Tech : Journal of Engineering Science
Vol. 1, No. 1, 2025, P. 14-25
e-ISSN: __________ (online)
https://jurnal.sinesia.id/index.php/tech/index

Impact of Feature Selection on the Performance of KNN and SVM in Heart Disease Prediction

Dhiyaussalam[1], M. Helmy Noor[2], Herlinawati[3], Isna Wardiah[4]
E-mail Korespondensi : salam@poliban.ac.id
[1,3,4]Teknik Informatika, Politeknik Negeri Banjarmasin, Banjarmasin, Indonesia
[2]Sistem Informasi Kota Cerdas, Politeknik Negeri Banjarmasin, Banjarmasin, Indonesia

Info Article
| **Submitted**: 5 June 2025 | **Revised**: 25 June 2025 | **Accepted**: 26 June 2025

How to Cite : Dhiyaussalam, etc., “Impact of Feature Selection on the Performance of KNN and SVM in Heart Disease Prediction”, *Tech : Journal of Engineering Scince*, Vol. 1, No. 1, 2025, P. 14-25.

ABSTRACT

Feature selection plays a vital role in enhancing the performance of machine learning models by eliminating irrelevant or redundant attributes. This study investigates the impact of feature selection on the classification accuracy of K-Nearest Neighbors (KNN) and Support Vector Machine (SVM) in predicting heart disease. Using the UCI Heart Disease data, which initially includes 13 input features, feature importance scores were calculated using a Random Forest model. A threshold-based method was then applied to identify and retain the most informative features. Through iterative testing of importance thresholds, a value of 0.03 yielded the best results, reducing the feature set from 13 to 9 attributes. Classification models were trained and evaluated using full and reduced feature sets. Performance was assessed using accuracy, precision, recall, and F1-score and validated with 5-fold cross-validation. The results demonstrate significant performance gains after feature selection. The KNN classifier improved accuracy from 83% to 92%, with notable gains in recall and F1-score for the positive class. Similarly, SVM achieved 92% accuracy, with improved precision and overall performance stability. These findings suggest that data-driven feature reduction simplifies the model and enhances its predictive power. This study systematically compares feature selection effects on two distinct machine learning algorithms and offers practical insights for optimizing medical prediction models in clinical decision support systems.

Keyword: *feature selection, heart disease, KNN, machine learning, SVM.*

ABSTRAK

Feature selection memainkan peran penting dalam meningkatkan kinerja model *machine learning* dengan mengurangi atribut yang tidak relevan atau berlebihan. Studi ini menyelidiki dampak pemilihan fitur pada akurasi klasifikasi K-Nearest Neighbors (KNN) dan Support Vector Machine (SVM) dalam memprediksi penyakit jantung. Menggunakan data UCI Heart Disease, yang awalnya mencakup 13 fitur, skor *feature importance* dihitung menggunakan model Random Forest. Metode berbasis ambang batas kemudian diterapkan untuk mengidentifikasi dan mempertahankan fitur yang paling informatif. Melalui pengujian iteratif ambang batas kepentingan, nilai 0,03 menghasilkan hasil terbaik, mengurangi set fitur dari 13 menjadi 9 atribut. Model klasifikasi dilatih dan dievaluasi menggunakan set fitur lengkap dan tereduksi. Kinerja dinilai menggunakan akurasi, presisi, recall, dan skor F1 dan divalidasi dengan *5-fold cross-validation*. Hasilnya menunjukkan peningkatan kinerja yang signifikan setelah *feature selection*. Pengklasifikasi KNN meningkatkan akurasi dari 83% menjadi 92%, dengan peningkatan yang nyata dalam recall dan skor F1 untuk kelas positif. Demikian pula, SVM mencapai akurasi 92%, dengan presisi yang lebih baik dan stabilitas kinerja secara keseluruhan. Temuan ini menunjukkan bahwa pengurangan fitur berdasarkan data menyederhanakan model dan meningkatkan daya prediktifnya. Studi ini secara sistematis membandingkan efek pemilihan fitur pada dua algoritma pembelajaran mesin yang berbeda dan menawarkan wawasan praktis untuk mengoptimalkan model prediksi medis dalam sistem pendukung keputusan klinis.

