Turkish Journal of Electrical Engineering and Computer Sciences

Volume 32 | Number 5 Article 5

9-23-2024

Finger Movement Recognition using Machine Learning Algorithms with Tree-Seed Algorithm

MUHAMMED SAMİ KARAKUL

AHMET GÖKÇEN

Turkish Journal of Electrical Engineering & Computer Sciences

http://journals.tubitak.gov.tr/elektrik/

Research Article

Turk J Elec Eng & Comp Sci
(2024) 32: 718 – 731
© TÜBİTAK
doi:10.55730/1300-0632.4097

Finger movement recognition using machine learning algorithms with tree-seed algorithm

Muhammed Sami KARAKUL*, **Ahmet GÖKÇEN**
Department of Computer Engineering, Faculty of Engineering and Natural Sciences,
Iskenderun Technical University, Hatay, Turkiye

Received: 19.12.2023 • **Accepted/Published Online:** 03.08.2024 • **Final Version:** 23.09.2024

Abstract: Electromyography (EMG) signals have been used to recognize various actions of hand movements, finger movements, and hand gestures. This paper aims to improve the classification accuracy of EMG signals while decreasing the number of features using the tree-seed algorithm. The dataset containing EMG signals utilized in this investigation is derived from a publicly accessible source. The rationale for selecting the tree-seed algorithm centers on its ability to enhance classification accuracy while minimizing the dimensionality of feature sets. The object function and tree-seed algorithm's nature avoids the results to have low accuracy with fewer features. The aim is not just to use a smaller number of features but also to achieve a higher accuracy rate. To ensure that selecting a smaller number of features does not decrease classification accuracy, the performance of all feature subsets was evaluated using the objective function. As a result, the number of selected features decreased, while the accuracy rate increased. The best accuracy improvement was observed, with the rate rising from 84.78% to 90.21% using the k-nearest neighbor (kNN) classifier with 50 out of 80 features. The maximum classification accuracy achieved was 99.75%, also using the kNN classifier. In this study, two different feature sets were compared using two different optimization algorithms in conjunction with four traditional machine learning algorithms to evaluate changes in classification accuracy. The classification accuracy and the improvements in accuracy, along with the number of selected features at the end of the iterations, have been reported.

Key words: Surface electromyography, tree-seed algorithm, feature selection

1. Introduction

In the literature, numerous studies have addressed the hand gesture recognition problem using various methods and datasets. These studies have demonstrated that muscle activity data can be recorded using noninvasive surface electromyogram (sEMG) electrodes placed on the skin [1].

In a previous study, the authors aimed to control a radio-controlled (RC) car using EMG sensors [2]. The EMG sensors were positioned on the user's forearm to record the data. They proposed a structure based on a decision tree and combined the k-nearest neighbor (kNN) and Bayes classifiers in two different combinations. One of these combinations achieved the best result, with an accuracy rate of 94.33%. In another study, Zhou et al. compared the classification accuracy of nine extracted features from a publicly available EMG dataset containing data from ten different subjects [3]. They aimed to classify finger motions, and they evaluated the classification accuracy of individual features as well as different combinations of features using the random forest

*Correspondence: msami.karakul@iste.edu.tr

(RF) algorithm. The best result achieved from a single extracted feature was 92.94%. Wahid et al. compared single-window voting with a novel scheme from a large publicly available dataset [4]. In their study, they varied window sizes between 50 ms and 500 ms and set overlapping sizes ranging from 0% to 80%. They applied six different machine learning algorithms and achieved the best accuracy of 80.70% using the random forest algorithm. Zhang et al. proposed a wearing-independent gesture recognition system [5]. While the subjects performed predefined gestures, EMG signals were collected from their forearms using the Thalmic Labs MYO armband. Subjects were instructed to perform gestures for a set period of time. Later, the MYO armband was rotated, and the subjects were asked to repeat the same movements. The signals recorded at different angles were classified using a random forest classifier, achieving an accuracy rate of 91.47%.

Wahid et al. used various machine learning algorithms to compare the classification results of different hand gestures [6]. They collected EMG signals from five healthy subjects as they performed three hand movements while wearing the MYO armband. They normalized the EMG features using the area under the curve root mean square (AUC-RMS) value and achieved a 98.75% accuracy rate with the support vector machine (SVM) algorithm. In Subasi and Qaiser's paper, the proposed method utilized various machine learning algorithms with tunable Q-Factor Wavelet Transform features derived from sEMG values [7]. They used an EMG dataset from the University of California Irvine (UCI) Machine Learning Repository, which includes hand grasp movements such as palmar grasp. They compared the performances of Bagging and Boosting ensemble classifiers using these features. They achieved a 100% accuracy rate from the SVM classifier using their method on the UCI dataset. Geng et al. introduced an instantaneous gesture recognition method using sEMG images [8]. In their study, they compared results with and without windowing sEMG signals. Without windowing, the accuracy rate from a single-frame sEMG image was 89.3%. When using 40 frames with a 1000 Hz sampling rate, they achieved an accuracy rate of 99.0%. In another study, Tepe and Erdim used the MYO armband with artificial neural networks (ANN) [9]. They utilized sEMG values from 8 EMG sensors along with gyroscopic signals acquired by the MYO armband. They recorded the EMG and gyroscopic signals while the subjects performed six finger gestures: thumb, index finger, middle finger, little finger, ring finger, and rest position. The subjects were asked to perform each gesture thirty times. They achieved a 94.40% accuracy rate by only using sEMG values, which increased to 96.30% when combining gyroscopic signals with sEMG values. Tavakoli et al. used two channels in their study to classify five hand movements [10]. They acquired the EMG signals from 7 subjects while they performed the five hand movements. The tests were conducted with both experienced users and beginners of the system. Additionally, the authors developed a system to classify movements as noise if they fell outside the predefined five classes. The accuracy rates for experienced users ranged between 95% and 100%.

