Classification of EMG Signals Between Healthy and Stroke Subjects in Upper Limb Muscle Activities

Pallab Das[1], Eashita Chowdhury[1], Vidit Gedam[2], Cheruvu Siva Kumar[2], and Manjunatha Mahadevappa[1]

[1] School of Medical Science and Technology, IIT Kharagpur, Kharagpur, India, pdas_smst@kgpian.iitkgp.ac.in, eashita06@kgpian.iitkgp.ac.in, mmaha2@smst.iitkgp.ac.in

[2] Dept of Mechanical Engineering, IIT Kharagpur, Kharagpur, India, vidit@iitkgp.ac.in, kumar@mech.iitkgp.ac.in

Abstract. Stroke is a quick loss of brain activity due to the disruption in the blood supply system. Among stroke survivors, losing hand function and finger spasticity are the main ailment. The purpose of this study is to investigate the electromyography (EMG) features in time domain and frequency domain during the shoulder complex and upper arm movement of stroke-affected and healthy muscles, and build a classification model using machine learning algorithm. An online database named "Mendeley" is used in this paper that contains EMG signals of twelve healthy subjects and thirteen stroke survivors from their six muscles (biceps brachii, triceps brachii, anterior deltoid, medial deltoid, posterior deltoid, infraspinatus) of the upper arm and shoulder complex. In this analysis, four time-domain features, namely mean absolute value, variance, root mean square, zero crossing, and three frequency-domain features, namely mean frequency, median frequency, and power spectrum ratio are extracted. A machine learning-based model is developed, which uses those extracted features to classify healthy and stroke groups. Hence, EMG features of upper limb muscles are trained by deploying linear support vector machine (SVM) classifier. Acceptable classification accuracy is noticed while generating the model. For the six muscles listed above, the achieved accuracy percentages are 81.2, 79.8, 77.8, 82.7, 79.4, and 80.6, respectively.

Keywords: Stroke · Electromyography · Feature extraction · Classification · Linear SVM.

1 Introduction

Stroke is a leading reason of death and disability in many countries. After a stroke, brain tissue gets damaged and motor units stop working properly, resulting in difficulties with muscular control and activation that cause the loss of hand function and finger stiffness. In this scenario, EMG signals are extensively used

R. Murugan et al. (eds.), *Proceedings of the International Conference on Signal Processing and Computer Vision (SIPCOV 2023)*, Advances in Engineering Research 239,
https://doi.org/10.2991/978-94-6463-529-4_8

as a control input in many human-machine interfaces and are also incorporated in various therapeutic applications [1]. EMG signal provides information on the timing of muscle activation to estimate the force generated by the muscle contraction irrespective of healthy or damaged condition. The present study aims at the disability outcome of stroke survivors during hand reaching movements which involves voluntary upper limb muscle activities using surface EMG signal analysis and compare the same with healthy groups.

The majority of surface EMG signal analysis are carried out under a static environment because it is considered that an EMG signal's frequency analysis is accurate if the signal is practically stationary [2]. Researchers have shown that the methods used for assessing stationary EMG data may be successfully employed in dynamic task-oriented activities [3] and also development of EMG based exoskeleton control for upper limb rehabilitation [4]. The current work undertakes a thorough investigation of changes in numerous time domain and frequency domain EMG features during shoulder complex and upper arm muscle movement to distinguish the difference between the healthy and post-stroke individuals. And this study also intends to build a machine learning based classification model to classify the different groups.

2 Methodology

2.1 Data Acquisition

In this work, the "Mendeley" database [5] was used which includes data acquired from twelve healthy subjects and thirteen stroke survivors. Here, participants used a surface EMG device with six electrodes attached to six muscles in the shoulder and upper arm while performing hand reaching activities for various target directions. The investigated muscles were: (1) Biceps brachii, (2) Triceps brachii, (3) Anterior deltoid, (4) Medial deltoid, (5) Posterior deltoid, (6) Infraspinatus. All participants from both groups were targeted in the same general direction.

