Proceeding of International Conference on Advanced Technologies, Innovations and Science
Vol 1, No 1, June 2026, Page 501-509
ISSN XXXX-XXXX (Media Online)
https://journals.adaresearch.or.id/icatis

Feature Selection-Based Support Vector Machine Optimization for Heart Disease Risk Classification

Putri Andin, Nazla Mutya Ramadhani, Darman Sahputra Harefa*

Informasion Systems Study Program, Sekolah Tinggi Ilmu Komputer Tunas Bangsa, Pematangsiantar, Indonesia
Jl. Jendral Sudirman Blok A No. 1/2/3, Siantar Barat 21127, Kota Pematangsiantar, Indonesia
Email: [1]putrianisiantar@gmail.com, [2] nazlamutya05@gmail.com, [3,*]darmansahputraharefa@gmail.com
Correspondence Author Email: darmansahputraharefa@gmail.com
Submitted: **01/06/2026**; Accepted: **12/06/2026**; Published: **30/06/2026**

Abstract−Heart disease is a serious health issue that requires rapid and accurate detection and risk classification. This study aims to build a heart disease risk classification model using a Support Vector Machine optimized with Chi-Square Feature Selection. The dataset used is the EarlyMed Heart Disease Risk Dataset, which consists of 70,000 initial data points, 18 input features, and one classification target, namely Heart_Risk. The research stages include data preprocessing, removal of duplicate data, splitting the data into training and testing sets with an 80:20 ratio, building a Standard SVM model, and building an Optimized SVM model using MinMaxScaler, Chi-Square SelectKBest, and GridSearchCV. The results show that the Standard SVM model achieved an accuracy, precision, recall, and F1-score of 99.14% using all 18 features. Meanwhile, the Optimized SVM model achieved an accuracy, precision, recall, and F1-score of 98.17% using 11 selected features. Although the performance of the optimized model is slightly lower, it reduces the number of features by 38.89% while maintaining high classification performance. Thus, Chi-Square Feature Selection can be used to produce a more compact and efficient model for classifying heart disease risk.

Keywords: Heart Disease; Support Vector Machine; Chi-Square Feature Selection; GridSearchCV; Classification

1. INTRODUCTION

The heart is the primary organ that plays a vital role in the circulatory system, and disorders of this organ can lead to serious health problems associated with risk factors and efforts to classify heart disease (Dion et al., 2023; Nasution et al., 2025; K. I. F. Pratama & Kusnawi, 2023). Heart disease is a complex health issue because it can be influenced by non-modifiable factors, such as age, gender, and family history, as well as factors that can be controlled, such as high blood pressure, high cholesterol, obesity, diabetes, lack of physical activity, smoking, and excessive alcohol consumption (Galih et al., 2022; Nasution et al., 2025). These findings indicate that heart disease is not only related to impaired organ function but is also linked to patients' lifestyles and clinical characteristics. Diagnosing heart disease in its early stages is also challenging due to the complex interplay of various medical factors that must be taken into account (Bintoro, 2024). Therefore, the classification of heart disease risk is necessary to help categorize patients' conditions based on patterns in medical data, so that potential risks can be identified earlier and to support decision-making in the healthcare sector. (Kurniadi et al., 2025) also explained that heart disease is a key focus because it is a priority in efforts to detect and classify diseases.

As the use of medical data continues to grow, machine learning (ML) has become one of the most widely used approaches to aid in the classification of heart disease. Various methods, such as Naïve Bayes (NB), K-Nearest Neighbor (k-NN), Random Forest (RF), Logistic Regression, Decision Tree, Gradient Boosting, Linear Discriminant Analysis, XGBoost, and Support Vector Machine (SVM), have been used to build classification models based on patient data (Aldila et al., 2026; Rashad et al., 2023; Rayadhani & Rahardi, 2025). These methods have the advantage of being able to recognize patterns in medical data and generate predictions based on available clinical attributes. However, model performance results can still vary due to the influence of dataset characteristics, the number of features, data quality, and the optimization techniques used. SVM is a relevant method for classifying heart disease because it can distinguish classes based on patient data patterns and works by constructing the optimal hyperplane to separate the data into different classes (Faradisia & Pakereng, 2025; Farida & Bahri, 2025). Given these characteristics, SVMs are considered suitable for heart disease risk classification problems, especially when the dataset contains multiple clinical attributes that need to be analyzed simultaneously.

