Journal of Scientific & Industrial Research
Vol. 79, November 2020, pp. 1011-1016

Machine Learning based Electromyography Signal Classification with Optimized Feature Selection for Foot Movements

Preeti Khera[1,2], Neelesh Kumar[1,2]* and Parul Ahuja[2]

[1]Academy of Scientific and Innovative Research, Ghaziabad 201 002, India

[2]Biomedical Instrumentation Unit, CSIR-CSIO, Sector-30 C, Chandigarh 160 030, India

Received 29 May 2020; revised 2 September 2020; accepted 18 September 2020

Electromyography (EMG) signals are bioelectric signals generated by the electrical activities of muscle fibers during contraction or relaxation. Detailed analysis and classification of the complex nature of the signal when related to movements is complicated. However, these are useful for controlling prosthesis and orthosis control systems. In this paper the relevant set of features and the classifier that maps these features to carry out EMG signal classification for four different foot movements is proposed. These movements such as plantar-flexion (PF), dorsi-flexion (DF), inversion (IV) and eversion (EV) are chosen, since these are useful for rehabilitation of persons having a lower limb ankle joint injury which results in gait abnormality. EMG signals are acquired using BIOPAC System (MP 150). The features for EMG signals, in time and frequency domain have been extracted to find optimal features. Further these are classified using support vector machine (SVM), neural network (NN) and logistic regression (LR). From the results, it is depicted that the time domain features reflected better performance. The maximum classification accuracy achieved is 99.69% and average classification accuracy being 94.92 ± 3.03 % using linear SVM for root mean square (RMS) as optimal feature.

Keywords: Classification accuracy, EMG, Foot movements, SVM, Time and frequency domain features

Introduction

The most useful method to analyze muscular condition, a technique to capture and measure the electrical activity during voluntary and involuntary activities is EMG.[1] EMG is the most prevalent technique as it has the least response time from the initiation of muscle activity till the measurement of the electrical activity by sensors directly. Moreover, the use of EMG allows predicting movements in advance as compared to inertial sensors.[2] The ability to detect movement intention[3] as electrical activity in muscles exists even if its power is insufficient to initiate the movement, as a result of nervous connection injury. It is also considered to be an analytical way that evaluates the strength of muscles and motor neurons termed as nerve cells that controls them. Moreover, the electrical signal generated can be used to analyze the activation level of the muscles. Each movement of muscle produces a specified pattern owing to its activation; this EMG signal recording can be useful for the identification of movements.[1] The EMG signals are non-linear and complex in nature, easily interrupted by environmental noise and other motion artifacts[4] that produces complexity in its analysis and classification. When a muscle contracts or relaxes, the muscle fiber action potential of an individual can be determined by two approaches viz. using needle electrodes (an invasive process) and through the use of skin surface electrodes (non-invasive process). The latter is increasingly used for recording, from superficial muscles due to its non-invasive nature and no need for surgical intervention.[5,6] It is used in a wide range of applications such as human-machine interfaces, robotic assisted rehabilitation therapies and, prosthetics and orthotic devices.[1,4,7–11] However, most of the studies optimized the redundant information from recording sites i.e. EMG channels to enhance system performance. These incorporated applications of dimensionality reduction techniques mainly for upper[12–14] limb movement. Also, pattern recognition of EMG signal for upper limb[15,16] has been widely explored compared to lower limb. The lower limb EMG analysis, especially in isometric contraction are required for therapeutic interventions.[17] Thus, there is a wide scope in this area. Nevertheless, the multifactorial analysis determines that the performance of a particular classifier depends upon the choice of feature vector. There is a need to

*Author for Correspondence
E-mail: neel5278@csio.res.in

determine relevant features and an accurate EMG signal classifier, which aids to map selected features to a target class for recognition of foot movements. As the proliferation in the number of features with least or no relevance and more channels[18], produces high computational time leading to increased computational complexity. Thus, this paper describes the optimal classifier by comparing the classification accuracy of different classifiers with time and frequency domain feature set. EMG signals from two most prominent superficial muscles were acquired. Section II describes the experimental paradigm along with data acquisition protocol and extraction of time and frequency domain features. The section III presents the results of our research work. Lastly, conclusion of the paper is provided in section IV.

