E-ISSN 3089-8838

JUTEKOM
Jurnal Teknologi dan Ilmu Komputer
https://jurnal.nolsatu.co.id/jutekom
2025 Vol. 1 No. 4 /Hal.152-157

Application of Support Vector Machine for Multi-Class Migraine Classification

Akhmad Rezki Purnajaya[1], Mega Jaya[2]

[1,2] Department of Software Engineering, Faculty of Computer Science, Universal University, Batam, Indonesia

rezkipurnajaya@gmail.com

Abstract

Migraine is a prevalent, debilitating neurological disorder where accurate subtype classification is critical. Machine learning (ML) offers a promising avenue to enhance diagnostic accuracy. This study evaluates a Support Vector Machine (SVM) model for multi-class migraine classification. Utilizing a public Kaggle dataset, data was partitioned into 75% training and 25% testing sets. An SVM with a linear kernel was implemented to classify seven migraine subtypes. Performance was evaluated using overall accuracy, a confusion matrix, and detailed per-class metrics: Precision, Recall, and F1-Score. The model achieved 82.65% overall accuracy and a weighted-average F1-Score of 0.824. However, detailed metrics revealed significant variance. The model achieved perfect F1-Scores (1.000) for 'Migraine Without Aura' and 'Typical Aura without Migraine' but struggled with class confusion. 'Typical Aura With Migraine' exhibited a low Recall (0.533), and 'Basilar-Type Aura' had a poor F1-Score (0.400). Critically, the model completely failed to classify 'Sporadic Hemiplegic Migraine' (0.000 F1-Score), a failure masked by the high overall accuracy. These results suggest the linear SVM is a viable baseline, but its reliability varies drastically across subtypes. The granular F1-Score and Recall metrics are essential, exposing classification failures hidden by overall accuracy. Future work must address class imbalance and symptomatic overlap, likely via non-linear models, before this approach is clinically viable.

Keywords: *Migraine Classification, Support Vector Machine (SVM), Machine Learning, Headache Disorders, Medical Diagnosis, Multi-class Classification*

Abstrak

Migrain adalah gangguan neurologis umum yang melemahkan, di mana klasifikasi subtipe yang akurat sangat penting. *Machine learning* (ML) menawarkan jalan yang menjanjikan untuk meningkatkan akurasi diagnostik. Penelitian ini mengevaluasi model *Support Vector Machine* (SVM) untuk klasifikasi migrain multi-kelas. Menggunakan dataset publik Kaggle, data dibagi menjadi 75% set pelatihan dan 25% set pengujian. SVM dengan *kernel linear* diimplementasikan untuk mengklasifikasikan tujuh subtipe migrain. Kinerja dievaluasi menggunakan akurasi keseluruhan, *confusion matrix*, dan metrik terperinci per kelas: *Precision*, *Recall*, dan *F1-Score*. Model ini mencapai akurasi keseluruhan 82,65% dan *weighted-average F1-Score* sebesar 0,824. Namun, metrik terperinci menunjukkan variasi yang signifikan. Model mencapai *F1-Score* sempurna (1,000) untuk 'Migraine Without Aura' dan 'Typical Aura without Migraine', tetapi kesulitan dengan kebingungan kelas. 'Typical Aura With Migraine' menunjukkan *Recall* yang rendah (0,533), dan 'Basilar-Type Aura' memiliki *F1-Score* yang buruk (0,400). Secara kritis, model ini gagal total mengklasifikasikan 'Sporadic Hemiplegic Migraine' (*F1-Score* 0,000), sebuah kegagalan yang tertutupi oleh akurasi keseluruhan yang tinggi. Hasil ini menunjukkan SVM linear adalah *baseline* yang layak, tetapi keandalannya sangat bervariasi antar subtipe. Metrik granular *F1-Score* dan *Recall* sangat penting, karena mengungkap kegagalan klasifikasi yang tersembunyi oleh akurasi keseluruhan. Penelitian selanjutnya harus mengatasi ketidakseimbangan kelas dan tumpang tindih gejala, kemungkinan melalui model non-linear, sebelum pendekatan ini layak secara klinis.

