Jurnal Kecerdasan Buatan dan Teknologi Informasi
Vol. 5, No. 2, May 2026, pp. 270-278
e-ISSN: 2964-2922, p-ISSN: 2963-6191
DOI : https://doi.org/10.69916/jkbti.v5i2.477

COMPARATIVE ANALYSIS OF PERFORMANCE OF MACHINE LEARNING FEATURE SELECTION (GINI DECREASE AND RELIEF-F) IN HEART DISEASE DATASET

Chindu Lintang Bhuana[*1], Rico Pramestiawan[2], Lilik Joko Susanto[3], Arry Verdian[4]

[1,2,3,4]Department of Education Informatics, STKIP Rosalia, Metro, Indonesia
Email: [1]chindu.lintangbhuana@gmail.com, [2]ricosimple25@gmail.com, [3]lilikjoko09@gmail.com, [4]verdian.2637@gmail.com

(Received: May 3, 2026; Revised: May 10, 2026; Published: May 12, 2026)

Abstract

Heart disease remains one of the leading causes of mortality worldwide and presents a major challenge in healthcare systems. Early detection plays an essential role in improving survival rates and minimizing complications through timely intervention. Recent advances in Machine Learning (ML) have provided new opportunities for developing accurate and efficient prediction systems for heart disease detection. However, one of the major challenges in ML-based prediction is identifying the most relevant features to improve classification performance while reducing computational complexity and noise. This study aims to evaluate the effectiveness of two feature selection techniques, namely Gini Decrease (GD) and ReliefF, combined with several ML models, including Support Vector Machine (SVM), Tree, Naïve Bayes, and Random Forest, for heart disease classification. The study employed the UCI Heart Disease Dataset consisting of 303 records and 14 attributes. Data preprocessing included handling missing values using mean imputation, followed by feature selection and classification using 10-fold cross-validation with an 80:20 training-testing ratio. Experimental results showed that ReliefF outperformed GD, achieving the highest average accuracy of 0.796, compared to GD with 0.767 and all features with 0.771. The SVM model achieved the highest accuracy using GD (0.833), while Random Forest demonstrated optimal performance with ReliefF (0.817). Furthermore, the Tree model exhibited the fastest computational time among all evaluated models. These findings indicate that integrating suitable feature selection methods with ML models significantly enhances heart disease classification performance, particularly in improving predictive accuracy and computational efficiency for early medical diagnosis applications.

Keywords: heart; machine learning; feature selection; gini decrease; relief.

1 INTRODUCTION

The heart is a serious global health problem, ranking at the top of cases of internal disease and the main cause of death worldwide [1]. Heart disease is one of the main causes of heart deaths among women globally. Recent data shows that the incidence of heart disease continues to increase, making it the most frequently detected type of heart [2]. The main cause of death in heart disease is blood sugar complications and the presence of other accompanying diseases, especially cardiovascular disease [3]. Early detection is a fundamental key in increasing survival rates and treatment success. When early detection is carried out, the patient's chance of recovery is significantly higher compared to cases detected in acute heart disease.

In recent years, rapid advances in the field of Machine Learning (ML) have paved the way for the development of more accurate and efficient early cardiac detection systems [4]. ML models can analyse medical data, including genetic data, medical images, and clinical history, to identify complex patterns that may not be visible to the human eye. The potential of ML is able to increase the accuracy of detection and predict disease risk in various studies [5].

However, the application of ML in early heart detection faces two main challenges, namely feature selection. Heart Dataset is often used as a basis for evaluating early heart detection models [6]. This dataset has fourteen features, where all features need to be analysed, whether they are relevant or not to model accuracy. The presence of irrelevant features can lead to decreased model performance, increased computational complexity, and overfitting [7]. Therefore, effective feature selection techniques are essential to identify the most informative subset of features that have an influence on model performance and reduce noise. Integration between feature selection and ML models can detect heart disease failure. However, there is still a need for a comprehensive evaluation of the combination of various feature selection techniques and ML models to identify the best approach in this context