Several previous studies have shown that the classification of heart disease has been extensively developed using various ML algorithms. The first study, conducted by (Rayadhani & Rahardi, 2025) compared RF, Support Vector Machines, and NB in predicting cardiovascular disease risk using 1,000 patient records and 14 clinical attributes. The study has the advantage of comparing several algorithms to assess model performance, but it does not specifically emphasize SVM optimization through the selection of relevant features. Furthermore, research conducted by (Aldila et al., 2026) evaluated LR, SVM, KNN, Decision Tree, RF, and Gradient Boosting on the Cleveland Heart Disease dataset, thereby expanding the comparison of classification methods; however, the focus of the study remained on evaluating the performance of several algorithms. Finally, the study conducted by (Humaira & Ningsih, 2026) has combined classification with feature selection using the ANOVA F-value in SVM and RF, whereas (Hirmayanti & Utami, 2025) comparing CFS, Chi-Square, and Information Gain in RF and SVM for heart disease diagnosis. Both studies indicate that feature selection is increasingly being used as an optimization step, but the methods and performance results obtained still vary. Other studies have also developed optimization and feature selection techniques, such as the use of ANOVA, Chi-Square, AdaBoost, RFECV, SHAP, Optuna, Robust SVM, and Dynamic Weighted Particle Swarm Optimization to

improve the quality of medical classification models (Fania et al., 2026; Nababan, 2024; A. Pratama et al., 2026; Ramdhan et al., 2026; Syaidah et al., 2024). However, research that specifically focuses on feature selection as the basis for SVM optimization in heart disease risk classification still needs to be further developed.

Based on the findings of previous research, the specific challenge in this study lies in the use of features in the process of classifying heart disease risk (Jiang et al., 2024). Medical datasets typically have many attributes, but not all attributes contribute equally to the classification target (Pudjihartono et al., 2022). Using all features without a selection process does not necessarily result in optimal performance, as irrelevant features can increase model complexity, slow down computations, and potentially reduce the model's ability to identify patterns of heart disease risk (Noroozi et al., 2023). In SVM, feature quality is a critical factor because the construction of the hyperplane is heavily influenced by the data representation used. If the features used are not sufficiently relevant, class separation may be suboptimal, and classification results may suffer as a result (Alcaraz et al., 2022). This represents a research gap because previous studies have primarily focused on comparing algorithms, while the optimization of SVMs through the selection of truly relevant features has not been a primary focus. Therefore, an approach is needed that can identify key attributes before the classification process begins, so that the resulting model is more efficient and performs better (Azzam et al., 2024).

To address this issue, this study proposes a Support Vector Machine optimization method based on Chi-Square feature selection for heart disease risk classification (Sarra et al., 2022). The feature selection method used is the Chi-Square method with the SelectKBest approach, a feature selection technique that assigns a score to each attribute based on its statistical relationship with the classification target (Alabrah, 2022). This method is used to select the most relevant attributes so that the SVM model does not rely solely on all available features, but instead uses features that actually contribute to the classification target (Bolón-canedo, 2024). The proposed solution is expected to produce a classification model that is more compact and efficient while maintaining high classification performance. The novelty of this research lies in the application of Chi-Square feature selection as an optimization step in SVM to improve the quality of heart disease risk classification (Nayl et al., 2026). Therefore, this study aims to develop a heart disease risk classification model using an SVM optimized with Chi-Square feature selection and to evaluate its performance based on the metrics of accuracy, precision, recall, and F1-score.

2. RESEARCH METHODS

2.1 Dataset

The dataset used in this study is the heart_disease_risk_dataset_earlymed, which contains data on heart disease risk based on symptoms, risk factors, and patient characteristics. This dataset consists of 70,000 records with 19 attributes: 18 attributes as input features and 1 attribute as the target label. The input features consist of clinical symptoms, such as chest pain, shortness of breath, fatigue, palpitations, dizziness, swelling, pain in the arms/jaw/back, as well as cold sweats or nausea. Additionally, the dataset includes risk factors such as high blood pressure, high cholesterol, diabetes, smoking habits, obesity, sedentary lifestyle, family history, chronic stress, gender, and age.

The target label in the dataset is Heart_Risk, which is used to classify the risk of heart disease. A value of 0 indicates a patient with low or no risk, while a value of 1 indicates a patient at high risk for heart disease. Most attributes in the dataset are binary with values of 0 and 1, while the Age attribute is numerical. Based on an initial review, the dataset contains no missing values and can therefore be used in the data processing and classification model development stages.