Methods and Materials

Experimental Paradigm

EMG signals were collected from 10 healthy subjects (5 females, 5 males) aged between 20–33 years with no past lower limb injuries and prior verbal consent. The data was acquired from the superficial muscles viz. fibularis longus (FL) and hallucis longus (HL) that were identified using the anatomical landmark system using BIOPAC Systems (MP 150). Every participant was given verbal instructions to perform four different foot movements consisting of PF, DF, IV and EV. Ten isometric maximum voluntary contractions for a period of 10 seconds corresponding to each movement were acquired. The collected EMG signal requires pre-processing, filtering and rectification. The features in time and frequency domain were extracted for evaluation. The EMG signal was band pass filtered at Hz with a notch set at 50 Hz. The use of the forward reverse digital filter provides no time shift compared to unidirectional filters.[19] In order to analyse the EMG signal, time and frequency domain features were extracted. These were used for classification of foot movements.

Data Acquisition Protocol and Feature Extraction

The right foot was chosen for all the subjects as being dominant side. The four trials for static sitting conditions were acquired as a baseline data. Another four trials for dynamic conditions i.e. PF, DF, IV and EV were taken for each subject for a period of 10 seconds. The data was acquired by BIOPAC (MP150) system using highly conductive disposable solid gel electrodes. For bipolar configuration, two electrodes were placed in close proximity to each other on the selected muscles. The signals were captured at a set gain of 2000 and sampling frequency 1000 Hz. The Fig. 1 shows placement of electrodes on one of the subjects at selected muscles. Thereafter, feature extraction for pre-processed EMG signal includes time domain (TD) and frequency domain (FD) characteristics. Even time-frequency domain characteristics also exist but require signal transformation for classification with not much added accuracy. However, it increases the signal complexity. The EMG signal is time dependant so TD features can be easily extracted without any transformation. But these types of signals are influenced by noise due to amplitude dependability and electrode shifting. So, extra care is taken at the acquisition time to record noise free signal. FD features are accessed using periodogram of the signal. Being insensitive to noise such signals are complex. So, this study includes four TD and only one FD feature. The Table 1 highlights various TD and FD features extracted with their definition.

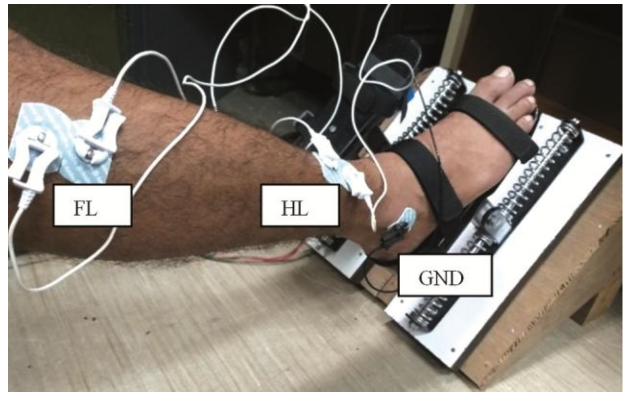

Fig. 1 — EMG Electrode Placement

Results and Discussion

Five features were extracted in time and frequency domain (RMS, MAV, WA, PSR, ARC) for each foot movement from all the subjects. The classifiers used in this study are support vector machine (SVM), neural network (NN) and logistic regression (LR). To determine the optimal classifier in our work, all kernel versions of SVM[24–26] and various activation functions of NN[27] were analyzed and discussed using leave-one-out cross validation.