Several studies regarding the comparison of feature selection in heart disease datasets have been carried out. Research conducted by Bashir et al, 2019 used the model Decision Tree, Logistic Regression, Logistic Regression SVM, Naïve Bayes, and Random Forest; algorithms are used as feature selection techniques and improvement is shown in the results by showing the accuracy [8]. Research Sreekumari et all., 2025 we employed an extension of ARM called Weighted ARM (WARM) with 70,000 observations of heart disease data made available through Kaggle. Using an ensemble voting method combining KNN, decision tree, and Naive Bayes, an accuracy of 78.42% is achieved for the selected set of attributes [9]. Next, Khan et all., 2024 carried out feature selection for heart disease prediction using modified Artificial Bee colony (M-ABC) and K-nearest neighbours (KNN). In our study, we have proposed a framework based on Modified Artificial Bee Colony (M-ABC) and k-Nearest Neighbours (KNN) for predicting the optimal feature selection to obtain better accuracy. Using a modified bee algorithm, this paper focuses on identifying the optimal subset of attributes from the dataset [10]

This research aims to carry out performance-based analysis of various feature selection methods such as Gini Decrease (GD), and ReliefF on ML models such as SVM, Tree, Naive Bayes, and Random Forest. By comparing the performance of various combinations of these methods, it is hoped that the most optimal approach can be found to build an accurate heart disease failure detection model, thus contributing to increasing patient survival.

2 RESEARCH METHODS

This stage outlines the procedures followed when conducting research. The research stage begins with data collection. Before the data is entered into the model, it is necessary to pre-process the data. At this stage, the data is cleaned and adjusted to be processed to the next step of feature selection using GD, and ReliefF techniques. Next, the selected features are used for Heart classification using ML models such as SVM, Tree, Naive Bayes, Random Forest. The framework design is illustrated in Figure 1.

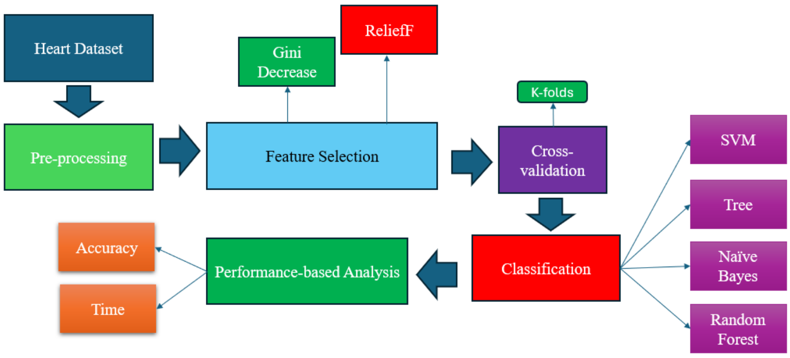

Figure 1. Framework design

After the heart classification is carried out, a comparative analysis is then carried out by calculating the accuracy value and the computing time required for the ML model to build the model

2.1. Hearth Dataset

The study utilized the Heart Disease Dataset from the UCI Machine Learning Repository. This dataset has undergone preprocessing procedures, comprising 303 rows and 14 columns (attributes). Each row represents patient data used to detect the likelihood of heart disease based on 13 features or variables. As a classification problem employing supervised learning, the dataset also includes one target column that classifies whether patients are indicated to have heart disease or not. A detailed description of the data is provided in Table 1. The features used in this dataset are presented in Table 1.

Table 1. Features, and attribute dataset

ID	Feature	Information
f1	Age	Patient's age in years
f2	Gender	Patient's gender
f3	Chest paint	Type of chest pain
f4	Rest SBP	Resting blood pressure (in mm Hg on admission to the hospital)
f5	Cholesterol	Serum cholesterol in mg/dl
f6	Fasting blood sugar	Fasting blood sugar > 120 mg/dl
f7	Rest ECG	Resting electrocardiographic results
f8	Max HR	Maximum heart rate achieved
f9	Exer ind ang	Exercise-induced angina
f10	ST by exercise	ST depression induced by exercise relative to rest

ID	Feature	Information
f11	Slope peak	The slope of the peak exercise ST segment
f12	Major vessels color	Number of major vessels (0-3) colored by fluoroscopy
f13	Thal	Thalassemia
f14	Num	Diagnosis of heart disease (severity of narrowing of blood vessels)

2.2. Pre-processing

The crucial process includes data cleaning to deal with missing values or outliers. This dataset has 0.2% missing values, so improvements need to be made. We apply charging using the technique of using average values. Design lost data repair show in Figure 2.

