ELINVO (Electronics, Informatics, and Vocational Education), November 2023; vol 8(2):264-274
ISSN 2580-6424 (printed), ISSN 2477-2399 (online) DOI:https://doi.org/10.21831/elinvo.v8i2.76811

Performance Analysis of EMG Signal Classification Methods for Hand Gesture Recognition in Stroke Rehabilitation

Anggun Winursito[1]*, Fatchul Arifin[1], Muslikhin[1], Herjuna Artanto[1], Femilia Hardina Caryn[1]
[1]Department of Electronics and Informatics Engineering Education, Universitas Negeri Yogyakarta
*E-mail: anggunwinursito@uny.ac.id

ABSTRACT

This study evaluates the performance of different classification methods in classifying healthy individuals and stroke patients. The hand gesture variations of the subjects were also analyzed based on electromyography (EMG) signals. Several classification methods were tested in this analysis to find out which method had the most suitable performance. The results showed that Decision Tree and Naive Bayes classifiers achieved the highest performance in classifying EMG signals from healthy individuals and stroke patients, with both methods showing high accuracy, precision, recall, and F1 score. Specifically, Decision Tree excelled in overall accuracy and recall, while Naive Bayes showed superior precision. For hand gesture recognition, SVM, KNN, and Random Forest classifiers showed similarly high performance, achieving accuracy, precision, recall, and F1 score above 82%. Naive Bayes also performed well, especially in precision, while Decision Tree performed poorly compared to other methods. This insight can form the basis for the development of more effective and personalized rehabilitation systems for stroke patients, by utilizing reliable and accurate EMG signal classification.

Keywords*:* EMG signal, classifier, stroke rehabilitation, hand gesture recognition

INTRODUCTION

An electromyogram is a technique for evaluating and recording contraction and relaxation activity of arm/leg muscle tissue. The electrical activities produced by skeletal muscles constitute the fundamental Electromyography (EMG) signal. Because EMG signals arise from neuromuscular activity, they can be used to diagnose several medical conditions caused by neurological and muscle disorders [1]. This allows for the efficient examination of specific muscle activation and has consequently been applied in various medical investigations, including orthopedics, surgical procedures, nervous system research, and gait and postural assessments [2], [3], [4], [5]. In recent years, EMG signal analysis has gained significant attention, particularly in the context of differentiating between healthy individuals and patients with neurological disorders, such as stroke. Stroke is a disease with the third highest rate of death and disability in the world [6]. Stroke often results in impaired motor control and reduced ability to perform daily activities. Understanding the differences in EMG signals between healthy individuals and stroke patients is essential for developing effective rehabilitation strategies and assistive technologies. Analysis of differences in hand or foot muscle gestures based on EMG signals is also widely used for medical purposes or control systems.

Research on EMG signal analysis has been widely conducted with various applications [7], [8], [9], [10]. Cao et al [7] analyzed the detection system algorithm to classify hand gestures based on EMG signals. The experiment was conducted using the OPENBCI method to collect data. Four features were extracted from each part of the movement activity and generated a feature vector for classification. In the classification process, researchers conducted a comparison between the K-nearest-neighbors (KNN) and support vector machine (SVM) methods using a relatively small sample size. The comparison focused specifically on classification methods deemed appropriate for detecting hand gestures, utilizing the KNN and SVM techniques. The experimental findings revealed that the SVM algorithm achieved an average recognition rate that was 1.25% higher than that of KNN and completed the task 2.03 seconds faster than KNN. Another researcher, Khan [8] conducted a study to classify six types of eye movements from extraocular muscle signals using Fourier–Bessel series-based empirical wavelet transform (FBSE-EWT) with time and frequency domain (TAFD) features. A combined approach was employed to choose significant FBIMFs, followed by feature extraction based on statistical and signal complexity measures. Additionally, a metaheuristic optimization algorithm was applied to minimize the feature space dimensionality. The discriminative capability of the reduced feature set was confirmed using the Kruskal–Wallis statistical test. For classification, a multiclass support vector machine (MSVM) was utilized. The combination of selected methods produces good performance with an accuracy of approximately 99%. The development of an EMG signal classification algorithm has also been carried out by Azhiri et al [9]. Researchers compared the KNN and SVM methods with the Extreme Value Machine (EVM) method. The Autoregressive (AR) reflection coefficient was used to train a set of classification data. The results showed that the SVM method had better accuracy than the conventional method. However, the accuracy of the hand gesture detection that was tested still requires further development, especially in terms of recognition accuracy. Other researchers [10] conducted EMG signal analysis for the design and validation of an accurate automated diagnostic system to classify intramuscular electromyography (iEMG) signals into healthy, myopathy, or neuropathy categories to aid in the diagnosis of neuromuscular diseases. The MLPNN-BMMV classifier was tested with 250 iEMG signals from three categories.

Based on the research that has been done previously related to EMG signal analysis, it can be concluded that the development of an EMG signal detection system, especially for applications on hand movements of healthy individuals and stroke patients, still needs a lot of development. Some algorithms that have been developed do produce increased accuracy compared to conventional methods, but this is still not enough to be applied in the very crucial medical field. In addition, the use of EMG signals to help the rehabilitation process of stroke patients directly has not been widely used. Seeing this, the analysis of the EMG signal detection system algorithm that is focused on helping the rehabilitation process of stroke patients is very important. The results of the analysis of the performance of the detection algorithm can be used to monitor the development of patient hand movements during the rehabilitation process using physiotherapy methods.