The Table 2 demonstrated the classification accuracy statistics using training and test data from all the subjects by using each of the features independently and in combination with other features. The results presented in Table 2 depicted the comparable accuracies as devised for upper limbs.[15,16] As

Table 1 — Time and Frequency domain features

S. No.	Feature and its definition	Formula
1.	Root Mean Square (RMS):A most robust feature representing the square root of the mean of squares of individual values.[16,19-21]	$RMS = (\sum_{i=1}^{N} x_i^2)^{1/2}$
2.	Mean Absolute Value (MAV): Defines the onset of movement and is used to measure the absolute amplitude of EMG signal in isometric contractions for given segment length.[15,19-21]	$MAV = \frac{1}{N}\sum_{i=1}^{N} x_i$
3.	Wilson Amplitude (WA): Determines the number of counts that depicts difference between the values of two adjacent scales in a time frame of a signal exceeding the given threshold. The threshold value used is $20\,\mu V$.[10] It is a measure of frequency information in TD.	$WA = \sum_{i=1}^{N} f(\lvert x_i - x_{i+1}\rvert)$ where, $f(x) = \begin{cases} 1, x > Threshold \\ 0, Otherwise \end{cases}$
4.	Auto-regressive coefficients (ARC): The time-series model in which the signal samples are predicted from their previous samples by using the linear combination of samples is the AR model.[10,16,21,22] ARC provides information about muscle contraction state and are used as features.	$C_{kn}(m,n) = \sum_{j=1}^{p} C_{kn}(m-j,n)a_j$ where, p=2 depicts the order of the model, m and n are taken as lag elements in two-dimensional plane and the coefficient k is the interval index.
5.	Power Spectral Ratio (PSR): It is defined as the ratio of maximum of the power (segment-wise) from given subset of EMG signal and the aggregate of the power of that EMG signal.[23]	$PSR = \frac{P(\max)}{P\,(total)}$

x_i is the EMG signal and N is taken as the length of EMG signal.

Table 2 — Time and frequency domain features accuracy statistics

FEATURES	TRAINING ACCURACY (Mean and Standard Deviation)	TESTING ACCURACY (Mean and Standard Deviation)
RMS	91.3 ± 6	94.92 ± 3.03
MAV	91.73 ± 6.13	92.72 ± 6.62
WA	84.63 ± 9.51	85.64 ± 9.77
PSR	48.01 ± 6.37	48.1 ± 5.43
ARC	67.95 ± 10.79	67.08 ± 7.04
RMS and MAV	92.17 ± 6.15	94.64 ± 3.19
RMS and WA	92.33 ± 5.79	93.87 ± 4.53
RMS and PSR	91.52 ± 5.93	91.49 ± 7.27
RMS and ARC	91.91 ± 5.39	95.16 ± 3.07
MAV and WA	92.55 ± 5.93	93.61 ± 5.5
MAV and PSR	92.05 ± 6.58	92.53 ± 6.7
MAV and ARC	92.74 ± 5.28	91.15 ± 7.7
WA and PSR	85.41 ± 9.15	85.50 ± 8.84
WA and ARC	90.69 ± 5.26	92.4 ± 5.64
PSR and ARC	68.9 ± 10.35	67.6 ± 6.23
RMS, MAV and WA	92.71 ± 5.87	92.28 ± 6.3
RMS, MAV and PSR	92.45 ± 5.97	92.87 ± 6.75
RMS, MAV and ARC	93.23 ± 5.41	94.98 ± 2.75
MAV, WA and PSR	92.84 ± 5.46	92.11 ± 6.99
MAV, WA and ARC	92.95 ± 5.51	86.35 ± 10.58
WA, PSR and ARC	90.82 ± 5.05	92.53 ± 5.57
RMS, MAV, WA and PSR	93.14 ± 5.37	93.43 ± 6.15
RMS, MAV, WA and ARC	93.15 ± 5.45	93.56 ± 4.8
MAV, WA, PSR and ARC	93.1 ± 5.15	94.22 ± 5.62
RMS, MAV, WA, PSR and ARC	92.97 ± 5.04	94.3 ± 5.55