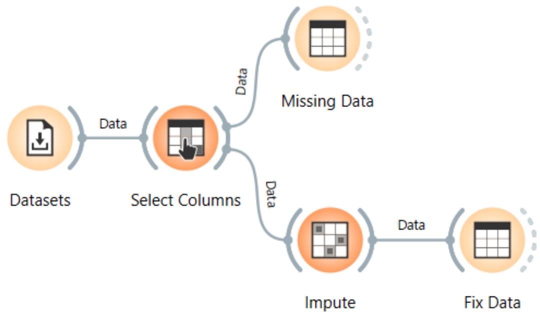

Figure 2. Lost data repair design

2.3. Feature Selection (FS)

One of the most important things in classification is determining features to get the best performance results [11]. Datasets used in ML processes usually contain redundant and irrelevant features, and do not improve accuracy [12], has no effect on the learning model, and may even degrade the performance of the learning model [8], therefore it is important to perform relevant feature analysis. In this case we apply the Gini Decrease, and ReliefF feature selection models. These two techniques have good capabilities in selecting features by giving weight to each feature and displaying them in a ranking[13].

a. Gini Decrease (GD)

Gini decrease feature selection or (Mean Decrease Gini Impurity) is a model-based feature selection method, mainly used in the Random Forest algorithm, to measure the level of importance of a feature based on its ability to reduce data impurity in the Decision Tree[14]. This method is used to reduce feature dimensions, deal with high-dimensional data (such as microarrays), and improve classification performance (for example, in KNN or Naive Bayes) by only using crucial features. provides a multivariate feature importance score that is relatively inexpensive computationally and very effective for removing noise from less relevant features [15].

b. ReliefF

Relief feature selection is a filter-based feature selection algorithm that is used to measure the quality or relevance of features in a dataset by assigning weight to each feature[16]. The main goal is to simplify the model by identifying the most influential features, reducing irrelevant features, and improving classification performance. ReliefF is a development of the basic Relief algorithm, which is more flexible because it can handle datasets with many classes (multiclass) and missing data. Relief is very sensitive to irrelevant features, easy to implement, and can handle interactions between features well [17].

2.4. Performance Machine Learning Evaluation

This study examines how the Confusion Matrix can be used to measure accuracy and error rates. Confusion matrix is a table used to evaluate the performance of classification models in ML. This table compares the model's predicted results with the actual (actual) values from the data, so it can provide an in-depth picture of the advantages and disadvantages of the mode [18]. The confusion matrix variable is displayed in Table 2.

Table 2. Confusion Matrix

Class	Positive Prediction	Negative Prediction
Positive Actual	Number of True Positive (TP)	Number False Negative (FN)
Negative Actual	Number of False Positive (FP)	Number True Negative (TN)

Accuracy is a method for evaluating the performance of an ML model. These variables can be obtained from the Confusion Matrix in Table 3 and calculated using equation (1).

$$Accuracy = \frac{TP + TN}{TP + FP + FN + TN} \tag{1}$$

2.5. Cross-validation (k-fold)

Classification is a form of data mining technique that is currently popular [19]. This strategy uses various methods to assess available data to produce Heart predictions [11]. The classification model will be validated using k-fold cross-validation. The cross-validation method is generally used for training sets [20]. Figure 3 displays the cross-validation procedure

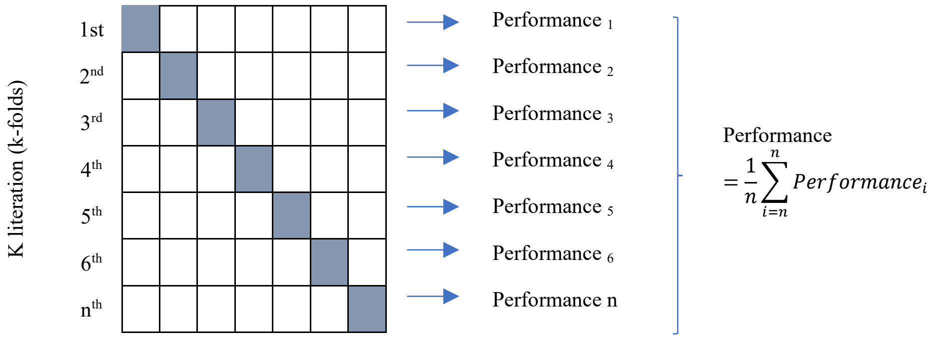

Figure 3. Procedure k-fold Validation

3 RESULTS AND DISCUSSION

3.1. Feature Selection Evaluation

The process of evaluating feature selection techniques and testing model performance was carried out using software (Orange 3.39.0). We implemented data separation into two parts, namely training set 80%, and testing set 20%, and used fold validation (k=10). Application of feature selection techniques using GD and ReliefF, and model ML as Naïve Bayes, SVM, Tree, and Random Forest. Model SVM uses a Sigmoid kernel configuration, number tolerance 0.00 10, iteration limit 10, Cost 1.00, and Regression loss epsilon 0.10. Meanwhile, for the Tree model, we use the configuration Number of instances level 2, split subset 5, maximum tree depth limit 100, and majority 95%. Our Random Forest model uses a configuration of number of trees 10, limit depth 3, number of attributes considered 5. The ML model test design is shown in Figure 4.

