Defining the Features of EMG Signals on the Forearm of the Hand Using SVM, RF, k-NN Classification Algorithms

Conference Paper · May 2020

DOI: 10.1109/ICTC49638.2020.9123287

Kudratjon Zohirov
Tashkent University of Information Technology

Defining the Features of EMG Signals on the Forearm of the Hand Using SVM, RF, k-NN Classification Algorithms

1st Adilbek Turgunov
Information technology
Karshi branch of the Tashkent University of Information Technologies named after Muhammad Al-Khwarizmi
Karshi, Uzbekistan
odilt@mail.ru

2nd Kudratjon Zohirov
Computer Systems
Tashkent University of Information Technologies named after Muhammad Al-Khwarizmi
Tashkent, Uzbekistan
qzohirov@gmail.com

3rd Alisher Ganiyev
Information technology
Karshi branch of the Tashkent University of Information Technologies named after Muhammad Al-Khwarizmi
Karshi, Uzbekistan
zaxirov91@mail.ru

4th Barno Sharopova
Information technology
Karshi branch of the Tashkent University of Information Technologies named after Muhammad Al-Khwarizmi
Karshi, Uzbekistan
lazizzeksen@gmail.com

***Abstract*—EMG (electromyogram) signal-based control systems that provide important information on muscle activity are now in development. In this article, through some of the features of the EMG signal, three hand movements are classified. Experiments show that by using with the combination of Simple Square Integral (SSI), Average Amplitude Change (ACC), Integrated EMG (IEMG), Waveform Length (WL), Root Mean Square (RMS), Mean Absolute Value (MAV), Zero Crossing (ZC), Wilson Amplitude (WAMP) and Log-Detect (log-D) of EMG signal, 99% result of the process of signal recognition used in the classifier algorithms in k-NN (K -nearest neighbor) and RF (Random Forest), 96% result in SVM (Support vector machine) is achieved.**

Keywords— rehabilitation, electromyograms, filtering, invasive, features, classification, knn, random forest, support vector machine.

I. INTRODUCTION

Today, people who have lost legs or arms (congenital or acquired organ failure) are the majority. They may have been born that way, or they may have lost their motion organs due to diabetes, various infections, trauma, cancer, and vascular complications. Research shows that in the United States in 2005, 1.6 million people were disabled by losing their legs or arms, and by 2050 it would be 3.6 million [1]. The use of prosthesis around the world is somewhat limited. According to the World Health Organization estimates, there are now 30 million needing prostheses on earth [2].

The earliest prosthetic legs and arms date back to Egypt, but many civil wars and world wars are just an advanced period in this field [3].

Congenital or acquired abnormalities of the legs and arms are associated with several diseases and situations. That is why the number of people with disabilities who have movement problems is increasing every year.

There are several types of prostheses: wood or iron prostheses, body-shaped prostheses, myoelectric prostheses, and biological transplant prostheses.

Nowadays myoelectric robot prostheses are developing rapidly. The main principle of these prostheses, namely myoelectric control systems (MCS), is the classification and transmission of EMG signals from the muscles. The classification accuracy of the MCS depends on many factors: the location and number of electrodes responsible for EMG signal collection, the correct selection of the EMG signal differentiation algorithm, and the distinguishing important features of the EMG signal. Proper selection of EMG signal criteria is a key criterion for classification accuracy. By extracting the signal from the EMG signal, the useful information from the initial signal is extracted and used to classify different behavioral pathologies.

There have been several scientific types of research on MCS. The muscle interface which helped car drivers has been developed [4]. LIBSVM (A Library for Support Vector Machines), LDA (Linear discriminant analysis) and Naïve Bayes algorithms were used to classify 14 different finger states.

Other scientists have classified eight different types of hand movements [5]. The ants' colony algorithm was used for classification. They used time and frequency markers of six EMG signals. The average classification accuracy was 98.08%.

Classification work of muscle by exposure loads in different weighs during four seconds was also performed [6]. The signal was analyzed by instantaneous Fourier modification (IFM), discrete wavelet modification (DWM), and wavelet packet modification (WPT).