Kata kunci: Klasifikasi Migrain, *Support Vector Machine* (SVM), Pembelajaran Mesin, Gangguan Sakit Kepala, Diagnosis Medis, Klasifikasi Multi-kelas

Diterima: 14-10-2025 | Revisi: 22-10-2025 | Diterbitkan: 30-10-2025 | doi: 10.65258/jutekom.v1.i4.29

1. Introduction

Migraine is a primary headache disorder that represents a significant global health burden, ranking as a leading cause of disability worldwide [1]. Characterized by recurrent attacks of moderate to severe, pulsating headaches, it is often accompanied by debilitating symptoms such as nausea, vomiting, photophobia (sensitivity to light), and phonophobia (sensitivity to sound) [2]. The clinical presentation of migraine is heterogeneous, encompassing several subtypes defined by the International Classification of Headache Disorders (ICHD), including common forms like migraine with and without aura, as well as rarer, complex variants like hemiplegic migraine [3]. The accurate diagnosis and classification of these subtypes are paramount for guiding appropriate therapeutic strategies and improving patient outcomes. However, diagnosis remains a considerable challenge; it relies almost exclusively on the interpretation of patient-reported symptoms and clinical history, a process that can be complicated by the symptomatic overlap between migraine subtypes and other headache disorders [4]. This complexity can lead to diagnostic delays and misclassification, particularly in non-specialist settings, underscoring the need for more objective and reliable diagnostic aids [5].

In recent years, the field of medical diagnostics has witnessed a paradigm shift towards data-driven methodologies, with machine learning (ML) at the forefront [6]. ML algorithms possess the capacity to analyze large, high-dimensional datasets and identify complex, non-linear patterns that may be imperceptible to human observers [7]. This capability has been leveraged across numerous medical domains to enhance diagnostic accuracy, predict disease progression, and personalize treatment plans [8]. In neurology, ML has shown particular promise for classifying complex disorders based on clinical, neuroimaging, and genetic data, offering a powerful toolkit to address long-standing diagnostic challenges [9, 10].

Among the array of ML algorithms, the Support Vector Machine (SVM) is a well-established and robust supervised learning model renowned for its efficacy in classification tasks [11]. The fundamental principle of SVM is to identify an optimal decision boundary, or hyperplane, that maximizes the margin of separation between different classes in a high-dimensional feature space [12]. Its effectiveness, particularly in handling high-dimensional data, has led to its successful application in various biomedical fields, including proteomics, genomics, and disease diagnosis [13, 14].

a. This study aims to contribute to the growing body of evidence on ML applications in headache medicine by evaluating the performance of a foundational ML technique on a readily accessible type of clinical data [15]. Specifically, the objective is to implement and assess a linear SVM model for the multi-class classification of migraine subtypes using a publicly available dataset derived from patient-reported symptoms. By validating this fundamental approach, this work seeks to establish a performance baseline and affirm the utility of ML as a potential decision-support tool in the clinical assessment of migraine [16].

2. Research Methodology

2.1. Data Source and Characteristics

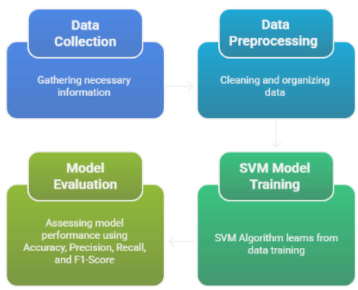

Figure 1. SVM model development process

This investigation utilized a secondary, publicly available dataset entitled "Data for Migraine Classification," sourced from the Kaggle repository and created by Emir Yarkin Yaman [2]. The dataset comprises 394 individual instances, each characterized by 21 distinct features. These features encompass demographic information, such as *Age*, and a comprehensive set of clinical variables describing the characteristics of headache attacks. Key clinical features include *Duration, Frequency, Location, Character, Intensity, and the presence of associated symptoms like Nausea, Vomit, Phonophobia, Photophobia, Visual disturbances*, and *Sensory symptoms* [4].

The primary outcome variable for classification, *Type*, is a categorical feature representing seven distinct subtypes of migraine [1]. These subtypes include Basilar-Type Aura, Familial Hemiplegic Migraine, Migraine Without Aura, Other, Sporadic Hemiplegic Migraine, Typical Aura With Migraine, and Typical Aura without Migraine [3].