Table 1. Description of Attributes in the Heart Disease Risk (EarlyMed) Dataset

No.	Feature Name	Data Types	Description
1	Chest_Pain	Biner (0/1)	Gejala nyeri dada
2	Shortness_of_Breath	Biner (0/1)	Sesak napas
3	Fatigue	Biner (0/1)	Kelelahan berlebihan
4	Palpitations	Biner (0/1)	Jantung berdebar
5	Dizziness	Biner (0/1)	Pusing/vertigo
6	Swelling	Biner (0/1)	Pembengkakan kaki/tangan
7	Pain_Arms_Jaw_Back	Biner (0/1)	Nyeri lengan, rahang, atau punggung
8	Cold_Sweats_Nausea	Biner (0/1)	Keringat dingin dan mual
9	High_BP	Biner (0/1)	Tekanan darah tinggi
10	High_Cholesterol	Biner (0/1)	Kolesterol tinggi
11	Diabetes	Biner (0/1)	Riwayat diabetes
12	Smoking	Biner (0/1)	Kebiasaan merokok
13	Obesity	Biner (0/1)	Kondisi obesitas
14	Sedentary_Lifestyle	Biner (0/1)	Gaya hidup kurang gerak
15	Family_History	Biner (0/1)	Riwayat keluarga penyakit jantung
16	Chronic_Stress	Biner (0/1)	Stres kronis
17	Gender	Biner (0/1)	Jenis kelamin (0=Perempuan, 1=Laki-laki)
18	Age	Numerik (Kontinu)	Usia pasien (tahun)
19	Heart_Risk	Biner (0/1)	Label risiko penyakit jantung (Target)

Proceeding of International Conference on Advanced Technologies, Innovations and Science
Vol 1, No 1, June 2026, Page 501-509
ISSN XXXX-XXXX (Media Online)
https://journals.adaresearch.or.id/icatis

2.2 Research Framework

This research framework outlines the process of classifying heart disease risk by comparing two models: the Baseline SVM and the Optimized SVM. The study begins with a heart disease risk dataset containing patients' clinical features and the Heart_Risk target. Next, data preprocessing is performed, which includes checking for missing values, removing duplicate data, and separating features from the target. After preprocessing is complete, the data is split using an 80:20 ratio, with 80% as training data and 20% as testing data. The training data is used to build two models: the SVM Baseline, which uses all features with the help of StandardScaler; and the SVM Optimization, which uses MinMaxScaler, Chi-Square Feature Selection, and GridSearchCV to select relevant features and find the best parameters. The test data is used to evaluate both models. The evaluation results are then assessed using accuracy, precision, recall, F1-score, and a confusion matrix. The results are compared to determine the impact of optimization on model performance and analyzed to identify which model is more effective or efficient in classifying heart disease risk.

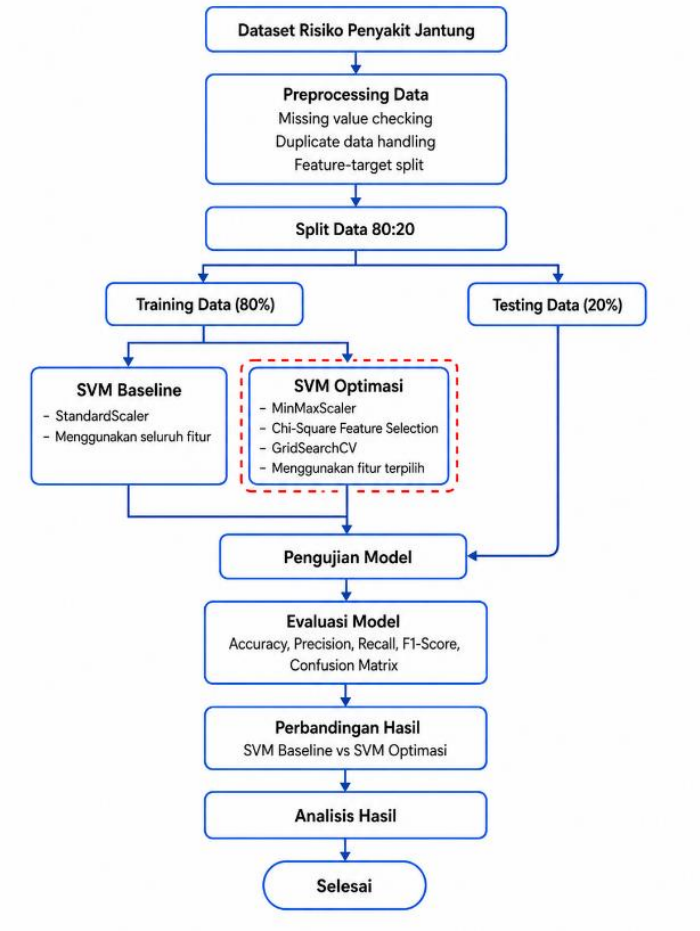

Figure 1. Research Framework

2.3 Proposed Method

As shown in Figure 2, the Baseline/Standard SVM process begins by using all 18 input features. These features are then processed using StandardScaler to standardize them, ensuring the data scales are more balanced before being fed into the model. Afterward, the data is processed using the Baseline SVM with an RBF kernel to generate a classification output labeled as Heart_Risk. This figure shows that the baseline model uses all available features without undergoing a feature selection process.