METHODS

A. Dataset

The dataset used in the performance analysis of the EMG signal classification method comes from Kaggle. The dataset preparation process itself begins with the collection of raw EMG signals from electrodes placed on the relevant muscles. After that, the signals go through a pre-processing stage to reduce unwanted noise and artifacts. The dataset used for the analysis of EMG signals from healthy individuals and stroke patients is the MUSED-I Surface Electromyography (sEMG) Dataset. The dataset includes information from 11 healthy subjects and 2 stroke patients, with a duration of about two months. In this study, ten healthy subjects (without a history of upper limb pathology) and two upper limb stroke patients participated. The dataset used consists of raw surface electromyography (sEMG) signals collected for six types of movements (wrist flexion, wrist extension, hand closing, wrist radial deviation, wrist ulnar deviation, and resting position) using the Myo armband. All eight channels of the Myo armband were used for data acquisition [11]. Fig. 1 shows the characteristics of the EMG signal resulting from recording using an 8-channel sensor.

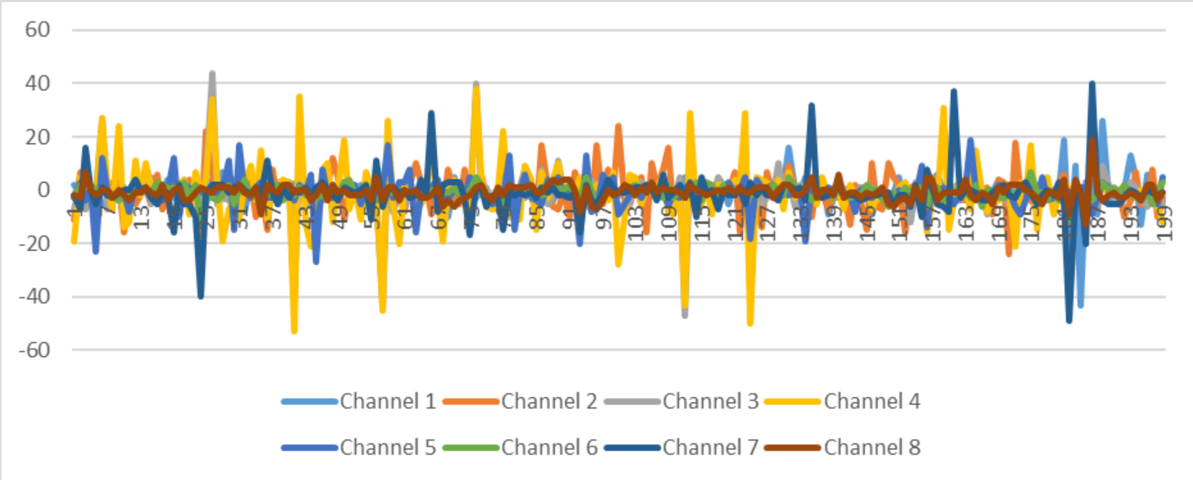

Figure 1. EMG signal with 8 channels

In addition, for hand gesture recognition classification, this study also used a dataset from the UCI Machine Learning Repository. The EMG signal dataset was obtained from 36 subjects performing a series of static hand movements. Each subject performed two series, each consisting of six basic movements. Each movement was performed for 3 seconds with a 3-second break between movements. The variations of hand movements in this dataset include hand at rest, hand clenched in a fist, wrist flexion, wrist extension, radial deviations, and ulnar deviations [12]. A complete description of the dataset used can be seen in Table 1 below.

Table 1. Details of the EMG signal dataset used

Dataset	Number of Subjects	Number of Classes
MUSED-I Surface Electromyography (sEMG)	11 healthy subjects 2 stroke patients	6 types of movements and 2 health conditions
UCI Machine Learning Repository	36 healthy subjects	6 types of movements

B. *Classification*

The EMG signal classification system basically works by analyzing EMG signals based on the characteristics of human hand muscle movements. The EMG signal classification process has several stages. The first step is initial processing which aims to standardize EMG signal data. Furthermore, important features of the EMG signal are extracted using feature aggregation techniques that include minimum value, maximum value, integral function, root mean square, absolute value, wavelength, and skewness. After the features are extracted, they are then used as input for the classification algorithm. The final step is testing by dividing the dataset into training data and testing data with a certain ratio. The stages of the EMG signal classification process are shown in Fig. 2 below.

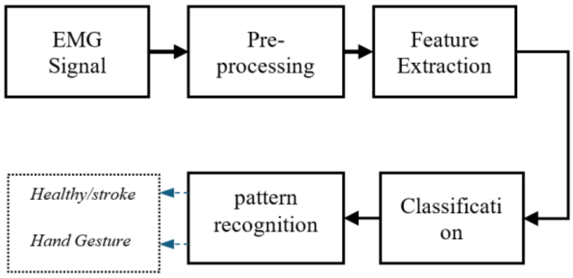

Figure 2. Stages of EMG Signal Classification System

Several classification algorithms tested include Support Vector Machine (SVM), K-Nearest Neighbour (KNN), Random Forest (RF), Decision Tree (DT), and Naive Bayes (NB). Each algorithm has a unique approach and characteristics in handling EMG signal feature data. SVM, for example, is known for its ability to handle classification problems with maximum margin, while KNN uses a distance-based approach to determine the class of the tested sample. RF and DT use a tree-based approach to build a classification model, while Naive Bayes uses probability to determine the class based on the distribution of features. The results of this classification process are used to recognize and distinguish hand movements performed, both in healthy individuals and in stroke patients. A comparative analysis of the performance of each classification method was carried out to determine the most effective and accurate algorithm in the context of EMG signal-based hand movement detection. Several parameters used in the comparison of the performance of the classification methods are the level of accuracy, precision, recall, and F1 Score.