In other research, different wavelet families were compared. Two types of radial neural networks and probable neural network algorithms were used. In this paper, four wavelet families (bior, coif, db, and sym) were tested to classify five hand movements. Accordingly, the Biorthogonal and Coiflets wavelet families yielded better results than the other families.

The real myoelectric prostheses have also been used to investigate and compare EMG signal recognition processes [7]. In this research, two types of forearm movements were most commonly implemented to identify the forearm movement and forearm rotation. The accuracy achieved was 84%.

In this paper, three different classification algorithms derived from the EMG signal were analyzed using k-NN (K -

nearest neighbor), RF (Random forest), and SVM (Support vector machine). The features selected were SSI, ACC, IEMG, WL, RMS, MAV, ZC, WAMP, log-Detect. We combine some features to get accurate and high results. They are SSI-ACC, SSI-ACC-IEMG, SSI-ACC-IEMG-WL, SSI-WL-RMS, SSI-ACC-IEMG-WL-RMS, SSI-ACC-IEMG-WL-RMS-MAV, SSI-ACC -IEMG-WL-RMS-MAV-ZC, SSI-ACC-IEMG-WL-RMS-MAV-ZC-WAMP, SSI-ACC-IEMG-WL-RMS-MAV-ZC-WAMP – log-Detect.

This article consists of five parts. The first part is about the experimental protocol, the second part is devoted to the separation of features. The classification algorithms used in this research are presented in the third part, In the last two parts the results of the experiment and discussion points, as well as relevant recommendations for future work are given.

II. Materials and Methods

A. Experimental Design

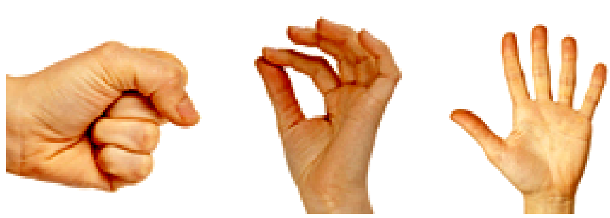
In this research, an EMG signal from 20 healthy individuals (males), aged 21 to 35 years is received.

Before starting the experiment, it is necessary to eliminate the factors that influence the results, that are, any stresses or muscle fatigue during the experiment. In addition, a good and appropriate environment is needed to obtain a clear and low noise EMG signal. Data were recorded within 10 days, taking into account muscle fatigue.

Before recording the signal, situations that could affect the results, that is, stress or fatigue during the experiment, were prevented. Besides, a good and appropriate environment has been created to get clear and low distortion signals. The data were gradually recorded within fifteen days. The portable and modern BITalino (Portugal, Plux - wireless biosignals Ins.) was used to record three movements on the forearm. The signal was recorded from the extensor carpi radialis muscle.

The EMG signal is influenced by physiological and anatomical factors. This in itself causes certain noise in the signal and causes a signal disturbance. The signal was therefore filtered through a bandpass filter. The signal data is recorded at a frequency of 1 kHz.

The disposable silver-silver / chloride (Ag/AgCl) electrodes were used to obtain the EMG signal from the skin surface. The electrodes were placed parallel to the muscle axis. Figure 1 shows the three-hand movements we used. The human arm area consists of several parts (Figure 2).



a) b) c)

Figure 1. Used hand movements. (a - hand punch, grip, b – touch finger, c - open hand)

Today, people in the world who need prostheses are people with disabilities that do not have 2% shoulder anterior, 3% shoulder, 16% arm, 1% elbow, 15% forearm, 2% wrist and 61% palm. This shows that there is a great demand for prostheses, mainly on the palms of the hand. The palm of the hand consists mainly of finger movements [8].

The movements that we are researching are the main functions of the palm, and these movements are considered to be the most important actions of affecting a hand movement to a subject [9].

