

Cardiovascular Disease Classification Using Support Vector Machine with Feature Selection

Yuwatsiqul Aqwam^{1*}, Irwan Budi Santoso², Nur Fitriyah Ayu Tunjung Sari³

^{1,2,3} Department of Informatics Engineering, Universitas Islam Negeri Maulana Malik Ibrahim Malang, Malang, Indonesia

Corresponding author: qwamm03@gmail.com

ARTICLE INFO

Article history:

24 December 2025

Received in revised form

13 October 2026

Accepted 17 October 2026

Available online 17 October 2026

Keywords: Cardiovascular Disease, Classification, Support Vector Machine, Chi-Square, Recursive Feature Elimination.

DOI:

<https://doi.org/10.26740/jieet.v10n01.p74-87>

ABSTRACT

Cardiovascular disease is one of the most serious health problems and remains the leading cause of death worldwide each year. This condition shows that early diagnosis and classification of cardiovascular disease are very important to increase the chances of recovery and prolong the life expectancy of patients. This study aims to develop a cardiovascular disease classification system using the Support Vector Machine (SVM) method and analyze the effect of Recursive Feature Elimination (RFE) and Chi-Square feature selection on model computational performance. The dataset used is the Heart Disease Dataset from IEEE Dataport, which consists of 1,190 data points. The RFE and Chi-Square methods were used to select the most relevant features, resulting in seven main features used in the classification process with various SVM kernels. The results showed that the application of feature selection significantly improved model stability and accelerated the computational process, especially for kernels with high complexity. The linear kernel provided the highest accuracy of 85.84%, while the seventh-degree polynomial kernel increased from 77.23% to 84.31%, and the RBF kernel increased from 83.66% to 84.86% after feature selection. The combination of the RFE and Chi-Square methods proved to be effective in improving stability and computational efficiency, making it a potential alternative to support the cardiovascular disease classification process.



This work is licensed under the Creative Commons Attribution Non-Commercial-Share Alike 4.0 International License.

INTRODUCTION

Cardiovascular disease refers to disorders of the heart and blood vessels, generally caused by unhealthy lifestyles and risk factors such as hypertension, high cholesterol, and diabetes. It is often detected late because it develops slowly (Natsir, 2024). According to World Health Organization (WHO) data, cardiovascular disease remains the leading cause of death globally. In 2019, this disease resulted in over 17.9 million deaths, accounting for approximately 32% of all global deaths. Around 85% of these cases were due to heart attacks and strokes (Rachmawati et al., 2021.).

Early diagnosis and classification of cardiovascular disease present a major challenge in the medical world, as symptoms and clinical indicators are often non-specific and vary across individuals. This condition requires medical professionals to conduct comprehensive clinical data analysis to accurately assess risk levels. This challenge is further amplified by limited medical

personnel and healthcare facilities that are not yet fully capable of providing fast and accurate diagnoses (Kumar et al., 2025). In recent years, advancements in artificial intelligence, particularly machine learning-based classification techniques, have opened new opportunities to assist medical professionals in identifying data patterns and classifying patient conditions based on clinical examination results. With this approach, the process of analysis and risk prediction for cardiovascular disease can be performed more efficiently and objectively to support precise medical decision-making (Ortega-Martorell et al., 2022).

In a technological context, one of the widely developed approaches to support medical diagnosis is the application of machine learning, specifically in classifying diseases based on patients' clinical data (Carolina Wibowo et al., 2024). One of the approaches used by the author in this study is the Support Vector Machine (SVM) algorithm. SVM has the capability to handle high-dimensional data and provide accurate classification results (Jiang, 2025). SVM works by forming an optimal hyperplane that separates data classes with a maximum margin and can handle non-linear cases through the application of kernels. Various previous studies have proven the effectiveness of this method, particularly the SVM algorithm. Research comparing the performance of SVM with Neural Networks in heart disease classification found that SVM performed better with a higher level of accuracy (Tino et al., 2023).

However, the performance of a classification algorithm is heavily influenced by the quality of the features used. The presence of irrelevant or redundant features can lead to suboptimal models, increasing the risk of overfitting and slowing down the computational process (M.A. & Thomas, 2019). Therefore, the feature selection process is a vital stage in simplifying data representation. Two widely used feature selection methods are Chi-Square and Recursive Feature Elimination (RFE). Chi-Square evaluates the relationship between features and the target using statistical tests, while RFE iteratively removes features based on their contribution to the model (Korial et al., 2024). While the separate use of these two methods has been proven to improve model accuracy, the application of a combination of both is still rarely found in cardiovascular disease classification studies (Bayuaji et al., 2024).

