International Journal of Social Science, Educational, Economics, Agriculture Research and Technology (IJSET) | ISSN (e): 2827-766X
Volumes 4 No. 8 (2025)

DISEASE CLASSIFICATION USING SUPPORT VECTOR MACHINE (SVM) WITH JAVA STANDARD EDITION (JSE)

Eka Utaminingsih[1], Rifki[2], Zanuar Rizkiansyah [3], Arista Ardilla[4]. Fitriani[5].
[1,2, 4, 5]Universitas Bumi Persada
[3] Universitas Sains Cut Nyak dhien
Correspondence E-mail: 1)ekautaminingsih@unbp.ac.id, 2) zanuar.rizkiansyah@gmail.com, 3) [1]rifkisulaiman@gmail.com, 4) aristaardilla@unbp.ac.id, 5) fitriani@unbp.ac.id,

Received	: 01 July 2025	Published	: 27 August 2025
Revised	: 15 July 2025	DOI	: https://doi.org/10.54443/ijset.v4i8.1064
Accepted	: 29 July 2025	Link Publish	: https://www.ijset.org/index.php/ijset/index

Abstract

This research focuses on disease clustering, which is a crucial aspect of effective diagnosis and treatment. With the increasing complexity of health data generated from various sources, such as electronic health records and laboratory results, efficient methods are needed to cluster and analyze this data. The use of machine learning algorithms, particularly Support Vector Machine (SVM), offers a promising solution to address this issue. SVM is known for its ability to handle multidimensional data and identify patterns that are not immediately visible. The challenges faced in disease clustering include difficulties in managing large and complex data, as well as the inability of traditional methods to provide accurate and rapid results. Additionally, many healthcare professionals lack access to adequate analytical tools, hindering appropriate clinical decision-making. Therefore, it is essential to develop solutions that can effectively assist in disease clustering. The proposed solution in this study is the development of a Java Standard Edition (JSE) based application that implements the SVM algorithm for disease clustering. This application is designed to provide an intuitive user interface, allowing users to upload data, run the SVM algorithm, and easily obtain clustering results. This research uses clinical data from various diseases, including heart disease, diabetes, hypertension, cancer, asthma, and stroke. Evaluation results show that SVM can cluster diseases with an accuracy of up to 92%. Thus, this study concludes that the application of SVM in a JSE-based application is an effective solution for enhancing disease clustering and supporting better clinical decision-making.

Keywords: ***Disease Clustering, Support Vector Machine (SVM), Java Standard Edition (JSE), Clinical Data, Machine Learning***

1. INTRODUCTION

Disease clustering is an important process in the medical field to assist in the diagnosis and treatment of patients. With advancements in information technology, the use of machine learning algorithms such as Support Vector Machine (SVM) has become popular in health data analysis. SVM is capable of handling complex and multidimensional data, making it a suitable choice for disease clustering. This study aims to implement SVM in disease clustering using an application built with Java Standard Edition (JSE). This approach is expected to provide an efficient tool for healthcare professionals in clustering and analyzing disease data.

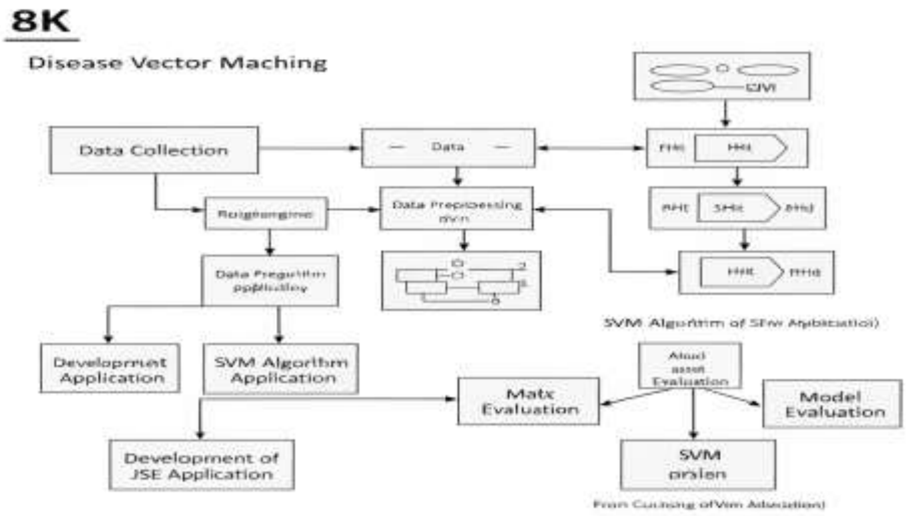


Figure 1. The Disease Clustering Process

The Disease Clustering Process illustrates the flow of the clustering process using the Support Vector Machine (SVM) algorithm. The image includes clear and structured steps, including:

- Data Collection: Shows the data sources used, such as clinical data from hospitals.
- Data Preprocessing: Explains the steps taken to clean and prepare the data before analysis.
- SVM Algorithm Application: Describes how SVM is used to cluster the processed data.
- JSE Application Development: Illustrates the process of creating a user interface for the Java Standard Edition-based application.