Based on Figure 3, the Optimized SVM process also begins with all 18 input features. However, unlike the baseline model, in the optimized model, the data is first processed using MinMaxScaler for feature normalization. After that, Chi-Square Feature Selection with SelectKBest is performed to select the most relevant features, resulting in 11 selected features. Next, parameter optimization is performed using GridSearchCV to find the best parameters such as C, kernel, and gamma. The results of this process are then used in the Optimized SVM with a Poly kernel to produce a Heart_Risk classification output.

Thus, the main difference between the two figures lies in the feature processing and model optimization stages. Figure 2 shows the Baseline SVM, which uses StandardScaler and all features, while Figure 3 shows the Optimized SVM, which uses MinMaxScaler, Chi-Square Feature Selection, and GridSearchCV. The optimized model is more targeted because it uses only selected features and the best parameters, whereas the baseline model is used as a comparison because it still uses all input features.

Proceeding of International Conference on Advanced Technologies, Innovations and Science
Vol 1, No 1, June 2026, Page 501-509
ISSN XXXX-XXXX (Media Online)
https://journals.adaresearch.or.id/icatis

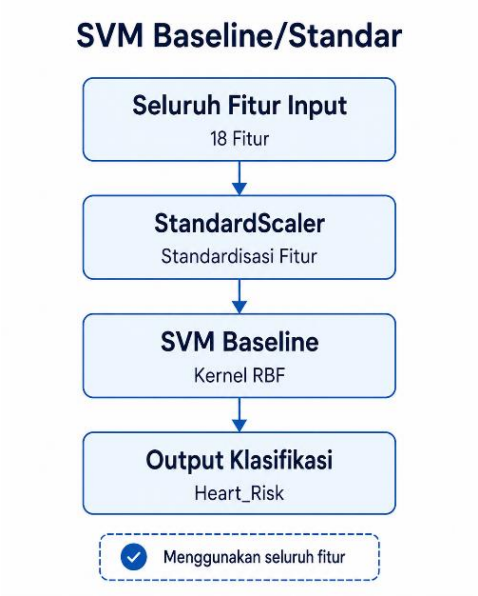

Figure 2. SVM Baseline

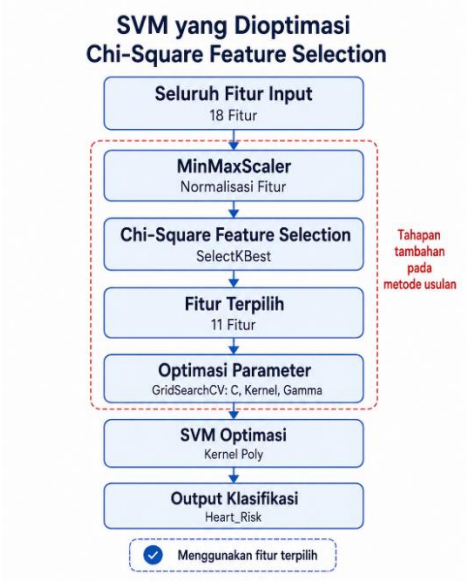

Figure 3. Optimized SVM

3. RESULTS AND DISCUSSION

This chapter discusses the results of the research phases that have been carried out. The discussion focuses on the results of data preprocessing, the results of testing the standard SVM model, and the results of testing the SVM model that has been optimized using Chi-Square Feature Selection and GridSearchCV. An evaluation was conducted to determine the ability of each model to classify heart disease risk based on the Heart_Risk target.

3.1 Data Preprocessing Results

The data preprocessing stage was conducted to ensure that the dataset was ready for use in the modeling process. The dataset used was a heart disease risk dataset with 18 input features and 1 classification target, namely Heart_Risk. In the initial stage, a check for missing values was performed to determine whether there were any empty values in each attribute. Based on the results of the check, the dataset had no missing values. Additionally, a check for duplicate data was performed, and a number of records with repeated values were found; these duplicate records were removed to prevent the model from learning the same patterns repeatedly. After the duplicate data removal process, the number of records decreased from 70,000 to 63,755. Next, a feature-target split was performed, which involves separating the input features from the classification target. After that, the dataset was divided into training and testing data using an 80:20 ratio, meaning 80% of the data was used for model training and 20% for model testing.

Table 2. Data Preprocessing Results

Preprocessing Steps	Results
Initial data volume	70,000 records
Number of input features	18 features
Classification target	Heart_Risk
Missing value	0
Duplicate data has been deleted	6.245 records
Amount of data after preprocessing	63.755 records
Feature-target split	Input and target features are separated
Split data	80% training and 20% testing

Based on Table 2, the dataset used contains no missing values, so no handling of missing data was required. However, 6,245 duplicate records were removed during the preprocessing stage. After the duplicate removal process was completed, the dataset consisted of 63,755 records. Next, the features and the target were separated, and the dataset was split into training and testing sets in an 80:20 ratio.

