Proceeding of the IEEE
International Conference on Robotics and Biomimetics
Dali, China, December 2019

Real-Time Surface EMG Pattern Recognition for Hand Gestures Based on Support Vector Machine*

Zhen Zhang, Yugui Tang, Sichen Zhao, Xinpeng Zhang
School of Mechatronic Engineering and Automation
Shanghai University
Shanghai, China
{zhangzhen_ta & tangyugui & George_Z & zhang605}@shu.edu.cn

***Abstract* - In this paper, a real-time hand gesture recognition model based on surface electromyography signals is proposed. The Myo armband is applied to acquire surface electromyography (sEMG) signals generated by hand movement and we extract the features by using a sliding window to split the data. Then we select support vector machine (SVM) model for gesture recognition, which is trained by the training set and tested by testing set. We count the labels returned by classifier and assign the label whose times reach the threshold of recognition to the tested gesture. In order to evaluate this model, we set up an experiment with five target gestures which conducted by 12 subjects. Experimental results show that the average accuracy of proposed model can reach 97.8% and the recognition process is completed before the continuous action is finished, which means that this model can achieve real-time identification with high accuracy.**

Index Terms - Hand gesture recognition; Surface electromyography; Support vector machine; Real-time

I. INTRODUCTION

Hand gesture is one of the earliest communications and has become the most powerful way to express the emotions due to its large amount of information and high flexibility. In recent years, gesture recognition is widely used in upper-limb prosthetics [1], human computer interfaces, virtual game control [2], and intelligent robotics [3]. Nowadays, gesture recognition has become an attractive field.

Data acquisition is the first step for a gesture recognition system. A lot of sensors can be used for this step, such as data glove [4], which sends data about hand movements to the computer, depth camera [5], which builds 3D model of the hand posture and surface electromyography (sEMG) [6, 7]. However, data glove is inconvenient to wear and there are many restrictions on users. Depth camera needs suitable conditions of light and position to acquire information. In comparison with these sensors, surface electromyography (sEMG) sensor can be effectively applied for the recognition of hand gesture because they are not influenced by the variations of light, position and orientation of the hand.

sEMG signals are generated from the contraction of muscles and consist of lots of information about muscles activities. Additionally, the sEMG signals can be easily detected by sensors that wear on the skin surface of the muscles.

Feature vectors are the most important data in the process of hand gesture recognition. The feature vectors are composed of the features extracted from the following domains: frequency domain, such as power spectrum ratio and median frequency [8]; time domain like mean absolute value (MAV) [9], and time-frequency domain like wavelets [10]. The recognition classifiers based on these features include artificial neural network (ANN) [11], convolutional neural networks [12], random forest [13], k-nearest neighbors (k-NN) [14], and support vector machines (SVM) [15].

The aim of this article is to propose a real-time model based on support vector machines and sEMG signals. This recognition system is divided into five modules: data acquisition, preprocessing, feature extraction, classification, and post-processing. For data acquisition, we use Myo armband to collect the sEMG signals. For preprocessing, these signals are processed by the low pass filter. For feature extraction, a sliding window approach is applied to segment the signals, the feature vector is formed by preprocessed signals and five time-domains features in each sliding window, then a SVM model is applied to classification and the classifier returns the identification result in real time.

Follow this introduction, this paper is divided into three sections. In section 2, we describe the materials used in the experiment. In section 3, we explain five modules of the proposed model in detail. In section 4, we present the results of the experiment and make analysis and discussion. In section 5, we summarize this article.

II. MATERIALS

The materials used in this work is the Myo armband that designed by Thalmic Labs in Canada. As shown in Fig. 1, the Myo armband is comprised of eight muscle pulse detection modules and Bluetooth module. These sEMG sensors on the armband detect electrical changes in the forearm muscles and transmit signals to computer via Bluetooth. Data from Myo armband are collected at a sampling frequency of 200Hz and consisted of 8 bits from eight channels.