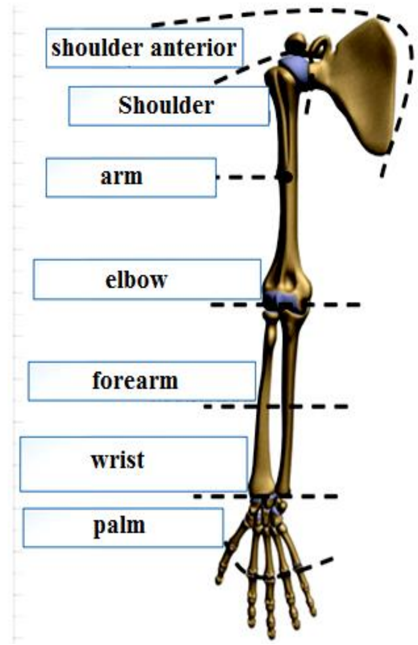


Figure 2. Split of the arm by movement functions

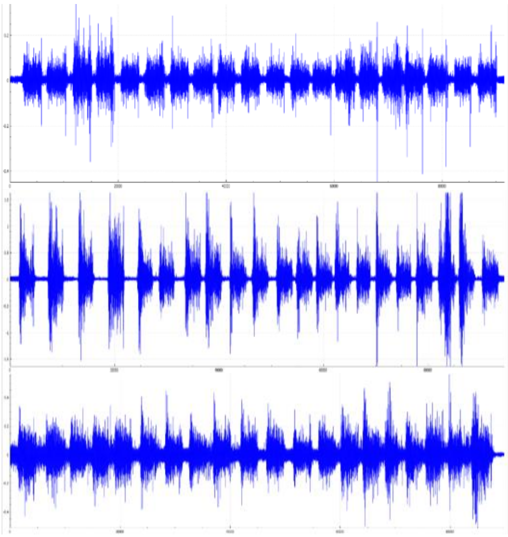
B. Experimental Protocol

We used three moves. Each movement was repeated 100 times and held for three seconds. In addition, muscles were given three to five seconds of rest after each movement. As shown in Figure 3, we can see that each movement is different in the amplitude and shape of the EMG signal. This is because muscles are affected by different forces.

III. Feature Extraction

In MCSs, some of the highlighted features show better EMG signals than other functions. It gives us the convenience of separating features, that is, reduces signal values, and provides us with the valueswe need, which reduces processing and classification time.

It is important to choose the optimal characteristics of the EMG signal. Some scientific researches show that some features of the EMG signal frequency domain do not perform well in EMG signal classification [10]. The timing characteristics of the EMG signal are very popular in signal recognition [11]. The timing features are easy to understand and precise as well.



a) – open hand

b) – hand punch

c) – touch finger

Figure 3. Initial view of the signal from three movements of the hand

In this paper, different features of the EMG signal and its analysis were analyzed through three classification algorithms. We used a combination of features to increase the accuracy, correctness, and reliability of the results.

Here are the features used in the article:

A. Feature parameters

These features include Simple Square Integral (SSI), Average Amplitude Change (ACC), integrated EMG (IEMG), wavelength (WL), root mean square (RMS), mean absolute value (MAV), Zero Crossings (ZC), Wilson amplitude (WAMP), the force of Log-Detect (log-Detect).

- Simple Square Integral (Simple Square Integral - SSI).

Square Integral (SSI) can be used as an EMG signal energy [7].

$$SSI = \sum_{i=1}^{N} |x_i|^2 \tag{1}$$

where xi is the ith sample and N is the number of samples in each segment.

- Average Amplitude Change (ACC)

It helps to determine the average of wavelength in a N-long window.

$$ACC = \frac{1}{N} \sum_{i=1}^{N-1} |x_{i+1} - x_i| \tag{2}$$

- Integrated EMG signal (IEMG)

It helps to determine the sum of the signal values along the EMG signal length.

$$IEMG = \sum_{i=1}^{N} |x_i| \tag{3}$$

- Waveform Length (WL)

The wavelength is intuitively the sum of the waveform along the segment. The calculation of WL can measure the amplitude, frequency, and duration of waveforms [7].

$$WL = \sum_{i=1}^{N} (x_i - x_{i-1}) \tag{4}$$

- Root Mean Square (RMS)