Table 3. Data Distribution Results

Dataset	Data Volume	Persentase
Training Data	51.004	80%
Testing Data	12.751	20%
Total Data	63.755	100%

According to Table 3, the training dataset consists of 51,004 data points, while the testing dataset consists of 12,751 data points. The training data is used to build the SVM model, while the testing data is used to evaluate the model's performance. With this separation, the model can be evaluated more objectively because testing is performed using data that was not used during the training process.

3.2 Standard SVM Results

The first model built was the standard SVM, or baseline SVM. This model was used as a baseline because it utilized all available input features in the dataset without any feature selection. For this model, the data was first processed using StandardScaler to standardize the features. Standardization was performed to ensure that the scales of the feature values were more balanced before being fed into the SVM model. After that, the SVM model was trained using an RBF kernel.

Table 4. Standard SVM Model Configuration

Components	Description
Model	Baseline SVM / Standard SVM
Number of features	18 features
Scaling	StandardScaler
Feature selection	Not in use
Kernel	RBF
Target	Heart_Risk

As shown in Table 4, the standard SVM model uses all 18 input features. This model does not employ feature selection, so all available attributes are directly used in the classification process after undergoing standardization using StandardScaler. This model serves as a baseline to assess the SVM's initial performance before optimization.

Table 5. Results of the Standard SVM Evaluation

Evaluation Metrics	Value
Accuracy	99,14%
Precision	99,14%
Recall	99,14%
F1-Score	99,14%

According to Table 5, the standard SVM model achieved accuracy, precision, recall, and F1-score values of 99.14%. These results indicate that the standard SVM is highly effective at classifying heart disease risk. The high evaluation scores indicate that the model can identify patterns in the data with a low error rate.

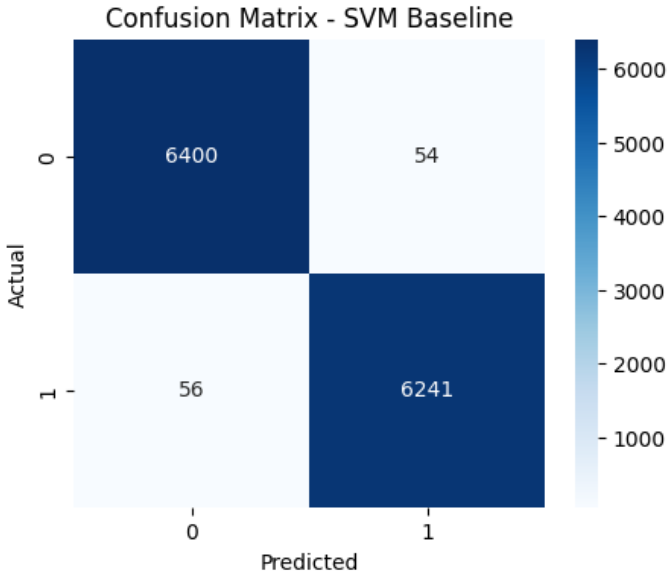

Figure 4. Standard SVM Confusion Matrix

As shown in Figure 4, the Standard SVM model successfully classified 6,400 class 0 data points and 6,241 class 1 data points correctly. Meanwhile, classification errors consisted of 54 class 0 data points predicted as class 1 and 56 class 1 data points predicted as class 0. These results indicate that the Standard SVM model has excellent classification capabilities with a relatively small number of errors. The total classification errors were 110 out of 12,751 test data points, so it can be concluded that the baseline model is capable of distinguishing heart disease risk classes with a high level of accuracy.

3.3 Optimized SVM Results

The second model built is an SVM optimized using Chi-Square Feature Selection and GridSearchCV. Unlike a standard SVM, this model does not immediately use all features in the classification process. In the optimization model, the data is first normalized using MinMaxScaler. This normalization is performed because the Chi-Square method requires non-negative feature values. After that, feature selection is performed using Chi-Square Feature Selection with SelectKBest to select the features most relevant to the Heart_Risk target.

Table 6. Optimized SVM Model Configuration

Components	Description
Model	SVM Optimization
Number of initial features	18 features
Scaling	MinMaxScaler
Feature selection	Chi-Square Feature Selection
Selection method	SelectKBest
Optimasi parameter	GridSearchCV
Target	Heart_Risk

As shown in Table 6, the optimized SVM model includes several additional steps compared to the standard SVM. These steps include normalization using MinMaxScaler, feature selection using Chi-Square, and parameter optimization using GridSearchCV. These steps aim to produce a more efficient model by selecting the most relevant features and using the optimal SVM parameters.

Table 7. Best Parameters for the SVM Optimization Model

Parameters	Best Value
selector__k	11
svm__kernel	poly
svm__C	10
svm__gamma	scale

Based on Table 7, the results of GridSearchCV indicate that the optimal number of features to use is 11. The best kernel obtained is the poly kernel, with a C value of 10 and a gamma value of scale. These parameters were then used to build the optimized SVM model.