C. *Support Vector Machine (SVM)*

SVM is one of the classification methods used in EMG signal classification. SVM is a machine learning method that works by finding the optimal hyperplane that can separate data into two or more classes. This hyperplane is positioned equidistant between the two classes, with the margin being the distance to the nearest data point from each class. The data points located exactly at this margin distance from the hyperplane are referred to as support vectors [13]. Equation (1) is the hyperplane found in the SVM algorithm, with $\vec{w}$ is the weighting vector, $\vec{x}$ is the input vector and b is the intercept or bias.

$$\vec{w} \cdot \vec{x} + \mathrm{b} = 0 \qquad (1)$$

In the context of hand gesture recognition, features extracted from EMG signals are used as input to an SVM, with the aim of classifying the type of hand gesture performed by the subject. Training data is used to build an SVM model, where the algorithm will determine the optimal hyperplane that separates the hand gesture classes. SVM can work with either a linear or non-linear kernel, depending on the complexity of the data. A linear kernel is used if the data can be separated by a straight line in the feature space, while a non-linear kernel such as RBF (Radial Basis Function) is used if the data is more complex and cannot be separated linearly. Once the SVM model is trained, its performance is evaluated using test data. The results of the SVM model are then compared with other classification methods to determine the effectiveness and accuracy of recognizing hand gestures based on EMG signals. An illustration of the formation of a hyperplane in SVM is shown in Fig. 3.

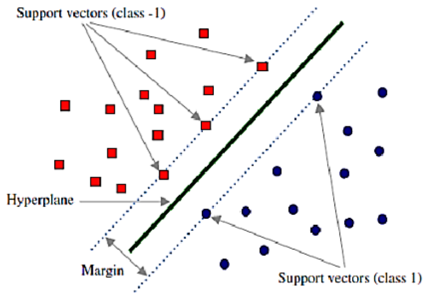

Figure 3. Hyperplane SVM illustration [13]

D. *K-Nearest Neighbors (KNN)*

KNN is a non-parametric, instance-based learning algorithm that classifies a sample by examining the K nearest neighbors within the feature space. These features are obtained from EMG signals that have gone through a feature extraction process, covering various aspects such as temporal, frequency, and time-frequency. The process of using KNN begins by determining the value of K, which is the number of nearest neighbors that will be considered to determine the class of the test sample. The selection of the K value is very important because it can affect the accuracy of the model. After determining the value of K, the KNN algorithm calculates the distance between the test sample and all data in the training set. Some commonly used distance metrics include Euclidean, Manhattan, and Minkowski distances. The Euclidean distance is the most commonly used, which measures the straight-line distance between two points in a multidimensional space. KNN then sorts all training data based on the nearest distance and selects the K nearest neighbors. The class of the test sample is determined by the majority class among the K nearest neighbors. One of the methods that is widely used for the process of calculating neighbor distances is using the Euclidean algorithm (2), where $a = a_1, a_2, \dots, a_n$, and $b = b_1, b_2, \ldots, bn$ represents the n attribute values of the two records [14].

$$euc = \sqrt{((a_1 - b1)^2 + \cdots + a_n - b_n)^2} \quad (2)$$

E. Random Forest (RF)

This method is known for its ability to handle varied and complex data by providing more stable and accurate results compared to a single decision tree model. Many researchers have conducted research related to the effectiveness and development of the Random Forest algorithm [15], [16], [17], [18]. Random forest (RF) regression is widely regarded as an effective tool for high-dimensional data analysis. Nevertheless, its efficacy may be compromised in sparse contexts due to the presence of weak predictors and the necessity for a preliminary dimensionality reduction (targeting) step [16]. Random forest is an ensemble classification method that combines the results of multiple decision trees by taking a vote from each tree's prediction [17]. The process of implementing Random Forest begins with the extraction of features from the EMG signals, which are then used as input to build a large number of decision trees, known as a "forest." Each tree in the forest is trained using a different subset of the training data through a bootstrap sampling technique, where a portion of the data is randomly selected with replacement. This creates diversity among the trees and reduces the risk of overfitting. When forming each tree, only a random subset of features is considered for splitting at each node. This approach helps reduce the correlation between trees and enhances the model's overall accuracy. At testing time, each tree in the forest provides a class prediction for the test sample. Random Forest then combines these predictions with a majority voting method to determine the final class of the sample. The mechanism of the random forest algorithm is shown in Fig. 4.