There are many studies that used optimization and machine learning algorithms together on different EMG datasets. Huang et al. presented new ant colony optimization for feature selection [11]. The authors acquired EMG signals from ten different subjects while they performed eight different requested gestures for 80 s, repeating the process three times. They aimed to use ant colony optimization to achieve higher classification accuracy with the fewest number of features. The authors extracted two different feature sets from their recorded sEMG signals: time-domain features combined with autoregressive model coefficients (TDAR) and wavelet transform (WT) features. The average accuracy results were 95.45 $\pm$ 2.2% for TDAR features and 96.08 $\pm$ 3.3% for WT features. Too et al. proposed a new competitive binary grey wolf optimizer (CBGWO) [12]. They used their proposed optimizer on the NinaPro4 dataset with the kNN classifier and achieved a

92.69% accuracy rate, which is the highest average accuracy rate among the other optimization algorithms they compared. Subasi classified EMG signals using the proposed PSO-SVM method [13]. The EMG data were collected from 27 subjects, including healthy, myopathic, and neurogenic individuals. The results were compared with different machine learning algorithms, and the highest accuracy of 97.41% was achieved using the PSO-SVM method. Another study focused on feature selection using optimization algorithms and proposed a new method called binary particle swarm optimization differential evolution (BPSODE) [14]. The authors compared their method with different feature selection algorithms and their proposed method achieved the highest accuracy rate with 92.5%. Too et al. proposed a new personal best guide binary particle swarm optimization (PBPSO) [15]. They used their proposed optimizer on the NinaPro3 dataset with the kNN classifier and achieved the highest classification accuracy rate compared to other optimization algorithms they used and reported 85.20% accuracy rate. In another study, Sahu et al. used the kNN for classification, incorporating an improved feature selection approach [16]. They proposed an algorithm to solve global optimization problems. Their proposed method was called global best guided Gaussian ABC (GGABC). They used this method with kNN classifier and achieved 94.13% average accuracy and 97.06% maximum accuracy.

The dataset used in this study have been taken from the Kaggle website [17]. There are several studies on the same dataset. Jain and Garg used the genetic algorithm (GA) with artificial neural networks (ANN) to classify the EMG data and achieved a 95% accuracy rate. [18]. Farag et al. used the same dataset to classify EMG signals with convolution neural network (CNN) for bionic arm control and achieved 90.8% classification accuracy [19]. Jain and Garg used the GA to select the rows of potential electrical activities [20] on the same dataset. They trained and used the SVM for classification. The classification accuracy rate for normal EMG signal was 91.3% and for pain EMG signal, they achieved 92.4% classification accuracy rate. Bittibssi et al. also used the same dataset on recurrent neural networks (RNN) based on long-term short-term memory (LSTM) [21]. The architecture the authors introduced is a convolution LSTM and gated recurrent unit using RNN architecture which improved the prediction accuracy rate to 99.6%.

In this study, improving the finger movement classification using sEMG signals is aimed. To accomplish this, tree-seed algorithm has been used with traditional machine learning algorithms. In previous research, the tree-seed algorithm has been employed with various machine learning algorithms for detection or classification across different domains. For instance, Chen et al. proposed a feature selection method for network intrusion detection [22]. In their proposed method, they used the tree-seed algorithm in combination with kNN classifier to select the proper features. The methodological difference between their study and this study is that Chen et al. used randomly generated numbers for feature selection, while this study employs increasing constants in the objective function. Additionally, to the best of our knowledge, the proposed object function differs from those in their and previous studies. In this study, using the tree–seed algorithm for feature selection on a publicly available dataset that includes finger movements from different subjects has been aimed [17]. The main goal is to increase the classification accuracy and achieve the maximum classification rate while using the smallest number of features possible. This approach aims to classify data with reduced complexity and lower dimensionality. The tree-seed algorithm, combined with the proposed objective function, is used to discover optimal feature subsets, while the machine learning algorithms are responsible for evaluating the performance of these selected feature subsets.