Mean square root - this feature characterizes signal amplitude through Gaussian distribution [7]. This method is very common when designing a feature. The computational speed and efficiency are very high.

$$RMS = \sqrt{\frac{1}{N} \sum_{i=0}^{N} x_i^{\,2}} \tag{5}$$

- Mean Absolute Value (MAV)

Mean Absolute Value is a feature used to detect muscle contraction and to recognize the signal. This is the average of the adjusted signal of the EMG signal [6, 7].

$$MAV = \frac{1}{N} \sum_{i=1}^{N} |x_i| \tag{6}$$

- Zero Crossings (ZC)

This setting determines the number of zero-pass signal values in the selected area. This simple parameter depends on the signal frequency. In order not to take account of the low noise resulting from interference, a threshold value is given to k.

$$ZC = \sum_{i=0}^{N-1} f(x_i * x_{i+1}), \tag{7}$$

f(x) = {1 if x >= k; 0 otherwise}

- Wilson amplitude (Wilson amplitude - WAMP)

This feature determines the amount of time the change in the EMG signal amplitude exceeds the input value. This parameter is an indicator of movement potential spread and, is, therefore, a key measure of muscle contraction.

$$WAMP = \sum_{n=1}^{N-1} f(|x_{i+1} - x_i|) \tag{8}$$

f(x) = {1 if x > ε; 0 otherwise}

where ε is the value of the incoming signal.

This feature is an indicator of the fi ring motor unit of action potentials (MUAP) and therefore an indicator of muscle contraction level.

- Log-Detect (log-Detect - logD)

This feature represents the force of muscle action. This is a non-linear feature.

$$\log D = e^{\frac{1}{N}\sum_{i=1}^{N}\log(|x_i|)} \quad (9)$$

The size of the EMG signal data we collect: three classified movements of twenty people, whose data comes from one channel, and each movement is repeated one hundred times, it is formed with a total of 6.000 sets of values (20 * 3 * 1 * 100 = 6000).

The data were divided into two sets, 85% for training and 15% for testing. For the classification step, three classifiers were implemented: SVM, RF, and k-NN. The coming section provides a brief introduction to the tested classifiers.

IV. Classification algorithms

An efficient means of classifying electromyography (EMG) signal patterns has been of interest to many researchers in the modern era. There are different types of classifiers that are effectively used for various EMG applications such as Artificial Neural Network (ANN), fuzzy classifier, Linear Discriminant Analysis (LDA), Self-Organizing Map (SOM) and Support Vector Machines (SVM).

A. *k-Nearest Neighbor (k-NN)*

The k-nearest neighbor (k-NN) classification method is the classical method used for EMG signal recognition [12]. The accuracy of this classifier can be deteriorated in noisy environments. The classifier is based on three basic mechanisms: 1- determining the distance between each of the initial signal values, 2- choosing the value of k-distance, and 3 - testing the value of k by multiple tests. In other words, the function of this classifier is to determine the probability of the best value of K, based on the accuracy of the test results, and to make the classification based on that.

B. *Support vector machine*

The SVM was originally designed for two-class classication. It aims to find the optimal hyperplane between two classes by mapping the sample space into a high-dimensional feature space, called the Hilbert space, to change the nonlinearly separable problem in the primary space to a linearly separable problem in the feature space.The classication problem in this paper is specically a nonlinearly separable problem for which the SVM has a great advantage. The SVM can be extended to multiclass classication [13]. The nonlinear transformation that transforms the input space into a high-dimensional space is realized by dening an appropriate inner product function.

C. *Random Forest*

Random forest is a tree algorithm that performs the prediction function using several trees throughout the forest. This classification method summarizes the selected results of each tree independently and determines the most reliable class data by summing and summarizing each result. [14]. This is one of the most successful classification methods available for many types of data.