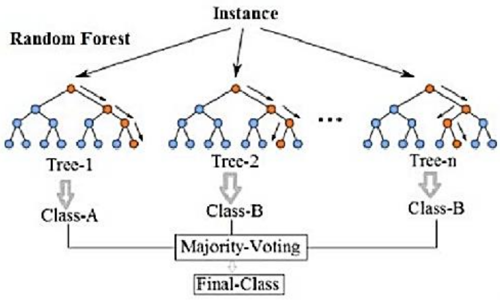

Figure 4. Random Forest Algorithm mechanism [19]

F. Decision Tree (DT)

Decision Tree is a learning model that uses a tree-like structure to make decisions based on input features, with each branch representing a choice between alternatives and each leaf representing a final decision or classification. Decision tree belongs to the simplest and most intuitive machine learning algorithms. Easy representation, low cost, and high quality make decision trees one of data science's most powerful and popular approaches [20]. Many studies have analysed the effectiveness and development of decision tree classification algorithms [18], [20], [21], [22], [23], [24]. The process of implementing a Decision Tree begins with the extraction of features from the EMG signal, which are then used to construct a decision tree. Each node in the tree represents a feature that is selected as the basis for the split, and each branch represents a possible value or range of values that the feature can take. This process of feature selection and node division is governed by an algorithm that aims to maximize the information gained or minimize impurities, such as by using metrics such as the Gini index or entropy (in the case of using the Information Gain method). Feature selection at each node is done by considering the features that are most effective in separating the data into different classes. The Decision Tree algorithm continues to divide the dataset at each node based on the values of the selected features until no more significant divisions can be made or until the node reaches the minimum amount of data that can be grouped as shown in Fig. 5.

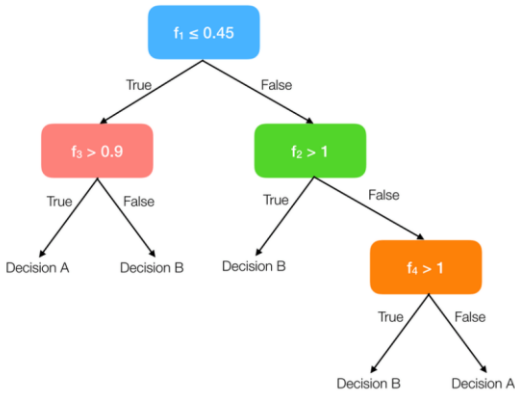

Figure 5. A simple decision tree classifier with 4 features [24]

G. Naive Bayes

Naïve Bayes model is widely used in classification because it is simple, efficient, and easy to understand [25]. Naive Bayes is a simple yet effective classification method, based on Bayes' theorem with the assumption of "naive" or independence between features. According to the Bayesian decision theorem, a Bayesian classifier predicts the test instance as the class with the highest membership probability. This means that learning a Bayesian classifier involves estimating the prior and posterior distributions of the training data [26]. In the Bayesian classification framework, equation (3) defines the posterior probability [27].

$$P(C|X) = \frac{P(X|C)*P(C)}{P(X)} \quad (3)$$

with the variable $P(C|X)$ is the probability of class C given feature X (posterior probability), $P(X|C)$ is the probability of feature X given class C (likelihood), $P(C)$ is the probability of class C (prior probability) and $P(X)$ is the probability of feature X (marginal probability). This method assumes that all features are independent of each other given the class. Thus equation (3) can be changed into a simpler equation (4).

$$P(C|X) \propto P(C)\Pi_{i=1}^{n} P(x_i|C) \quad (4)$$

with x_i is the i-th feature of the feature vector *X*.

RESULT AND DISCUSSION

The analysis process is carried out by testing the performance of several methods in classifying healthy people and stroke patients and recognizing subject hand gestures based on EMG signals. The analysis uses a comprehensive EMG signal dataset and various analysis methods, including machine learning and signal processing techniques, to assess the ability of each method to classify the appropriate class. The results of the analysis can be used to provide in-depth insights into the effectiveness of different approaches and provide guidance for the development of more effective rehabilitation systems based on EMG signals. This section includes a comparison of the results between the methods used, a discussion of the strengths and weaknesses of each method (accuracy, precision, recall, F1 Score), and practical implications in the context of stroke rehabilitation.

A. Classification of EMG signals of healthy and stroke patient

The characteristics of EMG signals produced by healthy individuals and stroke patients have significant differences. In healthy individuals, EMG signals tend to be more regular and consistent, reflecting better motor control and effective muscle coordination. Meanwhile, in stroke patients, EMG signals often show irregular patterns and larger amplitude fluctuations, caused by neurological damage that disrupts nerve signal transmission and muscle control. These differences reflect impaired motor function due to stroke, leading to less stable and uncoordinated movements. The differences in EMG signal characteristics between healthy individuals and stroke patients are shown in Fig. 6 and Fig. 7 below.