Several previous studies have shown promising results. Research by Nidhi Sinha demonstrated that using Chi-Square and RFE can improve SVM accuracy in heart disease classification on the Z-Alizadeh Sani dataset (Sinha et al., 2022). Research by Fahrudin & Suroso also proved that proper feature selection could increase the accuracy of the SVM model from 87.65% to 96.56% (Fahrudin et al., 2024). Meanwhile, research by Sarra and colleagues showed that integrating Chi-Square into an ensemble learning system could increase heart disease classification accuracy from 85.5% to 90.8% (Sarra et al., 2024.).

Based on these studies, it can be concluded that the combination of Chi-Square and RFE feature selection methods have the potential to provide a significant increase in classification performance. Therefore, this study will discuss the classification of cardiovascular disease using

the SVM algorithm based on a combination of features selected through Chi-Square and RFE. This research is expected to produce a more efficient and stable cardiovascular disease classification model, contributing to the development of machine learning-based medical decision support systems.

METHOD

The methodology used in this research is essential to ensure that the study is systematic and follows established procedures. With a correct methodology, it is expected that the research will produce satisfactory results to achieve the research objectives. The research steps are illustrated in Figure 1.

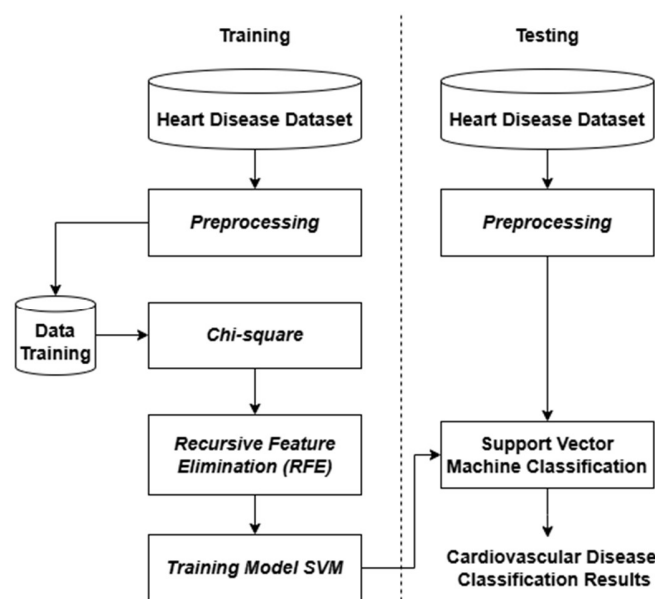


Figure 1. Research Workflow

The research workflow consists of two main stages: training and testing. In the training stage, the dataset undergoes preprocessing, which includes handling missing values, removing duplicates, and normalization. Subsequently, feature selection is performed using Chi-Square and RFE to obtain the most relevant attributes before the SVM model is trained. In the testing stage, the trained model is used to predict the test data, and its performance is evaluated using a confusion matrix, accuracy, precision, recall, and F1-score.

1. Dataset

This study utilizes the Heart Disease Dataset obtained from IEEE Dataport, containing a total of 1,190 patient records. The dataset consists of several clinical attributes covering cardiovascular disease risk factors such as blood pressure, cholesterol, blood sugar levels, heart rate, and other supporting variables. Each record is assigned a binary label indicating the presence

or absence of cardiovascular disease (Siddhartha, 2020). A complete explanation of the dataset attributes is presented in Table 1.

Table 1. Dataset Attribute Description

Feature Name	Description
Age	Patient's age in years.
Sex	Patient's gender (1: Male, 0: Female).
Chest Pain Type	Type of chest pain (1: Typical angina, 2: Atypical angina, 3: Non-anginal pain, 4: Asymptomatic).
Resting BP S	Resting blood pressure in mmHg.
Cholesterol	Serum cholesterol level in mg/dl.
Fasting Blood Sugar	Fasting blood sugar > 120 mg/dl (1: Yes, 0: No).
Resting ECG	Resting electrocardiogram results (0: Normal, 1: ST-T wave abnormality, 2: Left ventricular hypertrophy).
Max Heart Rate	Maximum heart rate achieved during exercise (ranging from 71-202 bpm).
Exercise Angina	Exercise-induced angina (1: Yes, 0: No).
Oldpeak	ST depression compared to rest.
ST Slope	The slope of the peak exercise ST segment (0: Up-sloping, 1: Flat, 2: Down-sloping).
Target	Disease classification (1: Heart disease, 0: No heart disease).

