ISSN: 2623-1662 (online), 2460-092X (printed)
DOI: https://doi.org/10.19109/jusifo.v11i1.28113
Vol. 11, No. 1 (2025), p.53-60

JUSIFO
JURNAL SISTEM INFORMASI

Research Article [OPEN ACCESS]

Support Vector Machine for Classifying Heart Failure, Hypertension, and Normal Heart Condition

Surya Amando Bangun*, Elvis Sastra Ompusunggu, Wilson Wilson, Eppi Kriawati Harefa

ABSTRACT

Cardiovascular diseases, particularly heart failure and hypertension, remain among the leading causes of global mortality, underscoring the urgent need for accurate early diagnosis. This study proposes a classification model based on the Support Vector Machine (SVM) algorithm to distinguish among heart failure, hypertension, and normal heart conditions using real-world clinical data. The dataset was preprocessed through normalization and nominal-to-numerical conversion and validated by medical experts to ensure data quality. K-Fold Cross Validation (K=10) was employed to ensure model robustness and mitigate overfitting. The SVM classifier utilized a linear kernel and achieved high performance in terms of accuracy, precision, and recall. The results demonstrate the effectiveness of the proposed model in classifying multiple cardiovascular conditions with clinically relevant input features. This research contributes to the advancement of intelligent diagnostic tools and supports the integration of machine learning into clinical decision-making processes.

Keyword: Heart disease prediction, machine learning in healthcare, support vector machine

Received: May 01, 2025; Revised: May 25, 2025; Accepted: June 29, 2025

Corresponding Author: Surya Amando Bangun, Department of Information System, Universitas Prima Indonesia, Indonesia, suryaamandos03@gmail.com
Authors: Elvis Sastra Ompusunggu, Department of Information System, Universitas Prima Indonesia, Indonesia, elvissastraompusunggu@unprimdn.ac.id; Wilson Wilson, Department of Information System, Universitas Prima Indonesia, Indonesia, wilsonkokmen456@gmail.com; Eppi Kriawati Harefa, Department of Information System, Universitas Prima Indonesia, Indonesia, harefaeppykriawati@gmail.com

1. INTRODUCTION

Cardiovascular diseases (CVDs) remain the leading cause of mortality worldwide, with hypertension and heart failure identified as two of the most common and life-threatening conditions (Chong et al., 2024). According to the World Health Organization (2021), these two diseases collectively account for approximately 17.9 million global deaths each year, underscoring the substantial burden they impose on public health systems. The increasing prevalence—affecting an estimated 1.13 billion individuals with hypertension (Galindra et al., 2024) and approximately 64.3 million with heart failure (Yuniarti, 2022)—exerts considerable pressure on healthcare infrastructure and significantly diminishes patients' quality of life. As such, early detection is crucial to reducing mortality rates and enabling timely clinical intervention.

One of the major challenges in managing hypertension and heart failure lies in their asymptomatic nature during the early stages. Many patients remain undiagnosed until complications arise, by which time therapeutic options become limited and outcomes worsen. This diagnostic delay underscores the urgency of developing analytical tools capable of identifying early cardiovascular abnormalities. Furthermore, the ability to accurately distinguish pathological conditions from normal cardiac states is essential for supporting clinical decision-making, minimizing unnecessary interventions, and optimizing the allocation of medical resources (Carey et al., 2018; Cifu & Davis, 2017; Suryadi et al., 2024).

In response to this challenge, machine learning has emerged as a powerful approach for improving medical classification accuracy. Among various algorithms, the Support Vector Machine (SVM) has demonstrated high performance in processing high-dimensional clinical data and identifying complex disease patterns (Ghasemi & Sharifi, 2025; Plati et al., 2021; Sandhya, 2020). SVM is highly effective in constructing optimal decision boundaries, generalizing across diverse input data, and supporting robust classification tasks (Abdullah & Abdulazeez, 2021; Amarappa & Sathyanarayana, 2014; Liu, 2025; Patle & Chouhan, 2013; Shihong et al., 2003; Shmilovici, 2023; Srivastava & Bhambhu, 2010; Yu & Kim, 2012). While SVM has been widely applied in the prediction of diseases such as diabetes and kidney failure, its application in classifying cardiovascular conditions—specifically heart failure, hypertension, and normal cardiac states—remains relatively underexplored.