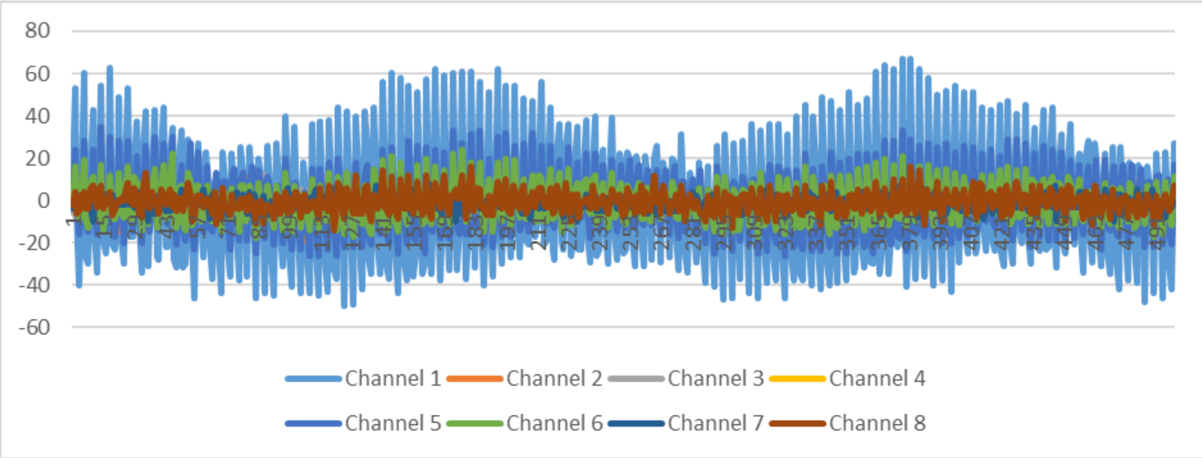

Figure 6. EMG signals of healthy people

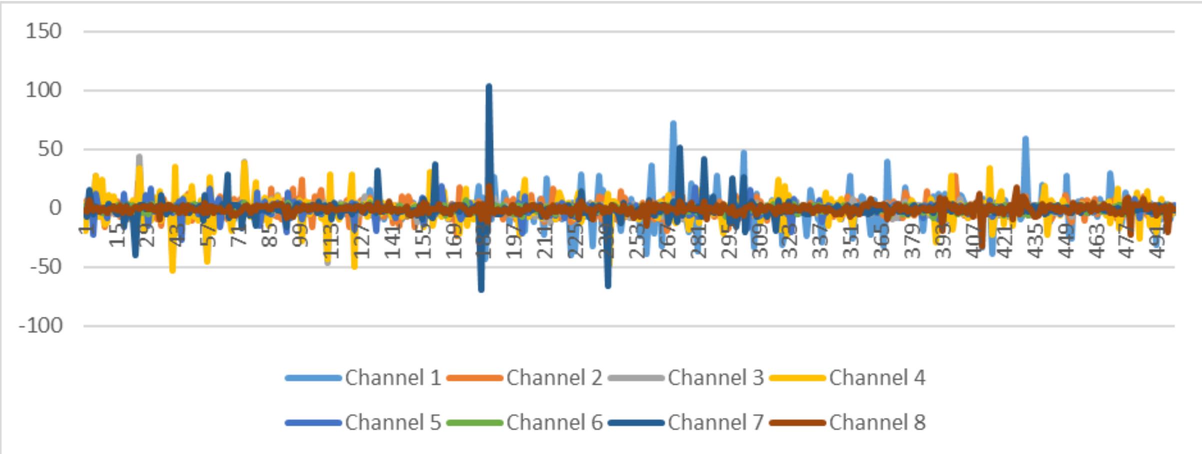

Figure 7. EMG signals of stroke patients

Based on the characteristics of the EMG signal in Fig. 6 and Fig. 7, the most dominant difference is related to the signal amplitude value between healthy individuals and stroke patients. In healthy individuals, the signal amplitude on channel 1 reaches an average value of 40, while in stroke patients it only has an average value of 10. Likewise, in other channels that have quite significant differences in amplitude values between healthy individuals and stroke patients. To find out the characteristics in more depth, the EMG signal is then subjected to a data extraction process with feature aggregation. The data from the feature extraction results are then inputted into several classification algorithms to carry out the classification process between EMG signals in healthy individuals and stroke patients. The performance results of several classification methods are shown in Table 2.

Table 2. Performance of classification methods on detecting EMG signals of healthy individuals and stroke patients

Model Classifier	Accuracy (%)	Precision (%)	Recall (%)	F1 Score (%)
SVM	76.47	79.58	76.47	76.97
KNN	76.47	85.88	76.47	76.80
Random Forest	82.35	82.16	82.35	81.93
Decision Tree	88.24	88.24	88.24	88.24
Naïve Bayes	88.24	90.05	88.24	87.55

The performance of the classification method was tested with several parameters, namely accuracy, precision, recall, and F1 score. Based on Table 2 above, it can be seen that all methods tested have quite good performance with accuracy, precision, recall, and F1 score values above 75%. To see a clearer comparison, it is shown in Fig. 8 below.