V. Results and discussion

A. *Individual application of features*

Following the procedure of the proposed classification algorithms, the three movements were examined several times in succession. The results of the three classification algorithms were compared using the nine features we initially used. The WL (87%) had the highest score in the k-NN classifier (k = 3). The lowest accuracy was present in the ZC (33%) feature (Figure 4). It can be seen that the WL, RMS, and SSI characteristics of the EMG signal are the main parameters.

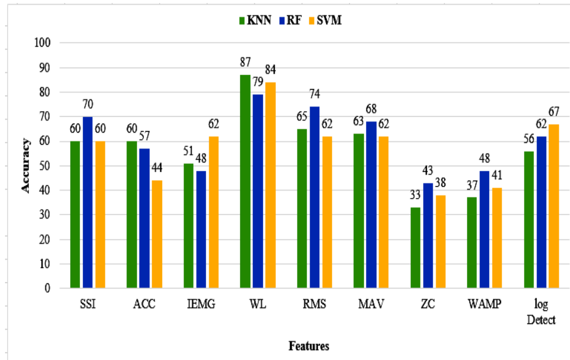


Figure 4. The degree of accuracy of the individual application of the features

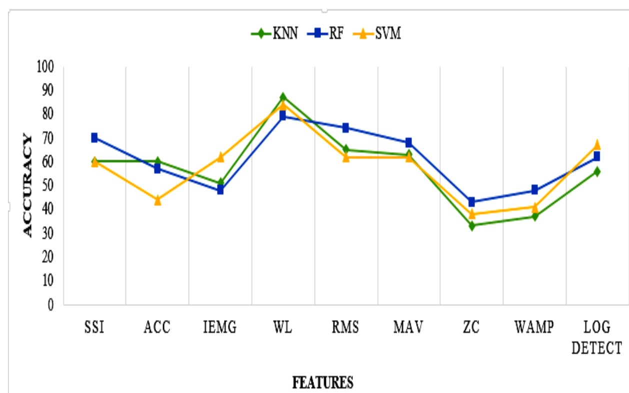


Figure 5. The degree of accuracy of the classification algorithms

The results show that the classification algorithms achieve almost the same results when applying individual features (Figure 5).

B. *Applying features in combination*

It is proved that the results of individual use of features are on average 40% -60%. Of course, this result is low. To obtain more accurate results, we compared the results by combining features that were used sequentially (Table 1).

In this case, it can see that the accuracy of the results obtained by a combination of two or more features is much higher than the initial (individual) results. This certainly improved the efficiency of EMG signal classification. The results improved dramatically since a combination of five features. The best results appeared in the combination of SSI, ACC, IEMG, WL, RMS, and MAV features.

TABLE I. RESULTS OBTAINED AS A COMBINATION OF PARAMETERS

Combination	k-NN (%)	RF(%)	SVM(%)
SSI+ACC	71	73	56
SSI+ACC+IEMG	86	87	70
SSI+ACC+IEMG+WL	94	97	89
+SSI+WL+RMS	97	96	91
SSI+ACC+IEMG+WL +RMS	95	97	92
SSI+ACC+IEMG+WL+ RMS+MAV	99	99	95
SSI+ACC+IEMG+WL +RMS+MAV+ZC	97	98	87
SSI+ACC+IEMG+WL+RMS+MAV+ZC+WAMP	98	95	89
SSI+ACC+IEMG+WL+RMS+MAV+ZC+WAMP +log-Detect	98	97	90

It is achieved an accuracy of 99% on the k -NN classifier, 99% in the RF and 95% in the SVM classifier.

k-NN classifier				
	1	2	3	Accuracy
1	21	0	0	100%
2	0	25	0	100%
3	0	1	23	97%
Average				99%

SVM classifier				
	1	2	3	Accuracy
1	19	0	1	95%
2	1	20	0	95%
3	0	1	23	96%
Average				95%

RF classifier				
	1	2	3	Accuracy
1	18	0	0	100%
2	0	20	0	100%
3	0	1	22	96%
Average				99%

Figure 6. The confusion matrix

The k-NN algorithm found all 21 and 25 tests conducted in the first and second classes respectively, and 23 out of 24 tests in the third class. The RF algorithm found all 18 and 20 tests conducted in the first and second classes respectively, and 22 out of 23 tests in the third class. The SVM algorithm found 19 out of 20 tests in the first class, 20 out of 21 tests in the second class, and 23 out of 24 tests in the third class (Figure 6).