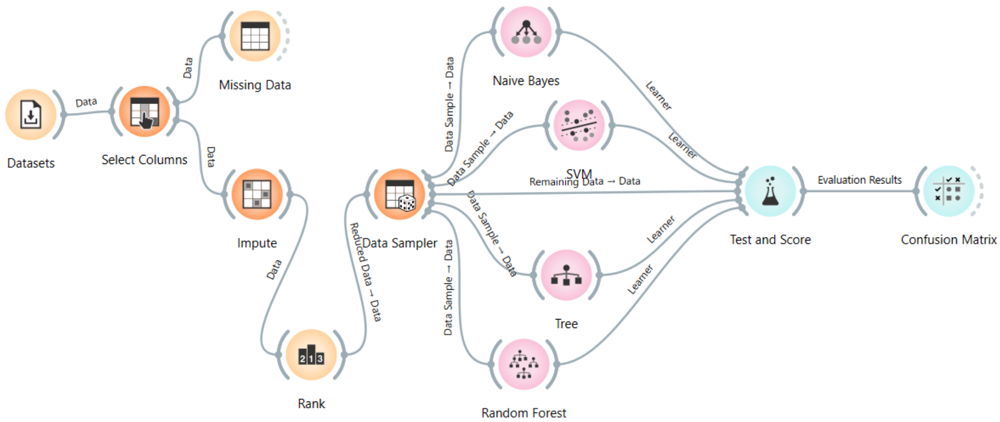

Figure 4. Design for testing feature selection techniques and ML models

Based on Figure 4, we get the features that have the most influence on the class and have the highest weight (W>0.076). The feature selection results are shown in Table 3.

Table 3. Ranking weighting (w) feature selection techniques (f)

Rank	Gini Decrease (GD)		Relief-F	
	Features	**Weight**	**Features**	**Weight**
1	chest pain	0.134	chest pain	0.202
2	major vessels colored	0.114	gender	0.148
3	ST by exercise	0.095	slope peak exc ST	0.106
4	exerc ind ang	0.093	major vessels colored	0.095
5	max HR	0.081	exerc ind ang	0.092
6	slope peak exc ST	0.075	rest ECG	0.078
7	chest pain	0.134	thal	0.076

3.2. Performance Model Evaluations

The selection results using two feature selection techniques produce features with different rankings. The GD and ReliefF techniques produce seven features that have the highest weight. Next, we carry out a performance comparison using all features and feature selection. The comparison test results are shown in Table 4.

Table 4. Performance comparison of ML models using feature selection

No	Model	All Features	GD	Relief-F	Average
1	Tree	0.700	0.683	0.767	0.717
2	Random Forest	**0.817**	0.750	**0.817**	0.794
3	Naive Bayes	0.750	0.800	0.800	0.783
4	SVM	**0.817**	**0.833**	0.800	**0.817**
	Average	**0.771**	**0.767**	**0.796**	

We found differences in the accuracy of each ML model's performance. In this case, the model performance is more optimal using the ReliefF feature selection technique with an average value of 0.796. The application of the GD technique shows the best performance on the SVM model, while the application of the ReliefF technique shows the best performance on the Random Forest model. Detailed ML model performance based on average values is shown in Figure 5.