Kata Kunci: *seleksi fitur, penyakit jantung, KNN, machine learning, SVM.*

Introduction

Cardiovascular diseases, particularly heart disease, remain the primary cause of death globally, accounting for an estimated 17.9 million deaths annually according to the World Health Organization (Bianco et al., 2021). Early diagnosis and effective risk stratification are therefore crucial in reducing mortality and improving patient outcomes. In recent years, the application of machine learning techniques in medical diagnostics has gained considerable attention due to their ability to analyze complex patterns in clinical datasets and support decision-making in a data-driven manner (Nagavallika, 2022).

Among various classification techniques, models such as K-Nearest Neighbors (KNN) and Support Vector Machine (SVM) are widely used because of their effectiveness in classifying nonlinear and high-dimensional data (Eswar & Karthick, 2022). However, the predictive performance of these models heavily depends on the quality and relevance of the features used during training. In real-world medical datasets, it is common to encounter irrelevant, redundant, or weakly informative features, which can degrade model accuracy, increase computational cost, and even introduce biases in prediction (Ma et al., 2022).

Despite their proven utility, the performance of KNN and SVM is highly dependent on the quality and relevance of the input features. Irrelevant or redundant features can degrade classification accuracy, increase computational cost, and lead to overfitting (Andono et al., 2021). To address this, feature selection is commonly applied to retain only the most informative attributes. One effective approach is to use feature importance scores generated by ensemble models such as Random Forest, which quantify each feature's contribution to the classification task (Liang et al., 2022).

Feature selection has emerged as a key preprocessing step to address this issue. By identifying and retaining only the most informative features, one can simplify the model, improve generalization, and reduce overfitting. A practical and interpretable method of performing feature selection is through the use of feature importances, particularly those obtained from ensemble-based models such as Random Forests. This approach ranks features based on their contribution to the model's decision-making process, enabling the identification of low-impact features that can be removed without significantly compromising accuracy (Aziz et al., 2021).

Recent years have witnessed a surge of interest in applying machine learning techniques for heart disease prediction, particularly using K-Nearest Neighbors (KNN) and Support Vector Machine (SVM). Kodete (2022) proposed a robust hybrid framework that integrates SVM and KNN through a voting ensemble classifier, optimized using Lasso regression and genetic algorithms. By employing

GridSearchCV for hyperparameter tuning, their model achieved a remarkable accuracy of 94.87% and an AUC score of 0.9926, demonstrating the strength of combining models with tailored feature selection. Similarly, Abidin et al. (2025) focused on the classification of cardiovascular diseases using SVM on a standard clinical dataset, highlighting its effectiveness in identifying heart disease despite the lack of additional optimization techniques.

Further supporting this trend, Nasution et al. (2025) conducted a comparative evaluation of Logistic Regression, Random Forest, SVM, and KNN using the UCI Heart Disease data. Their findings showed that both SVM and KNN performed competitively, especially when paired with appropriate feature selection strategies. In another study, Osei-Nkwantabisa & Ntumy (2024) compared several classifiers and reported that KNN produced reliable predictions in classifying heart disease, reinforcing its utility in clinical diagnostics. Lastly, Ingole et al. (2024) emphasized early detection and risk stratification using various algorithms, where SVM and KNN stood out due to their balance between interpretability and performance. These studies consistently highlight that feature relevance and parameter tuning are key factors in enhancing classification accuracy in heart disease prediction.

While numerous studies have successfully applied KNN and SVM for heart disease classification, most focus primarily on model comparison or parameter optimization. However, few have explored how feature selection, particularly based on feature importance from ensemble models, directly affects the performance of individual classifiers. Moreover, the varying impact of such feature reduction on distance-based models like KNN versus margin-based models like SVM remains underexplored.

This highlights a research gap regarding the systematic analysis of how feature selection contributes to classifier performance in medical prediction tasks. This study aims to investigate the impact of feature selection (derived from Random Forest importance scores) on the classification performance of two widely used algorithms, K-Nearest Neighbors (KNN) and Support Vector Machine (SVM), in predicting heart disease. Using the UCI Heart Disease dataset, this research compares model performance before and after feature reduction, employing cross-validation to ensure robustness. The results are expected to provide insights into how data-driven feature reduction can enhance predictive modeling in clinical decision support systems.