Table 8. Fitur Selected Using Chi-Square Feature Selection

No	Featured Features
1	Chest_Pain
2	Shortness_of_Breath
3	Fatigue
4	Palpitations
5	Dizziness
6	Swelling
7	Pain_Arms_Jaw_Back
8	Cold_Sweats_Nausea
9	High_BP
10	High_Cholesterol
11	Sedentary_Lifestyle

Based on Table 8, Chi-Square Feature Selection selected 11 features out of the initial 18. The selected features are primarily related to clinical symptoms and risk factors for heart disease, such as Chest_Pain, Shortness_of_Breath, High_BP, and High_Cholesterol. These features were then used as inputs for the optimized SVM model.

Table 9. Results of the Optimized SVM Evaluation

Evaluation Metrics	Value
Accuracy	98,17%
Precision	98,17%
Recall	98,17%
F1-Score	98,17%

According to Table 9, the optimized SVM model achieved accuracy, precision, recall, and an F1-score of 98.17%. These results indicate that the optimized model was still able to deliver high performance despite using only 11 features out of the initial 18. Thus, the feature selection process successfully reduced the number of features used without causing a significant drop in performance.

Proceeding of International Conference on Advanced Technologies, Innovations and Science
Vol 1, No 1, June 2026, Page 501-509
ISSN XXXX-XXXX (Media Online)
https://journals.adaresearch.or.id/icatis

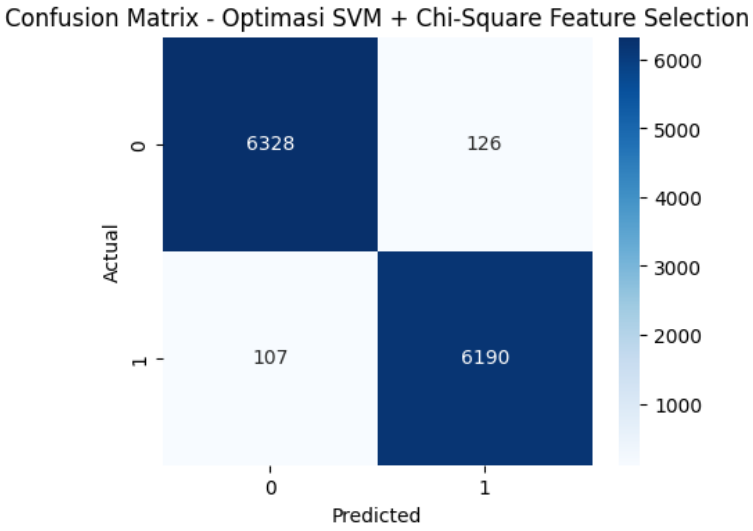

Figure 5. Optimized SVM Confusion Matrix

As shown in Figure 5, the optimized SVM model successfully classified 6,328 class 0 data points and 6,190 class 1 data points correctly. Meanwhile, classification errors consisted of 126 class 0 data points predicted as class 1 and 107 class 1 data points predicted as class 0. These results indicate that the optimized model remains capable of performing classification well even when using only selected features. The total classification error for this model was 233 data points out of 12,751 test data points. Although the number of errors was higher than that of the Standard SVM, the optimized model still demonstrated high performance with an evaluation score above 98%, making it sufficiently good and more efficient due to its use of fewer features.

3.4 Comparison of Results from Standard SVM and Optimized SVM

After testing both models, the next step is to compare the evaluation results between the Standard SVM and the Optimized SVM. This comparison is conducted to determine the impact of using Chi-Square Feature Selection and GridSearchCV on the performance of the heart disease risk classification model. The Standard SVM is used as the baseline model because it uses all features, while the Optimized SVM is used as the proposed model because it uses only selected features.

Table 10. Comparison of Results from Standard SVM and Optimized SVM

Model	Number of Features	Accuracy	Precision	Recall	F1-Score
Standard SVM	18 features	99,14%	99,14%	99,14%	99,14%
SVM Optimization	11 features	98,17%	98,17%	98,17%	98,17%

Based on Table 10, the standard SVM model using all 18 features achieved accuracy, precision, recall, and F1-score values of 99.14%. These results indicate that the SVM algorithm is capable of classifying heart disease risk with a very high level of accuracy when all attributes are used in the model training process. Using all features allows the model to utilize all available information in the dataset, thereby achieving optimal classification performance.

Meanwhile, the optimized SVM model using 11 selected features achieved accuracy, precision, recall, and F1-score values of 98.17%. Although there was a 0.97% decrease in performance compared to the baseline model, the evaluation scores obtained are still considered very good, as all evaluation metrics are above 98%. These results indicate that the features selected through the feature selection process are able to retain most of the important information required for the classification process.