Although research in cardiovascular disease prediction using machine learning continues to grow, many existing studies are limited to binary classification and rely on synthetic datasets with narrow clinical features. In addition, access to real-world datasets remains a challenge that hinders the development of practical diagnostic models. Therefore, there is a pressing need for robust, data-driven classification systems that incorporate comprehensive physiological indicators—such as systolic and diastolic blood pressure, heart rate, and comorbidities—to enable accurate and context-aware clinical diagnoses. This research gap highlights the necessity of developing predictive models that are both clinically relevant and scalable.

To address this gap, the present study develops a classification model using the Support Vector Machine algorithm to distinguish among heart failure, hypertension, and normal cardiac conditions. The model utilizes clinically relevant input variables and is designed to support diagnostic accuracy and medical decision-making. The findings of this study contribute to the integration of artificial intelligence in healthcare systems, particularly in facilitating early detection and risk management of cardiovascular diseases.

2. MATERIALS AND METHODS

2.1 Materials

This study employs a quantitative research approach, selected for its effectiveness in addressing problems through the collection and statistical analysis of numerical data to construct predictive models. This approach emphasizes the use of measurable variables and systematic methodologies to generate objective and replicable findings. In this context, the Support Vector Machine (SVM) algorithm is applied to classify cardiovascular conditions based on relevant clinical parameters (Khan et al., 2023; Kumar et al., 2025; Wan et al., 2025).

The dataset used in this study comprises secondary numerical data related to cardiovascular disease cases, obtained from an open-access data repository. It includes several clinically relevant features such as systolic and diastolic blood pressure, heart rate, presence of comorbidities (e.g., diabetes and kidney disease), and the administration of cardiovascular medications. All data were processed using RapidMiner, a general-purpose data science platform widely utilized for building predictive models across various domains, including healthcare. Prior to model training, the open-access dataset was reviewed and validated by a certified medical professional to ensure the accuracy, consistency, and appropriateness of the assigned labels throughout the dataset. This validation was conducted to enhance the reliability of the classification model and ensure the clinical relevance of the research outcomes.

2.2 Methods

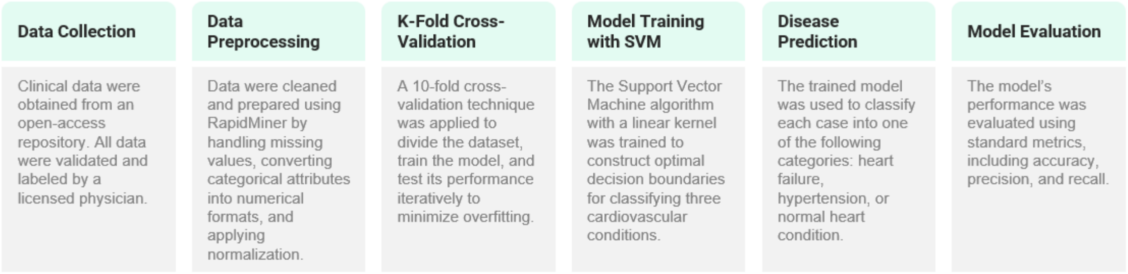
This study follows a structured research workflow consisting of six sequential stages (Figure 1). The process begins with data collection, where clinical records related to cardiovascular conditions were retrieved from a publicly accessible dataset. These records include essential medical features such as systolic and diastolic blood pressure, heart rate, and comorbidities such as diabetes and kidney disease. To ensure clinical validity and ethical compliance, all data were anonymized and subsequently reviewed by a licensed medical doctor to assign categorical labels identifying whether each instance corresponds to heart failure, hypertension, or a normal heart condition.