In our experiment, five hand gestures are targets of recognition: Fist, Spread, Wave In, Wave On and Double Tap (as shown in Fig. 2), which are common in virtual control and are preset in device by default.

* This work is jointly supported by Shanghai science and technology commission under grant No18JC1410402

978-1-7281-6321-5/19//$31.00 ©2019 IEEE

Fig. 1 Myo armband

We collected the data of 12 healthy subjects (eight men and four women, aged from 21 to 25) as training set and testing set. In this experiment, the signal acquisition goes through the following steps: relaxes arm, performs gestures and returns arm to the relaxed position. Five repetitions of the five gestures measured during 2 seconds were used to form the training set. Additionally, to find a reference about preprocessing, we also acquire five sEMG signal samples recorded in the relaxed position, which is labeled as "No action" (as shown in Fig. 2). The testing set is formed by 30 repetitions of five hand gestures recorded during 2 seconds.

Fig. 2 Five target gestures and relaxed position

III. MODEL

The proposed model is composed of five sections: Data Acquisition, Data Preprocessing, Feature Extraction, Classification, and Data Post-processing (as shown in Fig. 3).

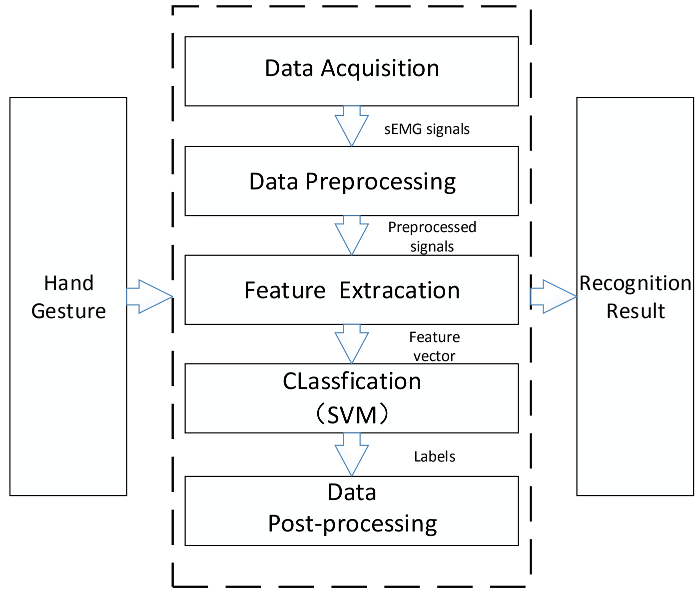

Fig. 3 Five sections of proposed model

A. Data Acquisition

In data acquisition section, the raw signals about muscle activity from materials are recorded. The original signals measured by 8 sEMG sensors are saved with form of matrix $\mathbf{T}=[\mathbf{T}_1, \ldots \mathbf{T}_8]^{N\times 8}$, where sampling length is denoted by N and 8 denotes the number of channels. Meanwhile, value of each element in matrix are normalized in range $[-1, 1]$.

B. Data Preprocessing

In preprocessing section, the raw signals include some additional noise, which makes difficult to extract the features we needed. In order to avoid extracting invalid features and disturbing the identification process, the signals are rectified using an absolute value function, then we get new signals $abs[\mathbf{T}] = \left[abs[\mathbf{T}_1], \ldots, abs[\mathbf{T}_8]\right] \in [0, 1]^{N\times 8}$. Additionally, a filter is be used to remove the noise doped in the original signals, we obtain the amplitude of rectified signals by using the Fourier transform and realize that the most reasonable cut-off frequency is 5 Hz. Therefore, we select the 4th order digital Butterworth filter ψ whose cut-off frequency is 5Hz to remove invalid noise and obtain the sEMG envelop of each channel (as shown in Fig. 4). We denote the results from this filter as $\mathbf{L}=\psi\left[abs[\mathbf{T}]\right] \in [0, 1]^{N\times 8}$.