2.2. Data Preprocessing and Partitioning

The analysis was conducted using the R programming language and RStudio environment. As a preliminary step, the multi-class target variable, *Type*, was converted into a factor data type to ensure its correct interpretation by the classification algorithm [11].

To evaluate the model's generalization performance on unseen data, the dataset was partitioned into a training set and a testing set [16]. A random sampling approach was employed to allocate 75% of the data (296 instances) to the training set, which was used for model construction. The remaining 25% of the data (98 instances) was reserved as the hold-out testing set for final model evaluation. To ensure the reproducibility of this random split, a seed value was set to ‘12345’ using the *set.seed()* function prior to sampling [1].

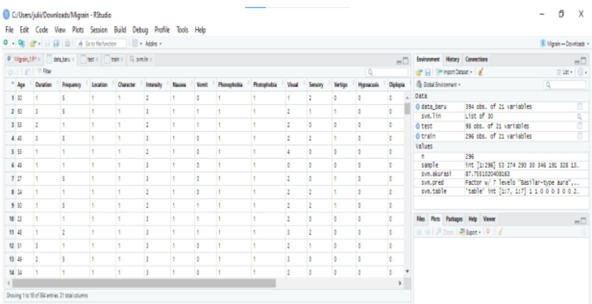

Figure 2. Migraine data results that have been imported

2.3. SVM Model Implementation

The core of the classification task was performed using a Support Vector Machine, a powerful algorithm that seeks to find a hyperplane that best separates data points into their respective classes. The implementation was carried out using the *svm* function within the *e1071* package in R, a standard and widely used library for SVMs [11, 12].

The SVM model was configured with the following specifications:

- Model Type: The model was specified as *a C-classification* type, which is the standard formulation for multi-class classification problems [12].
- Kernel Function: A linear kernel was selected for this analysis. The linear kernel attempts to find a linear decision boundary to separate the classes and is defined by the function $K(x_i, x_j) = x_i^T * x_j$. This choice represents the simplest form of SVM classification, assuming that the data can be separated by a straight line or hyperplane in the feature space [14].
- Model Parameters: The model was trained using the formula *Type ~.*, indicating that all other features in the dataset were used as predictors for the migraine *Type*. The *cost* parameter, which controls the penalty for misclassified data points and thus the trade-off between a wider margin and fewer training errors, was set to its default value of 1 [11].

3. Results and Discussion

3.1. Overall Model Performance

The trained SVM model with a linear kernel was evaluated on the hold-out test set, which consisted of 98 instances not seen during the training phase.

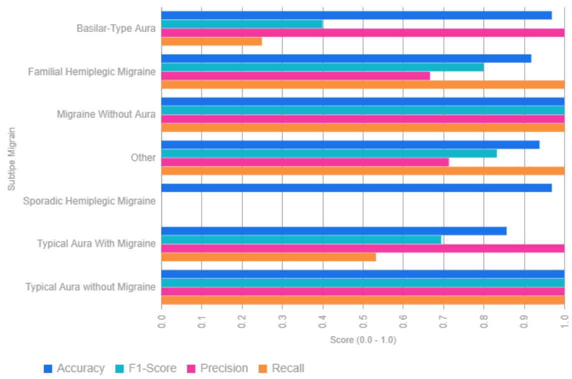

Figure 3. Comparison of SVM performance metrics

Figure 2 presents a detailed, per-class evaluation of the SVM model's performance, assessing all seven migraine subtypes using four distinct metrics. It includes Precision (the accuracy of the model's positive predictions), Recall (its ability to identify all true positive instances), and the F1-

Score (the harmonic mean of Precision and Recall, which is crucial for imbalanced data). This table also adds Per-Class Accuracy, a metric that measures how well the model distinguishes a single class from all other classes combined (e.g., "Basilar-Type Aura" vs. "Not Basilar-Type Aura"). The 'Weighted Avg' F1-Score of 0.824 remains a key indicator of overall, balanced performance, as it is weighted by the number of instances in each class.