The second stage is data preprocessing, conducted using the RapidMiner platform. In this phase, missing values were addressed using the "Replace Missing Values" operator, categorical data were converted into numerical format using the "Nominal to Numerical" operator, and continuous variables were normalized to ensure feature comparability. These procedures were essential to prepare a consistent and machine-readable dataset suitable for modeling with the Support Vector Machine (SVM) algorithm.

The third step involves applying K-Fold Cross-Validation to assess model robustness. Using RapidMiner's "X-Validation" operator with K=10, the dataset was divided into ten subsets. The model was trained on nine subsets and tested on the remaining one, and this process was repeated ten times to ensure that every data point was included in the validation set exactly once. This approach mitigates overfitting and enables the calculation of averaged performance metrics.

Next, the model training phase was conducted using the Support Vector Machine algorithm, selected for its proven ability to construct optimal hyperplanes in high-dimensional spaces. A linear kernel (dot product) was used with standard hyperparameters (e.g., C=1.0, epsilon=0.001). The SVM model was trained on the prepared data to learn discriminative patterns that separate the three target classes. Following training and validation, the disease prediction phase involved classifying patient records into one of three diagnostic categories: heart failure, hypertension, or normal heart condition. These predictions were made by applying the trained SVM model to the test folds during cross-validation.

Finally, the evaluation and analysis phase assessed model performance using standard metrics including accuracy, precision, and recall. These metrics provided insights into the model's classification capabilities across all classes.



Data Collection

Clinical data were obtained from an open-access repository. All data were validated and labeled by a licensed physician.

Data Preprocessing

Data were cleaned and prepared using RapidMiner by handling missing values, converting categorical attributes into numerical formats, and applying normalization.

K-Fold Cross-Validation

A 10-fold cross-validation technique was applied to divide the dataset, train the model, and test its performance iteratively to minimize overfitting.

Model Training with SVM

The Support Vector Machine algorithm with a linear kernel was trained to construct optimal decision boundaries for classifying three cardiovascular conditions.

Disease Prediction

The trained model was used to classify each case into one of the following categories: heart failure, hypertension, or normal heart condition.

Model Evaluation

The model's performance was evaluated using standard metrics, including accuracy, precision, and recall.

Figure 1. Research workflow

3. RESULTS AND DISCUSSION

3.1 Data Collection

This study utilizes a secondary dataset comprising cardiovascular patient records, categorized into three clinical conditions: heart failure, hypertension, and normal heart status. The dataset was initially formatted in .xlsx using Microsoft Excel and later imported into RapidMiner for analysis. Each record was annotated based on predetermined class attributes to support classification modeling. Out of a total of 1,119 entries, only 731 records were considered complete and valid. The remaining 388 were excluded due to missing or incomplete values. This selection process ensured the quality and consistency of the input data, which is essential for reliable model training. The class distribution of the dataset is summarized in Table 1.

Table 1. Summary of data collection and class distribution

No.	Class	Count
1	Heart Failure	133
2	Hypertension	345
3	Normal	253
4	Blank Data	388
	Total	1119

