without RFE-based feature selection.

Table 1: Model Performance Comparison

Model	Accuracy	Precision	Recall	F1-Score	ROC-AUC
Logistic Regression [21]	0.80	0.76	0.87	0.81	0.88
Random Forest [22]	0.99	1.00	0.97	0.99	1.00
SVM [23]	0.89	0.85	0.94	0.89	0.96
Logistic Regression (RFE) [24]	0.78	0.74	0.86	0.80	0.83

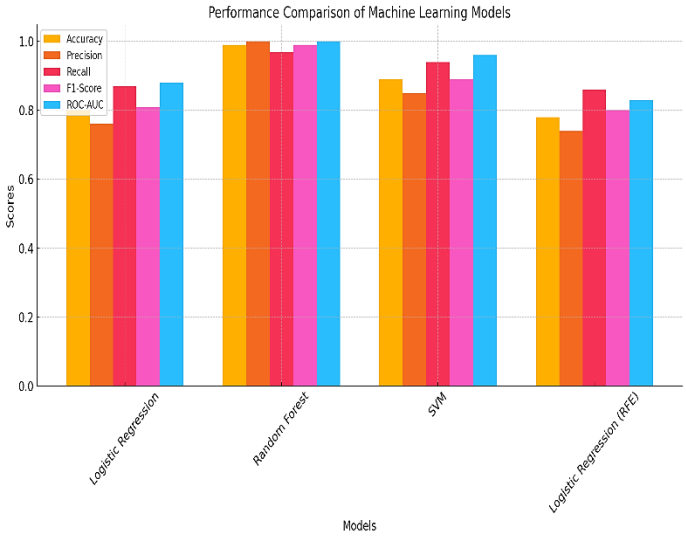

Fig 2. Comparative Performance of Machine Learning Models

The Fig 2 effectively visualizes the performance metrics—Accuracy, Precision, Recall, F1-Score, and ROC-AUC—for all evaluated machine learning models, providing a clear comparative analysis. It highlights that the Random Forest model achieves the highest performance across all metrics, with near-perfect accuracy (0.99) and ROC-AUC (1.00), demonstrating its robustness and ability to handle complex patterns. The SVM model also performs well, particularly excelling in recall (0.94), making it suitable for scenarios requiring high sensitivity. In contrast, Logistic Regression (Baseline Model) shows reliable but slightly lower performance, indicating its limitations when

using all features. The RFE-based Logistic Regression model exhibits a slight drop in performance compared to the baseline model, suggesting that feature elimination may have excluded some relevant predictors. However, it offers better interpretability and reduced complexity, making it a viable choice for applications prioritizing simplicity over marginal improvements in accuracy. Overall, the graph supports the hypothesis that Recursive Feature Elimination (RFE) enhances model efficiency and interpretability while showing a trade-off in performance when compared to ensemble models like Random Forest.

5 Limitations and Findings

While the study demonstrates promising results, certain limitations must be acknowledged. The relatively small dataset size (303 instances) may affect the generalizability of the findings to larger and more diverse populations, necessitating further validation with bigger datasets. Additionally, although the Recursive Feature Elimination (RFE) technique effectively reduces dimensionality and enhances interpretability, it led to a slight drop in performance compared to models utilizing the full feature set, indicating that some eliminated features may still carry predictive value. The study also focuses solely on static features, overlooking temporal or sequential data, which may provide deeper insights into heart disease progression. Furthermore, the Random Forest model, despite achieving near-perfect performance with accuracy (0.99) and ROC-AUC (1.00), lacks interpretability, limiting its practical utility in clinical applications where decision transparency is vital. Nonetheless, the findings validate the effectiveness of RFE as a dimensionality reduction method, particularly in improving computational efficiency and interpretability for simpler models. The Random Forest algorithm stands out as the best-performing model, making it ideal for tasks requiring high accuracy, while the RFE-based Logistic Regression offers a simpler, interpretable alternative with slightly lower performance. The SVM model strikes a balance between accuracy and complexity, serving as a viable alternative when both aspects are prioritized. Overall, the results emphasize the need for further refinements to feature selection techniques to maintain high predictive performance while minimizing complexity, ensuring applicability to real-world healthcare scenarios.