2. Preprocessing Data

Data preprocessing is conducted to ensure that the data used in the algorithm is in good condition and suitable for processing in the next stage. The main stages include handling missing values, removing duplicate data, and normalization. The complete preprocessing workflow is shown in Figure 2

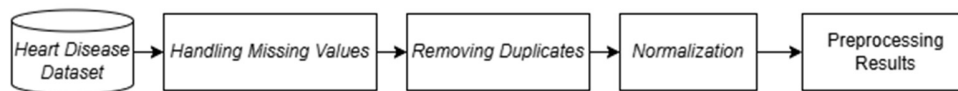


Figure 2 Data Preprocessing Workflow

First, the data is examined to identify missing values in critical attributes that could potentially decrease model accuracy; rows with null values are removed so that only complete data is used in the training process. Next, duplicate data removal is performed to prevent bias and ensure each sample provides unique information. After these two stages, the preprocessing concludes with data normalization using the MinMaxScaler method to standardize the scales between features, preventing any single attribute from dominating the modeling process. Normalization via MinMaxScaler scales all features into a uniform range between 0 and 1.

3. Chi-Square

The Chi-Square method evaluates the relationship between each feature and the target variable based on a statistical test (Wang & Zhou, 2021). A larger Chi-Square value indicates a higher level of correlation. The formula for calculating Chi-Square is:

$$\chi^2 = \sum \frac{(O-E)^2}{E} \quad (1)$$

Where O is the observed value and E is the expected value. The results of the Chi-Square selection produce features with the highest significance levels based on the test score values.

4. Recursive Feature Elimination (RFE)

Recursive Feature Elimination using SVM (SVM-RFE) as a ranking criterion is an embedded-type feature selection method that utilizes the weight values from the SVM model to determine the importance of each feature. The main principle is to iteratively eliminate features with the lowest squared weights. This mechanism ensures that only the features with the most significant contributions are retained in the classification model (Rustam & Kharis, 2020). Technically, the SVM-RFE algorithm mechanism is described as follows:

- 1) Initialization: Prepare the initial feature set (S) and an empty feature ranking list (R).
- 2) Weight Estimation: Train the SVM model at each iteration to obtain the weight vector (w) through the equation:
$$w = \sum_k \alpha_k y_k x_k \quad (2)$$
- 3) Ranking Criteria: Calculate the feature ranking score (c_k) based on the squared weight value $c_k = w_k^2$.
- 4) Recursive Elimination: Identify the feature with the smallest c_k value, add it to the ranking list, and remove it from the active feature set.
- 5) Output: The process is repeated until all features are ordered in the list based on their level of relevance.

5. Support Vector Machine (SVM)

The Support Vector Machine (SVM) is a supervised classification algorithm that works by finding the optimal hyperplane that separates two classes with a maximum margin. To handle both linear and non-linear data patterns, SVM uses kernel functions to transform the original feature space into a higher-dimensional space where the data is more easily separable (Wahyuni & Kusumodestoni, 2024). Generally, the SVM model uses a kernel function as written in the following equation:

$$K(x_i, x_j) = \phi(x_i) \cdot \phi(x_j) \quad (3)$$

In this study, five types of kernels are used to evaluate model performance: Linear, Radial Basis Function (RBF), and Polynomial kernels with degrees 3, 5, and 7. The Linear kernel is used for data that tends to be linearly separable, while RBF is used to handle non-linear patterns.

Polynomial kernels of degrees 3, 5, and 7 are applied to evaluate how the complexity of the polynomial order affects the performance and stability of the model on cardiovascular disease data.