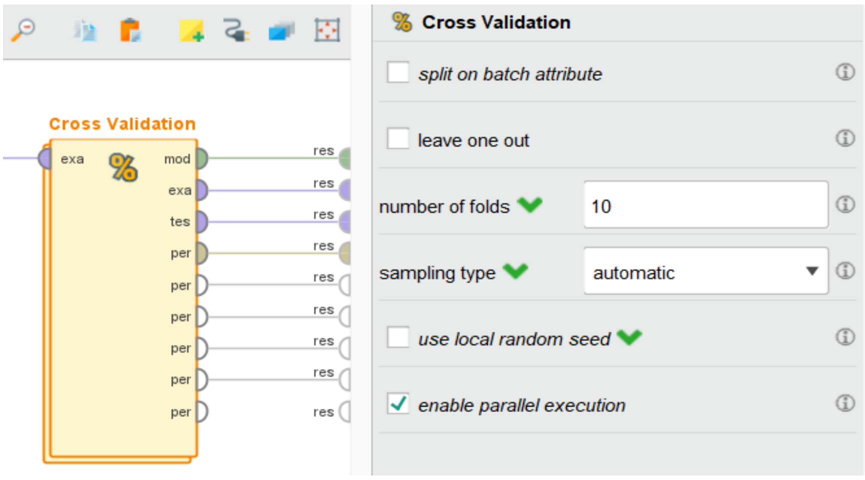
3.2 Data Preprocessing

Data preprocessing is a critical stage in the machine learning pipeline, aimed at ensuring that the dataset is clean, consistent, and suitable for model training (Hidayaturrohman & Hanada, 2024; Saleem et al., 2014). In this study, the preprocessing phase began with an inspection for missing values within the collected patient dataset. Several records were found to be incomplete; therefore, the initial step involved handling these missing values using automated replacement techniques to prevent negative impacts on model accuracy. Subsequently, data type conversion was performed to meet the requirements of the Support Vector Machine (SVM) algorithm. Since SVM operates on numerical input, categorical or nominal attributes—such as types of medication and comorbidity status—were converted into numerical representations. This transformation enables the algorithm to interpret and process categorical variations meaningfully within the feature space.

In addition, normalization was applied to ensure that all features operated on a similar scale. This step is crucial to avoid feature dominance due to differing value ranges, thereby promoting stable and efficient model training. Together, these procedures ensured the dataset was of high quality and fully prepared to support accurate and reliable classification. By thoroughly completing the preprocessing stage, the initially heterogeneous and unstructured clinical data was transformed into a standardized dataset appropriate for predictive model development using machine learning techniques.

3.3 K-Fold Cross Validation

K-Fold Cross Validation, particularly the 10-fold variant (K=10), is a widely used method in machine learning to evaluate model performance, especially when employing algorithms such as Support Vector Machine (SVM) within platforms like RapidMiner. This approach aims to systematically test the generalizability of a model by partitioning the dataset and conducting repeated training and testing cycles to produce more robust predictive results. The configuration of this step is depicted in Figure 2.



Cross Validation

☐ split on batch attribute

☐ leave one out

number of folds: 10

sampling type: automatic

☐ use local random seed

☑ enable parallel execution

Figure 2. Cross-validation parameter configuration in rapidminer

The fundamental structure of K-Fold Cross Validation involves dividing the entire dataset into K equally sized subsets or "folds". In this study, the data were split into ten folds (K=10). The model is trained on nine folds and tested on the remaining one; this process is repeated iteratively until each fold has served as the test set once. This technique ensures that all observations are used for both training and validation, providing a more comprehensive and unbiased estimation of model accuracy. By averaging the performance across all folds, this method reduces the likelihood of overfitting and enhances model robustness.

Within the cross-validation framework, the training subprocess plays a crucial role. During this stage, the SVM algorithm is trained on the designated training folds. The objective is to identify the most discriminative features and establish optimal decision boundaries that can effectively classify cardiovascular conditions. This stage is visually represented in Figure 3, which shows the SVM training configuration.

After the training phase, the testing subprocess is implemented to evaluate the model's predictive performance using the reserved fold, which has not been involved in the training phase. This independent testing is vital to measure how accurately the model generalizes to unseen data, thereby simulating real-world application scenarios. The structure of the testing stage is presented in Figure 4.