Methodology

This study was conducted to assess the impact of feature selection, based on feature importance scores, on the performance of two classification algorithms: K-Nearest Neighbors (KNN) and Support Vector Machine (SVM). These two algorithms were selected due to their contrasting learning mechanisms and frequent usage in clinical classification tasks. KNN is a simple, instance-based learning algorithm that classifies samples based on the majority label of their nearest neighbors in the feature space. It is known for its interpretability and effectiveness when the number of relevant features is properly optimized (Jiang et al., 2022). On the other hand, SVM is a margin-based classifier that constructs an optimal hyperplane to separate classes and performs well in high-dimensional spaces, especially with kernel functions (Kadhim & Kamil, 2023).

Both algorithms have been widely used in heart disease prediction studies due to their solid baseline performance and low computational cost. Comparing the effect of feature selection on these two fundamentally different models provides valuable insights into how data dimensionality interacts with model structure. Moreover, as one is sensitive to feature scaling and noise (KNN), and the other is relatively more robust (SVM), evaluating both enables a more comprehensive understanding of the benefits of feature selection in clinical data analysis.

The experiment was carried out using the Heart Disease dataset from the UCI Machine Learning Repository, which contains clinical features relevant to cardiovascular risk, such as age, sex, resting blood pressure, cholesterol levels, and others (Janosi et al., 1989).

2.1 Dataset

The dataset comprises 303 records and 14 attributes, including 13 input features and one target variable. The features represent a mix of demographic, clinical, and test result data relevant to heart disease diagnosis. Table 1 summarizes each attribute along with its description and data type.

Table 1. Description of Features in the Heart Disease Dataset

No.	Feature	Description	Data Type
1	age	Age of the patient (in years)	Integer
2	sex	Gender (1 = male; 0 = female)	Binary
3	cp	Chest pain type (0–3)	Categorical
4	trestbps	Resting blood pressure (in mm Hg)	Integer
5	chol	Serum cholesterol (in mg/dl)	Integer
6	fbs	Fasting blood sugar > 120 mg/dl (1 = true; 0 = false)	Binary
7	restecg	Resting electrocardiographic results (0–2)	Categorical
8	thalach	Maximum heart rate achieved	Integer
9	exang	Exercise-induced angina (1 = yes; 0 = no)	Binary
10	oldpeak	ST depression induced by exercise	Float
11	slope	Slope of peak exercise ST segment (1–3)	Categorical
12	ca	Number of major vessels (0–3) colored by fluoroscopy	Integer
13	thal	Thalassemia (3 = normal; 6 = fixed defect; 7 = reversible defect)	Categorical
14	num/target	Presence of heart disease (0 = no; 1 = yes)	Binary

For the purpose of binary classification, the target variable was transformed into two classes:

- 0: Absence of heart disease
- 1: Presence of heart disease (any value from 1 to 4)

Missing values in features such as ca and thal were handled by removing rows containing null entries, resulting in a cleaned data of 297 instances. All features were then used for the initial modeling phase.

2.2 Experimental Workflow

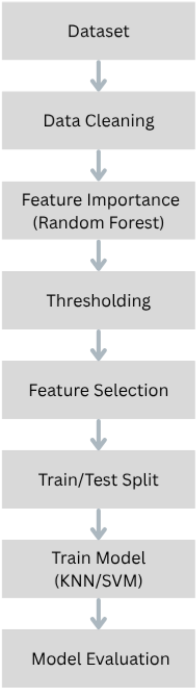

Figure 1. Experimental workflow of the heart disease classification process.