2.2 Signal Processing

The database was recorded in a *.csv* file format which was used as an input to a customized script written in MATLAB software. After computing the power spectral density of raw EMG signal it was seen that the usable energy of capturing EMG signal mostly lies in between (30-300)Hz. So, the signals were filtered between 30Hz to 300Hz using a fourth order Butterworth bandpass filter to remove noise and artifacts. Then the muscle signals were rectified to cancel out the negative side of the spectrum and integrated to accumulate the absolute value. Now the root mean square (RMS) calculation was assumed to give the most information about the EMG signal's amplitude because it provides a measure of the signal's strength. and the envelope means boundary which contained the information of the signal. Thus the EMG envelope extraction was done for stroke patients and healthy groups which is seen in Fig. 1.

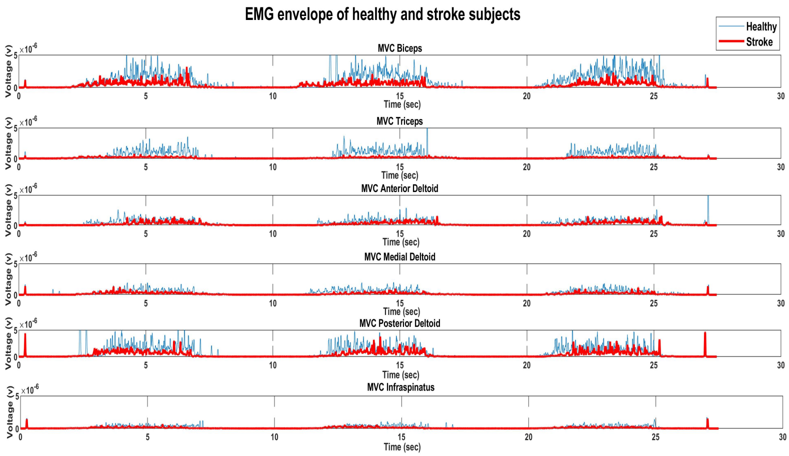

Fig. 1. EMG envelope extraction for healthy groups and stroke patients.

2.3 Feature Extraction

Feature extraction and selection is an important strategy to bring out the valid information unseen in the surface EMG signals. It also eliminates the interference and redundant part of the signals. Three major types of features namely time domain, frequency domain, and time-frequency domain are commonly used to analyse EMG signals [6]. This study examined the time domain and frequency domain features of the EMG signal.

Time domain features Time-domain features, which are calculated using raw EMG time series, are typically rapid and simple to construct because they don't require any manipulation [7]. A large variety of time domain features are discussed in the previous literature [1]. In this paper a comprehensive investigation was done on mean absolute value (MAV), variance (VAR), root mean square (RMS), and zero crossing (ZC) features of EMG signal. These features were calculated using the formula below.

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

$$VAR = \frac{1}{N-1}\sum_{i=1}^{N}e_i^2 \tag{2}$$

$$RMS = \frac{1}{N}\sqrt{\sum_{i=1}^{N}e_i^2} \tag{3}$$

$$ZC = \sum_{i=1}^{N} sgn(-e_i e_{i+1}) \tag{4}$$

where e_i denotes the i^{th} sample of EMG signal and N indicates the length of the signal. MAV is one of the most often utilized techniques for analysing EMG signals [7]. It is close to the IEMG feature, particularly when sensing the surface EMG signal for the control of prosthetic limb. VAR is another power indicator [8]. The variance of a variable is often described as the average of its square values of deviation. Another widely used tool for analysing the EMG signal is RMS [9]. It is described as a constant force, non-fatiguing contraction amplitude modulated Gaussian random process. ZC is a time-domain specification of the EMG signal's frequency information [7]. It measures how many times the EMG signal's amplitude values cross the zero level, which signifies the activation of muscles. Table I illustrates and compares the four time domain features (MAV, VAR, RMS, and ZC) of EMG signals between healthy groups and stroke patients during the movement of different muscles.