The relatively small drop in performance indicates that some of the removed features made a minor contribution to the model's decision-making process. By reducing the number of features from 18 to 11, the model's complexity can be reduced without causing a significant decline in performance. This reduction in the number of features also has the potential to improve computational efficiency, speed up the model training process, and reduce the risk of overfitting on larger datasets.

The results of this study demonstrate a trade-off between classification performance and model complexity. The standard SVM model delivers the best performance with an accuracy of 99.14%, while the optimized SVM model offers a simpler structure with fewer features yet still achieves an accuracy of 98.17%. The performance difference of less than 1% indicates that the application of feature selection successfully maintains classification quality despite a significant reduction in the number of features.

Overall, the application of feature selection proved effective in simplifying the heart disease risk classification model by reducing the number of features by 38.89%, from 18 to 11. Although there was a slight decrease in accuracy, precision, recall, and F1-score, the resulting performance remained very high, making the optimized SVM model a more efficient alternative for use in heart disease risk prediction systems.

4. CONCLUSION

Based on the results of the study, heart disease risk classification using Support Vector Machines demonstrated excellent performance on the EarlyMed Heart Disease Risk dataset. The Standard SVM model, which uses all 18 features, yields accuracy, precision, recall, and F1-score values of 99.14%, while the SVM model optimized using Chi-Square Feature Selection and GridSearchCV yields accuracy, precision, recall, and F1-score values of 98.17% using 11 selected features. These results indicate that the Standard SVM still offers the highest accuracy, but the Optimized SVM model reduces the number of features by 38.89% with a relatively small performance drop of 0.97%. Thus, the application of Chi-Square Feature Selection can serve as a solution to reduce feature complexity and produce a more efficient model without significantly compromising classification performance. The limitations of this study lie in the use of a single dataset and testing that focused solely on the SVM method with Chi-Square Feature Selection. Future research could expand this approach by using other medical datasets, incorporating different feature selection methods, comparing more machine learning algorithms, and applying the model to an app-based early detection system so that the research results can be utilized more practically.

ACKNOWLEDGMENT

The author would like to express gratitude to STIKOM Tunas Bangsa Pematangsiantar, particularly the Information Systems Study Program, for the support and facilities provided during the conduct of this research. The author also extends gratitude to the academic advisor and all parties who provided guidance, feedback, and motivation throughout the process of drafting this article. Additionally, the author thanks the providers of the dataset used in this study, which enabled the successful implementation of heart disease risk classification using a Support Vector Machine with Chi-Square Feature Selection.

REFERENCES

Alabrah, A. (2022). *applied sciences A Novel Study : GAN-Based Minority Class Balancing and Machine-Learning-Based Network Intruder Detection Using Chi-Square Feature Selection*.

Alcaraz, J., Labbé, M., & Landete, M. (2022). *Support Vector Machine with feature selection : A multiobjective approach*. *204*(September 2021).

Aldila, A. S., Supriyono, L. A., Science, D., Program, S., Engineering, S., Program, S., & Internasional, U. J. (2026). *Evaluating Logistic Regression , Svm , Knn , And Ensemble Models For Accurate Heart Disease Risk Prediction*. *11*(3), 949–957. https://doi.org/10.33480/jitk.v11i3.6738.EVALUATING

Azzam, S. M., Emam, O. E., & Abolaban, A. S. (2024). An improved Differential evolution with Sailfish optimizer (DESFO) for handling feature selection problem. *Scientific Reports*, 1–27. https://doi.org/10.1038/s41598-024-63328-w

Bintoro, R. A. J. W. A. E. S. P. (2024). Machine Learning untuk Klasifikasi Penyakit Jantung. *Aisyah Journal of Informatics and Electrical Engineering*, *6*(1), 145–150. http://jti.aisyahuniversity.ac.id/index.php/AJIEE

Bolón-canedo, L. M. V. (2024). *Finding a needle in a haystack : insights on feature selection for classification tasks*. 459–483. https://doi.org/10.1007/s10844-023-00823-y

Dion, M., Tino, F., Hasanah, H., Santosa, T. D., Informatika, T., Komputer, F. I., Duta, U., & Surakarta, B. (2023). *Perbandingan Algoritma Support Vector Machines (Svm) Dan Neural Network Untuk Klasifikasi Penyakit Jantung*. *3*(1), 232–235.

Fania, D., Waspada, I., & Wibawa, H. A. (2026). *Cardiovascular Disease Risk Prediction Using Random Forest , RFECV Feature Selection , and SHAP with Multisource Clinical Data Integration*. *7*(1), 691–705.

Faradisia, A., & Pakereng, M. A. I. (2025). *Comparative Analysis of Linear , Polynomial , RBF , and Sigmoid Kernels in Support Vector Machine for Heart Disease Classification Analisis Komparatif Kernel Linear , Polynomial , RBF , dan Sigmoid pada Support Vector Machine untuk Klasifikasi Penyakit Ja*. *5*(October), 1531–1537.