The output from these steps includes some relaxing signals that generated from relaxed position. Therefore, a muscle detection function is used to remove the head and tail of signal and extract the region of muscle activity by referring to the relaxing sEMG signals. The steps are same as in [14].

C. Feature Extraction

In feature extraction section, a sliding window approach is applied to obtain features of a segment of the preprocessed signal. We use a sliding window and set its length is l, which divides the signal as shown in Fig. 5.

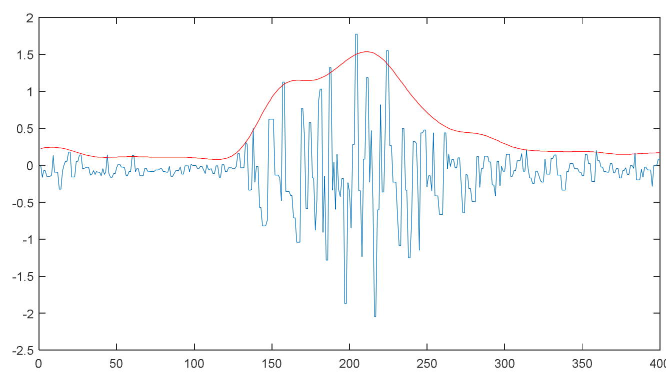

Fig. 4 The raw sEMG signal (blue line) and envelop (red line)

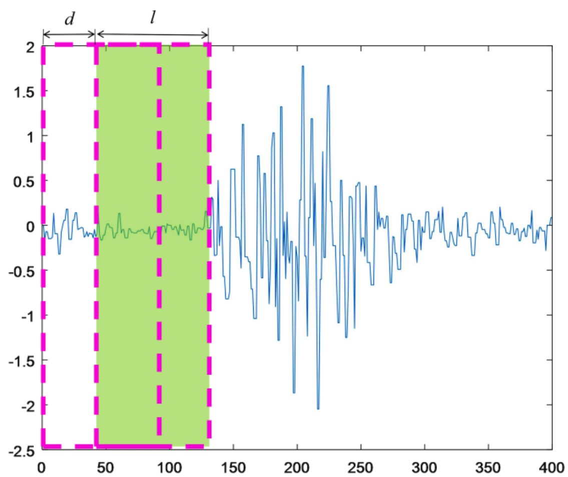

Fig. 5 The sliding window extracts feature on one channel

To find optimal solution of window length, the t-distributed stochastic neighbor embedding (t-SNE) is used to visualize how the feature vectors from each type of gesture are clustered in the feature space [6]. Data from one subject were selected and the results from t-SNE as shown in Fig. 6. We can realize that the distance between gestures of the same type decreases as the size of window increases in a certain range. However, the number of feature vector reduces and the size of feature vector increases when the length is above this range, which leads to the occurrence of overfitting. Therefore, the most reasonable size of window is set as $l = 40$ by analyzing the results.

During the process of extracting features, the window moves from the left end to the right end with a uniform fixed stride size, obtaining the features in each window. To make sure the model has high precision and short respond time, the stride size of two adjacent windows is set as 5ms, which corresponding to one point in the signal.

In our approach, the five time-domain features are extracted by a bag of functions, including mean absolute value (MAV), waveform length (WL), root mean square (RMS), slope sign change (SSC), and Hjorth parameter (HP), which includes complexity parameter, activity parameter, and mobility parameter.

MAV is one of the most widely used features in signal processing, represents the level of the muscle contraction. The calculation formula of MAV is (1):

$$M = \frac{1}{\mathrm{N}}\sum_{k=1}^{\mathrm{N}}|\mathrm{A}(k)| \tag{1}$$

Where M denotes mean absolute value, N denotes the sample size, k denotes indices of the samples and A(k) denotes the kth amplitude. The WL is the total length of the waveform curve and can be influenced by many factors, such as amplitude, frequency and total time. This feature represents the complexity of the sEMG signals and can be calculated as (2).

$$W = \sum_{k=2}^{\mathrm{N}}\left|\mathrm{A}(k)\text{-}\mathrm{A}(k\text{-}1)\right| \tag{2}$$