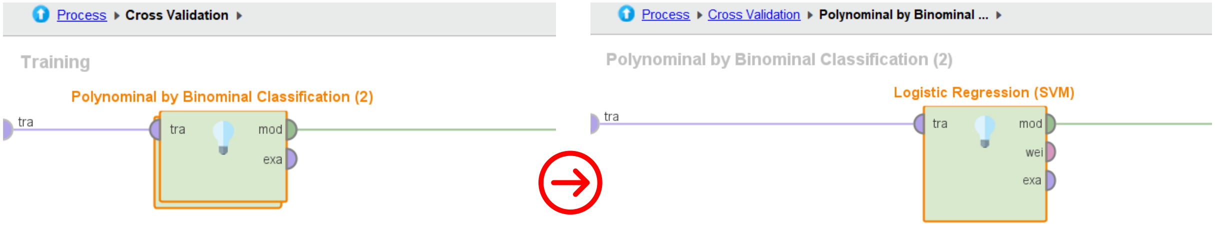


Figure 3. Training subprocess using svm algorithm

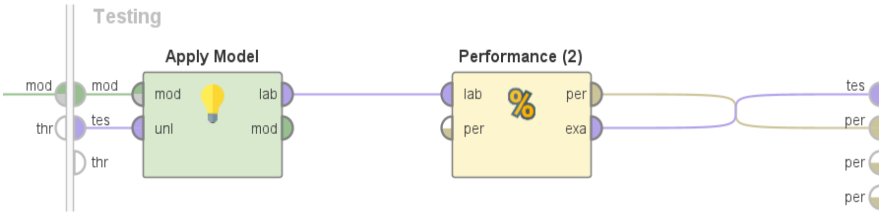


Figure 4. Testing subprocess using unseen data

The entire process, from partitioning the data to model evaluation, is encapsulated in the RapidMiner workflow as shown in Figure 5. The results derived from this cross-validation process—such as accuracy, precision, and recall—are subsequently analyzed to determine the model's overall effectiveness in classifying heart failure, hypertension, and normal heart condition.

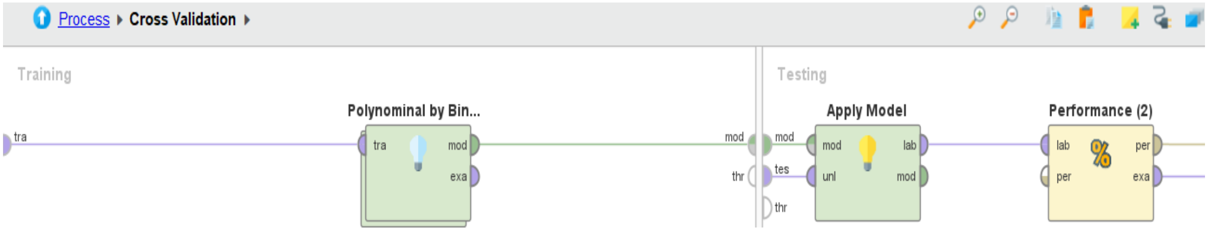
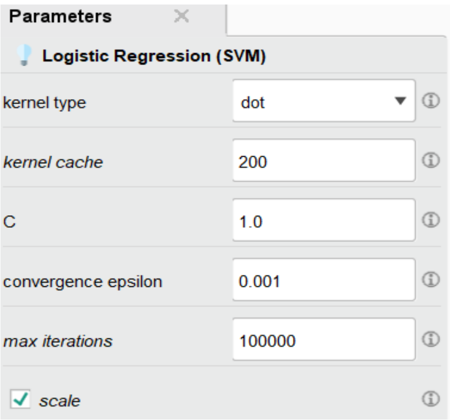


Figure 5. Full cross-validation pipeline in rapidminer

3.4 SVM Classification

The Support Vector Machine (SVM) algorithm is employed in this study due to its robust capability in handling classification tasks, especially in high-dimensional clinical data. SVM is a supervised learning method that identifies the optimal decision boundary to separate distinct classes based on feature vectors. It has been widely used in various domains, including medical diagnostics, owing to its high generalizability and performance on both linear and non-linear data structures.

To address potential non-linearity in the dataset, the SVM model utilizes a linear (dot) kernel, which is computationally efficient and suitable when data is linearly separable in the original feature space. The kernel function defines the geometry of the decision boundary, influencing the model's classification behavior. In this research, the model parameters were configured as follows: a linear kernel with a kernel cache size of 200, regularization parameter (C) set to 1.0 to control the trade-off between classification accuracy and margin width, a convergence threshold (epsilon) of 0.001, and a maximum iteration limit of 100,000. These parameter settings were selected based on experimental tuning using the RapidMiner platform. The configuration process is illustrated in Figure 6.