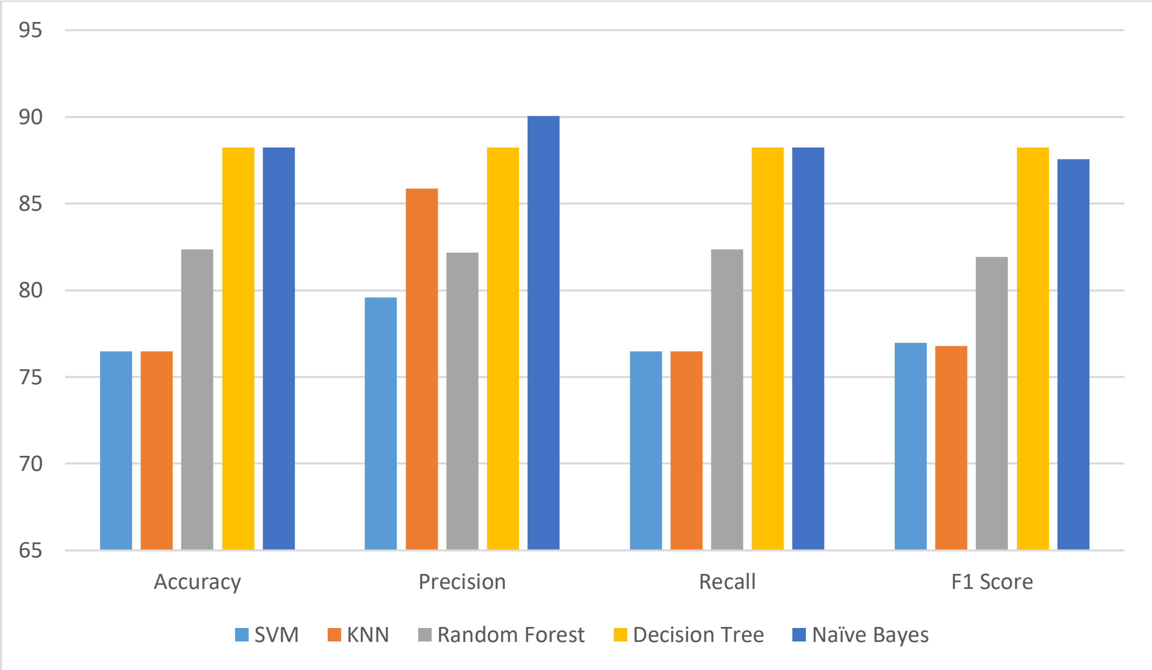

Figure 8. Comparison of classifier model performance based on several parameters

Based on the comparison of the performance of the classification methods shown in Fig. 8, it can be seen that the Decision tree and Naive Bayes methods have the best performance compared to other methods. The Decision tree and Naive Bayes methods have the highest percentage of accuracy, precision, recall, and F1 score in classifying EMG signals of healthy individuals and stroke patients, compared to other methods. The difference between the Decision tree and Naive Bayes methods lies only in the level of precision with a better value being Naive Bayes and in the F1 score with a better value in the Decision tree method.

B. *Hand Gesture Recognition*

The next analysis process is related to hand gesture recognition. Hand gesture recognition is needed in the development of a stroke patient rehabilitation monitoring system. The dataset used consists of 36 subjects while they performed a series of static hand gestures. The subject performs two series, each of which consists of six basic gestures (hand at rest, hand clenched in a fist, wrist flexion, wrist extension, radial deviations, and ulnar deviations). Each gesture was performed for 3 seconds with a pause of 3 seconds between gestures. The performance of several classification methods was tested to determine the most optimal method. The feature extraction process was carried out first using feature aggregation. Table 3 shows the results of hand gesture classification with several classification methods that were tested.

Table 3. Comparison of classifier performance on hand gesture detection

Model Classifier	Accuracy	Precision	Recall	F1 Score
SVM	82.81	82.73	82.81	82.47
KNN	82.81	83.58	82.81	83.06
Random Forest	82.81	83.97	82.81	83.16
Decision Tree	67.19	73.50	67.19	68.70
Naïve Bayes	79.69	84.14	79.69	80.72

Based on the test results data shown in Table 3, it can be seen that almost all classification methods have good performance. The decision tree method has poor performance with the lowest value on all tested parameters compared to other methods. A comparison of classification method performance related to hand gesture recognition is shown in Fig. 9 below.