The SVM training process is conducted using the Sequential Minimal Optimization (SMO) algorithm, developed by Platt (1998) to simplify the resolution of Quadratic Programming (QP) problems in SVM. SMO operates by optimizing two Lagrange multipliers in each iteration through a working set approach of size two. Each sub-problem can be solved analytically, making the computation time more efficient even as the number of iterations increases (Noviyanto et al., 2016). The main stages of SMO consist of:

- 1) Selecting a multiplier that violates the Karush–Kuhn–Tucker (KKT) conditions.
- 2) Determining a second multiplier to be optimized as a pair.
- 3) Repeating the process until all KKT conditions are met.

Once converged, the model produces an optimal hyperplane to be used in the classification process.

6. K-Fold Cross-Validation (CV)

K-Fold Cross-Validation (CV) is a model evaluation technique used to measure the performance of a machine learning algorithm more objectively and stably. This technique works by dividing the dataset into k equal-sized parts (folds). Generally, one fold is used as test data, while the other $(k - 1)$ folds are used as training data in each iteration (Nti et al., 2021). This study uses a 5-Fold Cross-Validation scheme.

7. Evaluation

Model evaluation is performed using a Confusion Matrix to assess the classification performance in this study. The Confusion Matrix is used to measure the rate of correct predictions through True Positive (TP) and True Negative (TN) values, as well as incorrect predictions through False Positive (FP) and False Negative (FN) values (Fahmy Amin, 2022). This method helps identify four main categories of model prediction results, as shown in Table 2.

Table 2. Confusion Matrix

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

Based on Table 2, True Positive (TP) represents patients correctly classified as having cardiovascular disease, while True Negative (TN) refers to healthy patients correctly identified by the model. False Positive (FP) occurs when a healthy patient is incorrectly predicted as having the

disease, whereas False Negative (FN) represents cases in which the model fails to detect patients who actually have cardiovascular disease.

The overall performance of the model is then assessed using the metrics of Accuracy, Precision, Recall, and F1-Score. Accuracy measures the general correctness of predictions, precision shows the proportion of correct positive predictions, recall assesses the model's ability to detect all positive cases, and the F1-Score provides a balance between precision and recall, which is particularly useful for imbalanced datasets (Riyanto et al., 2023). To calculate accuracy, precision, and recall from the confusion matrix table, the following equations are used.

$$Accuracy(\%) = \frac{TP+T}{TP + TN + FP + FN} \times 100\% \quad (4)$$

$$Precision(\%) = \frac{TP}{TP+F} \times 100\% \quad (5)$$

$$Recall(\%) = \frac{TP}{TP+F} \times 100\% \quad (6)$$

$$F1\ Score(\%) = \frac{2 \times Precision \times Recall}{Precision + Recall} \times 100\% \quad (7)$$

RESULT

1. Result of Data Pre-processing

The data pre-processing stage was conducted to ensure the dataset was in an optimal condition before being used in the modeling process. At this stage, 172 zero values were identified in the *cholesterol* feature, which are clinically irrelevant. These values were subsequently handled through imputation using the median value to maintain the data characteristics. Additionally, 272 duplicate records were removed, reducing the dataset size from 1,190 to 918 unique records. Furthermore, Min-Max Scaling normalization was applied to transform the value ranges of all numerical features into a scale of 0 to 1, preventing the dominance of specific features during the distance calculation process in the SVM algorithm.

2. Result of Feature Selection

Dimensionality reduction was performed in two stages to obtain a subset of features that most significantly influence classification accuracy. The first stage utilized the **Chi-Square** statistical test to measure the dependency of features on the target label. Based on the significance criteria (p-value < 0.05), 8 features were declared relevant, while the *resting ecg*, *cholesterol*, and *resting bp s* features were eliminated. The complete statistical results of the Chi-Square test are presented in Table 3.

Table 3. Chi-Square Significance Test Results

No.	Feature	Chi-Square	p-value	Signifikan (< 0.05)
1	exercise angina	133.6401	6.55×10^{-31}	Ya
2	Fasting blood sugar	50.2969	1.32×10^{-12}	Ya
3	chest pain type	26.1441	3.17×10^{-7}	Ya
4	ST slope	21.2479	4.04×10^{-6}	Ya
5	sex	18.0062	2.20×10^{-5}	Ya
6	max heart rate	8.7383	3.12×10^{-3}	Ya
7	oldpeak	5.5465	1.85×10^{-2}	Ya
8	age	5.1919	2.27×10^{-2}	Ya
9	resting ecg	1.8370	0.1753	Tidak
10	cholesterol	0.2196	0.6393	Tidak
11	resting bp s	0.1374	0.7109	Tidak

The second stage implemented Recursive Feature Elimination (RFE) with a linear SVM estimator to iteratively identify the optimal number of features. Cross-validation results indicated that the model performance peaked at 8 features with an accuracy of 85.51%. The trend of accuracy relative to the number of features is visualized in Figure 3.