Table 1. Comparison of four time domain features between healthy and stroke subjects

Muscle name	Healthy groups				Stroke subjects			
	MAV(Volt)	VAR(Volt)	RMS(Volt)	Zero Crossing	MAV(Volt)	VAR(Volt)	RMS(Volt)	Zero Crossing
Biceps brachii	9.97x10^-5 ±2.21x10^-5*	7.53x10^-9 ±4.45x10^-9	1.32x10^-6 ±2.88x10^-7	1934±500.05	8.61x10^-5 ±3.37x10^-5	7.25xx10^-9 ±3.92x10^-9	9.51x10^-7 ±2.79x10^-7	2075±522.03
Triceps brachii	6.52x10^-5 ±2.44x10^-5	5.31x10^-9 ±4.96x10^-9	9.03x10^-7 ±3.49x10^-7	1730±551.37	2.79x10^-5 ±2.37x10^-5	6.03x10^-10 ±3.21x10^-10	2.9x10^-7 ±2.1x10^-7	1159±469.44
Anterior deltoid	8x10^-5 ±2.63x10^-5	1.61x10^-8 ±5.41x10^-9	7.2x10^-7 ±3.66x10^-7	2907±580.05	4.78x10^-5 ±3.64x10^-5	2.9x10^-9 ±1.26x10^-9	6.06x10^-7 ±2.2x10^-7	1518±532.61
Medial deltoid	9.28x10^-5 ±3.19x10^-5	9.24x10^-9 ±3.15x10^-9	5.97x10^-7 ±4.04x10^-7	2259±454.46	3.75x10^-5 ±2.15x10^-5	1.13x10^-9 ±0.87x10^-9	5.57x10^-7 ±2.51x10^-7	1242±538.46
Posterior deltoid	1.15x10^-4 ±4.45x10^-5	7.36x10^-9 ±3.01x10^-9	1.12x10^-6 ±3.39x10^-7	2296±569.22	1.07x10^-4 ±5.82x10^-5	9.71x10^-9 ±6.62x10^-9	6.17x10^-7 ±3.26x10^-7	1965±500.15
Infraspinatus	5.21x10^-5 ±2.42x10^-5	3.1x10^-9 ±1.21x10^-9	6.39x10^-7 ±2.84x10^-7	1491±482.57	2.47x10^-5 ±2.42x10^-5	3.42x10^-10 ±2.31x10^-10	1.56x10^-7 ±7x10^-8	692±191.62

*indicates (mean±SD)

Frequency domain features Frequency domain features are significantly studied for analysing the fatigue of the muscle. In this context, power spectral density (PSD) plays a significant role in the frequency domain. PSD is referred to as a Fourier transform of the autocorrelation function of the EMG signal. This is computed using the periodogram method. In this study, three frequency-domain features, namely mean frequency (MNF), median frequency (MDF), and power spectrum ratio (PSR), were extracted. Each feature was computed by using the statistical parameters of the signal's PSD. These features were calculated as follows.

$$MNF = \frac{\sum_{i=1}^{N} f_i P_i}{\sum_{i=1}^{N} P_i} \tag{5}$$

$$MDF = \frac{1}{2} \sum_{i=1}^{N} P_i \tag{6}$$

$$PSR = \frac{\sum_{i=f_0-n}^{f_0+n} P_i}{\sum_{i=-\infty}^{\infty} P_i} \quad (7)$$

where f_i and P_i are the frequency and PSD at frequency bin i respectively, and N is the frequency bin's length. MNF is determined by summing the products of the EMG power spectrum and the frequency divided by total sum of the spectrum intensity [10]. MDF is a frequency where the spectrum is split into two equally amplitude areas [10]. PSR is referred to as the ratio of the EMG power spectrum's energy close to its maximal value to its total energy [11]. Table II compares and demonstrates the three frequency domain features (MNF, MDF, and PSR) of EMG signals between healthy groups and stroke patients.