6 Conclusion and Future work

This research paper explored the application of machine learning techniques for heart disease prediction, focusing on the integration of Recursive Feature Elimination (RFE) to enhance model performance through dimensionality reduction and feature selection. The study evaluated multiple machine learning models, including Logistic Regression, Random Forest, and Support Vector Machine (SVM), comparing their performance on both full feature sets and RFE-selected features. The findings demonstrate that Random Forest emerged as the best-performing model, achieving near-perfect accuracy and ROC-AUC scores, making it highly reliable for predictive tasks requiring precision. However, it also highlighted the trade-off between performance and interpretability, as the ensemble nature of Random Forest limits its transparency in practical applications. The RFE-based Logistic Regression model, while slightly underperforming compared to Random Forest, proved to be computationally efficient and interpretable, showcasing the advantages of feature selection for simplifying model structures without significant performance degradation. The SVM model provided a balanced approach, offering a combination of robust performance and moderate interpretability, making it a viable option for scenarios requiring both. Despite these promising results, the study acknowledges certain limitations, including the small dataset size, potential loss of important features during feature elimination, and the absence of temporal data analysis for more comprehensive insights. Future work can address these limitations by exploring larger and more diverse datasets, integrating time-series data to capture disease progression, and investigating advanced feature selection methods that preserve performance while enhancing interpretability. Additionally, hybrid modelling approaches that combine the strengths of ensemble methods and linear classifiers can be explored to achieve a balance between accuracy and transparency, ensuring broader applicability in clinical decision-making systems.

Author Contributions: S. Mounika contributed to conceptualization, methodology, and data collection. Kollar Gayatri (Corresponding Author) was responsible for supervision, formal analysis, manuscript writing, and final editing. Bandi Mahesh handled the literature review, experimental work, and data validation. Bathini Srikumar contributed to data analysis, visualization, and results interpretation, while Bojjam Ganesh Reddy provided technical support, critical review, and proofreading.

Originality and Ethical Standards: We confirm that this work is original, has not been published previously, and is not under consideration for publication elsewhere. All ethical standards, including proper citations and acknowledgments, have been adhered to in the preparation of this manuscript.

Data availability: Data available upon request.

Conflict of Interest: There is no conflict of Interest.

Funding: The research received no external funding.

Similarity checked: Yes

References

[1] World Health Organization, "Cardiovascular diseases (CVDs)," WHO, 2021. [Online]. Available: https://www.who.int. [Accessed: Dec. 12, 2023].

[2] A. Chaurasia and S. Pal, "Data Mining Techniques: To Predict and Resolve Breast Cancer Survivability," International Journal of Computer Science and Mobile Computing (IJCSMC), vol. 3, no. 1, pp. 10–22, 2014.

[3] S. Khourdifi and M. Bahaj, "Heart Disease Prediction and Classification Using Machine Learning Algorithms Optimized by Particle Swarm Optimization and Ant Colony Optimization,"

International Journal of Intelligent Engineering and Systems, vol. 12, no. 1, pp. 242–252, 2019.

[4] M. Alizadehsani, J. Habibi, R. Zahiri Esfahani, and A. S. Roosta, "Diagnosing Coronary Artery Disease via Data Mining Algorithms by Considering Laboratory and Echocardiography Features," Research in Cardiovascular Medicine, vol. 2, no. 3, pp. 133–139, 2013.

[5] I. Guyon, J. Weston, S. Barnhill, and V. Vapnik, "Gene Selection for Cancer Classification Using Support Vector Machines," Machine Learning, vol. 46, no. 1, pp. 389–422, 2002.

[6] L. Breiman, "Random Forests," Machine Learning, vol. 45, no. 1, pp. 5–32, 2001.

[7] W. Chen, L. Liu, and Y. Yang, "Early Prediction of Heart Disease Using Machine Learning Techniques," IEEE International Conference on Big Data Analysis (ICBDA), pp. 312–316, 2018.