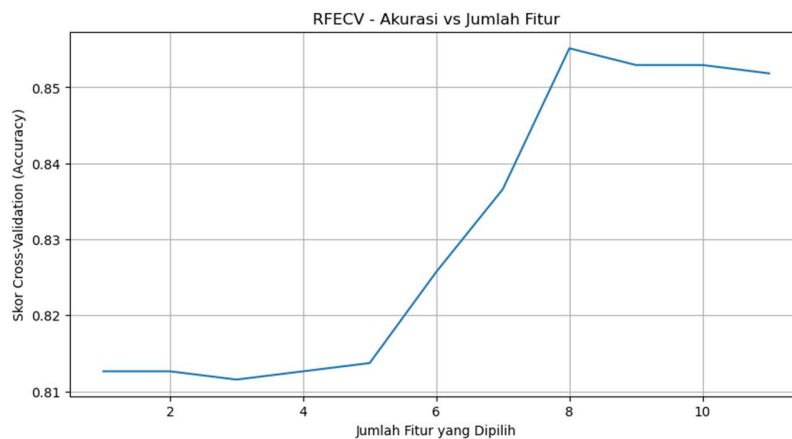


Figure 3. RFECV Accuracy Curve vs. Number of Features

The final feature selection result was obtained through the intersection of both methods. This process yielded 7 primary features consistently deemed important: sex, chest pain type, fasting blood sugar, max heart rate, exercise angina, oldpeak, and ST slope. This subset of features was then used as input for the subsequent SVM classification model training phase.

3. Model Testing Results

Initial testing was conducted using the Support Vector Machine (SVM) algorithm with five kernel types: Linear, Radial Basis Function (RBF), and Polynomial kernels with degrees of 3, 5, and 7, without applying the feature selection process. Performance evaluation was carried out using a 5-Fold Cross-Validation scheme. The metric values displayed in the table represent the

mean values of accuracy, precision, recall, and F1-score obtained from all testing folds. The performance results for each kernel are presented in Table 4.

Table 4. Average SVM Classification Performance Without Feature Selection

Kernel	Akurasi	Presisi	Recall	F1-Score
Linear	0.8584	0.8620	0.8868	0.8730
RBF	0.8366	0.8518	0.8539	0.8519
Polynomial d = 3	0.8540	0.8626	0.8771	0.8686
Polynomial d = 5	0.8191	0.8429	0.8334	0.8355
Polynomial d = 7	0.7723	0.8080	0.7756	0.7892

Based on these results, the Linear kernel yielded the highest average accuracy of 85.84%, followed by the Polynomial kernel (degree 3) with an average accuracy of 85.40%. Conversely, the Polynomial kernel (degree 7) showed the lowest performance with an average accuracy of 77.23%. These results indicate that using all features without a reduction process tends to decrease performance in high-complexity kernels due to increased data dimensionality and potential overfitting.

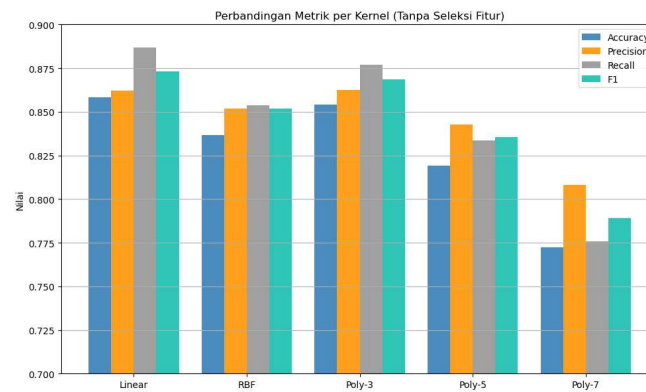


Figure 4. Evaluation Metric Comparison Without Feature Selection

In the next stage, testing was performed using the subset of features resulting from the combination of Chi-Square and Recursive Feature Elimination (RFE) methods. The model evaluation consistently used the 5-Fold Cross-Validation scheme. The performance evaluation results after feature selection are shown in Table 5.