Table 2. Comparison of three frequency domain features between healthy and stroke subjects

Muscle name	**Healthy groups**			**Stroke subjects**		
	MNF (Hz)	MDF (Hz)	PSR	MNF (Hz)	MDF (Hz)	PSR
Biceps brachii	60.2±1.67*	52.53±2.06	6.51x10^-4 ±8.16x10^-6	58.45±1.2	54.79±2.04	0.65x10^-3 ±6.32x10^-6
Triceps brachii	61±1.85	55.57±2.22	6.41x10^-4 ±9.13x10^-6	61.02±1.31	52.69±1.75	0.64x10^-3 ±5.55x10^-6
Anterior deltoid	61.12±2.13	52.97±2.31	6.52x10^-4 ±7.42x10^-6	57.89±0.86	50.66±1.89	0.65x10^-3 ±6.22x10^-6
Medial deltoid	60.15±2.58	53.09±2.81	6.63x10^-4 ±9x10^-6	59.22±0.88	53.02±2.26	0.662x10^-3 ±7.57x10^-6
Posterior deltoid	63.25±3.36	56.61±3.2	6.54x10^-4 ±6.36x10^-6	59.45±2.37	52.65±1.38	0.652x10^-3 ±9.9x10^-6
Infraspinatus	58.13±2.8	51.06±2.48	6.45x10^-4 ±6.12x10^-6	57.83±2.58	48.94±1.26	0.644x10^-3 ±7.89x10^-6

*indicates (mean±SD)

3 Results and Discussion

A sound increase in the amplitude of biceps brachii, triceps brachii, and posterior deltoid muscles of healthy subjects was observed as compared to stroke affected patients, which depicts that certain muscles were used more actively in hand reaching movements than other muscles. Again it is observed that the mean absolute value, variance, and zero crossing of healthy groups are more than the stroke survivors. It signifies that the muscles of healthy subjects more actively participated in the motor task than stroke patients. To differentiate EMG variability of the mentioned muscles, an RMS comparison was done among healthy and stroke subjects' data demonstrated in Fig. 2.

Fig. 3 compares the mean absolute value of EMG signal. It is observed that MAV and RMS of EMG signal have a similar correlation with the contraction

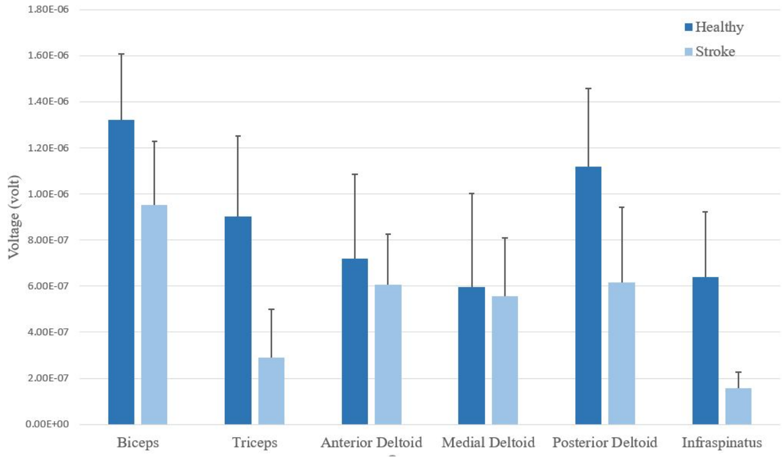

Fig. 2. RMS value (of EMG signal) comparison between healthy and stroke subjects.

of muscles. Mean values of healthy subjects are larger than stroke patients for all of the muscles suggesting that the stiffness level of the muscles is higher for stroke individuals.

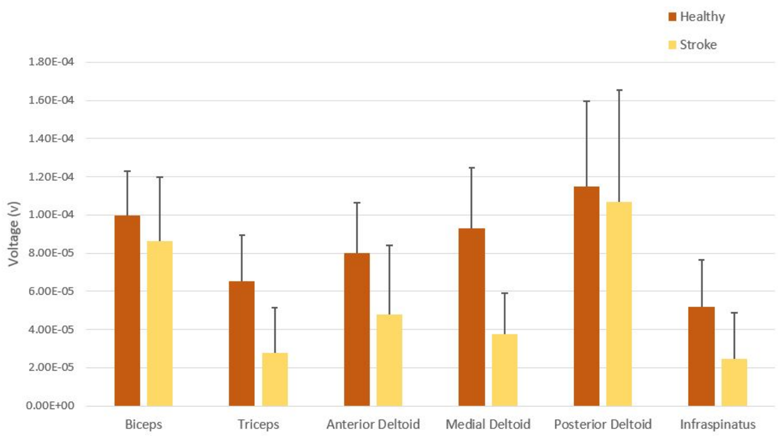

Fig. 3. MAV value (of EMG signal) comparison between healthy and stroke subjects.