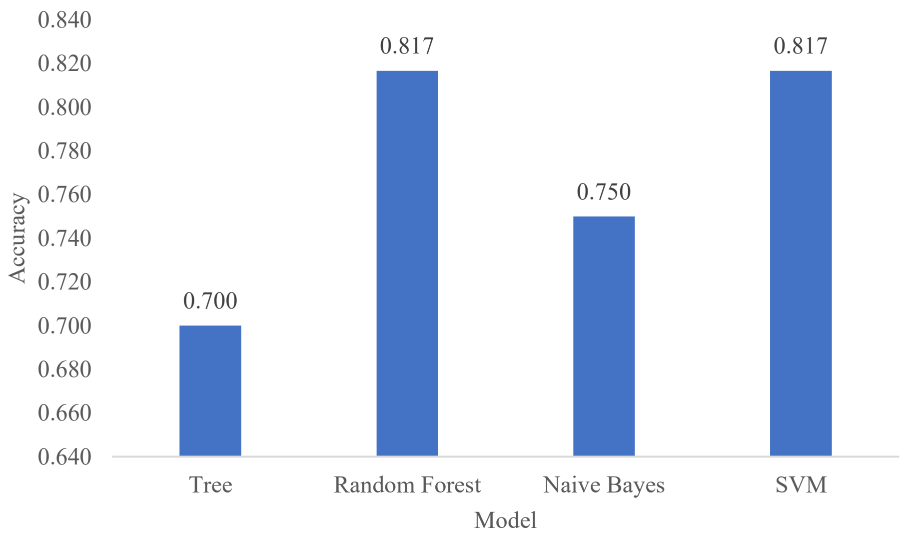

Figure 5. Accuracy model machine learning

This is because SVM can handle high-dimensional data. SVM works by finding the optimal hyperplane to maximize the margin between classes, making it a powerful model for classification and regression. Thanks to its ability to find the maximum separation limit (margin), SVM often produces more accurate predictions than other ML models on certain datasets. The Random Forest model can handle large data with high dimensions, tolerance for outliers and missing values, as well as flexibility for classification and regression tasks. Next, we carried out a comparative analysis of the F1-Score variable shown in Figure 6.

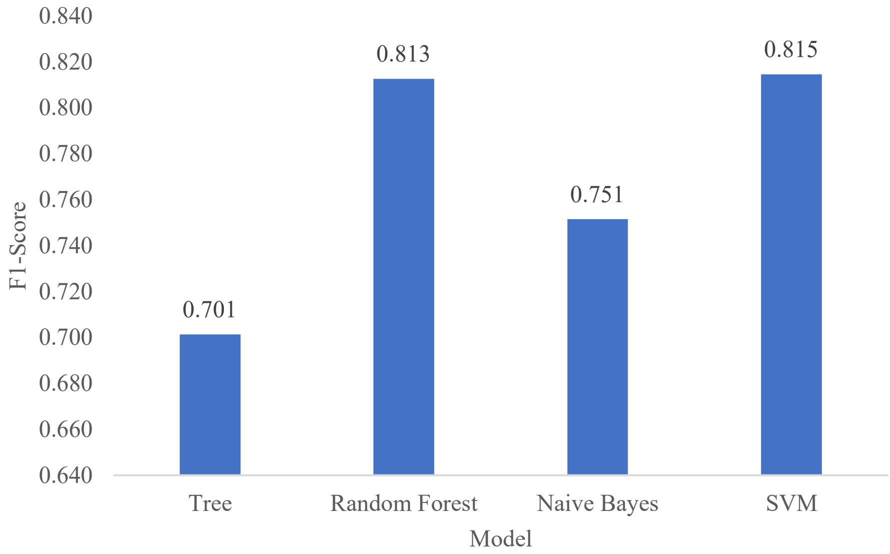

Figure 6. F1-Score model machine learning

Next, we conducted a comparative analysis of the performance of the two feature selection techniques using the average accuracy values shown in Figure 7.