Table 5. Average SVM Classification Performance With Feature Selection

Kernel	Akurasi	Presisi	Recall	F1-Score
Linear	0.8529	0.8613	0.8747	0.8671
RBF	0.8486	0.8651	0.8619	0.8625
Polynomial d=3	0.8442	0.8407	0.8876	0.8626
Polynomial d=5	0.8529	0.8625	0.8785	0.8678
Polynomial d=7	0.8431	0.8613	0.8590	0.8579

The testing results demonstrate that the application of feature selection improved the average performance across nearly all kernels. The RBF kernel experienced an increase in average accuracy from 83.66% to 84.86%, while the Polynomial kernel (degree 7) showed the most significant improvement, rising from 77.23% to 84.31%. This improvement indicates that feature selection effectively reduces model complexity and assists the SVM in forming a more optimal hyperplane through a more relevant feature representation.

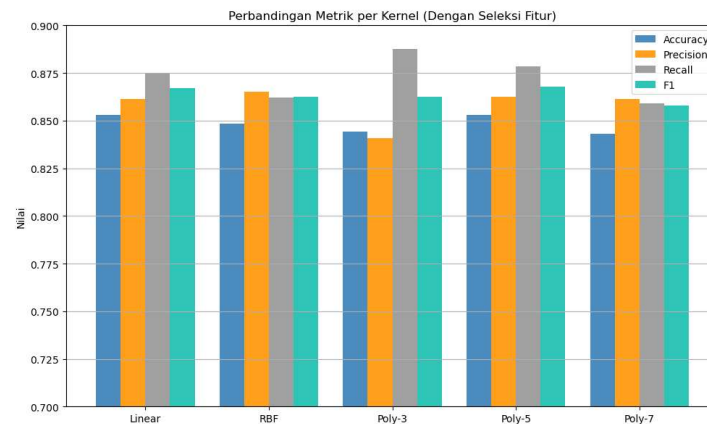


Figure 5. Evaluation Metric Comparison With Feature Selection

A comparison of model performance before and after feature selection shows that the combination of Chi-Square and RFE methods provided an increase in the average values of evaluation metrics, particularly for high-complexity kernels. The Polynomial kernels (degrees 5 and 7) experienced more consistent improvements across all evaluation metrics after the feature selection process was applied. A summary of the average accuracy comparison before and after feature selection is displayed in Figure 6.

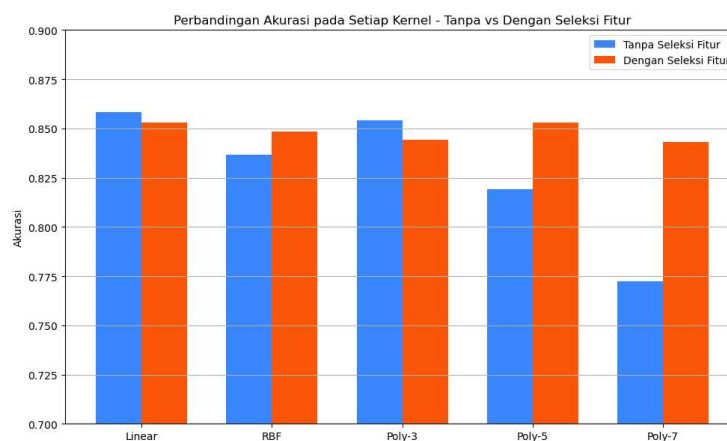


Figure 6. Accuracy Comparison for Each Kernel

Meanwhile, the Linear kernel showed relatively stable performance both before and after feature selection. This suggests that the data pattern tends to be close to linear. Although the

increase in average accuracy was not significant, feature selection still provides benefits in terms of computational efficiency as fewer features are processed.

DISCUSSION

This study demonstrates that integrating Chi-Square and Recursive Feature Elimination (RFE) feature selection effectively enhances the performance and stability of the Support Vector Machine (SVM) classifier for cardiovascular disease prediction. The improvement observed across most kernel functions indicates that reducing data dimensionality plays a crucial role in minimizing noise and improving class separability during hyperplane construction. By eliminating irrelevant and redundant features, the SVM model can focus on the most informative attributes, resulting in a more robust classification boundary.