RMS is a common feature in process of signal to illustrate the mean power of the signal. We can obtain this value as R by using (3).

$$R = \sqrt{\frac{1}{\mathrm{N}}\sum_{k=1}^{\mathrm{N}}\mathrm{A}(k)^2} \tag{3}$$

SSC is used to describe the number of change in signal slope sign, which represents the frequency information of the signal. It is denoted as S and the formula for this feature is (4):

$$S = \sum_{k=2}^{\mathrm{N}\text{-}1}\left|\left(\mathrm{A}(k)\text{-}\mathrm{A}(k\text{-}1)\right)\times\left(\mathrm{A}(k)\text{-}\mathrm{A}(k\text{+}1)\right)\right| \tag{4}$$

HP is composed of three parameters: activity (A_{HP}) and mobility (M_{HP}) and complexity (C_{HP}). The formulas about these parameters are (5), (6), and (7):

$$A_{\mathrm{HP}}=\mathrm{VAR}\left(\mathrm{A}(k)\right)=\frac{1}{\mathrm{N\text{-}1}}\sum_{k\text{-}1}^{\mathrm{N}}\mathrm{A}(k)^2 \tag{5}$$

$$M_{\mathrm{HP}}=\sqrt{\frac{\mathrm{VAR}(\frac{d\mathrm{A}(k)}{dk})}{\mathrm{VAR}(\mathrm{A}(k))}} \tag{6}$$

$$C_{\mathrm{HP}}=\frac{\mathrm{Mobility}(\frac{d\mathrm{A}(k)}{dk})}{\mathrm{Mobility}(\mathrm{A}(k))} \tag{7}$$

The feature vector applied in our proposed model consists of two parts: the processed signals and five time domain features. Firstly, the matrix of processed signals in sliding window are denoted as $\mathbf{F}=\psi\left[abs[\mathbf{T}]\right]^{l\times 8}$, we take the row vectors of $\mathbf{F}$ end to end and obtain a vector $\mathbf{x}_i\in[0, 1]^{1\times 320}$, where i represents the ith point of the sliding window movement. Secondly, five features in time domain of signal are extracted and formed matrix $\mathbf{W}=[\mathbf{w}_1, \ldots, \mathbf{w}_8]\in\mathrm{R}^{7\times 8}$, where $\mathbf{w}_j=(M, W, R, S, A_{\mathrm{HP}}, M_{\mathrm{HP}}, C_{\mathrm{HP}})^{\mathrm{T}}\in\mathrm{R}^{7\times 1}$ and we take the matrix $\mathbf{W}$ into a vector $\mathbf{z}_i\in\mathrm{R}^{1\times 56}$ by concatenating rows, where i represents the ith point of the sliding window movement. Finally, feature vector $\mathbf{v}_i$ is generated by splicing $\mathbf{z}_i$ and $\mathbf{x}_i$ as row vector. The length of the $\mathbf{v}_i$ can be calculated as $|\mathbf{v}_i|=|\mathbf{x}_i|+|\mathbf{z}_i|=320+56=376$, which indicates that each feature vector has 376 feature to be used for classification. These steps are performed in each sliding window of iterative process.

D. *Classification*

In this section, we select SVM as the classifier of gesture recognition. The data used in this experiment are divided into training part and testing part. The training set includes standardized training raw data $\mathbf{T}=[\mathbf{T}_1, \ldots\mathbf{T}_8]^{\mathrm{N}\times 8}\in\left[\text{-}1, 1\right]^{\mathrm{N}\times 8}$ and label $\mathrm{y}_i\in\{1, 2, \ldots, 6\}$ that represents the hand gesture corresponding to signal. The training set of each subject is composed of 30 samples and these samples include 5 repetitions of hand gestures in Fig. 2. After preprocessing and extracting feature, we use the matrix $\mathbf{D}=\left[\mathbf{v}_i, \mathrm{y}_i\right]^{s\times 377}$ to train the SVM classifier, where s depicts the number of feature vector.