concluded from the analysis that RMS is the dominant feature which outperforms the classification accuracy of all other features using linear SVM classifier with deviation of ± 3.03%. The performance of MAV in combination with RMS (94.64 ± 3.19 %) is also comparable to RMS with almost similar deviation. Another feature i.e. RMS and ARC provides accuracy of 95.16 ± 3.07 % even though ARC independently provides only 67.08 ± 7.04% accuracy. The PSR in combination with RMS is able to achieve CA of 91.49 ± 7.27 % even though its CA independently was 48.1 ± 5.43%. Thus, its use with other features helps to improve the accuracy. Another feature which is of significant interest is the combination of RMS, MAV and ARC with CA of 94.98 ± 2.75 %. This combination has reported the least deviation of magnitude 0.28% even less than RMS alone. This can be considered as the most stable factor with additional increase in computational time and producing almost comparable results. Thus, it can be concluded from above mentioned statistics that combination of features with RMS is able to achieve comparable results but with little more deviations as when only dominant feature RMS has been used. Moreover, the average individual contribution of various features taken independently clearly depicts that RMS is the most relevant feature for EMG signals to classify foot movements as shown in Fig. 2(a).

Further, to determine the optimal classifier, the classification accuracies of three classifiers viz. LR, NN and SVM were compared using the test dataset. From the results, it can be noted that SVM

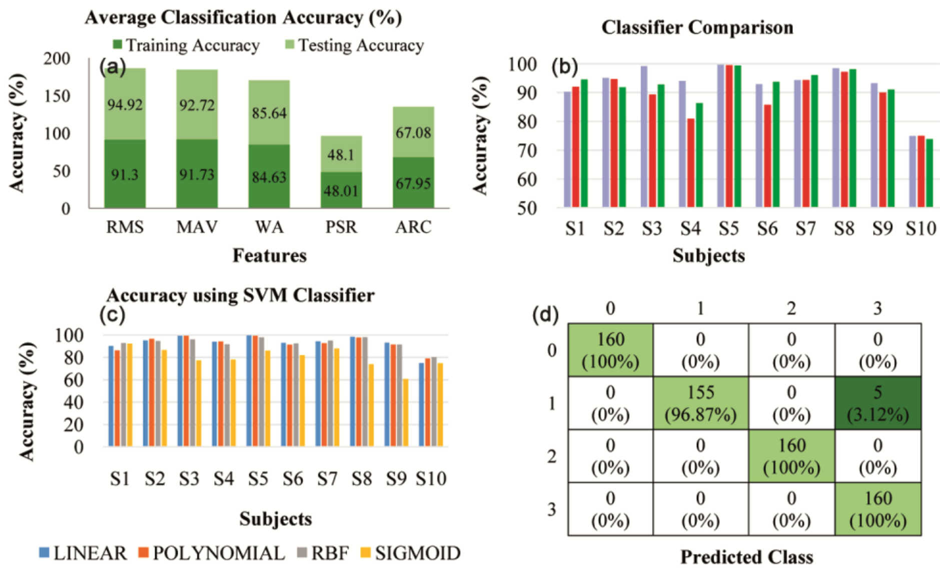

Fig. 2 — Classification Accuracy (a) Features using SVM (b) using RMS with all classifiers (c) different kernel functions of SVM classifier (d) Confusion Matrix for SVM using RMS for subject ‘S5’

outperforms the other classifiers for majority of the subjects as shown in Fig. 2(b). The average classification accuracy of 10 healthy subjects using various classifiers; SVM is 93.23%, NN is 89.86% and LR is 91.78%. As SVM is a kernel based approach for comparative analysis, the original EMG data was transformed into high dimensional plane and tested for different kernel functions of SVM classifier. The linear SVM classifier depicted highest average classification accuracy of 93.23% which was greater than all the other functions such as polynomial (by 0.44%), RBF (by 0.16%) and sigmoid (by 13.25%) as highlighted in Fig. 2(c) for all the subjects. The classification performance of a particular dataset depends upon the signal characteristics. The features extracted in this study are TD and FD (RMS, MAV, WA, PSR and ARC). Thus, the hyperplane of SVM with optimization can accurately classify four different foot movements for the selected features. However, the same trend was reported in previous studies[13,16,18] with SVM classifiers for upper limb movement.