The most substantial performance gain was achieved by the polynomial kernel with degree 7. This result suggests that high-complexity kernels are more sensitive to irrelevant features and benefit significantly from feature selection. Without dimensionality reduction, complex kernels tend to overfit the data, leading to unstable performance across different validation folds. The application of Chi-Square and RFE mitigates this issue by preserving features with strong statistical relevance and high model contribution, thereby improving generalization capability.

In contrast, the linear kernel exhibited relatively stable performance before and after feature selection. This finding indicates that the underlying data distribution is close to linearly separable, allowing the linear kernel to perform well even without extensive feature reduction. Nevertheless, feature selection still contributes to improved computational efficiency by reducing the number of processed features without sacrificing classification accuracy. Overall, these results confirm that the proposed feature selection strategy is particularly beneficial for complex kernel configurations and supports the development of efficient machine learning-based decision support systems in medical diagnosis.

CONCLUSION

This study concludes that the integration of Chi-Square and Recursive Feature Elimination (RFE) feature selection improves the classification performance and stability of the Support Vector Machine (SVM) model for cardiovascular disease prediction. The application of feature selection significantly benefits high-complexity kernels, particularly the polynomial kernel with degree 7 and the radial basis function kernel, by reducing overfitting and enhancing generalization capability. Meanwhile, the linear kernel maintains strong performance, indicating that the dataset exhibits near-linear separability, while still gaining advantages in computational efficiency due to reduced feature dimensionality. These findings demonstrate that the proposed approach effectively

balances accuracy and efficiency, making it suitable for implementation in machine learning-based medical decision support systems.

REFERENCES

- Bayuaji, L., Kusnadi, Amzah, M. Y., & Pebrianti, D. (2024). Optimization of Feature Selection in Support Vector Machines (SVM) Using Recursive Feature Elimination (RFE) and Particle Swarm Optimization (PSO) for Heart Disease Detection. *2024 9th International Conference on Mechatronics Engineering (ICOM)*, 304–309. <https://doi.org/10.1109/ICOM61675.2024.10652561>
- Carolina Wibowo, A., Ardi Lestari, S., & Nurchim, N. (2024). Analisis Penggunaan Machine Learning dalam Klasifikasi Penentuan Penyakit Jantung. *Simtek: jurnal sistem informasi dan teknik komputer*, 9(2), 97–101. <https://doi.org/10.51876/simtek.v9i2.395>
- Fahmy Amin, M. (2022). Confusion Matrix in Binary Classification Problems: A Step-by-Step Tutorial. *Journal of Engineering Research*, 6(5), 0–0. <https://doi.org/10.21608/erjeng.2022.274526>
- Fahrudin, G. F., Suroso, S., & Soim, S. (2024). Pengembangan Model Support Vector Machine untuk Meningkatkan Akurasi Klasifikasi Diagnosis Penyakit Jantung. *Jurnal Teknologi Sistem Informasi dan Aplikasi*, 7(3), 1418–1428. <https://doi.org/10.32493/jtsi.v7i3.42254>
- Jiang, Y. (2025). Research on the Application of Support Vector Machines in High-Dimensional Data Mining. *2025 8th International Conference on Advanced Algorithms and Control Engineering (ICAACE)*, 2088–2091. <https://doi.org/10.1109/ICAACE65325.2025.11020080>
- Korial, A. E., Gorial, I. I., & Humaidi, A. J. (2024). An Improved Ensemble-Based Cardiovascular Disease Detection System with Chi-Square Feature Selection. *Computers*, 13(6), 126. <https://doi.org/10.3390/computers13060126>
- Kumar, R., Garg, S., Kaur, R., Johar, M. G. M., Singh, S., Menon, S. V., Kumar, P., Hadi, A. M., Hasson, S. A., & Lozanović, J. (2025). A comprehensive review of machine learning for heart disease prediction: Challenges, trends, ethical considerations, and future directions. *Frontiers in Artificial Intelligence*, 8, 1583459. <https://doi.org/10.3389/frai.2025.1583459>
- M.A., A., & Thomas, P. A. (2019). Comparative Review of Feature Selection and Classification modeling. *2019 International Conference on Advances in Computing, Communication and Control (ICAC3)*, 1–9. <https://doi.org/10.1109/ICAC347590.2019.9036816>
- Natsir, F. M., Bakti, R. Y., & Wahyuni, T. (2024). Analisis deteksi dini penyakit jantung dengan pendekatan Support Vector Machine pada data pasien. *Arus Jurnal Sains dan Teknologi*, 2(2), 437–446.