Previous studies in the literature demonstrate that effective analysis of electromyography (EMG) signals is crucial for various applications. Despite progress, these analyses still require significant improvements. In the literature, many different datasets and many different classification methodologies have been used to achieve better results on the EMG classification. In this study, a freely accessible EMG dataset is used to evaluate the impact of an optimization algorithm, which has not been previously applied to this dataset, on changes in accuracy results using a proposed objective function.

Main contributions of this study are as follows:

- It achieved the highest classification accuracy on the used dataset compared to previous studies.
- It demonstrated that the tree-seed algorithm can be used to reduce the number of features while improving sEMG classification accuracy.

2. Methods

2.1. Tree-seed algorithm

Tree-seed algorithm (TSA) is a population-based algorithm designed to solve continuous optimization problems, as proposed by Kiran [23]. It mimics the behavior of trees and their seeds in nature, assuming that trees search for the optimal solution within the search space for optimization problems. The locations of trees and their seeds represent possible solutions to the problem [22]. The search for solutions occurs around the trees, making the initial positions of the trees crucial. The dimensions of the trees and seeds correspond to the dimensions of the optimization problem [24]. Initially, trees are created using Equation (1).

$$Tree_{i,j} = Low_j + rnd_{i,j}\, X\, (High_j - Low_j), \tag{1}$$

where $Tree_{i,j}$ is the jth dimension of ith tree, and $rnd_{i,j}$ is NxD matrix filled with uniformly random numbers in the range of [0,1] [25]. Low_j and $High_j$ are lower and higher bounds of the jth dimension.

In TSA, the search process involves exploration or exploitation. Exploration is used to avoid the local minimum by ensuring that all areas of the search space are explored equally [26]. When tree and seed locations are evaluated, the best seed becomes a new tree, and the best tree represents the optimal solution to the problem [23]. Exploitation involves analyzing the search space to find the optimal solution [26].

Exploration and exploitation in TSA are controlled by a parameter called search tendency (ST), which ranges from [0,1]. ST determines which equation will be used to generate seed locations. If ST is less than a randomly generated number, Equation (2) will be used to generate the seed location; otherwise, Equation (3) will be used [27]. Seeds are generated in every iteration for each tree.

$$Seed_{i,j} = Tree_{i,j} + (Best_j - Tree_{r,j})x(rnd - 0.5) * 2. \tag{2}$$

$$Seed_{i,j} = Tree_{i,j} + (Tree_{i,j} - Tree_{r,j})x(rnd - 0.5) * 2. \tag{3}$$

where i is the index of the seeds and trees, and j is the dimension index. $Seed_{i,j}$ is ith seeds' jth dimension that will be generated with ith tree. $Best_j$ is jth dimension of current best tree location, rnd is a randomly generated number between [-1,1], and r is an index of $Tree_{r,j}$ which is jth dimension of a randomly selected tree where r is not equal to i.

In both equations, seed locations are generated based on the tree locations. Equation (2) utilizes the current best tree location and a randomly generated index in addition to the seed's tree location, enhancing the intensification of the algorithm [23]. In contrast, Equation (3) relies on the seed's tree location and a randomly generated index, applied to both the seed's and the tree's locations.

2.2. Dataset

The dataset used in this study is publicly available [17][1]. It includes signal values of 7 different finger movements from 10 different subjects. However, in this study, only 6 finger movements were utilized to compare the classification accuracy results with those reported in the dataset's original study. The finger movements used in this study are index finger, middle finger, ring finger, little finger, thumb, and rest motion. The 7th movement was not included in the feature set used in this study because it is a victory gesture, which involves multiple finger movements simultaneously. Ten subjects were asked to perform the specified gestures for 20-30 trials, with a 1-s pause between each trial, while wearing the MYO armband [17].

2.3. Feature extraction

In this study, features were extracted both with and without using the overlapping method. The overlapping method involves sliding a window before the previous window ends. For this study, 1000 ms windows were used with a 50 ms sliding increment. The initial window covers the 0–1000 ms range. In the second iteration of feature extraction, the window slides by 50 ms, covering the 50–1050 ms range. Using both overlapping and nonoverlapping windows provides different results, allowing for a comparison of accuracy changes. The overlapping windows are illustrated in Figure. Signals were recorded within a 1000 ms window range with a 50 ms sliding increment. This method enables more detailed feature extraction from the dataset, as features are obtained from a relatively wide window range. Additionally, the 50 ms window movements allow for a more detailed perspective on feature discovery. The results section discusses the differences in classification accuracy between the features extracted using overlapping and nonoverlapping windows.

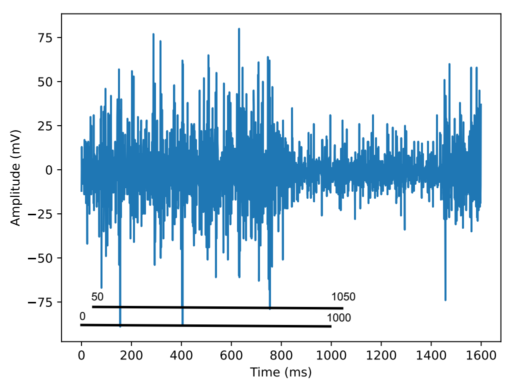

Figure: 1000 ms windows with 50 ms sliding and 950 ms overlapping drawing of the EMG signal.