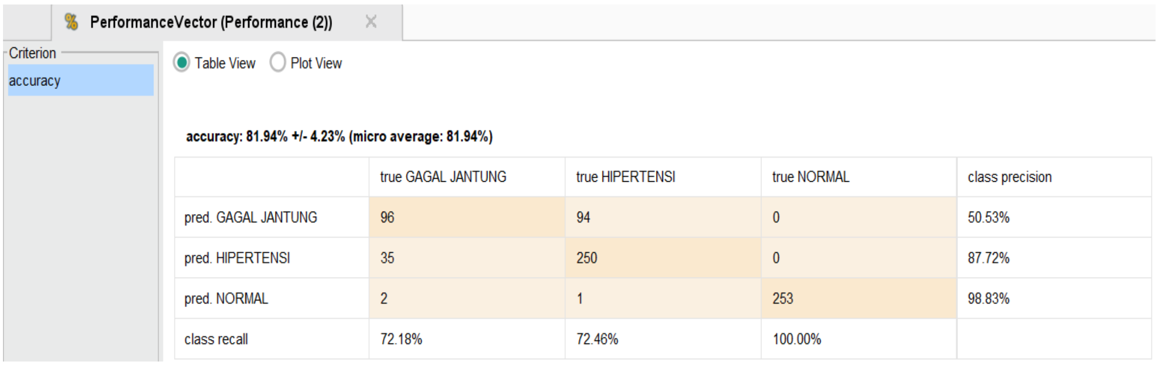
kernel type: dot
kernel cache: 200
C: 1.0
convergence epsilon: 0.001
max iterations: 100000
☑ scale

Figure 6. Parameter settings of the svm model

3.5 Model Evaluation

Model evaluation is a critical step to assess the predictive performance of the SVM classifier. In this study, evaluation was conducted using a confusion matrix, which quantifies the effectiveness of the classification model in categorizing cardiovascular conditions—heart failure, hypertension, and normal heart condition. The evaluation focused on several key performance metrics: accuracy (the proportion of total correct predictions), precision (the proportion of correctly predicted positive cases out of all positive predictions), and recall (the proportion of correctly identified positive cases out of all actual positives). These metrics provide a comprehensive assessment of the model's reliability in distinguishing between the three targeted clinical categories.

To enhance the reliability and generalizability of the model, a 10-Fold Cross-Validation procedure was applied. This method divides the dataset into ten equal partitions, where the model is iteratively trained on nine partitions and tested on the remaining one. The final evaluation metrics are derived by averaging the results across all folds. The outcomes of this evaluation, including the confusion matrix and performance scores, are presented in Figure 7.



accuracy: 81.94% +/- 4.23% (micro average: 81.94%)

	true GAGAL JANTUNG	true HIPERTENSI	true NORMAL	class precision
pred. GAGAL JANTUNG	96	94	0	50.53%
pred. HIPERTENSI	35	250	0	87.72%
pred. NORMAL	2	1	253	98.83%
class recall	72.18%	72.46%	100.00%	

Figure 7. Evaluation results of the svm classification model

3.6 Discussion

This study demonstrates the effectiveness of the Support Vector Machine (SVM) algorithm in classifying cardiovascular conditions—specifically heart failure, hypertension, and normal heart condition—using real-world patient data. The results suggest that SVM can accurately distinguish among these classes when trained with appropriately preprocessed and labeled clinical datasets. Compared to many previous studies that focused solely on binary classification or used synthetic datasets, this research adds value by addressing a multi-class classification problem in a real clinical context.