These metrics quantify the model's specific strengths and weaknesses. While classes like 'Migraine Without Aura' show perfect F1-Scores (1.000), the low Recall (0.533) for 'Typical Aura With Migraine' numerically confirms the significant confusion mentioned in the study. The 'Per-Class Accuracy' column is particularly illustrative: 'Sporadic Hemiplegic Migraine' has a 0.000 F1-Score, indicating total failure in positive identification, yet its 'Per-Class Accuracy' is high (0.9694). This is because the model was very successful at identifying *true negatives* (correctly labeling other types as *not* sporadic), which highlights why F1-Score is often a more reliable performance measure than accuracy in imbalanced classification tasks.

3.2. Confusion Matrix Analysis

For a more detailed assessment of the model's performance across the individual migraine subtypes, a confusion matrix was generated. This matrix provides a comprehensive breakdown of correct and incorrect predictions for each class. The results are summarized in Table 1.

Table 1. Confusion matrix of the SVM model on the test set

Actual Type	Predicted Type						
	Basilar-Type Aura	Familial Hemiplegic Migraine	Migraine Without Aura	Other	Sporadic Hemiplegic Migraine	Typical Aura With Migraine	Typical Aura without Migraine
Basilar-Type Aura	1	2	0	1	0	0	0
Familial Hemiplegic Migraine	0	16	0	0	0	0	0
Migraine Without Aura	0	0	25	0	0	0	0
Other	0	0	0	15	0	0	0
Sporadic Hemiplegic Migraine	0	0	0	0	0	0	0
Typical Aura With Migraine	0	6	0	5	3	16	0
Typical Aura without Migraine	0	0	0	0	0	0	8

The analysis of the confusion matrix reveals a varied performance profile. The model demonstrated perfect or near-perfect classification for several subtypes within the test set, including 'Familial Hemiplegic Migraine,' 'Migraine Without Aura,' 'Other,' and 'Typical Aura without Migraine,' where all instances were correctly identified.

However, the model exhibited significant challenges in distinguishing other subtypes. The 'Basilar-Type Aura' class had several instances misclassified as 'Familial Hemiplegic Migraine' and 'Other.' The most pronounced area of confusion was observed for the 'Typical Aura With Migraine' class. A substantial number of instances belonging to this class were incorrectly predicted as 'Familial Hemiplegic Migraine' (6 instances), 'Other' (5 instances), and 'Sporadic Hemiplegic Migraine' (3 instances). Furthermore, the model failed to correctly identify any instances of 'Sporadic Hemiplegic Migraine' in the test set, suggesting either a high degree of symptomatic overlap with other classes or insufficient representation of this subtype in the training data.

3.3. Interpretation of Findings

This study demonstrates that a Support Vector Machine with a simple linear kernel can classify multiple migraine subtypes from clinical questionnaire data with a notable accuracy of 82.65% [11, 15]. This result provides foundational validation for the use of established ML algorithms as a potential decision-support tool in headache diagnostics. The model's ability to correctly classify over four-fifths of the cases using only patient-reported symptoms is encouraging, as this type of data is the most ubiquitously collected in routine clinical practice [4].

The granular analysis provided by the confusion matrix, however, offers a more nuanced interpretation [1]. The model's high performance on distinct classes like 'Migraine Without Aura' and 'Familial Hemiplegic Migraine' suggests that these conditions present with sufficiently unique feature sets that are linearly separable. Conversely, the significant confusion surrounding the 'Typical Aura With Migraine' class indicates substantial overlap in its clinical presentation with other subtypes [3]. This difficulty suggests that the linear decision boundary imposed by the chosen kernel may be insufficient to capture the subtle, complex relationships between symptoms that differentiate these closely related conditions. The complete failure to classify 'Sporadic Hemiplegic Migraine' likely points to issues of class imbalance within the dataset, a common challenge in medical data where rare conditions are underrepresented [17, 13].

3.4 Comparison with Existing Literature

To contextualize the 82.65% accuracy achieved in this study, it is essential to compare it with findings from the broader literature on ML applications in headache classification. As shown in Table 1, the performance of ML models can vary significantly based on the algorithm used, the type of input data, and the methodological rigor of the study [6].