Features were extracted from the EMG signals received on each sensor. The extracted features are root mean square (RMS), slope sign change (SSC), waveform length (WL), mean absolute value (MAV), average amplitude change (AAC), maximum fractal length (MFL), average power (AP), zero crossing count (ZC), variance (VAR), and standard deviation (STD). After extracting features, the total feature number becomes 80 (8 sensors x 10 features). The formulas and the used notation explanation of the extracted features have been given in the related subsections.

[1]Kaggle(2018). Electro-Myography-EMG-Dataset [online]. Website https://www.kaggle.com/datasets/nccvector/electromyography-emg-dataset [accessed 22.07.2023]

2.3.1. Root mean square

Root mean square is a time domain feature and is calculated using Equation (4).

$$RMS = \sqrt{\frac{1}{N}\sum_{n=1}^{N} X_n^2} \tag{4}$$

where x is the voltage value at ith sampling and N is the range of the window [28].

2.3.2. Slope sign change

Slope sign change feature indicates how many times the slope of the EMG signal changes sign. It is calculated using Equation (5) [29]

$$SSC = \sum_{n=2}^{N-1} [f[(X_n - X_{n+1})\ x\ (X_n - X_{n-1})]],$$
$$f(x) = \begin{cases} 1, & x \geq threshold \\ 0, & \text{otherwise} \end{cases} \tag{5}$$

2.3.3. Wavelength

Wavelength is the cumulative length of the EMG signal over time and is a measure of complexity of the EMG signal [29]. It is calculated using Equation (6).

$$WL = \sum_{n=1}^{N-1} |x_{n+1} - x_n| \tag{6}$$

2.3.4. Mean absolute value

Mean absolute value feature is an average of the absolute value of the EMG signal voltage value in a time and is calculated using Equation (7) [29].

$$MAV = \frac{1}{N}\sum_{n=1}^{N-1} |x_n| \tag{7}$$

2.3.5. Average amplitude change

Average amplitude change is the averaged value of the waveform length feature and is calculated using Equation (8) [29].

$$AAC = \frac{1}{N}\sum_{n=1}^{N-1} |x_{n+1} - x_n| \tag{8}$$

2.3.6. Maximum fractal length

Maximum fractal length is a measure about low level muscle activation and is calculated using Equation (9) [30].

$$MFL = \log_{10}(\sqrt{\sum_{n=1}^{N-1}(x_{n+1} - x_n)^2}) \tag{9}$$

2.3.7. Average power

Average power is a statistical feature used to calculate the energy distribution. It is calculated using Equation (10) [15].

$$AP = \frac{1}{N}\sum_{n=1}^{N} x_n^2 \tag{10}$$

2.3.8. Zero crossing

Zero crossing feature is used to achieve an approximate estimation about frequency domain. It is calculated using Equation (11) [31].

$$\begin{gathered} ZC = \sum_{n=1}^{N-1} sign(S), S = |x_n - x_{n+1}| \\ sign(S) = \begin{cases} 0, & s_n \; x \; s_{n+1} > 0 \; and \; S > 0 \\ 1, & \text{otherwise} \end{cases} \end{gathered} \tag{11}$$

2.3.9. Variance

Variance is calculated based on the power of the EMG signal. Since the mean value of the EMG signal is close to zero, it is excluded from the calculation [32]. The VAR feature is calculated using Equation (12) [33].

$$VAR = \frac{1}{N-1}\sum_{n=1}^{N} x_n^2 \tag{12}$$

2.3.10. Standard deviation

Standard deviation represents interferences, such as noise, and measures the spread of the signal values relative to the mean by comparing individual values to the mean [34]. It is calculated using Equation (13) [33]:

$$STD = \sqrt{\frac{1}{N-1}\sum_{n=1}^{N}(x_n - \overline{x})} \tag{13}$$

2.4. Feature selection using machine learning algorithms and TSA

In this study, SVM, kNN, RF, and DT classifiers were used to classify the sEMG signals. These classifiers were chosen to allow for comparisons with traditional classifiers and to compare the results with previous studies. The classifier parameters were determined after testing various parameters. A linear kernel with a one-vs-one approach was selected for SVM, the k parameter was chosen for kNN, and 50 trees were selected for both RF and DT classifiers.

Feature selection is a crucial step in data processing, as its proper application can enhance the dataset by eliminating irrelevant features or noise [35]. In the feature selection step of this study, the aim was to identify the optimal combination of features to achieve the highest classification accuracy. The tree-seed algorithm was chosen for its strong exploitation and exploration capabilities. TSA explores the search space from multiple points, which is advantageous for feature selection [23]. The motivation behind using TSA in this study was to leverage its exploration ability to test different feature subsets and improve the selection process. Since the aim is to improve classification accuracy, the fitness value of each tree and seed was calculated based on the classification accuracy of the selected features and the number of features chosen. The objective function in this study was calculated using the classification accuracy rate and the number of selected features, as explained in Section 2.5.