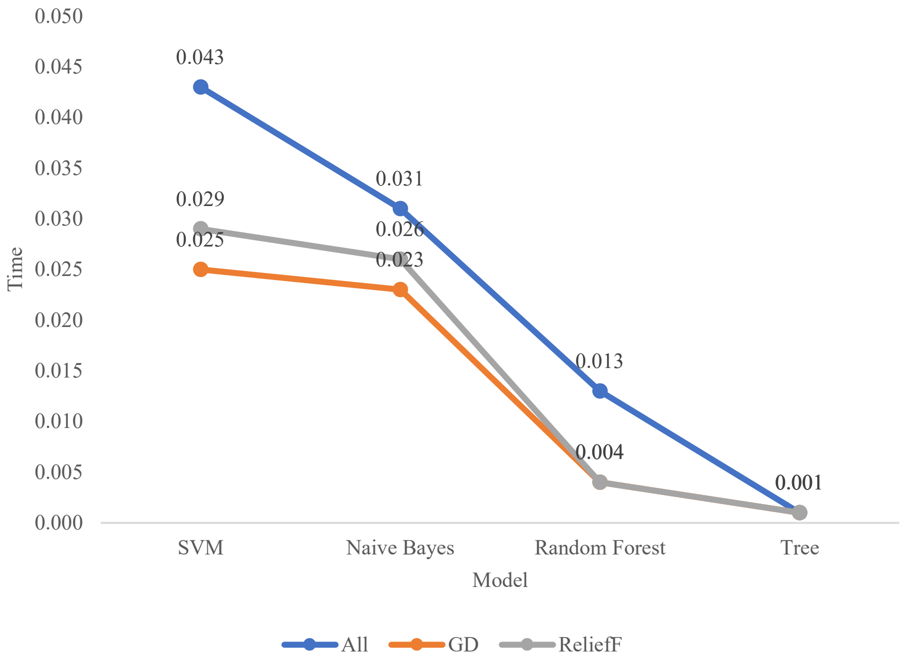

Figure 7. Comparison of model computing times

Figure 7 shows that the Tree model shows the lowest or fastest computing time followed by Random Forest, while the SVM model takes the longest time. The results of calculating the average time required by the feature selection technique, the ReliefF technique shows the fastest time. This is because when making predictions, the model only needs to follow the path from the root node to the leaf node based on the if-else rule. Computation time for prediction is generally logarithmic to the amount of training data, i.e. O (\log n), which is very fast even for large datasets. Next, we carried out a performance comparison based on the average accuracy value of the feature selection techniques shown in Figure 8.

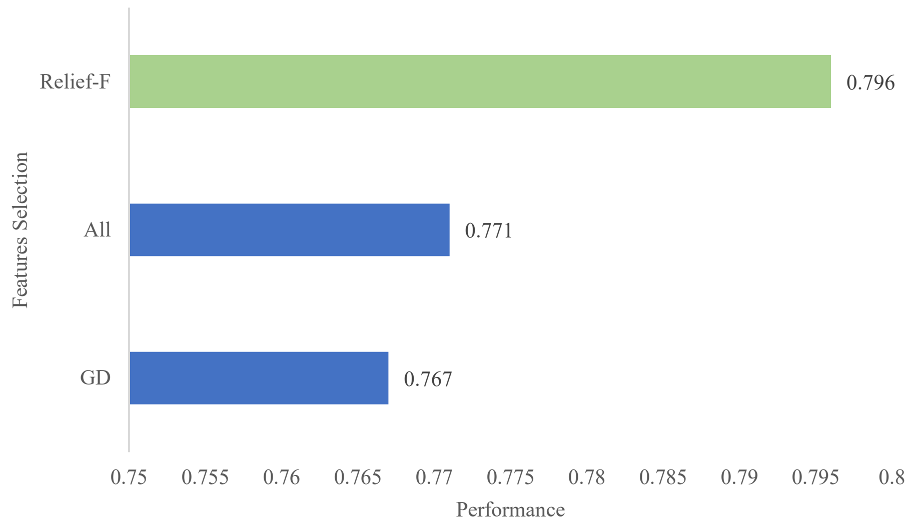

Figure 8. Feature selection performance comparison

Figure 8 shows the performance of different selection techniques. The ReliefF technique performs best in this case. This is due to its ability to efficiently identify the most relevant features, handle noisy data, and consider interactions between features without the need for high computational complexity. Relief works by evaluating feature quality based on how well the feature can differentiate samples from different classes that are within proximity. These findings indicate that the choice of feature selection techniques and ML models is an important step in the case of medical classification.

3.3. Discussion

The findings of this study demonstrate that feature selection plays a crucial role in improving the performance of ML models for heart disease classification. The results reveal that different feature selection methods produce distinct subsets of influential features, which directly affect classification performance. Both GD and ReliefF identified several important features, such as chest pain, exercise-induced angina, major vessels colored, and slope peak exercise ST, indicating that these variables significantly contribute to heart disease prediction.

The experimental results indicate that ReliefF achieved better overall performance than GD, with an average accuracy of 0.796. This suggests that ReliefF is more effective in identifying relevant features while minimizing the impact of irrelevant or noisy attributes. ReliefF evaluates feature importance by considering the relationship between neighboring instances of different classes, allowing it to better capture complex interactions among variables. This result is consistent with previous studies that reported ReliefF as an efficient method for improving classification performance in medical datasets.