The experimental workflow adopted in this study is illustrated in Figure 1. The process begins with obtaining and preprocessing the UCI Heart Disease data, including handling missing values and verifying data consistency. Feature importance scores were then calculated using a Random Forest classifier to assess the contribution of each feature to the prediction task. Based on the resulting scores, a thresholding mechanism was applied to filter out low-importance features. The dataset, both in its original and reduced forms, was subsequently normalized using a standard scaler to ensure fair input to the learning algorithms. An 80:20 train-test split was performed with a fixed random state to ensure reproducibility. Classification models, namely K-Nearest Neighbors (KNN) and Support Vector Machine (SVM), were then trained and evaluated using both versions of the dataset. To ensure robustness of the results, 5-fold cross-validation was used in addition to test-set evaluation.

2.2.1 Feature Importance Analysis

A Random Forest classifier was trained on the cleaned data to extract feature importance scores. These scores were then used to evaluate the relative contribution of each feature toward predicting the target class.

2.2.2 Threshold-based Feature Selection

The thresholding process was conducted before splitting the dataset into training and testing sets. This decision was made to ensure that the same selected features were used consistently across both subsets and to enable a fair comparison between models trained on all features versus those trained on reduced feature sets. A fixed threshold value was applied to filter features based on their importance scores, and the resulting feature-reduced data was then normalized and split using an 80:20 ratio with a fixed random state for reproducibility.

A range of threshold values (from 0.01 to 0.10, incremented by 0.01) was used to eliminate features with importance scores below each threshold iteratively. Empirical precedents in feature selection literature guided the choice of this range. For instance, Prasetiyowati et al., (2021) applied incremental thresholds based on the standard deviation of feature importance scores (derived from Information Gain). They showed significant accuracy improvements compared to a fixed threshold of 0.05.

By adopting this exploratory threshold range, we systematically assessed how progressive feature reduction, from full (13) to minimal inputs, affects classification outcomes. For each threshold level, the number of retained features was recorded, facilitating a quantitative analysis of the relationship between feature count and model performance in subsequent stages.

2.2.3 Model Training and Evaluation

For every subset of selected features, the data was split into training and testing sets (80/20 split). The data splitting process was conducted using a fixed random_state=42 to ensure reproducibility. Additionally, models that include randomness, such as Random Forest, were initialized with the same random state. Features were standardized using StandardScaler since both KNN and SVM are sensitive to feature scale. Then:

- A K-Nearest Neighbors (KNN) classifier was trained with default parameters (k=5).
- A Support Vector Machine (SVM) classifier with an RBF kernel was trained using the C=1.0 and gamma='scale' parameters.

2.3 Performance Comparison

After evaluating all threshold levels, the feature subset that yielded the highest average accuracy was selected as the optimal configuration. Both KNN and SVM models were then trained and tested using only the features retained under the optimal threshold. Their classification performance was compared with models trained on the full feature set, using metrics such as accuracy, precision, recall, F1-score, and confusion matrix.

Results and discussion

This section presents the results of classification experiments using K-Nearest Neighbors (KNN) and Support Vector Machine (SVM) algorithms before and after applying feature selection based on Random Forest feature importances. The aim is to assess the impact of feature reduction on model performance, using the UCI repository as a case study.

3.1 Feature Importance Analysis and Optimal Threshold

A Random Forest classifier was trained on the full feature set to calculate the importance of each attribute. The resulting importance scores are visualized in Figure 1, which ranks the features based on their contribution to the classification task.