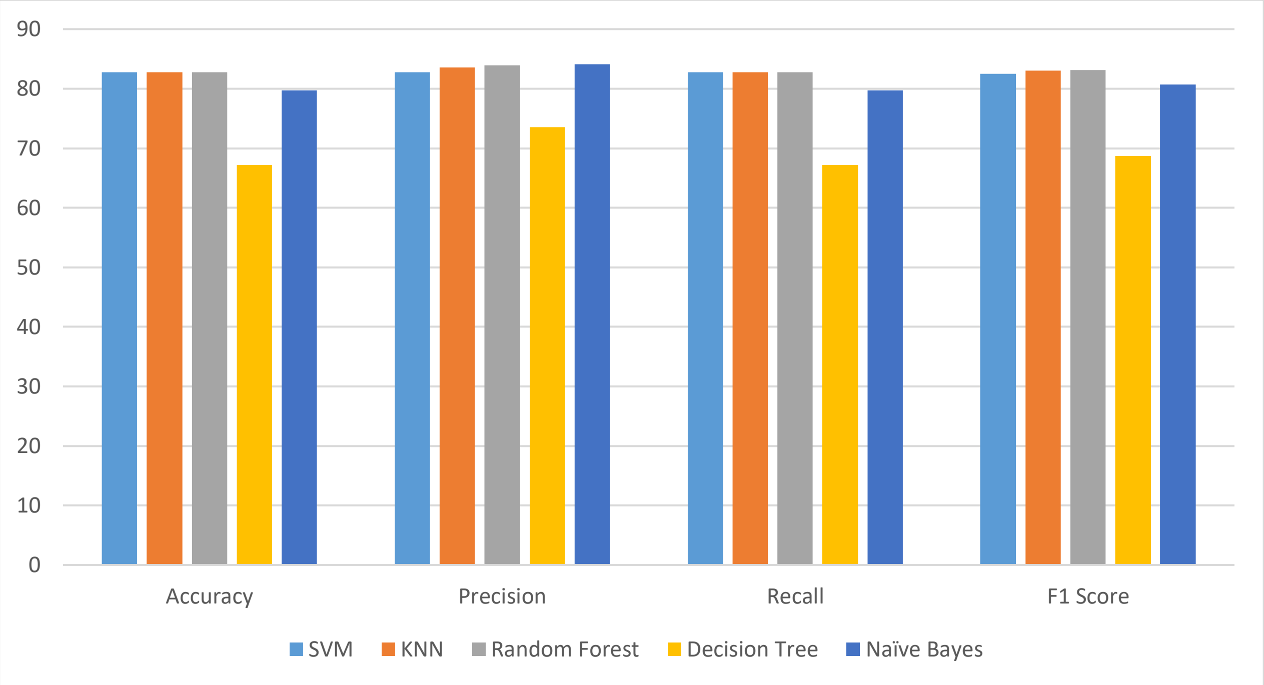

Figure 9. Comparison of classifier model performance

Based on the graph shown in Fig. 3, it can be seen that the SVM, KNN, Random Forest, and Naive Bayes methods have similar performance. At the accuracy level, the SVM, KNN, and Random forest algorithms have the same percentage value, which is 82.81%. At the precision level, the Naive Bayes algorithm has the best performance, with a percentage value of 84.14%. As for Recall, the SVM, KNN, and Random forest algorithms again have the same performance with a percentage value of 82.81%. In the f1 Score parameter, the Random Forest algorithm has the best performance with a percentage value of 83.16%.

CONCLUSION

Comprehensive datasets and the application of machine learning techniques allow for detailed evaluation of the performance of various classifiers. Based on the analysis results, there are differences in EMG signals from healthy individuals and stroke patients. The most dominant difference is related to the signal amplitude value between healthy individuals and stroke patients. In healthy individuals, the signal amplitude on all channels has a high value, while in stroke patients the average signal amplitude value is low. In the classification of EMG signals from healthy individuals and stroke patients, it was observed that the Decision Tree and Naive Bayes classifiers performed the best, with both achieving high accuracy, precision, recall, and F1 scores. Decision Tree had the highest overall accuracy and recall, while Naive Bayes showed superior precision. This indicates that both methods are very effective in distinguishing EMG signals from healthy individuals from stroke patients, making them suitable for application in stroke rehabilitation systems.

For hand gesture recognition, the results showed that the SVM, KNN, and Random Forest classifiers had similar high performance, with accuracy, precision, recall, and F1 scores all above 82%. Naive Bayes also performed well, especially in terms of precision. However, the Decision Tree classifier performed worse than the other methods. These findings suggest that SVM, KNN, and Random Forest are powerful methods for recognizing hand movements and can be effectively applied in a monitoring system for stroke rehabilitation. In the process of stroke patient rehabilitation, certain hand movements need to be monitored to observe the development of the training process over time. An appropriate classification algorithm will be able to assist the monitoring system in observing the optimal development of the patient's muscles. Overall, this study provides valuable insights into the strengths and weaknesses of various classification methods in the context of EMG signal analysis. These findings support the development of a more effective and personalized rehabilitation system based on reliable and accurate EMG signal classification.

REFERENCES

[1] V. Gohel and N. Mehendale, "Review on electromyography signal acquisition and processing," *Biophys. Rev.*, vol. 12, no. 6, pp. 1361–1367, Dec. 2020, doi: 10.1007/s12551-020-00770-w.

[2] C.-W. Wu *et al.*, "Optimal stimulation during monitored thyroid surgery: EMG response characteristics in a porcine model," *The Laryngoscope*, vol. 127, no. 4, pp. 998–1005, Apr. 2017, doi: 10.1002/lary.26141.

[3] D. C. Preston and B. E. Shapiro, *Electromyography and neuromuscular disorders e-book: clinical-electrophysiologic correlations (Expert Consult-Online)*, 4th ed. Amsterdam, The Netherland: Elsevier, 2020, doi: 10.1016/C2010-0-68780-3.

[4] G. I. Papagiannis *et al.*, "Methodology of surface electromyography in gait analysis: a review of the literature," *J. Med. Eng. Technol.*, vol. 43, no. 1, pp. 59–65, Jan. 2019, doi: 10.1080/03091902.2019.1609610.

[5] T. Hellig, V. Rick, A. Mertens, V. Nitsch, and C. Brandl, "Investigation of observational methods assessing workload of static working postures based on surface electromyography," *Work Read. Mass*, vol. 62, no. 2, pp. 185–195, 2019, doi: 10.3233/WOR-192854.

[6] G. Kwakkel *et al.*, "Motor rehabilitation after stroke: European Stroke Organisation (ESO) consensus-based definition and guiding framework," *Eur. Stroke J.*, vol. 8, no. 4, pp. 880–894, Dec. 2023, doi: 10.1177/23969873231191304.

[7] J. Cao, Z. Tian, and Z. Wang, "Hand Gestures Recognition Based on One-Channel Surface EMG Signal," *J. Softw. Eng. Appl.*, vol. 12, no. 09, pp. 383–392, 2019, doi: 10.4236/jsea.2019.129023.

[8] S. I. Khan and R. B. Pachori, "Automated Eye Movement Classification Based on EMG of EOM Signals Using FBSE-EWT Technique," *IEEE Trans. Hum.-Mach. Syst.*, vol. 53, no. 2, pp. 346–356, Apr. 2023, doi: 10.1109/THMS.2023.3238113.

[9] R. B. Azhiri, M. Esmaeili, and M. Nourani, "EMG-Based Feature Extraction and Classification for Prosthetic Hand Control," 2021, *arXiv*. doi: 10.48550/ARXIV.2107.00733.