Each dimension of the tree and seed location is compared using a parameter called the comparison parameter. This parameter is initialized with a small value and increased with each iteration. If the dimension of the tree or seed location is less than the given parameter, it is accepted as '0'; otherwise, it is accepted as '1'. The dimensions accepted as '1' correspond to the selected feature indexes for training.

$$feature_j = \begin{cases} 0, & V_{i,j} \leq CP \\ 1, & V_{i,j} > CP \end{cases} \tag{14}$$

where $feature_j$ is feature of jth index, CP is the comparison parameter, V is the representation of the tree or seed, selection value of jth feature ($feature_j$) will be 0 or 1 depending on the comparison result of jth dimension of ith tree or seed with comparison parameter.

After training with the training dataset, the classifier prediction accuracy of the test dataset becomes the fitness value of the tree or seed. Increasing the comparison parameter for each iteration helps the algorithm avoid possible false results, such as always selecting a similar number of features and repeatedly selecting the same features.

For each loop with different comparison parameters, new trees are generated. Since the parameter increases for each parameter loop, only the dimensions with values higher than the parameter are selected as feature indexes. This process directs the new trees to select a smaller number of features. If the results do not improve with the new comparison parameter, it indicates that selecting a smaller number of features does not enhance the classification accuracy. Increasing the comparison parameter is crucial because it helps reduce the number of features while still retaining the maximum number of relevant ones. In this application of TSA, seeds are relocated to dimensions with either higher or lower values. If the seed dimensions have values lower than the comparison parameter and yield the best classification accuracy, this seed is considered optimal. The other seeds then attempt to relocate closer to this optimal seed. This process encourages seeds to have lower values, leading to the selection of fewer features.

This study aims to get the maximum accuracy rate with the minimum number of features. This means features that are irrelevant, causing noise or not effective as other features are eliminating with this method.

And as a result, this method chooses the most effective feature subset. The number of features and selected feature subsets change the classification accuracy. For every tree and seed, it is possible to choose a different subset of features. The features that are giving the better performance get chosen mostly and other trees and seeds try to relocate their dimension near to the best trees' better performing feature index. With that, better-performing feature subsets with fewer features get selected and higher classification accuracy rates are gained with these feature subsets.

2.5. The object function

To apply the tree-seed algorithm to the feature selection problem, a novel objective function has been proposed, to the best of our knowledge. The objective function in Equation (15) is designed to achieve high accuracy with a smaller number of features. Simple geometric equations have been used to optimize the balance between accuracy and the number of features.

In the feature selection problem, there are two dimensions: the number of features and the accuracy rate. In a two-dimensional space, if a line is drawn from the origin to a point at the intersection of the feature number axis (X) and the accuracy rate axis (Y), an angle is formed between the line and the X axis. If this angle is high, it indicates that the point has a lower X value and a higher Y value. However, when two points have proportional X and Y values, the drawn line will create the same angle. To address this issue, the length of the drawn line is also incorporated into the objective function.

$$f(a) = \tan^{-1}(\frac{Y}{X}) + \sqrt{X^2 + Y^2} \tag{15}$$

where X is the number of selected features, Y is the classification accuracy rate, $f(a)$ is the fitness value.

2.6. The implementation and the pseudocode of TSA

The implementation and the pseudocode of the TSA for feature selection with machine learning algorithms are provided in Algorithm 1 [22]. The fitness values of trees and seeds are calculated using classifiers.

Algorithm 1 Feature selection of machine learning algorithm

```
for cp = 1 to comparisonParameters do
    Initialize N random tree with D dimension Using Equation 1
    Evaluate the trees calculating the fitness using Equation 15 using features selected with Equation 14
    while The iteration is not less than the maximum number of wanted iterations do
        for i=1 to N do
            Calculate the number of seeds to generate for this tree between 15% and 25% of the population
            Select a random tree index (rnd) that is not equal to the current tree index (i)
            Generate a calculated number of seeds with D dimension
            for j=1 to D do
                if random < Search Tendency (ST) then
                    Update this seeds jth dimension using Equation 2
                else
                    Update this seeds jth dimension using Equation 3
                end if
            end for
            Calculate the fitness of seed using Equation 15
            Select the seed with higher fitness value
            if the seed has better fitness than tree then
                the seed replaces the tree
            end if
        end for
    end while
end for
```

3. Results and discussion

Table 1 illustrates the changes in accuracy and the number of selected features after applying optimization algorithms. PSO was used for comparison with TSA. As shown in Table 1, TSA is capable of achieving higher accuracy with fewer features than PSO. The results from feature sets extracted using overlapping windows exhibit high accuracy even in the initial iterations, which is why changes in accuracy and feature numbers for these sets are observed less. Consequently, the results from feature sets extracted without overlapping windows highlight the potential performance benefits of the method used.