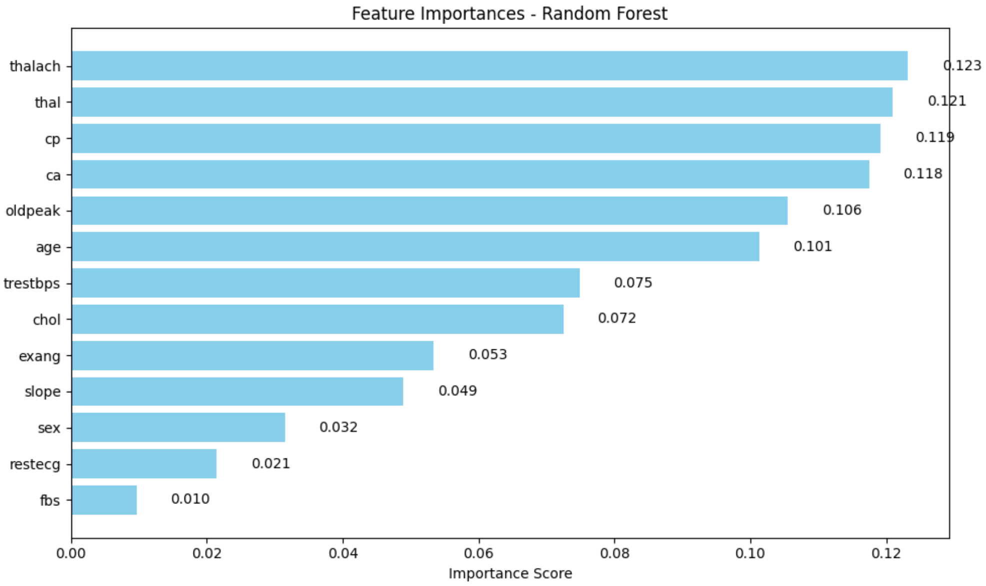

Figure 2. Feature importance scores for heart disease prediction based on Random Forest.

A series of cut-off values ranging from 0.01 to 0.10 were tested to determine the optimal threshold for feature selection. Features with importance scores below the threshold were excluded at each threshold level. Table 2 presents the number of features retained at each threshold. As expected, higher thresholds resulted in fewer features being selected. This allowed a trade-off analysis between model simplicity and prediction performance in subsequent classification stages.

The analysis identified the 0.03 threshold, which retained nine features, as optimal based on the highest average classification accuracy observed in both KNN and SVM models.

Table 2. Number of Features Retained at Various Importance Thresholds

Threshold	Number of Features Retained
0.01	13
0.02	12
0.03	9

0.04	7
0.05	6
0.06	5
0.07	4
0.08	3
0.09	2

The most influential features identified were thalach, thal, cp, and ca, each with an importance score above 0.11. Meanwhile, features such as fbs, restecg, and sex showed low contributions with scores below 0.04. Based on these results, a threshold of 0.03 was selected, resulting in the removal of four less informative features: fbs, restecg, sex, and slope. The reduced feature set thus retained 11 features deemed most relevant to heart disease prediction.

3.2 Classification Performance without Feature Selection

To evaluate baseline performance, both KNN and SVM were trained using all 13 original features. As shown in Figures 2, the models were assessed using a hold-out test set (80/20 split) consisting of 60 instances.

- The KNN classifier achieved an accuracy of 83%, with a precision of 0.82 and recall of 0.75 for class 1 (presence of heart disease), indicating some tendency toward false negatives.
- The SVM classifier performed better, achieving 88% accuracy, with class 1 precision and recall of 0.87 and 0.83, respectively.

These results suggest that while both models are effective, SVM demonstrates superior consistency and balance between precision and recall, making it more reliable in detecting heart disease.