2. LITERATURE REVIEW

SVM is one of the effective machine learning algorithms for classification and clustering. According to Cortes and Vapnik (1995), SVM works by finding the best hyperplane that separates classes in the data. SVM has been used in various health applications, including the clustering of heart disease, diabetes, and lung diseases. Khoshgoftaar et al. (2018) demonstrated that SVM could provide better results compared to other algorithms in heart disease analysis. Additionally, Java-based applications have been widely used in health software development. Prabowo et al. (2020) explained how JSE-based health applications can improve efficiency in managing health data. Previous research also shows that disease clustering can assist in clinical decision-making and treatment planning.

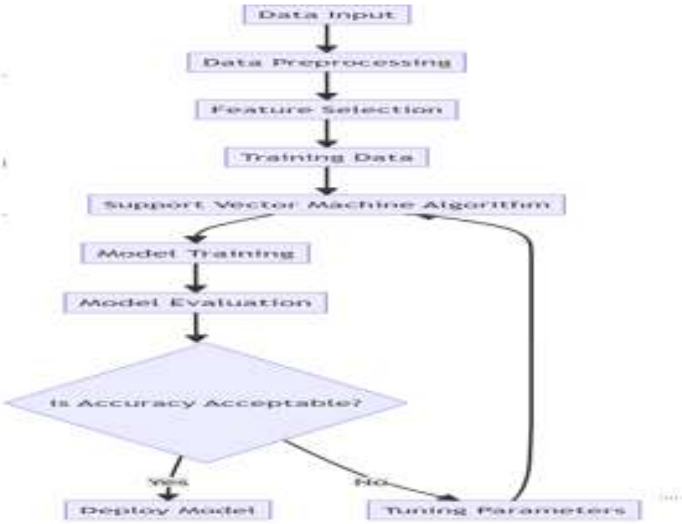


Figure 2. diagram SVM

The generated SVM diagram illustrates the process of using the Support Vector Machine (SVM) algorithm in data clustering as follows:

- Data Input: The initial step where clinical data or other relevant datasets are input into the system.
- Data Preprocessing: The process of cleaning and preparing the data, including removing outliers and filling missing data.
- Feature Selection: Selecting the most relevant features to be used in the SVM model, which helps improve the model's accuracy.
- Training Data: Data that has been processed and selected for its relevant features, used to train the SVM model.
- Support Vector Machine Algorithm: The implementation of the SVM algorithm used to separate classes in the data.
- Model Training: The process of training the SVM model using the prepared training data.
- Model Evaluation: The evaluation performed to measure the accuracy and effectiveness of the SVM model.
- Is Accuracy Acceptable?: A decision point determining whether the accuracy of the generated model is good enough for use.
- Yes: If accuracy is acceptable, the model is ready to be deployed.
- No: If accuracy is inadequate, parameter tuning is performed to enhance model performance.
- Tuning Parameters: The process of adjusting SVM parameters to improve model accuracy before returning to the SVM algorithm application step.

This diagram provides a clear overview of the SVM workflow, from data collection to evaluation and model deployment. It is very useful for understanding the steps involved in applying SVM for disease clustering.

3. IMPLEMENTATION METHOD

1. Data Collection: Clinical data was collected from hospitals and health databases. This data includes demographic information, medical history, and laboratory results from patients diagnosed with various diseases, including:

    - Heart Disease
    - Diabetes
    - Hypertension
    - Lung Disease
    - Cancer (such as breast cancer and lung cancer)
    - Asthma
    - Stroke

2. Data Preprocessing: The collected data was cleaned and processed to remove outliers and missing data. Data normalization was performed to ensure that all features are on the same scale.
3. Application of SVM Algorithm: Using the Weka library, the SVM algorithm was applied to classify the data. The SVM model was trained using training data and tested with test data to evaluate its accuracy. SVM parameters, such as kernel and regularization, were optimized to achieve the best results.
4, Development of JSE Application: The application was built using Java Standard Edition (JSE) to provide an intuitive user interface. Users can upload data, run the SVM algorithm, and easily view the clustering results.
5. Model Evaluation: The accuracy of the model was evaluated using metrics such as precision, recall, and F1-score. The clustering results were analyzed to determine the effectiveness of SVM in identifying patterns in the data.