Among the evaluated ML models, SVM showed the highest accuracy when combined with GD, achieving 0.833. This performance can be attributed to SVM's capability to handle high-dimensional data and identify optimal decision boundaries through maximum margin separation. Meanwhile, Random Forest demonstrated stable and competitive performance using ReliefF, achieving an accuracy of 0.817. The ensemble-based nature of Random Forest contributes to better robustness against noise and overfitting.

In terms of computational efficiency, the Tree model exhibited the shortest processing time due to its simple decision-rule structure. Although SVM produced high predictive accuracy, it required longer computational time than other models. Therefore, selecting an appropriate combination of feature selection methods and ML models should consider both predictive performance and computational efficiency, especially in real-world medical applications where fast and reliable diagnosis is essential.

4 CONCLUSION

Based on the results of the comparative analysis that has been carried out, it can be concluded that the GD and ReliefF feature selection techniques in the Heart Dataset have an impact on the performance of the classification model. The GD and ReliefF techniques produce seven features that have the highest weight. Based on the selected features, we evaluate the ML model to obtain comprehensive information on model performance. The results of the performance comparison test of the two FS techniques, we found that the ReliefF technique showed the best performance. Apart from that, the results of the evaluation of the ML model showed that the performance of the SVM model showed optimal performance, followed by Random Forest, while the Tree model showed the worst performance. This is because SVM can handle high-dimensional data. SVM works by finding the optimal hyperplane to maximize the margin between classes, making it a powerful model for classification and regression. Thanks to its ability to find the maximum separation limit (margin), SVM often produces more accurate predictions than other ML models on certain datasets. Even though the Tree model shows poor performance, in the computing time test, this model shows the fastest performance. This is because when making predictions, the model only needs to follow the path from the root node to the leaf node based on the if-else rule. Computation time for prediction is generally logarithmic to the amount of training data, i.e. O (\log n), which is very fast even for large datasets.

Feature selection strategies should be an integral part of the medical classification system development pipeline, as they can influence the overall interpretability and detection accuracy. Further research is recommended to evaluate the effectiveness of other FS handling techniques such as Information Gain, and Gain Ratio, as well as applying this approach to larger and more complex clinical datasets.

5 ACKNOWLEDGMENTS

The author would like to express his deepest gratitude to the STKIP Rosalia Metro, for helping validate the anonymous breast heart patient data used in this research. We also thank the Machine Learning (ML) Research Group for their valuable input during the model development and evaluation phase. Special awards are given to doctors whose clinical insight contributes significantly to the interpretation of breast heart prediction results

REFERENCES

[1] S. Dixit and R. Kala, "Early detection of heart diseases using a low-cost compact ECG sensor," *Multimed. Tools Appl.*, vol. 80, no. 21–23, pp. 32615–32637, 2021, doi: 10.1007/s11042-021-11083-9.

[2] I. Miramontes and P. Melin, "Sugeno-Based Fuzzy Systems for Accurate Heart Rate Level Classification BT - Intelligent System Design Based on Soft Computing Models," O. Castillo and P. Melin, Eds., Cham: Springer Nature Switzerland, 2025, pp. 3–14. doi: 10.1007/978-3-031-97309-3_1.

[3] A. Vianello *et al.*, "Medical Expert System for Intelligent Telemonitoring of Patients with Chronic Heart Failure: Preliminary Validation and Perspectives," 2025, *ahajournals.org*. doi: 10.1161/CIRCHEARTFAILURE.124.012478.

[4] N. Fazakis, O. Kocsis, E. Dritsas, S. Alexiou, N. Fakotakis, and K. Moustakas, "Machine Learning Tools for Long-Term Type 2 Diabetes Risk Prediction," *IEEE Access*, vol. 9, pp. 103737–103757, 2021, doi: 10.1109/ACCESS.2021.3098691.

[5] B. Mahesh, "Machine Learning Algorithms - A Review," *Int. J. Sci. Res.*, vol. 9, no. 1, pp. 381–386, 2020, doi: 10.21275/art20203995.

[6] W. Hanon, H. A. Al Essa, and S. H. Jihad, "A Novel Hybrid Model Combining Feature Selection and Imbalance Handling for Prediction of Heart Failure Survival," *Ing. des Syst. d'Information*, vol. 30, no. 6, pp. 1459–1467, 2025, doi: 10.18280/isi.300605.