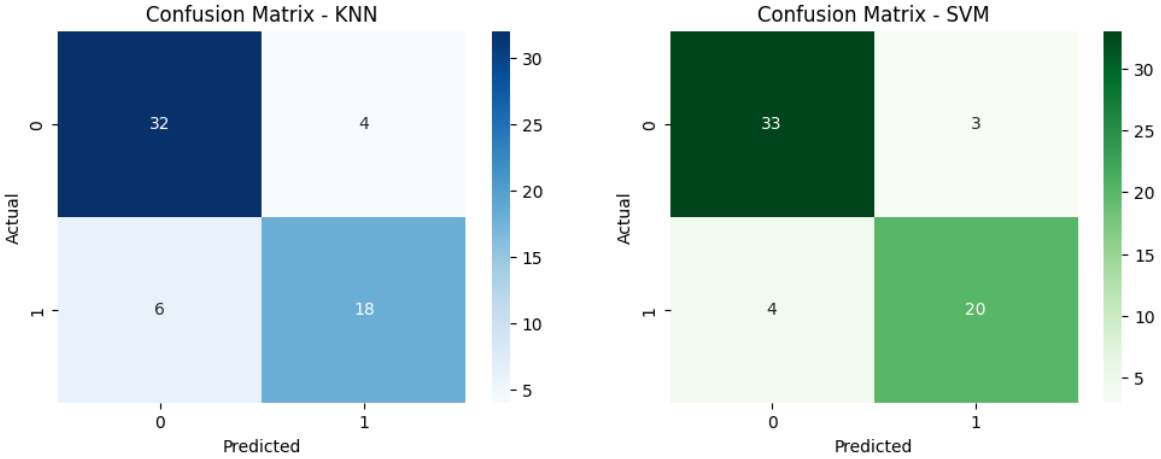

Figure 3 KNN and SVM classification report using all features

3.3 Classification Performance with Feature Selection

After applying feature selection, the same classification procedure was repeated using the reduced feature set. As seen in Figures 3, both KNN and SVM classifiers showed noticeable improvements:

- The KNN model achieved an accuracy of 92%, with significant gains in class 1 recall (from 0.75 to 0.92) and F1-score (from 0.78 to 0.90).

- The SVM model also reached 92% accuracy, with increased precision for class 1 (from 0.87 to 0.95) and improved F1-score, although recall slightly decreased (from 0.83 to 0.83).

These improvements highlight the effectiveness of eliminating less informative features. Not only does feature selection reduce dimensionality and simplify the model, but it also enhances classification performance by focusing learning on the most relevant predictors.

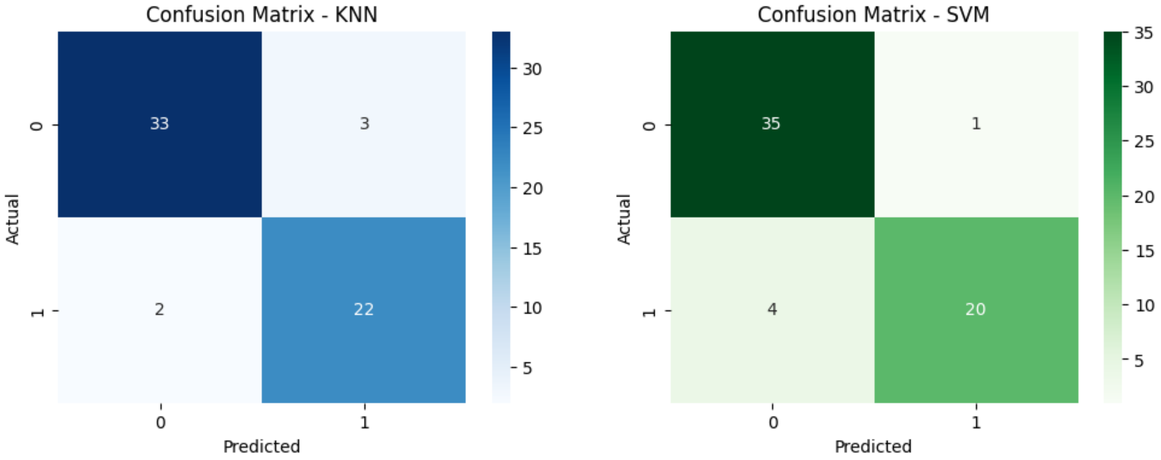

Figure 4 KNN and SVM classification report after feature selection

A summary of the classification performance of KNN and SVM before and after applying feature selection is presented in Table 3. The results indicate a consistent improvement in accuracy and F1-score, especially for the KNN model, which showed a substantial gain in class 1 recall.

Table 3 Performance Comparison of KNN and SVM Before and After Feature Selection

Model	**Feature Set**	**Accuracy**	**Precision (Class 1)**	**Recall (Class 1)**	**F1-Score (Class 1)**
KNN	All Features	83%	0.82	0.75	0.78
KNN	Selected Features	92%	0.88	0.92	0.90
SVM	All Features	88%	0.87	0.83	0.85
SVM	Selected Features	92%	0.95	0.83	0.89

3.4 Discussion

The performance comparison in Table X shows a notable improvement for both KNN and SVM classifiers after applying feature selection. The KNN model's accuracy increased from 83% to 92%, with a particularly strong gain in recall (from 0.75 to 0.92), indicating an enhanced ability to detect positive heart disease cases. Similarly, SVM achieved a precision increase from 0.87 to 0.95 and a higher F1 score, although its recall remained constant.