To evaluate the classification results, confusion matrix was used as a performance measure to depict the classification accuracy for each class has been presented in Fig. 2(d). It represented the classification performance computed using the test dataset of subject ‘S5’ with actual (true) and observed (predicted) class along the rows and columns, respectively. The predicted samples containing the percentage per class represent the classification accuracy. Here, four classes 0, 1, 2, 3 correspond to four different foot movements such as IV, EV, DF and PF respectively. As shown in Fig. 2(d) some data samples of class 1 were misclassified as belonging to class 3 resulting in degraded performance. The evaluation metrics i.e. precision, recall and f1-score were calculated for test dataset of subject ‘S5’ to determine the performance and reliability of linear SVM classifier. It can be further depicted from the statistical measures i.e. class 0 (IV) and class 2 (DF) can be predicted with 100% accuracy while some more consideration while collecting the data are required for other two foot movements that belongs to class 1 (EV) and class 3 (PF) for improvement. The f1-score for these two classes are 98%.

Conclusions

The study was successfully completed with the accomplishment of three main objectives. Firstly, to predict the relevant features for EMG signal classification related to different foot movements. Secondly, to determine the classifier that maps the selected feature to achieve maximum classification accuracy. Third, to reduce the computational time spent for training the classifier with huge amount of training data containing lots of features. The single feature i.e. RMS concluded in this study reports high accuracy as compared to multiple feature setfor EMG signal classification for different foot movements. The average classification accuracy of all the participants under study as predicted by linear SVM is

94.92 ±3.03% which outperforms the other classifiers i.e. LR by 3.14% and NN by 5.06 %. Moreover, the results presented showed lower variability as the deviation is just around ± 3% and can be further reduced by including more data at the time of training. When RMS is combined with other feature sets like MAV, WA, PSR and ARC its classification accuracy is almost comparable with the increase in deviation even though that being less than ± 7.5% but it leads to increase in computational time and complexity. Meanwhile, the maximum classification accuracy is achieved using linear SVM in original domain avoiding the need for transformation in high dimensional plane. The classification results of the applied classifier are noteworthy and thus can be used to classify EMG signals for prosthesis control studies. The present technique can be applied for rehabilitation treatment. It includes control of ankle therapeutic devices. This strategy will enhance muscle strength. The future perspective is to make this offline analysis to real-time. Moreover, the above set of features analysed in this study can be applied on patient group for more rigours' analysis. Also, the classification result can be tested with other classifiers with more number of subjects and using the same classifier in different fatigue conditions.

Acknowledgement

The authors would like to present their sincere gratitude for the support and contribution provided by Director, CSIR-Central Scientific Instrument Organisation, Chandigarh, India. Khera P acknowledges UGC, India for supporting her PhD through its national fellowship programme.

Conflict of Interest

The authors hereby declare that they have no conflict of interest.

References

1 Alkan A, Günay M,Identification of EMG signals using discriminant analysis and SVM Classifier, *Expert Syst Appl*, **39** (2012) 44–47.

2 Wentink E C, Schut V G, Prinsen E C Rietman J S, Veltink P H, Detection of the onset of gait initiation using kinematic sensors and EMG in transfemoral amputees, *Gait Posture*, **39(1)** (2014) 391–396.

3 Fleischer C, Reinicke C, and Hommel G, Predicting the intended motion with EMG signals for an exoskeleton orthosis controller, *2005IEEE/RSJ Int ConfIntell Robo Syst*, (2005) 2029–2034.

4 Too J, Abdullah A R, Mohd Saad N and Tee W, EMG feature selection and classification using a Pbest-guide binary particle swarm optimization, *Computation*, **7(1)** (2019) 12.

5 Toledo-Pérez D C, Martínez-Prado M A, Gómez-Loenzo R A, Paredes-García, W J, Rodríguez-Reséndiz J, A study of movement classification of the lower limb based on up to 4-emg channels, *Electronics*, **8(3)** (2019) 259.

6 Kleissen R F, Buurke J H, Harlaar J, Zilvold G, Electromyography in the biomechanical analysis of human movement and its clinical application, *Gait Posture*, **8(2)** (1998) 143–158.

7 Englehart K, Hudgin B, Parker P A, A wavelet-based continuous classification scheme for multifunction myoelectric control, *IEEE Trans Biomed Eng*, **48(3)** (2001) 302–311.