[7] S. S. Rautaray, S. Dey, M. Pandey, and M. K. Gourisaria, "Nuclei segmentation in cell images using fully convolutional neural networks," *Int. J. Emerg. Technol.*, vol. 11, no. 3, pp. 731–737, 2020, [Online]. Available: https://www.scopus.com/inward/record.uri?eid=2-s2.0-85087307723&partnerID=40&md5=4356c05335e3af18806af217ccfe5d55

[8] S. Bashir, Z. S. Khan, F. H. Khan, A. Anjum, and K. Bashir, "Improving Heart Disease Prediction Using Feature Selection Approaches," in *2019 16th International Bhurban Conference on Applied Sciences and Technology (IBCAST)*, 2019, pp. 619–623. doi: 10.1109/IBCAST.2019.8667106.

[9] S. S, R. Bhalla, and G. Singh, "Feature Selection and Model Evaluation for Heart Disease Prediction Using Ensemble Methods," *Procedia Comput. Sci.*, vol. 259, pp. 1282–1295, 2025, doi: https://doi.org/10.1016/j.procs.2025.04.083.

[10] M. A. Khan *et al.*, "Optimal feature selection for heart disease prediction using modified Artificial Bee colony (M-ABC) and K-nearest neighbors (KNN)," *Sci. Rep.*, vol. 14, no. 1, p. 26241, 2024, doi: 10.1038/s41598-024-78021-1.

[11] H. Sulistiani, A. Syarif, K. Muludi, and Warsito, "Performance evaluation of feature selections on some ML approaches for diagnosing the narcissistic personality disorder," *Bull. Electr. Eng. Informatics*, vol. 13, no. 2, pp. 1383–1391, 2024, doi: 10.11591/eei.v13i2.6717.

[12] J. Wang, S. Zhou, Y. Yi, and J. Kong, "An improved feature selection based on effective range for classification," *Sci. World J.*, vol. 2014, 2014, doi: 10.1155/2014/972125.

[13] W. Van Casteren, "The Waterfall Model And The Agile Methodologies : A Comparison By Project Characteristics-Short The Waterfall Model and Agile Methodologies," *Acad. Competences Bachelor*, no. February, pp. 10–13, 2017, [Online]. Available: https://www.researchgate.net/publication/313768860

[14] M. A. Elsadig, A. Altigani, and H. T. Elshoush, "Breast cancer detection using machine learning approaches: a comparative study," *Int. J. Electr. Comput. Eng.*, vol. 13, no. 1, pp. 736–745, 2023, doi: 10.11591/ijece.v13i1.pp736-745.

[15] Y. Zhang *et al.*, "Feature selection based on neighborhood rough sets and Gini index," *PeerJ Comput. Sci.*,

vol. 9, p. e1711, 2023, doi: 10.7717/peerj-cs.1711.

[16] N. Aggarwal *et al.*, “Mean based relief: An improved feature selection method based on ReliefF,” *Appl. Intell.*, vol. 53, no. 19, pp. 23004–23028, 2023, doi: 10.1007/s10489-023-04662-w.

[17] C. Karima and W. Anggraeni, “Performance Analysis of the Ada-Boost Algorithm For Classification of Hypertension Risk With Clinical Imbalanced Dataset,” *Procedia Comput. Sci.*, vol. 234, pp. 645–653, 2024, doi: https://doi.org/10.1016/j.procs.2024.03.050.

[18] M. Ohsaki, P. Wang, K. Matsuda, S. Katagiri, H. Watanabe, and A. Ralescu, “Confusion-matrix-based kernel logistic regression for imbalanced data classification,” *IEEE Trans. Knowl. Data Eng.*, vol. 29, no. 9, pp. 1806–1819, 2017, doi: 10.1109/TKDE.2017.2682249.

[19] F. Islam, R. Ferdousi, S. Rahman, and H. Y. Bushra, *Likelihood Prediction of Diabetes at Early Stage Using Data Mining Techniques*. London, 2019. doi: 10.1007/979-981-13-8798-2-12.

[20] T. Yan, S.-L. Shen, A. Zhou, and X. Chen, “Prediction of geological characteristics from shield operational parameters by integrating grid search and K-fold cross validation into stacking classification algorithm,” *J. Rock Mech. Geotech. Eng.*, vol. 14, no. 4, pp. 1292–1303, 2022, doi: https://doi.org/10.1016/j.jrmge.2022.03.002.