These findings can be explained by the fundamental differences in how KNN and SVM operate. KNN is a distance-based algorithm that heavily depends on the geometry of the feature space to determine neighbors. When irrelevant or noisy features are present, the Euclidean distance used in KNN becomes distorted, leading to poor classification. The classifier can measure similarity more accurately by eliminating less informative features, resulting in better performance. In contrast, SVM constructs an optimal separating hyperplane and employs regularization and kernel transformations to handle high-dimensional data. This makes SVM less sensitive to irrelevant features, and as a result, its performance gain from feature selection is present but not as dramatic.

The results align with previous research, such as Nasution et al., (2025), who observed that SVM and KNN performed better when paired with proper feature selection strategies. Ingole et al., (2024) also demonstrated the value of reducing input dimensionality for improving generalization, particularly in early detection systems. However, most prior works did not systematically isolate the effect of feature reduction per algorithm. This study fills that gap by quantitatively analyzing how feature selection impacts the individual behavior of KNN and SVM in the same experimental conditions.

Despite the promising findings, this study is subject to several limitations. The relatively small size of the dataset (297 records) may affect the generalizability of the results. The feature selection process was also applied using the entire dataset, which introduces the risk of information leakage. Future work should consider incorporating feature selection within the training folds of cross-validation. Moreover, only two classifiers were evaluated. Including other algorithms, such as ensemble methods, decision trees, or neural networks, would provide broader insight into the general effect of feature selection across learning paradigms.

In summary, the study confirms that feature selection can significantly enhance classifier performance, particularly for distance-sensitive models like KNN, while also offering improvements for SVM. These insights are valuable for developing more efficient and interpretable clinical decision support systems.

Conclusion

This study has demonstrated that applying feature selection based on Random Forest feature importance scores can significantly improve the performance of classification algorithms for heart disease prediction. When trained on a reduced feature set, k-Nearest Neighbors (KNN) and Support Vector Machine (SVM) showed improved accuracy, precision, and F1 scores. In particular, the KNN model benefited substantially, confirming the sensitivity of distance-based algorithms to irrelevant features in high-dimensional data.

This research's scientific contribution lies in its systematic comparison of the impact of feature selection on two fundamentally different classifiers (KNN and SVM) within a consistent and controlled experimental setting. Unlike prior studies that mainly compare algorithmic performance, this study isolates the effect of feature reduction. It reveals how different model types respond to dimensionality reduction, providing deeper insight into the interplay between data preprocessing and model behavior.

These findings have practical implications for developing more interpretable, efficient, and accurate clinical decision support systems. Medical AI applications can achieve better generalization while reducing computational complexity by focusing only on the most relevant features. For future work, this approach can be applied to larger and more diverse datasets, integrated with cross-validation-based feature selection to avoid data leakage, and extended to other classification algorithms such as ensemble methods, Naive Bayes, or deep learning. Real-world deployment scenarios may also be explored to assess usability and impact in clinical environments.

References

Abidin, M., Munzir, M., Imantoyo, A., Bintang Grendis, N. W., Hadi San, A. S., Mostfa, A. A., Furizal, F., & Sharkawy, A.-N. (2025). Classification of Heart (Cardiovascular) Disease using the SVM Method. *Indonesian Journal of Modern Science and Technology*, *1*(1), 9–15. https://doi.org/10.64021/ijmst.1.1.9-15.2025

Andono, P. N., Rachmawanto, E. H., Herman, N. S., & Kondo, K. (2021). Orchid types classification using supervised learning algorithm based on feature and color extraction. *Bulletin of Electrical Engineering and Informatics*, *10*(5), 2530–2538. https://doi.org/10.11591/eei.v10i5.3118

Aziz, F., Malek, S., Ibrahim, K. S., Shariff, R. E. R., Wan Ahmad, W. A., Ali, R. M., Liu, K. T., Selvaraj, G., & Kasim, S. (2021). Short- And long-term mortality prediction after an acute ST-elevation myocardial infarction (STEMI) in Asians: A machine learning approach. *PLoS ONE*, *16*(8 August), 1–23. https://doi.org/10.1371/journal.pone.0254894