The application of K-Fold Cross Validation (K=10) enhanced the robustness of the model by minimizing the risk of overfitting and improving the reliability of performance metrics. The evaluation results—

comprising accuracy, precision, and recall—indicated consistent classification performance across the three categories. These findings align with prior studies highlighting the robustness of SVM in handling high-dimensional medical data and complex pattern recognition tasks (Ghasemi & Sharifi, 2025; Plati et al., 2021).

Furthermore, the use of a linear kernel function proved effective in this study, given the structure of the dataset. Although non-linear kernels may offer better performance in certain complex scenarios, the linear kernel provided an efficient and interpretable model without compromising accuracy. This finding supports the notion that simpler kernels can be sufficient when data preprocessing is conducted rigorously, and feature engineering is clinically informed. Nevertheless, several limitations should be acknowledged. First, the study relies on a secondary dataset with limited contextual features, such as lifestyle or genetic predispositions, which may influence cardiovascular conditions. Second, the exclusion of incomplete records, while necessary for ensuring data quality, may introduce selection bias.

4. CONCLUSION

This study demonstrates that the Support Vector Machine (SVM) algorithm is effective for classifying three major cardiovascular conditions—heart failure, hypertension, and normal heart condition—based on secondary clinical data that has been medically validated. The classification process was supported by a systematic methodological framework, including data collection and preprocessing, K-Fold cross-validation, and model training using a linear kernel configuration. The evaluation results, measured by accuracy, precision, and recall metrics, indicate that the model achieved strong and reliable classification performance.

These findings highlight the potential of integrating machine learning algorithms, particularly SVM, into clinical decision support systems to facilitate early diagnosis of cardiovascular diseases. By employing an accurate and efficient data-driven approach, the proposed model contributes to enhancing the effectiveness of healthcare services and supporting evidence-based clinical decision-making.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

REFERENCES

Abdullah, D. M., & Abdulazeez, A. M. (2021). Machine learning applications based on svm classification a review. *Qubahan Academic Journal*, *1*(2), 81–90. https://doi.org/10.48161/QAJ.V1N2A50

Amarappa, S., & Sathyanarayana, S. V. (2014). Data classification using support vector machine (svm), a simplified approach. *International Journal of Electronics and Computer Science Engineering*, *3*(4).

Carey, R. M., Whelton, P. K., Aronow, W. S., Casey, D. E., Collins, K. J., Himmelfarb, C. D., DePalma, S. M., Gidding, S., Jamerson, K. A., Jones, D. W., McLaughlin, E. J., Muntner, P., Ovbiagele, B., Smith, S. C., Spencer, C. C., Stafford, R. S., Taler, S. J., Thomas, R. J., Williams, K. A., … Wright, J. T. (2018). Prevention, detection, evaluation, and management of high blood pressure in adults: synopsis of the 2017 american college of cardiology/american heart association hypertension guideline. *Annals of Internal Medicine*, *168*(5), 351–358. https://doi.org/10.7326/M17-3203

Chong, B., Jayabaskaran, J., Jauhari, S. M., Chan, S. P., Goh, R., Kueh, M. T. W., Li, H., Chin, Y. H., Kong, G., Anand, V. V., Wang, J.-W., Muthiah, M., Jain, V., Mehta, A., Lim, S. L., Foo, R., Figtree, G. A., Nicholls, S. J., Mamas, M. A., … Chan, M. Y. (2024). Global burden of cardiovascular diseases: projections from 2025 to 2050. *European Journal of Preventive Cardiology*. https://doi.org/10.1093/EURJPC/ZWAE281

Cifu, A. S., & Davis, A. M. (2017). Prevention, detection, evaluation, and management of high blood pressure in adults. *JAMA*, *318*(21), 2132–2134. https://doi.org/10.1001/JAMA.2017.18706

Galindra, Y., Astiah, A. A., & Nuralyfah, N. S. (2024). Hubungan antara derajat hipertensi dengan kualitas tidur pada pasien hipertensi di rumah sakit pmi kota bogor. *Zona Kedokteran: Program Studi Pendidikan Dokter Universitas Batam*, *14*(2). https://doi.org/10.37776/ZKED.V14I2.1533

Ghasemi, F., & Sharifi, S. (2025). Heart failure prediction using support vector machine. *International Journal of Novel Research in Life Sciences*, *25*(1).