Farida, L. N., & Bahri, S. (2025). *Komputika : Jurnal Sistem Komputer Klasifikasi Gagal Jantung Menggunakan Metode SVM (Support Vector Machine) Classification of Heart Failure using the SVM (Support Vector Machine) Method*. *13*, 0–7. https://doi.org/10.34010/komputika.v13i2.11330

Galih, D., Luthfi, M., & Farhan, M. (2022). *Klasifikasi Penyakit Jantung Menggunakan Metode Artificial Neural Network*. *3*(2), 55–60.

Hirmayanti & Utami, E. (2025). Enhanced Heart Disease Diagnosis Using Machine Learning Algorithms: A Comparison of Feature Selection. *Jurnal RESTI (Rekayasa Sistem Dan Teknologi Informasi)*, *9*(2), 385–392.

Jiang, Z., Liu, X., Mang, H. A., Zhang, J., & Pichler, B. (2024). *applied sciences Convergence-Related Serviceability Limit States of Segmental Tunnel Rings : Lessons Learned from Structural Analysis of Real-Scale Tests*.

Kurniadi, D., Pertiwi, A. I., & Mulyani, A. (2025). *Ensemble Voting Classifier Berbasis Multi-Algoritma dan Metode SMOTE untuk Klasifikasi Penyakit Jantung*. *14*, 145–153. https://doi.org/10.22146/jnteti.v14i2.17157

Nababan, E. B. (2024). *Performance Of Robust Support Vector Machine Classification Model On Balanced , Imbalanced And Outliers Datasets*. *10*(1), 208–215. https://doi.org/10.33480/jitk.v10i1.5272

Nasution, I. S., Rahmadani, A. D., Audina, W., & Purnama, D. (2025). *Systematic Review : Pengaruh Gaya Hidup dan Pengetahuan Masyarakat terhadap Risiko Penyakit Jantung Koroner*. *4*(2), 287–298. https://doi.org/10.54259/sehatrakyat.v4i2.4337

Nayl, H., Aly, E. S., Rezk, A., & Fares, M. E. (2026). *Improved chi-square feature selection for robust heart disease data classification*. *11*(September 2025), 2682–2701.

Noroozi, Z., Orooji, A., & Erfannia, L. (2023). Analyzing the impact of feature selection methods on machine learning algorithms for heart disease prediction. *Scientific Reports*, *0123456789*, 1–15. https://doi.org/10.1038/s41598-023-49962-w

Humaira & Ningsih. (2026). Perbandingan Algorit Ma Svm Dan Random Forest Untuk Prediksi Penyakit Jantung Menggunakan

Metode Feature Selection Anova F Value. *Jurnal Sistem Informasi Bisnis (JUNSIBI)*, *7*, 7–17.

Pratama, A., Assegaff, S., Jasmir, J., & Nurhadi, N. (2026). *Optimizing Heart Disease Classification Using C4 . 5 , Random Forest , and XGBoost with ANOVA , Chi-Square , and AdaBoost*. *7*(2), 1072–1090.

Pratama, K. I. F., & Kusnawi. (2023). Komparasi Algoritma Supervised Learning dan Feature Selection pada Klasifikasi Penyakit Gagal Jantung. *Indonesian Journal of Computer Science*, *12*(6), 3722–3733.

Pudjihartono, N., Fadason, T., & Kempa-liehr, A. W. (2022). *A Review of Feature Selection Methods for Machine Learning-Based Disease Risk Prediction*. *2*(June), 1–17. https://doi.org/10.3389/fbinf.2022.927312

Ramdhan, W., Hutahaean, J., Jollyta, D., & Karim, A. (2026). *Machine Learning Decision Support System for Heart Disease Prediction with Optuna and Threshold Optimization*. *7*(2), 1853–1870.

Rashad, I., Isnanto, R. R., & Edi, C. (2023). *Klasifikasi Penyakit Jantung Menggunakan Algoritma Analisis Diskriminan Linier*. *01*, 29–36. https://doi.org/10.21456/vol13iss1pp29-36

Rayadhani, W. A., & Rahardi, M. (2025). *Comparative Analysis of Random Forest , SVM , and Naive Bayes for Cardiovascular Disease Prediction*. *9*(6), 3234–3243.

Sarra, R. R., Dinar, A. M., & Mohammed, M. A. (2022). *Enhanced Heart Disease Prediction Based on Machine Learning and χ 2 Statistical Optimal Feature Selection Model*.

Syaidah, I. B., Surono, S., & Goh, K. W. (2024). *Dynamic Weighted Particle Swarm Optimization - Support Vector Machine Optimization in Recursive Feature Elimination Feature Selection*. *23*(3), 627–640. https://doi.org/10.30812/matrik.v23i3.3963