8 Shi W T, Lyu Z J, Tang S T, Chia T L, Yang, C Y, A bionic hand controlled by hand gesture recognition based on surface EMG signals: A preliminary study, *Biocybern Biomed Eng*, **38** (2018) 126–135.

9 Sapsanis C, Georgoulas G, Tzes A, EMG based classification of basic hand movements based on time frequency features, *21st Med Conf Cont& Auto*, (2013) 716–722.

10 Boostani R, Moradi M H, Evaluation of the forearm EMG signal features for the control of a prosthetic hand, *Physiol Meas*, **24(2)** (2003) 309–319.

11 Gupta R, Agarwal R, Electromyographic signal-driven continuous locomotion mode identification module design for lower limb prosthesis control, *Arab J Sci Eng*, **43** (2018) 7817–7835.

12 Wang D, Zhang X, Gao X, Chen X, Zhou P, Wavelet packet feature assessment for high-density myoelectric pattern recognition and channel selection toward stroke rehabilitation. *Front Neurol*, **7**(2016).

13 Mesa I, Rubio A, Tubía I, No J D, Diaz J, Channel and feature selection for a surface electromyographic pattern recognition task, *Expert Syst Appl*, **41** (2014) 5190–5200. 10.1016/j.eswa.2014.03.014.

14 Liu J, Li X, Li G, Zhou P, EMG feature assessment for myoelectric pattern recognition and channel selection: a study with incomplete spinal cord injury, *Med Eng Phys*, **36(7)** (2014) 975–980. doi: 10.1016/j.medengphy. 2014.04.003

15 Hudgins B, Parker P and Scott R N, A new strategy for multifunction myoelectric control, IEEE*Trans Biomed Eng*, **40(1)** (1993) 82–94.

16 Oskoei M A, Hu H, Support vector machine-based classification scheme for myoelectric control applied to upper limb, *IEEE Trans Biomed Eng*, **55(8)** (2008) 1956–1965.

17 Nazmi N, Abdul Rahman M A, Yamamoto S, Ahmad S A, Zamzuri H, Mazlan S A, A Review of classification techniques of EMG signals during isotonic and isometric contractions, *Sensors*, **16(8)** (2016) 1304.

18 Ceseracciu E, ReggianiM, Sawacha Z, Sartori M, Spolaor F, Cobelli C, Pagello E, SVM classification of locomotion modes using surface electromyography for applications in rehabilitation robotics,*19th Int Sym Robo Hum Int Comm*, (2010) 165–170.

19 Veer K, Sharma T, A novel feature extraction for robust EMG pattern recognition, *J Med EngTecchnol*, **40(4)** (2016) 149–154.

20 Kiran K, Uma Rani K, Analysis of EMG Signal to Evaluate Muscle Strength and Classification, *IRJET*, **4(7)** (2017) 177–182.
21 Spiewak C, Islam M R, Zaman M A, Rahman M H, A comprehensive study on EMG feature extraction and classifiers, *JBEB*, **1(1)** (2018).
22 Philipson L, Larsson P G, The electromyogram signal as measure for muscular force: a comparison of detection and quantification techniques, *Electromyogr Clin Neurophysiol*, **28(2)** (1988) 141–150.
23 Qingju Z and Zhizeng L, Wavelet de-noising of electromyography, 2006 *Int Conf Mech Auto*, Luoyang, Henan, (2006) 1553–1558, doi: 10.1109/ICMA.2006. 257406.
24 Osuna E, Freund R, Girosi F, Training support vector machines: an application to face detection, *Proc IEEE Comp Society Conf Comp Vision Pattern Recognit*, (1997) 130–136.
25 Liu C, Nakashima K, Sako H, Fujisawa H, Handwritten digit recognition: benchmarking of state-of-the-art techniques, *Pattern Recognit*, **36(10)** (2003) 2271–2285.
26 Pontil M, Verri A, Support vector machines for 3d object recognition, *IEEE Trans Pattern Anal Mach Intell*, **20(6)** (1998) 637–646.
27 Vapnik V N, An overview of Statistical Learning Theory, *IEEE TransNeural Netw*, **10(5)** (1999) 988–999.