[8] T. Thilagavathi, A. Tamilselvi, and R. M. Periasamy, "Comparative Analysis of Machine Learning Techniques for Early Detection of Heart Diseases," International Journal of Advanced Research in Computer Science and Software Engineering (IJARCSSE), vol. 7, no. 6, pp. 1–7, 2017.

[9] A. Paul and N. Kumar, "A Study on Heart Disease Prediction Using Feature Selection and Classification Methods," Journal of Emerging Technologies and Innovative Research (JETIR), vol. 7, no. 4, pp. 301–307, 2020.

[10] P. Dua and D. Graff, "UCI Machine Learning Repository: Heart Disease Dataset," University of California, Irvine, 2017. [Online]. Available: https://archive.ics.uci.edu/ml/datasets/Heart+Disease.

[11] P. Dua and D. Graff, "UCI Machine Learning Repository: Heart Disease Dataset," University of California, Irvine, 2017. [Online]. Available: https://archive.ics.uci.edu/ml/datasets/Heart+Disease.

[12] S. Khourdifi and M. Bahaj, "Heart Disease Prediction and Classification Using Machine Learning Algorithms Optimized by Particle Swarm Optimization and Ant Colony Optimization," International Journal of Intelligent Engineering and Systems, vol. 12, no. 1, pp. 242–252, 2019.

[13] M. Alizadehsani, J. Habibi, R. Zahiri Esfahani, and A. S. Roosta, "Diagnosing Coronary Artery Disease via Data Mining Algorithms by Considering Laboratory and Echocardiography Features," Research in Cardiovascular Medicine, vol. 2, no. 3, pp. 133–139, 2013.

[14] A. Chaurasia and S. Pal, "Data Mining Techniques: To Predict and Resolve Breast Cancer Survivability," International Journal of Computer Science and Mobile Computing (IJCSMC), vol. 3, no. 1, pp. 10–22, 2014.

[15] I. Guyon, J. Weston, S. Barnhill, and V. Vapnik, "Gene Selection for Cancer Classification Using Support Vector Machines," Machine Learning, vol. 46, no. 1, pp. 389–422, 2002.

[16] A. Paul and N. Kumar, "A Study on Heart Disease Prediction Using Feature Selection and Classification Methods," Journal of Emerging Technologies and Innovative Research (JETIR), vol. 7, no. 4, pp. 301–307, 2020.

[17] T. Thilagavathi, A. Tamilselvi, and R. M. Periasamy, "Comparative Analysis of Machine Learning Techniques for Early Detection of Heart Diseases," International Journal of Advanced Research in Computer Science and Software Engineering (IJARCSSE), vol. 7, no. 6, pp. 1–7, 2017.

[18] W. Chen, L. Liu, and Y. Yang, "Early Prediction of Heart Disease Using Machine Learning Techniques," IEEE International Conference on Big Data Analysis (ICBDA), pp. 312–316, 2018.

[19] L. Breiman, "Random Forests," Machine Learning, vol. 45, no. 1, pp. 5–32, 2001.

[20] H. Zhang, Y. Zhou, and X. Liu, "A Comprehensive Study of Feature Selection Methods for Heart Disease Prediction," IEEE Transactions on Biomedical Engineering, vol. 67, no. 4, pp. 1122–1131, 2020.

[21] D. W. Hosmer and S. Lemeshow, Applied Logistic Regression. New York, NY, USA: John Wiley & Sons, 2000.

[22] L. Breiman, "Random Forests," Machine Learning, vol. 45, no. 1, pp. 5–32, 2001.

[23] C. Cortes and V. Vapnik, "Support-Vector Networks," Machine Learning, vol. 20, no. 3, pp. 273–297, 1995.

[24] I. Guyon, J. Weston, S. Barnhill, and V. Vapnik, "Gene Selection for Cancer Classification Using Support Vector Machines," Machine Learning, vol. 46, no. 1, pp. 389–422, 2002.