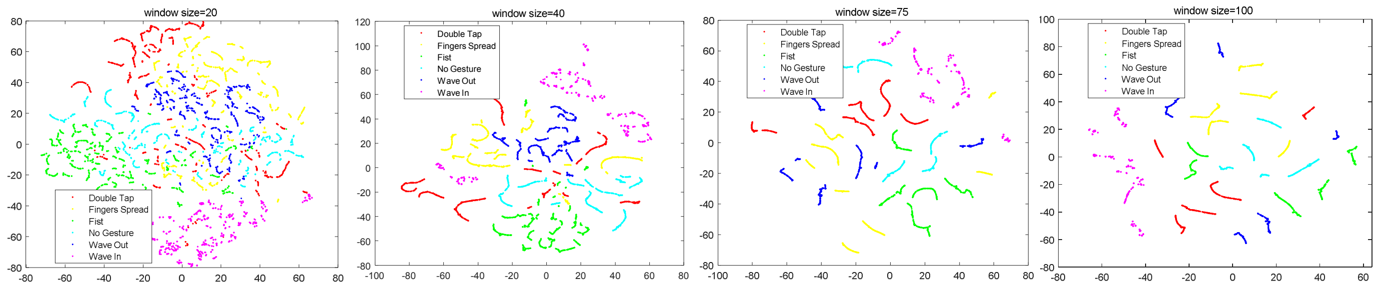

Fig. 6 The results from t-SNE in different sizes of sliding window

The testing set of each subject is formed by 150 samples, which includes 30 repetitions of five target gestures, and the signals detected in relaxed position are removed, because the goal of testing part is to recognize the five kinds of hand gestures that not include the relaxed state. The steps of data processing in testing part are same as training part, a sliding window with the same size is used to structure feature vector. The feature vector is put into the classifier as input and the trained SVM classifier returns a vector of labels based on the input feature vector, which each label corresponds to $\mathbf{v}_i$ of each sliding window observation in iterative process.

E. Post-processing

The purpose of this section is to analyze the label vector returned by the SVM classifier to get the predictive gesture of classifier. For this goal, an answering racer algorithm is applied to determine label of the test gesture [6], we define a vector $\mathbf{n}=[0, 0, 0, 0, 0]^{\mathrm{T}} \in \mathrm{R}^{5\times 1}$ at first and the labels returned by classifier are checked in the post-processing section. We record $\mathbf{n}_t := \mathbf{n}_t + 1$ when a label $y_i = t, t \in \{1, 2, 3, 4, 5\}$, where the t denotes the label of tested gesture. In order to improve the response speed while ensuring accuracy, the threshold h is set as reference to recognize gesture. In the post-processing, we predict the label of gesture is t as soon as the element $\mathbf{n}_t$ is equal to or greater than h. If the element that meets the above conditions does not exist, we assign the label 6 to this gesture, which means "No action". For determining the optimal solution of threshold h, we experiment the different value of h and set $h = 30$, which has the highest accuracy.

IV. Discussion

For evaluating the accuracy, we use the confusion matrix to describe results (as shown in Fig. 7). The overall accuracy of the proposed model is 97.8% and the highest sensitivity (99.4%) appears in the recognition of "Wave Out". The lowest sensitivity is 95.8%, which appears in the recognition of "Double Tap". Meanwhile, the gesture "Wave In" has the highest precision (99.4%) and "Fist" has the lowest precision with 95.7%. The most common mistake of this model is to assign the label of "Fist" to the gesture "Double Tap". Meanwhile, we plot a figure to depict accuracy value of all subjects (as shown in Fig. 8). We can realize that the accuracy of all subjects except two subjects (NO.3 and NO.7) is greater than 95% and there are three subjects have a perfect accuracy (100%).