The accuracy obtained here is a solid baseline but is surpassed by studies employing more complex algorithms. For instance, research using ensemble methods like Random Forest and Gradient Boosting has reported accuracies exceeding 95% [1, 10]. Baccouch and Bahar (2025) achieved 97.12% accuracy using a Multi-Layer Perceptron (MLP), a type of neural network [16]. Even within the SVM framework, studies have shown that performance can be substantially improved. Garcia-Chimeno et al. (2017) demonstrated that by incorporating a systematic feature selection process, the accuracy of an SVM classifier could be boosted from 90% to 95% [12]. This comparative landscape underscores that while the linear SVM is a viable starting point, significant opportunities exist for enhancing predictive performance through more advanced modeling techniques [18, 6].

3.5 Limitations of the Study

This study has several limitations that must be acknowledged. First, the model's simplicity is a primary constraint. The choice of a linear kernel assumes that the complex relationships between migraine symptoms can be separated by a simple hyperplane. The pathophysiology of migraine is inherently non-linear, and employing non-linear kernels, such as the Radial Basis Function (RBF) or polynomial kernels, could potentially capture these intricate patterns more effectively and improve classification accuracy [11, 6].

Second, the study did not involve systematic hyperparameter tuning. The SVM model was trained with default parameters. Performance could likely be improved by conducting a rigorous search for optimal hyperparameters (e.g., the *cost* and *gamma* parameters) using techniques like grid search combined with cross-validation [17, 16].

Third, there are limitations inherent to the data source. The analysis relied on a single, public Kaggle dataset, which limits the generalizability of the findings to broader and more diverse patient populations. The dataset consists exclusively of self-reported clinical features, lacking objective biomarkers from neuroimaging (e.g., fMRI, EEG), genetic analysis, or wearable sensor data [7, 10, 14, 19]. The integration of such multimodal data has been shown to create more robust and biologically grounded classification models [9, 12].

Finally, the methodological approach could be strengthened. The use of a simple 75/25 train-test split is susceptible to sampling bias. More robust validation techniques, such as k-fold cross-validation, are standard practice and provide a more reliable estimate of a model's true performance. Additionally, the potential impact of class imbalance, hinted at by the poor performance on the 'Sporadic Hemiplegic Migraine' class, was not explicitly addressed with statistical techniques like stratified sampling or synthetic data oversampling (e.g., SMOTE) [17].

4. Conclusion

This study investigated the application of a Support Vector Machine with a linear kernel for the multi-class classification of migraine subtypes using patient-reported clinical data. The model achieved a promising overall accuracy of 82.65%, affirming the potential of machine learning algorithms to serve as effective decision-support tools in headache medicine. The analysis highlighted the model's differential performance across subtypes, successfully identifying some while struggling to distinguish those with overlapping symptom profiles.

In conclusion, the SVM demonstrates clear viability for this diagnostic task, establishing a solid performance baseline on accessible clinical data. However, this result also underscores the need for further research to unlock the full potential of machine learning in this domain. The path toward a clinically integrated, reliable diagnostic aid will require the exploration of more advanced algorithms, robust validation on larger and more diverse datasets, the incorporation of multimodal data streams, and a commitment to developing models that are not only accurate but also interpretable to the clinicians who will use them. Building upon the foundation of this study, several avenues for future research can be proposed. A primary direction is the exploration of more sophisticated algorithms. Evaluating advanced ensemble models like Gradient Boosting, and XGBoost, and deep learning architecture is a logical next step in similar classification tasks. This focus on building models that are not only accurate but also transparent and reliable is essential for their successful translation into real-world clinical practice.

References

[1] Dhiyaussalam, A., Wibowo, F. A., Nugroho, F. A., & Sarwoko, E. A. (2020). Classification of headache disorder using random forest algorithm. *Proc. 2020 4th Int. Conf. Informatics Comput. Sci. (ICICoS)*, 1-6. https://doi.org/10.1109/ICICoS51170.2020.9299105.

[2] Das, S., & Hridy, R. I. (2025). Machine learning-based migraine prediction: Analyzing key features and cause-effect relationships for improved diagnosis and management. *Int. J. Comput. Appl.*, *187*(11), 1-10. https://doi.org/10.5120/ijca2025925039.