In the results for the second feature set, both TSA and PSO showed increasing accuracy rates as the number of features decreased. However, as shown in Table 1, despite PSO concluding its iterations with fewer features, it did not achieve a higher accuracy rate than TSA. TSA outperformed PSO in both achieving higher accuracy rates and improving accuracy rates as fewer features were selected. The maximum accuracy rate achieved in this study is 99.75% with kNN by using 12.5% less features. As seen in Table 1, this method decreases the feature numbers while increasing the accuracy rates. It was observed that the feature set extracted without using overlapping windows yielded the highest accuracy change results.

The feature set extracted using the overlapping method has more rows because the 1000 ms window slides by 50 ms. In contrast, the feature set extracted without using the overlapping method includes features from distinct 1000 ms windows, with each new window starting where the previous one ended. This is why the second feature set has fewer rows compared to the first one, resulting in a lower starting accuracy rate. Consequently, the change in accuracy rate in the second feature set is more pronounced. The lower starting accuracy makes the impact of using different feature subsets on accuracy changes more evident.

Table 1. Test results of the finger movement classification accuracy change and the feature count on the last iteration of the optimization algorithms.

Feature extraction method	Optimization algorithm	Machine learning algorithm	Starting accuracy	Accuracy	Feature count
With overlap	TSA	SVM	88.2	92.93	73
		kNN	99.67	99.75	70
		DT	99.1	99.53	71
		RF	99.53	99.64	74
	PSO	SVM	90.21	92.3	70
		KNN	99.59	99.72	78
		DT	99.01	99.20	72
		RF	99.31	99.48	77
Without overlap	TSA	SVM	93.43	93.5	77
		KNN	84.78	90.21	50
		DT	90.76	93.47	44
		RF	91.84	93.47	69
	PSO	SVM	92.4	92.46	75
		KNN	87.5	90.76	48
		DT	86.41	89.13	42
		RF	88.04	90.21	57

Studies in the literature are summarized in Table 2. Among the first five studies that used the same dataset as this study, the highest accuracy achieved was 99.6% using LSTM [21]. In other studies that used different datasets and varied classification and optimization algorithms, the highest accuracy recorded was 97.41% with PSO and SVM [13].

Some studies have focused on enhancing optimization algorithms to improve classification accuracy and reduce the number of features [12, 14, 15]. Other studies have aimed to reduce the number of features while improving classification accuracy [11, 13, 17, 18, 20]. To the best of our knowledge, this study is the first to apply the tree-seed algorithm to improve EMG classification accuracy.

In this study, the TSA was applied with different classification algorithms to increase accuracy rates and reduce the number of features. The results demonstrate that it is possible to achieve higher accuracy while using fewer features with TSA. The maximum accuracy achieved was 99.75% with the kNN classifier, using 70 out of 80 features

Table 2. Comparison of classification accuracy results from various studies in the literature, highlighting the highest achieved accuracy result of this paper.

Authors	Used dataset	Optimization algorithm	Classification algorithms	Highest accuracy
Naseer et al. [17]	[17]	Genetic algorithm (GA)	LDA, KNN, SVM, DNN	97.4 (KNN)
Jain et al. [18]	[17]	Genetic algorithm (GA)	ANN	95
Jain et al. [20]	[17]	Genetic algorithm (GA)	SVM and KNN	92.4
Bittibssi et al. [21]	[17]	-	LSTM	99.6
Farag et al. [19]	[17]	-	CNN	90.8
Too et al. [12]	NinaPro4	CBGWO	KNN	92.69
Huang et al. [11]	Their collected dataset	Ant colony opt. (ACO)	BPNN	96.08
Too et al. [14]	NinaPro4	BPSODE	KNN	92.5
Too et al. [15]	NinaPro3	PBPSO	KNN	85.20
Subasi [13]	[36]	PSO	RBFN, KNN, SVM	97.41 (PSO-SVM)
Sahu et al. [16]	NinaPro3	GGABC	KNN	97.06
This study	[17]	TSA	SVM, KNN, RF, DT	99.75 (TSA-KNN)

4. Conclusion

In this paper, the tree-seed algorithm is used with machine learning algorithms for increasing the accuracy rate while decreasing the number of features.

The goal of this approach is to select the minimum number of feature subsets from the feature set while achieving maximum accuracy. Feature subsets were selected using a parameter called the comparison parameter. This parameter was initialized with a small value and progressively increased during iterations. Increasing comparison parameters enabled the method to discover various feature combinations with different numbers of features. After selecting the features, their performance was tested with different machine learning algorithms. Since the objective function aims to find the optimal balance between fewer features and higher accuracy, the trees and seeds were directed to locations that yields the optimal result. In this paper, the performance of the TSA was compared with that of the PSO algorithm. TSA demonstrated a greater change in accuracy rate compared to PSO, achieving higher accuracy results overall. Although PSO occasionally selected fewer features, its accuracy was lower than that of TSA in those instances. In conclusion, TSA proves to be more effective in feature selection for EMG datasets, offering higher accuracy with fewer selected features.