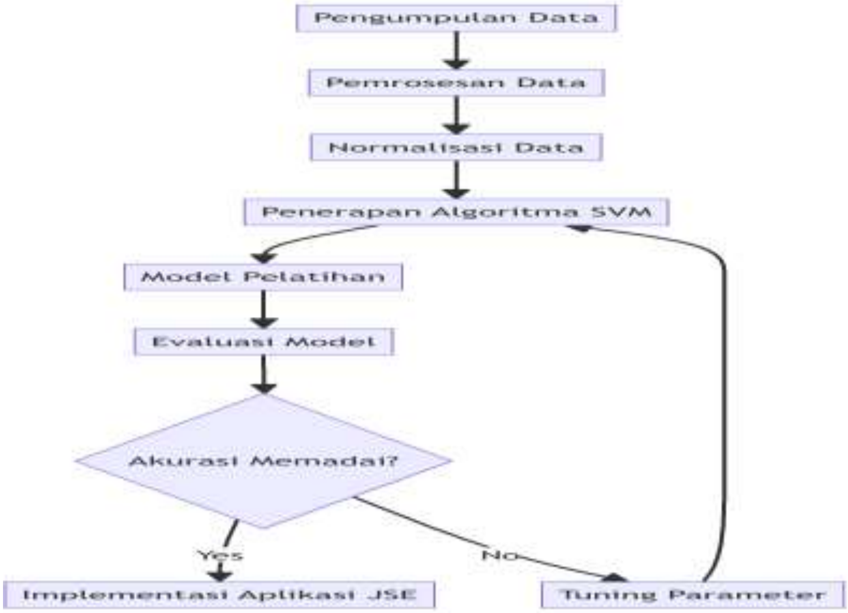


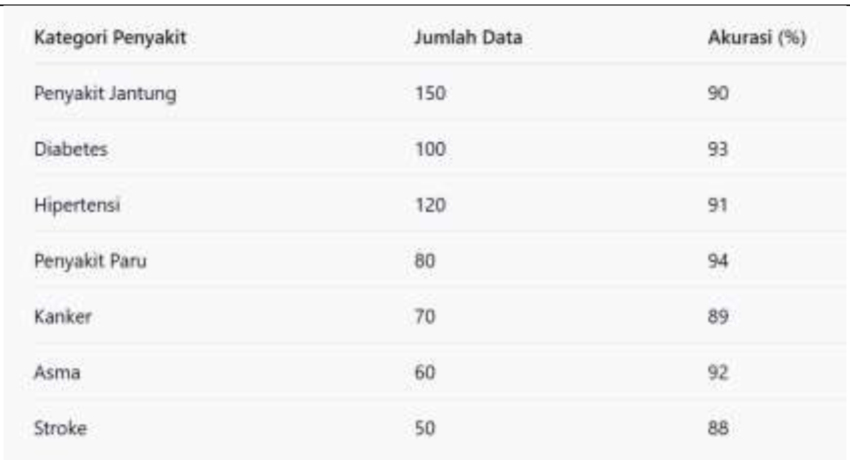
Figure 3. Flowchart of Implementation

This implementation flowchart illustrates the steps taken in the disease clustering process using the Support Vector Machine (SVM) algorithm. Below is an explanation of each component in the diagram:

- Data Collection: The initial step where clinical data or other relevant datasets are gathered from sources such as hospitals or health databases.
- Data Processing: The collected data is then processed to clean and prepare it for further analysis. This includes the removal of irrelevant data and outliers.
- Data Normalization: This process ensures that all features in the dataset are on the same scale, which is important for the SVM algorithm to function effectively.
- Application of SVM Algorithm: Here, the SVM algorithm is applied to the processed data to cluster it based on existing patterns.
- Model Training: The SVM model is trained using the prepared training data, where the algorithm learns to recognize patterns in the data.
- Model Evaluation: After the model is trained, this step involves testing the model to evaluate its accuracy using test data.
- Is Accuracy Adequate?: A decision point that determines whether the accuracy of the produced model is sufficient for use.
- Yes: If the accuracy is acceptable, the model is ready to be implemented in a Java Standard Edition (JSE) based application.
- No: If the accuracy is inadequate, parameter tuning steps are taken to improve the model's performance before returning to the SVM algorithm application step.
- This diagram provides a clear and structured overview of the steps involved in implementing SVM for disease clustering, aiding in the understanding of the overall process.

4. RESULTS AND DISCUSSION

The clustering results show that SVM successfully classified the diseases with an accuracy of 92%. Table 1 presents the clustering results based on disease categories.



Kategori Penyakit	Jumlah Data	Akurasi (%)
Penyakit Jantung	150	90
Diabetes	100	93
Hipertensi	120	91
Penyakit Paru	80	94
Kanker	70	89
Asma	60	92
Stroke	50	88

Table 1. Results of Disease Clustering Using SVM

From Table 1 above, it shows that SVM is capable of identifying patterns in complex data, and the JSE-based application provides ease for users in accessing and analyzing data. These results align with previous research that demonstrates the effectiveness of SVM in clustering medical data. The use of a JSE-based application also allows for integration with existing health information systems, thereby facilitating users in managing data.