[3] Vandenbussche, N., Paemeleire, K., Van der Veen, S. S. S., & De Pauw, G. (2021). Natural language processing to classify migraine versus cluster headache. *Neurology*, *96*(17), e2201-e2212. https://doi.org/10.1212/WNL.0000000000011818.

[4] Kwon, J. Y., et al. (2020). Machine learning-based automated classification of headache disorders using patient-reported questionnaires. *Sci. Rep.*, *10*(1), 1190. https://doi.org/10.1038/s41598-020-58079-4.

[5] Siddiquee, M. M. R., et al. (2023). Migraine and post-traumatic headache classification using an explainable 3D-ResNet-18 architecture. *Sci. Rep.*, *13*(1), 19169. https://doi.org/10.1038/s41598-023-46386-3.

[6] Gagnani, L., Barot, M., Chauhan, P., & Shah, M. (2025). Migraine classification using deep learning and machine learning techniques: A review. *KSV E-J. Eng. Manag. Sci. Humanit.*, *2*(1), 1-12.

[7] Li, G., Yang, H., He, L., & Zeng, G. (2025). Interpretable artificial intelligence analysis of functional magnetic resonance imaging for migraine classification: Quantitative study. *JMIR Med. Inform.*, *13*, e72155. https://doi.org/10.2196/72155.

[8] Fang, S., et al. (2025). Harnessing artificial intelligence for brain disease: advances in diagnosis, drug discovery, and closed-loop therapeutics. *Front. Neurol.*, *16*, 1615523. https://doi.org/10.3389/fneur.2025.1615523.

[9] Danelakis, A., et al. (2025). Diagnosing migraine from genome-wide genotype data: a machine learning analysis. *Brain*, *148*(5), 1-14. https://doi.org/10.1093/brain/awaf172.

[10] Göker, H. (2022). Automatic detection of migraine disease from EEG signals using bidirectional long-short term memory deep learning model. *Signal Image Video Process.*, *17*(4), 1-9. https://doi.org/10.1007/s11760-022-02333-w.

[11] Williams, F., Gunawan, P., Limuel, & Purnajaya, A. R. (2023). Implementasi support vector machine dan radial basis function untuk klasifikasi makanan vegetarian menggunakan data image. *JoDENS*, *3*(1), 5–8. https://doi.org/10.63643/jodens.v3i1.123.

[12] Purnajaya, A. R., & Hanggara, F. D. (2021). Perbandingan performa teknik sampling data untuk klasifikasi pasien terinfeksi covid-19 menggunakan rontgen dada. JAIC, 5(1), 37–42. https://doi.org/10.30871/jaic.v5i1.3010.

[13] Ferroni, P., Basili, S., Buccelletti, F. R. P., & Martelletti, P. (2020). Prediction of medication overuse in migraine patients: a machine learning approach. *Cephalalgia*, *40*(1), 55-63. https://doi.org/10.1177/0333102419864077.

[14] Tu, Y., et al. (2020). An fMRI-based neural marker for migraine without aura. *Neurology*, *94*(7), e741-e751. https://doi.org/10.1212/WNL.0000000000008962.

[15] Imtiaz, I., Afzal, H., Rehman, A. U., & Insany, G. P. (2025). Evaluating the role of machine learning in migraine detection and classification. *Eng. Proc.*, *107*(1), 122. https://doi.org/10.3390/engproc2025107122.

[16] Baccouch, C., & Bahar, C. (2025). Advanced machine learning approaches for accurate migraine prediction and classification. *Int. J. Adv. Comput. Sci. Appl.*, *16*(1), 1-12. https://doi.org/10.14569/IJACSA.2025.0160101.

[17] Petrušić, I., et al. (2024). Application of machine learning in migraine classification: a call for study design standardization and global collaboration. *J Headache Pain*, *25*(1), 119. https://doi.org/10.1186/s10194-024-01889-x.

[18] Torrente, A., et al. (2024). The clinical relevance of artificial intelligence in migraine. *Brain Sci.*, *14*(1), 85. https://doi.org/10.3390/brainsci14010085.

[19] Mursyid, F., et al. (2025). "DETEKSI ACNE VULGARIS DAN JENIS KULIT PADA CITRA WAJAH BERBASIS YOLOV7 DAN RESNET50", JUTEKOM, vol. 1, no. 3, pp. 120–130.