[10] S. Jose *et al.*, "Robust Classification of Intramuscular EMG Signals to Aid the Diagnosis of Neuromuscular Disorders," *IEEE Open J. Eng. Med. Biol.*, vol. 1, pp. 235–242, 2020, doi: 10.1109/OJEMB.2020.3017130.

[11] M. M. Khan, M. F. Ammar, H. S. Qarni, A. Waris, and O. Gillani, "USED-I: "Multi-Gesture Surface Electromyography (sEMG) Dataset for Stroke Rehabilitation." IEEE Dataport, Aug. 05, 2023. doi: https://dx.doi.org/10.21227/04zq-yz45.

[12] I. K. N. Krilova, "EMG Data for Gestures." UCI Machine Learning Repository, 2018. doi: 10.24432/C5ZP5C.

[13] Dania Siregar, Faroh Ladayya, Naufal Zhafran Albaqi, and Bintang Mahesa Wardana, "Penerapan Metode Support Vector Machines (SVM) dan Metode Naïve Bayes Classifier (NBC) dalam Analisis Sentimen Publik terhadap Konsep Child-free di Media Sosial Twitter," *J. Stat. Dan Apl.*, vol. 7, no. 1, pp. 93–104, Jun. 2023, doi: 10.21009/JSA.07109.

[14] M. W. Prihatmono, S. Arni, J. N. Iin, and D. Moeis, "Application of the KNN Algorithm for Predicting Data Card Sales at PT. XL Axiata Makassar," *Conf. Ser.*, vol. 4, pp. 59–64, Jan. 2022, doi: 10.34306/conference series.v4i1.692.

[15] V. Y. Kulkarni and P. K. Sinha, "Effective Learning and Classification using Random Forest Algorithm," vol. 3, no. 11, 2014.

[16] D. Borup, B. J. Christensen, N. S. Mühlbach, and M. S. Nielsen, "Targeting predictors in random forest regression," *Int. J. Forecast.*, vol. 39, no. 2, pp. 841–868, Apr. 2023, doi: 10.1016/j.ijforecast.2022.02.010.

[17] S. Bharathidason and C. Jothi Venkataeswaran, "Improving Classification Accuracy based on Random Forest Model with Uncorrelated High Performing Trees," *Int. J. Comput. Appl.*, vol. 101, no. 13, pp. 26–30, Sep. 2014, doi: 10.5120/17749-8829.

[18] B. Talekar, "A Detailed Review on Decision Tree and Random Forest," *Biosci. Biotechnol. Res. Commun.*, vol. 13, no. 14, pp. 245–248, Dec. 2020, doi: 10.21786/bbrc/13.14/57.

[19] M. Azhari, A. Alaoui, Z. Achraoui, B. Ettaki, and J. Zerouaoui, "Adaptation of the random forest method: solving the problem of pulsar search," in *Proceedings of the 4th International*

Conference on Smart City Applications, Casablanca Morocco: ACM, Oct. 2019, pp. 1–6. doi: 10.1145/3368756.3369004.

[20] J. Biedrzycki and R. Burduk, "Weighted Scoring in Geometric Space for Decision Tree Ensemble," *IEEE Access*, vol. 8, pp. 82100–82107, 2020, doi: 10.1109/ACCESS.2020.2990721.

[21] L. Canete-Sifuentes, R. Monroy, and M. A. Medina-Perez, "A Review and Experimental Comparison of Multivariate Decision Trees," *IEEE Access*, vol. 9, pp. 110451–110479, 2021, doi: 10.1109/ACCESS.2021.3102239.

[22] J. Liang, Z. Qin, S. Xiao, L. Ou, and X. Lin, "Efficient and Secure Decision Tree Classification for Cloud-Assisted Online Diagnosis Services," *IEEE Trans. Dependable Secure Comput.*, vol. 18, no. 4, pp. 1632–1644, Jul. 2021, doi: 10.1109/TDSC.2019.2922958.

[23] Z. Liu, T. Wen, W. Sun, and Q. Zhang, "Semi-Supervised Self-Training Feature Weighted Clustering Decision Tree and Random Forest," *IEEE Access*, vol. 8, pp. 128337–128348, 2020, doi: 10.1109/ACCESS.2020.3008951.

[24] I. Mollas, G. Tsoumakas, and N. Bassiliades, "LionForests: Local Interpretation of Random Forests through Path Selection", in *Proceedings of the First International Workshop on New Foundations for Human-Centered AI (NeHuAI) co-located with 24th European Conference on Artificial Intelligence (ECAI*), 2020, pp. 17-24, doi: 10.48550/arXiv.1911.08780.

[25] S. Ruan, H. Li, C. Li, and K. Song, "Class-Specific Deep Feature Weighting for Naïve Bayes Text Classifiers," *IEEE Access*, vol. 8, pp. 20151–20159, 2020, doi: 10.1109/ACCESS.2020.2968984.

[26] T. Kim and J.-S. Lee, "Exponential Loss Minimization for Learning Weighted Naive Bayes Classifiers," *IEEE Access*, vol. 10, pp. 22724–22736, 2022, doi: 10.1109/ACCESS.2022.3155231.

[27] S. Wang, J. Ren, and R. Bai, "A Regularized Attribute Weighting Framework for Naive Bayes," *IEEE Access*, vol. 8, pp. 225639–225649, 2020, doi: 10.1109/ACCESS.2020.3044946.