A categorical comparison is done on the calculated frequency domain. Fig. 4 illustrates two EMG signal's features (MNF and MDF) in the frequency domain. It is observed that the mean frequency (MNF) is significantly greater than the median frequency (MDF) for healthy as well as stroke patients, and mean, median frequencies are also higher in healthy groups for most of the muscles. Fig.

5 shows the EMG signal's power spectrum ratio (PSR) of both the groups. The calculated PSR were not found significant.

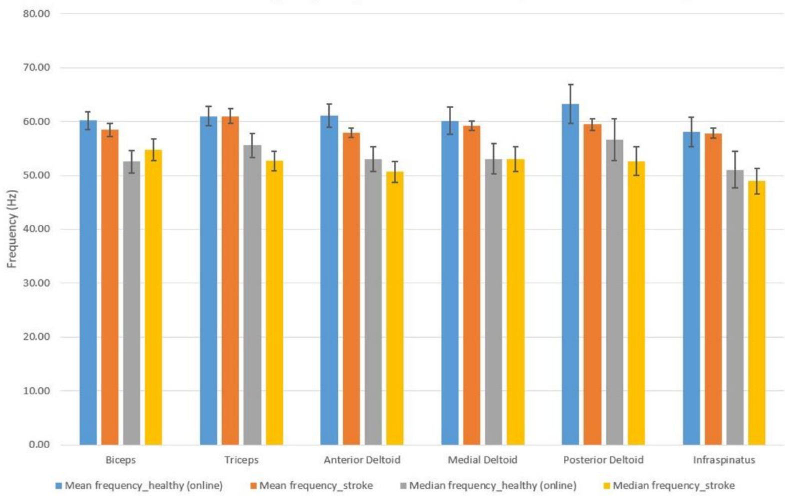

Fig. 4. Comparison of mean and median frequency between healthy and stroke subjects.

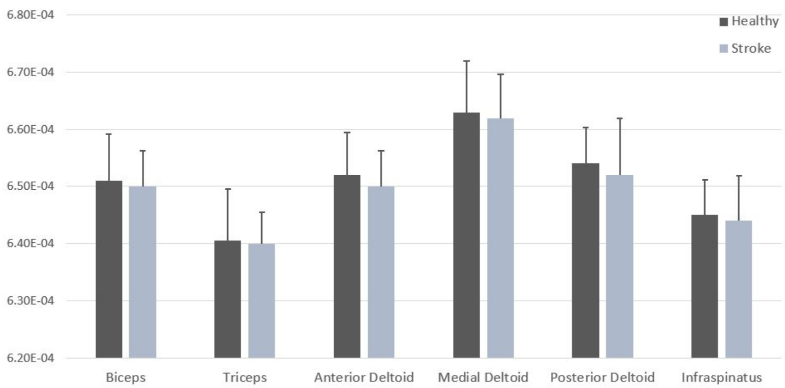

Fig. 5. Power spectrum ratio of healthy and stroke subjects.

A binary classification was performed between healthy and stroke subjects during the movement of six different muscles of the upper arm and shoulder summarized in Table III. In this classification technique, RMS values of EMG signals were considered as predictors and two groups (healthy and stroke) were categorized as response. Five fold cross-validation approach was used during the training phase. In this work, a linear SVM classifier was used for classification purpose. SVM is one type of supervised learning model, developed by Boser, Gyon and Vapnik [12] that are utilised in machine learning for data classification and regression analysis [13]. The objective of this SVM model was to create a prediction model using the provided training samples and to identify specific patterns in such samples. Post the training phase, this model was applied to a testing dataset that was used to categorise unknown input data. In this study, the used dataset was RMS features. However, from the classification results a satisfactory performance accuracy was observed that distinguishes the difference between the muscle activities of healthy and stroke subjects.