For the purpose of evaluating the quality of real time, the time for recognition is recorded by our model, the real-time response means that the response time required to complete the identification must be within a small range, which is typically 300ms after the action completion. The response time in this experiment starts from the beginning of the action and the comparison between action time and response time is described as shown in Fig. 9. The response time is around 500ms while active time of gestures is more than 1200ms. Obviously, the response time is much shorter than active time, which means that the recognition is completed before action is finished.

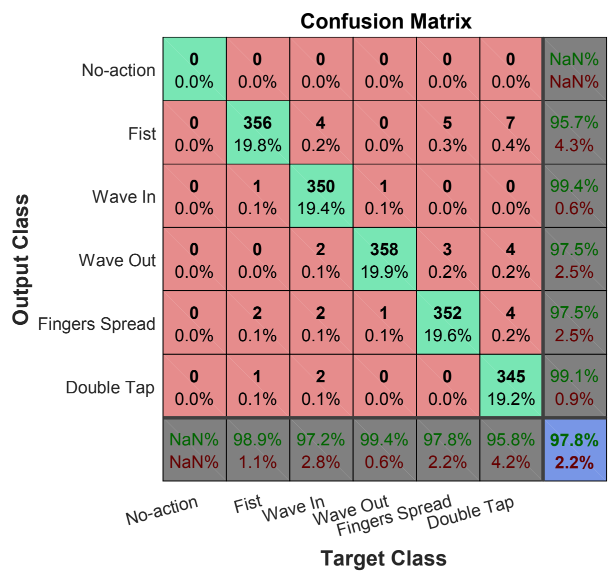

Fig. 7 Confusion matrix of the experimental result

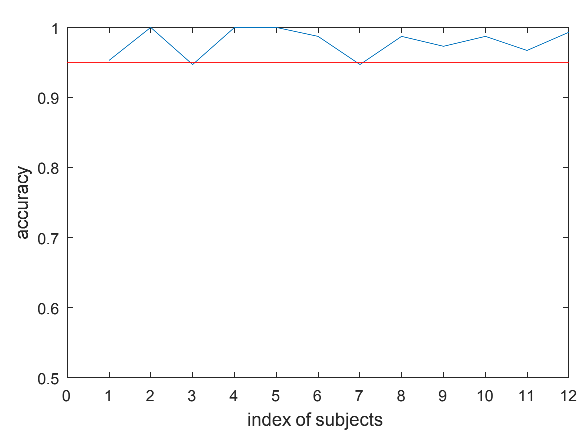

Fig. 8 The accuracy of each subject (red line denotes 0.95)

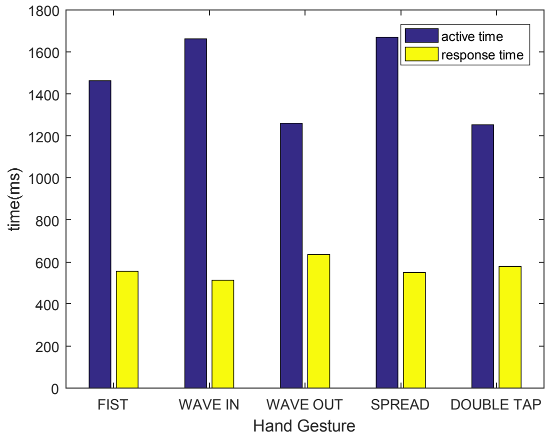

Fig. 9 The comparison between active time and response time

The proposed model is compared with other model designed by other paper. The comparison of accuracy and real-time response is shown in TABLE I. We can note that our model has the highest accuracy while meets the requirement of real-time response.

TABLE I
THE PROPOSED MODEL COMPARED OTHER MODELS

Model	Accuracy (%)	Real time
Proposed model	97.8	√
Model using k-NN with DTW [14]	89.5	×
Model only using SVM [6]	93.99	×
Model using Random Forest [6]	89.92	×
General model [14]	53.7	×

V. CONCLUSION

A real-time recognition model for hand gestures based on SVM and sEMG signals is proposed in this paper. The Myo armband is applied to acquire the signal generated by hand movement and the feature vector is extracted by using a sliding window. A classifier based on support vector machine is trained, we count the labels returned by classifier and assign the label whose times is equal to or greater than the threshold to the tested signal. The overall accuracy of our model can reach 97.8%, which is greater than other model. Meanwhile, the proposed model can recognize gesture before action is finished, which illustrates that our model has the advantage of real-time response.