- Noviyanto, H., Adji, T. B., & Setiawan, N. A. (2016). Seleksi fitur menggunakan metode kombinasi algoritme genetika dan sequential minimal optimization untuk klasifikasi halaman web. *Prosiding Konferensi Nasional Teknologi Informasi dan Komunikasi (KNASTIK 2016)*, 212–217.
- Nti, I. K., Nyarko-Boateng, O., & Aning, J. (2021). Performance of machine learning algorithms with different K values in K-fold cross-validation. *International Journal of Information Technology and Computer Science (IJITCS)*, 13(6), 61-71. <https://doi.org/10.5815/ijitcs.2021.06.05>
- Ortega-Martorell, S., Riley, P., Olier, I., Raidou, R. G., Casana-Eslava, R., Rea, M., Shen, L., Lisboa, P. J. G., & Palmieri, C. (2022). Breast cancer patient characterisation and visualisation using deep learning and fisher information networks. *Scientific Reports*, 12(1), 14004. <https://doi.org/10.1038/s41598-022-17894-6>
- Rachmawati, C., Martini, S., & Artanti, K. D. (2021). Analisis faktor risiko modifikasi penyakit jantung koroner di RSUD Haji Surabaya tahun 2019. *Media Gizi Kesmas*, 10(1), 47–55.
- Riyanto, S., Sitanggang, I. S., Djatna, T., & Atikah, T. D. (2023). Comparative Analysis using Various Performance Metrics in Imbalanced Data for Multi-class Text Classification. *International Journal of Advanced Computer Science and Applications*, 14(6). <https://doi.org/10.14569/IJACSA.2023.01406116>
- Rustam, Z., & Kharis, S. A. A. (2020). Comparison of Support Vector Machine Recursive Feature Elimination and Kernel Function as feature selection using Support Vector Machine for lung cancer classification. *Journal of Physics: Conference Series*, 1442(1), 012027. <https://doi.org/10.1088/1742-6596/1442/1/012027>
- Sarra, R. R., Gorial, I. I., Manea, R. R., Korial, A. E., Mohammed, M., & Ahmed, Y. (2024). Enhanced stacked ensemble-based heart disease prediction with chi-square feature selection method. *Journal of Robotics and Control (JRC)*, 5(6), 1753–1763. <https://doi.org/10.18196/jrc.v5i6.23191>
- Siddhartha, M. (2020). *Heart Disease Dataset (Comprehensive)* [Dataset]. IEEE DataPort. <https://doi.org/10.21227/DZ4T-CM36>
- Sinha, N., Jangid, T., Joshi, A. M., & Mohanty, S. P. (2022). *iCardo: A Machine Learning Based Smart Healthcare Framework for Cardiovascular Disease Prediction* (No. arXiv:2212.08022). arXiv. <https://doi.org/10.48550/arXiv.2212.08022>
- Tino, M. D. F., Herliyani Hasanah, & Tri Djoko Santosa. (2023). Perbandingan Algoritma Support Vector Machines (SVM) dan Neural Network untuk Klasifikasi Penyakit Jantung. *INFOTECH journal*, 9(1), 232–235. <https://doi.org/10.31949/infotech.v9i1.5432>
- Wahyuni, S. D., & Kusumodestoni, R. H. (2024). Optimalisasi algoritma Support Vector Machine (SVM) dalam klasifikasi kejadian data stunting. *Bulletin of Information Technology (BIT)*, 5(2), 56–64. <https://doi.org/10.47065/bit.v5i2.1247>

Wang, Y., & Zhou, C. (2021). Feature Selection Method Based on Chi-Square Test and Minimum Redundancy. In M. Tavana, N. Nedjah, & R. Alhajj (Eds.), *Emerging Trends in Intelligent and Interactive Systems and Applications* (Vol. 1304, pp. 171–178). Springer International Publishing. https://doi.org/10.1007/978-3-030-63784-2_22