References

[1] Hu Y, Wong Y, Wei W, Du Y, Kankanhalli M et al. A novel attention-based hybrid CNN-RNN architecture for sEMG-based gesture recognition. PLOS ONE 2018; 13 (10):1-18. https://doi.org/10.1371/journal.pone.0206049

[2] Kim J, Mastnik S, André E. EMG-based hand gesture recognition for realtime biosignal interfacing. In: Proceedings of the 13th international conference on Intelligent user interfaces; Gran Canaria, Spain; 2008. pp. 30–39.

https://doi.org/10.1145/1378773.1378778

[3] Zhou T, Omisore OM, Du W, Wang L, Zhang Y. Adapting random forest classifier based on single and multiple features for surface electromyography signal recognition. In: 2019 12th International Congress on Image and Signal Processing, BioMedical Engineering and Informatics (CISP-BMEI); Suzhou, China; 2019. pp. 1-6. https://doi.org/10.1109/CISP-BMEI48845.2019.8965719

[4] Wahid MF, Tafreshi R, Langari R. A Multi-window majority voting strategy to improve hand gesture recognition accuracies using electromyography signal. IEEE Transactions on Neural Systems and Rehabilitation Engineering 2020; 28 (2): 427-436. https://doi.org/10.1109/TNSRE.2019.2961706

[5] Zhang Y, Chen Y, Yu H, Yang X, Lu W et al. Wearing-independent hand gesture recognition method based on EMG armband. Personal and Ubiquitous Computing 2018; 22:511-524. https://doi.org/10.1007/s00779-018-1152-3

[6] Wahid MF, Tafreshi R, Sowaidi M, Langari R. An efficient approach to recognize hand gestures using machine-learning algorithms. In: 2018 IEEE 4th Middle East Conference on Biomedical Engineering (MECBME); Tunis, Tunisia; 2018. pp. 171-176. https://doi.org/10.1109/MECBME.2018.8402428

[7] Subasi A, Qaisar SM. Surface EMG signal classification using TQWT, Bagging and Boosting for hand movement recognition. Journal of Ambient Intelligence and Humanized Computing 2022; 13 (7): 3539–3554. https://doi.org/10.1007/s12652-020-01980-6

[8] Geng W, Du Y, Jin W, Wei W, Hu Y et al. Gesture recognition by instantaneous surface EMG images. Scientific Reports 2016; 6 (1):36571. https://doi.org/10.1038/srep36571

[9] Tepe C, Erdim M. Classification of surface electromyography and gyroscopic signals of finger gestures acquired by Myo armband using machine learning methods. Biomedical Signal Processing and Control 2022; 75: 103588. https://doi.org/10.1016/j.bspc.2022.103588

[10] Tavakoli M, Benussi C, Lopes PA, Osorio LB et al. Robust hand gesture recognition with a double channel surface EMG wearable armband and SVM classifier. Biomedical Signal Processing and Control 2018; 46: 121-130. https://doi.org/10.1016/j.bspc.2018.07.010

[11] Huang H, Xie HB, Guo JY, Chen HJ. Ant colony optimization-based feature selection method for surface electromyography signals classification. Computers in Biology and Medicine 2012; 42 (1): 30-38. https://doi.org/10.1016/j.compbiomed.2011.10.004

[12] Too J, Abdullah AR, Saad NM, Ali NM, Tee WA. New Competitive binary grey wolf optimizer to solve the feature selection problem in EMG signals classification. Computers 2018; 7 (4): 58. https://doi.org/10.3390/computers7040058

[13] Subasi A. Classification of EMG signals using PSO optimized SVM for diagnosis of neuromuscular disorders. Computers in Biology and Medicine 2013; 43 (5): 576-586. https://doi.org/10.1016/j.compbiomed.2013.01.020

[14] Too J, Abdullah AR, Saad NM. Hybrid binary particle swarm optimization differential evolution-based feature selection for EMG signals classification. Axioms 2019; 8 (3): 79. https://doi.org/10.3390/axioms8030079

[15] Too J, Abdullah AR, Saad NM, Tee W. EMG feature selection and classification using a Pbest-guide binary particle swarm optimization. Computation 2019; 7 (1):12. https://doi.org/10.3390/computation7010012

[16] Sahu P, Singh BK, Nirala N. An improved feature selection approach using global best guided Gaussian artificial bee colony for EMG classification. Biomedical Signal Processing and Control 2023; 80:104399. https://doi.org/10.1016/j.bspc.2022.104399

[17] Naseer N, Ali F, Ahmed S, Iftikhar S, Khan RA et al. EMG based control of individual fingers of robotic hand. In: 2018 International Conference on Sustainable Information Engineering and Technology (SIET); Malang, Indonesia; 2018. pp. 6-9. https://doi.org/10.1109/SIET.2018.8693177

[18] Jain R, Garg VK. EMG classification using nature-inspired computing and neural architecture. 2021 9th International Conference on Reliability, Infocom Technologies and Optimization (Trends and Future Directions) (ICRITO); Noida, India; 2021. pp. 1-5. https://doi.org/10.1109/ICRITO51393.2021.9596077