In this work, certain algorithms were used to classify the motion and a special device for recording the signal. The main difference between the work and the others is that it is considered to achieve the high results by achieving portability, utilizing a minimal amount of muscle and hybridizing signal characteristics.

VI. CONCLUSION

In this article, the main palm movements using EMG signal features are analyzed and the processes of classifying these signals are given. The following results were achieved in the article:

• Signal recognition procedures were performed using the k-NN, SVM and RF classification algorithms by the EMG signal parameters. In this case, some parameters of the EMG signal, when applied individually, produced better results than other parameters. These parameters are 87% in wavelength (WL) k-NN algorithm, energy (SSI) 70% in RF algorithm, and average Root Mean Square (RMS) 74% in RF algorithm. The average accuracy obtained by the individual application of the parameters was 57% in k -NN, 61% in RF and 56% in SVM.

• The results of the classification process using a combination of parameters were much higher than the use of individual parameters. The best combination of parameters SSI, ACC, IEMG, WL, RMS and MAV especially yielded the best results (99% in k-NN classifier, 99% in RF, and 95% in SVM classifier). The average accuracy obtained using the combination of parameters was 93% in the k -NN, 94% in RF and 84% in SVM.

REFERENCES

[1] https://www.sralab.org/facts-about-limb-loss

[2] http://www.who.int/medical_devices/publications/guide_prosthe_ortho_train/en/

[3] http://www.amputeecoalition.org/inmotion/nov_dec_07/history_prosthetics.pdf

[4] R. N. Khushaba, S. Kodagoda, D. Liu, and G. Dissanayake, "Muscle computer interfaces for driver distraction reduction", Computer Methods and Programs in Biomedicine, vol. 110, no. 2, pp. 137-149, 2012.

[5] H. Huang, J. Y. Guo, and H. J. Chen,"Ant colony optimization-based feature selection method for surface electromyography signals classification", Computers in Biology and Medicine, vol. 42, no. 1, pp. 30-38, 2012.

[6] J. Kilby, H. Gholam, "Wavelet analysis of surface electromyography signals," In: Annual International Conference of the IEEE Engineering in Medicne and Biology Society, San Francisco-USA, 2004.

[7] Fraiwan, L., Awwad, M., Mahdawi, M., and Jamous, S., "Real time virtual prosthetic hand controlled using EMG signals", pp. 225-227 (2011).

[8] Francesca Cordella, Anna Lisa Ciancio, Rinaldo Sacchetti, Angelo Davalli., et al. “Literature Review on Needs of Upper Limb Prosthesis”, Frontiers in Neuroscience journal, May 2016

[9] H. Lippert, Lehrbuch Anatomie: 184 Tabellen. Elsevier, Urban&FischerVerlag, 2006)

[10] M. A. Oskoei, and H. Hu. "A survey-Myoelectric control systems", Biomedical Signal Processing and Control, vol. 2, pp. 275-294, 2012.

[11] A. Phinyomark, P. Phukpattaranont, and C. Limsakul, "Feature reduction and selection for EMG signal classification," Expert Systems with Applications, vol. 39, no. 8, pp. 7420-7431, 2012.

[12] K. S. Kim, H. H. Moon, and C. S. Mun, "Comparison of k-nearest neighbor, quadratic discriminant and linear discriminant analysis in classification of electromyogram signals based on the wrist-motion directions", Current Applied Physics, vol. 11, no. 3, pp. 740-745, 2011.

[13] Gonzalez-Abril L, Velasco F, Angulo C, Ortega JA. A study on output normalization in multiclass SVMs. Pattern Recogn Lett 2013; 34: 344-348.

[14] BreimanLeo Random Forests inMachine Learning, 45, 5–32, 2001, Statistics Department, University of California, Berkeley, CA 94720 Kluwer Academic Publishers. 2001.