Hidayaturrohman, Q. A., & Hanada, E. (2024). Impact of data pre-processing techniques on xgboost model performance for predicting all-cause readmission and mortality among patients with heart failure. *BioMedInformatics*, *4*(4), 2201–2212. https://doi.org/10.3390/BIOMEDINFORMATICS4040118

Khan, A., Qureshi, M., Daniyal, M., & Tawiah, K. (2023). A novel study on machine learning algorithm-based cardiovascular disease prediction. *Health & Social Care in the Community*, *2023*(1), 1406060. https://doi.org/10.1155/2023/1406060

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

Liu, Z. “Leo.” (2025). Support vector machines. In *Artificial Intelligence for Engineers* (pp. 129–140). Springer Nature Switzerland. https://doi.org/10.1007/978-3-031-75953-6_5

Patle, A., & Chouhan, D. S. (2013). Svm kernel functions for classification. *2013 International Conference on Advances in Technology and Engineering, ICATE 2013*. https://doi.org/10.1109/ICADTE.2013.6524743

Plati, D. K., Tripoliti, E. E., Bechlioulis, A., Rammos, A., Dimou, I., Lakkas, L., Watson, C., McDonald, K., Ledwidge, M., Pharithi, R., Gallagher, J., Michalis, L. K., Goletsis, Y., Naka, K. K., & Fotiadis, D. I. (2021). A machine learning approach for chronic heart failure diagnosis. *Diagnostics 2021*, *11*(10), 1863. https://doi.org/10.3390/DIAGNOSTICS11101863

Saleem, A., Asif, K. H., Ali, A., Awan, S. M., & Alghamdi, M. A. (2014). Pre-processing methods of data mining. *Proceedings - 2014 IEEE/ACM 7th International Conference on Utility and Cloud Computing, UCC 2014*, 451–456. https://doi.org/10.1109/UCC.2014.57

Sandhya, Y. (2020). *Prediction of heart diseases using support vector machine*. *8*. https://doi.org/10.22214/ijraset.2020.2021

Shihong, Y., Ping, L., & Peiyi, H. (2003). Svm classification: its contents and challenges. *Applied Mathematics*, *18*(3), 332–342. https://doi.org/10.1007/S11766-003-0059-5

Shmilovici, A. (2023). Support vector machines. In *Machine Learning for Data Science Handbook: Data Mining and Knowledge Discovery Handbook, Third Edition* (pp. 93–110). Springer International Publishing. https://doi.org/10.1007/978-3-031-24628-9_6

Srivastava, D. K., & Bhambhu, L. (2010). Data classification using support vector machine. *Journal of Theoretical and Applied Information Technology*, *12*(1).

Suryadi, S., Solikin, S., & Uni, U. (2024). Analisa faktor risiko komplikasi gagal jantung pada pasien hipertensi di rsud ulin banjarmasin. *Jurnal Keperawatan Suaka Insan (JKSI)*, *9*(2), 142–148. https://doi.org/10.51143/JKSI.V9I2.708

Wan, S., Wan, F., & Dai, X. jian. (2025). Machine learning approaches for cardiovascular disease prediction: A review. *Archives of Cardiovascular Diseases*. https://doi.org/10.1016/J.ACVD.2025.04.055

World Health Organization. (2021). *Cardiovascular diseases (CVDs)*. https://www.who.int/news-room/fact-sheets/detail/cardiovascular-diseases-(cvds)

Yu, H., & Kim, S. (2012). Svm tutorial — classification, regression and ranking. In *Handbook of Natural Computing* (Vols. 1–4, pp. 479–506). Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-540-92910-9_15

Yuniarti, E. (2022). *Epidemiologi gagal jantung*. https://www.alomedika.com/penyakit/kardiologi/gagal-jantung/epidemiologi