[19] Farag HO, Awad MI, Baoumy AM, Shawky AA et al. Electromyography signal classification using convolution neural network architecture for bionic arm high level control. 2021 16th International Conference on Computer Engineering and Systems (ICCES); Cairo, Egypt; 2021. pp. 1-6. https://doi.org/10.1109/ICCES54031.2021.9686171

[20] Jain R, Garg VK. EMG signal feature extraction, normalization and classification for pain and normal muscles using genetic algorithm and support vector machine. Revue d'Intelligence Artificielle 2020; 34 (5). https://doi.org/10.18280/ria.340517

[21] Bittibssi TM, Zekry AH, Genedy MA, Maged SA. sEMG pattern recognition based on recurrent neural network. Biomedical Signal Processing and Control 2021; 70; 103048. https://doi.org/10.1016/j.bspc.2021.103048

[22] Chen F, Ye Z, Wang C, Yan L, Wang R. A feature selection approach for network intrusion detection based on tree-seed algorithm and k-nearest neighbor. In: 2018 IEEE 4th International Symposium on Wireless Systems within the International Conferences on Intelligent Data Acquisition and Advanced Computing Systems (IDAACSSWS); Lviv, Ukraine; 2018. pp. 68-72. https://doi.org/10.1109/IDAACS-SWS.2018.8525522

[23] Kiran MS. TSA: tree-seed algorithm for continuous optimization. Expert Systems with Applications 2015; 42 (19): 6686-6698. https://doi.org/10.1016/j.eswa.2015.04.055

[24] Cinar AC. Training feed-forward multi-layer perceptron artificial neural networks with a tree-seed algorithm. Arabian Journal for Science and Engineering 2020; 45 (12): 10915–10938. https://doi.org/10.1007/s13369-020-04872-1

[25] Liu J, Hou Y, Zhou A. A multi-strategy improved tree-seed algorithm for numerical optimization and engineering optimization problems. Scientific Reports 2023; 13 (1): 10768. https://doi.org/10.1038/s41598-023-37958-5

[26] Gharehchopogh FS. Advances in tree seed algorithm: a comprehensive survey. Archives of Computational Methods in Engineering 2022; 29 (5): 3281-3304. https://doi.org/10.1007/s11831-021-09698-0

[27] Aslan M, Beskirli M, Kodaz H, Kıran MS. An improved tree seed algorithm for optimization problems. International Journal of Machine Learning and Computing 2018; 8 (1):20-25. https://doi.org/10.18178/ijmlc.2018.8.1.657

[28] Lalitharatne TD, Hayashi Y, Teramoto K, Kiguchi K. A study on effects of muscle fatigue on EMG-based control for human upper-limb power-assist. 2012 IEEE 6th International Conference on Information and Automation for Sustainability, Beijing, China; 2012. pp. 124-128. https://doi.org/10.1109/ICIAFS.2012.6419892

[29] Phinyomark A, Phukpattaranont P, Limsakul C. Feature reduction and selection for EMG signal classification. Expert Systems with Applications 2012; 39 (8): 7420-7431. https://doi.org/10.1016/j.eswa.2012.01.102

[30] Phinyomark A, Phukpattaranont P, Limsakul C. Fractal analysis features for weak and single-channel upper-limb EMG signals. Expert Systems with Applications 2012; 39 (12): 11156-11163. https://doi.org/10.1016/j.eswa.2012.03.039

[31] Li W, Shi P, Yu H. Gesture recognition using surface electromyography and deep learning for prostheses hand: state-of-the-art, challenges, and future 2021; 15: 621885. https://doi.org/10.3389/fnins.2021.621885

[32] Phinyomark A, Hirunviriya S, Limsakul C, Phukpattaranont P. Evaluation of EMG feature extraction for hand movement recognition based on euclidean distance and standard deviation. ECTI-CON2010: The 2010 ECTI International Conference on Electrical Engineering/Electronics, Computer, Telecommunications and Information Technology; Chiang Mai, Thailand; 2010. pp. 856-860.

[33] Altın C, Er O. Comparison of different time and frequency domain feature extraction methods on elbow gesture's EMG. European journal of interdisciplinary studies 2016; 2 (3): 25-34. https://doi.org/10.26417/ejis.v2i3.p35-44

[34] Daud WMBD, Yahya AB, Horng CS, Sulaima MF, Sudirman R. Features extraction of electromyography signals in time domain on biceps brachii muscle. International Journal of Modeling and Optimization 2013; 3 (6): 515.

[35] Kumar V, Minz S. Feature selection: a literature review. SmartCR 2014; 4 (3): 211-229.

[36] Subasi A, Yilmaz M, Ozcalik HR. Classification of EMG signals using wavelet neural network. Journal of Neuroscience Methods 2006; 156 (1-2):360-367. https://doi.org/10.1016/j.jneumeth.2006.03.004