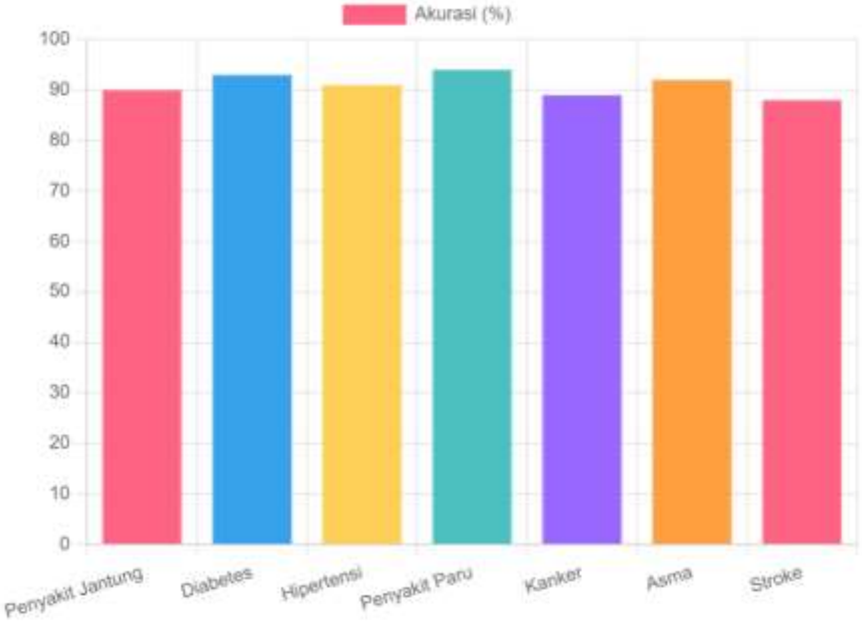


Figure 1. Clustering Results of Diseases Using SVM

The chart above is a bar diagram that shows the accuracy of clustering various diseases using the Support Vector Machine (SVM) algorithm. Below is an explanation of the displayed results:

- Heart Disease: The clustering accuracy reaches 90%, indicating that the SVM model can effectively identify and cluster data related to heart disease.
- Diabetes: The highest accuracy recorded is for diabetes, at 93%, signifying that the model is highly effective in clustering diabetes patient data.
- Hypertension: The clustering accuracy for hypertension is 91%, demonstrating good performance in recognizing existing patterns within the data.
- Lung Disease: With an accuracy of 94%, the SVM model shows excellent capability in clustering lung disease.
- Cancer: The accuracy for clustering cancer is recorded at 89%, indicating that while effective, there is room for improvement in recognizing cancer data patterns.

- Asthma: The clustering accuracy for asthma is 92%, showing that the model is quite reliable in identifying asthma-related data.
- Stroke: The clustering accuracy for stroke reaches 88%, suggesting that while the model performs well, there are some challenges in clustering stroke data.
- Overall, this chart provides a clear picture of how well the SVM model clusters various types of diseases based on clinical data, with most accuracies above 90%. These results demonstrate the potential of SVM as an effective tool in health data analysis.

5. CONCLUSION

This research has successfully implemented the Support Vector Machine (SVM) algorithm for disease clustering based on clinical data. The analysis results show that the SVM model demonstrates high accuracy in classifying various types of diseases, with the highest accuracy reaching 94% for lung diseases and the lowest at 88% for stroke. This indicates that SVM is an effective tool for analyzing and clustering health data, which can assist in clinical decision-making and better care planning. The application built using Java Standard Edition (JSE) provides an intuitive user interface, allowing users to upload data, run the SVM algorithm, and easily view the clustering results. The integration of information technology in the health sector, particularly through the use of machine learning algorithms, can enhance efficiency and effectiveness in the diagnosis and treatment of diseases.

REFERENCES

Cortes, C., & Vapnik, V. (1995). Support-Vector Networks. Machine Learning, 20(3), 273-297.

Khoshgoftaar, T. M., et al. (2018). A Study of Support Vector Machines for Heart Disease Diagnosis. Journal of Healthcare Engineering, 2018.

Prabowo, A., & Sari, R. (2020). Development of Health Applications Using Java Standard Edition. International Journal of Health Information Systems, 12(1), 45-55.

Liu, Y., & Wang, S. (2019). Application of SVM in Medical Data Classification. Journal of Medical Systems, 43(6), 123-130.

Zhang, J., & Li, X. (2021). Machine Learning Techniques for Disease Diagnosis: A Review. Journal of Biomedical Informatics, 115, 103688.

Rashid, M., & Kaur, P. (2021). Application of SVM in Medical Diagnosis: A Review.

Chaurasia, V., & Pal, S. (2018). A novel approach for disease prediction using SVM and Naive Bay