Bianco, E., Skipalskyi, A., Goma, F., Odeh, H., Hasegawa, K., Zawawi, M. Al, Stoklosa, M., Dalmau, R., Dorotheo, E. U., Berteletti, F., Mwangi, J., & Wang, Y. (2021). E-Cigarettes: A New Threat to Cardiovascular Health - A World Heart Federation Policy Brief. *Global Heart*, *16*(1), 1–10. https://doi.org/10.5334/gh.1076

Eswar, T. S., & Karthick, V. (2022). Realtime Visual Object Recognition using Support Vector Machine comparing with K- Nearest Neighbor algorithm for improving accuracy. *Journal of Pharmaceutical Negative Results*, *13*(SO4), 831–837. https://doi.org/10.47750/pnr.2022.13.s04.097

Ingole, S., Ramineni, V., Bangad, N., Ganeeb, K. K., & Patel, P. (2024). *Advancements In Heart Disease Prediction : A Machine Learning Approach For Early Detection And Risk Assessment*. *11*(4), 164–172. https://doi.org/https://doi.org/10.48550/arXiv.2410.14738

Janosi, A., Steinbrunn, W., Pfisterer, M., & Detrano, R. (1989). *Heart Disease [Dataset]*. https://doi.org/10.24432/C52P4X

Jiang, X., Zhou, J., Qiao, X., Peng, C., & Su, S. (2022). A Neighborhood Model with Both Distance and Quantity Constraints for Multilabel Data. *Computational Intelligence and Neuroscience*, *2022*. https://doi.org/10.1155/2022/9891971

Kadhim, R. R., & Kamil, M. Y. (2023). Comparison of machine learning models for breast cancer diagnosis. *IAES International Journal of Artificial Intelligence*, *12*(1), 415–421. https://doi.org/10.11591/ijai.v12.i1.pp415-421

Kodete, C. S. (2022). Improved Heart Disease Classification Using SVM-KNN Voting Classifier and GridSearchCV Optimization. *Journal of Health, Metabolism & Nutrition Studies*, *Vol.* *19*(April). https://www.cambridgenigeriapub.com/wp-content/uploads/2025/02/BJHMNS_Vol19_No3_March_2022-21.pdf

Liang, J. X., Zhao, J. F., Sun, N., & Shi, B. J. (2022). Random Forest Feature Selection and Back Propagation Neural Network to Detect Fire Using Video. *Journal of Sensors*, *2022*. https://doi.org/10.1155/2022/5160050

Ma, X. H., Shu, L., Jia, X., Zhou, H. C., Liu, T. T., Liang, J. W., Ding, Y. S., He, M., & Shu, Q. (2022). Machine Learning-Based CT Radiomics Method for Identifying the Stage of Wilms Tumor in Children. *Frontiers in Pediatrics*, *10*(May), 1–9. https://doi.org/10.3389/fped.2022.873035

Nagavallika, V. (2022). Prediction of Heart Disease Using Machine Learning Techniques. *Journal of Contemporary Medical Practice*, *4*(6), 195–198. https://doi.org/10.53469/jcmp.2022.04(06).39

Nasution, N., Hasan, M. A., & Bakri Nasution, F. (2025). Predicting Heart Disease Using Machine Learning: An Evaluation of Logistic Regression, Random Forest, SVM, and KNN Models on the UCI Heart Disease Dataset. *IT Journal Research and Development*, *9*(2), 140–150. https://doi.org/10.25299/itjrd.2025.17941

Osei-Nkwantabisa, A. S., & Ntumy, R. (2024). *Classification and Prediction of Heart Diseases using Machine Learning Algorithms*. 1–10. http://arxiv.org/abs/2409.03697%0Ahttp://dx.doi.org/10.48550/arXiv.2409.03697

Prasetiyowati, M. I., Maulidevi, N. U., & Surendro, K. (2021). Determining threshold value on information gain feature selection to increase speed and prediction accuracy of random forest. *Journal of Big Data*, *8*(1). https://doi.org/10.1186/s40537-021-00472-4