Table 3. Classification performance on EMG features of healthy and stroke subjects using linear SVM

Binary Classification	**Linear SVM Classifier**		
(Healthy vs Stroke)	***Accuracy (%)***	***Specificity (%)***	***Sensitivity (%)***
Biceps brachii	81.2	71.5	69.8
Triceps brachii	79.8	70.6	72.9
Anterior deltoid	77.8	75.3	70.7
Medial deltoid	82.7	77.3	74.6
Posterior deltoid	79.4	73.2	71.7
Infraspinatus	80.6	72.2	75.4

4 Conclusion

It was found that the activity of the biceps and triceps muscles had a close relation to the movement of the upper limb. When the shoulder complex moves, the anterior deltoid, medial deltoid, and posterior deltoid muscles are more prevalent than the infraspinatus muscles. And those active muscles are less functional for stroke-affected patients. The hand reaching movement tasks related to this work engage various activities of daily living similar to real-life situations, and also recover stroke survivors from their long-term motor disorder. To extend this work, an alternative classification method can be deployed to improve accuracy including the use of external devices can be considered in future as well.

Acknowledgment

The authors express their gratitude to S. Israely *et al.* the creator of the database "Mendeley" [5] for providing online data crucial for this study.

References

1. Phinyomark, A., Phukpattaranont, P. and Limsakul, C.: Feature reduction and selection for EMG signal classification. Expert systems with applications 39(8), pp. 7420-7431 (2012).
2. Hostens, I., Seghers, J., Spaepen, A. and Ramon, H.: Validation of the wavelet spectral estimation technique in biceps brachii and brachioradialis fatigue assessment during prolonged low-level static and dynamic contractions. Journal of electromyography and kinesiology 14(2), pp. 205-215 (2004).
3. Knaflitz, M. and Bonato, P.: Time-frequency methods applied to muscle fatigue assessment during dynamic contractions. Journal of Electromyography and Kinesiology 9(5), pp. 337-350 (1999).
4. Ranjan, S., Bakshi, K., Gaur, A., Manjunatha, M. and Kumar, C.S.: Case Study of Developing an Electromyogram-Based Exoskeleton Control for Upper Limb Rehabilitation. In Soft Computing for Problem Solving: Proceedings of SocProS 2020, vol. 1, pp. 171-184. Singapore: Springer Singapore (2021).
5. Israely, S., Leisman, G. and Carmeli, E.: EMG datasets form hand-reaching movements for multiple directions of healthy and post-stroke individuals, doi: 10.17632/f4hh43nd78.2, (2019).
6. Oskoei, M.A. and Hu, H.: Myoelectric control systems—A survey. Biomedical signal processing and control 2(4), pp. 275-294 (2007).
7. Hudgins, B., Parker, P. and Scott, R.N.: A new strategy for multifunction myoelectric control. IEEE transactions on biomedical engineering 40(1), pp. 82-94 (1993).
8. Park, S.H. and Lee, S.P.: EMG pattern recognition based on artificial intelligence techniques. IEEE transactions on Rehabilitation Engineering 6(4), pp. 400-405 (1998).
9. Boostani, R. and Moradi, M.H.: Evaluation of the forearm EMG signal features for the control of a prosthetic hand. Physiological measurement 24(2), pp. 309-319 (2003).
10. Oskoei, M.A. and Hu, H.: Support vector machine-based classification scheme for myoelectric control applied to upper limb. IEEE transactions on biomedical engineering 55(8), pp. 1956-1965 (2008).
11. Qingju, Z. and Zhizeng, L.: Wavelet de-noising of electromyography. In 2006 International Conference on Mechatronics and Automation, pp. 1553-1558. IEEE (2006).
12. Vapnik, V.: The nature of statistical learning theory. Springer science & business media, (1999).
13. Evgeniou, T. and Pontil, M.: Support vector machines: Theory and applications. In Machine Learning and Its Applications: Advanced Lectures pp. 249-257. Berlin, Heidelberg: Springer Berlin Heidelberg (2001).