In future work, the new features and the new algorithm models for feature extraction and classification will be experimented to find a model that performs better in real-time response.

REFERENCES

[1] B. He, S. Wang, and Y. Liu, "Underactuated robotics: A review," *International Journal of Advanced Robotic Systems*, vol. 16, no. 4, Jul 16 2019.
[2] X. Zhang, X. Chen, Y. Li, V. Lantz, K. Wang, and J. Yang, "A Framework for Hand Gesture Recognition Based on Accelerometer and EMG Sensors," *IEEE Transactions on Systems Man and Cybernetics Part a-Systems and Humans*, vol. 41, no. 6, pp. 1064-1076, Nov 2011.
[3] A. Ferreira, W. Celeste, F. Cheein, T. Bastos, M. Sarcinelli, and R. Carelli, "Human-machine interfaces based on EMG and EEG applied to robotic systems," *Journal of Neuroengineering and Rehabilitation*, vol. 5, Mar 26 2008.
[4] H. Wu et al, "Fabric-based self-powered noncontact smart gloves for gesture recognition," *Journal of Materials Chemistry A*, vol. 6, no. 41, pp. 20277-20288, Nov 7 2018.
[5] W. Lu, Z. Tong, and J. Chu, "Dynamic Hand Gesture Recognition With Leap Motion Controller," *IEEE Signal Processing Letters*, vol. 23, no. 9, pp. 1188-1192, Sep 2016.
[6] Z. Zhang, K. Yang, J. Qian, and L. Zhang, "Real-Time Surface EMG Pattern Recognition for Hand Gestures Based on an Artificial Neural Network," *Sensors*, vol. 19, no. 14, Jul 12 2019.
[7] M. Benalcazar, J. Zea, A. Jaramillo, C. Anchundia, P. Zambrano, and M. Segura, "Real-Time Hand Gesture Recognition Based on Artificial Feed-Forward Neural Networks and EMG," *2018 26th European Signal Processing Conference (Eusipco)*, pp. 1492-1496, 2018.
[8] 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, Jun 15 2012.
[9] H. Al-Angari, G. Kanitz, S. Tarantino, and C. Cipriani, "Distance and mutual information methods for EMG feature and channel subset selection for classification of hand movements," *Biomedical Signal Processing and Control*, vol. 27, pp. 24-31, May 2016.
[10] A. Phinyomark, C. Limsakul, and P. Phukpattaranont, "Application of Wavelet Analysis in EMG Feature Extraction for Pattern Classification," *Measurement Science Review*, vol. 11, no. 2, pp. 45-52, 2011.
[11] J. Singha and R. Laskar, "ANN-Based Hand Gesture Recognition Using Self co-articulated Set of Features," *IETE Journal of Research*, vol. 61, no. 6, pp. 597-608, Nov 2 2015.
[12] P. Molchanov, S. Gupta, K. Kim, and J. Kautz, "Hand Gesture Recognition with 3D Convolutional Neural Networks," *2015 IEEE Conference on Computer Vision and Pattern Recognition Workshops (Cvprw)*, 2015.
[13] A. Joshi, C. Monnier, M. Betke, and S. Sclaroff, "Comparing random forest approaches to segmenting and classifying gestures," *Image and Vision Computing*, vol. 58, pp. 86-95, Feb 2017.
[14] M. Benalcazar et al, "Real-Time Hand Gesture Recognition Using the Myo Armband and Muscle Activity Detection," *2017 IEEE Second Ecuador Technical Chapters Meeting (Etcm)*, 2017.
[15] D. Ghosh and S. Ari, "Static Hand Gesture Recognition using Mixture of Features and SVM Classifier," *2015 Fifth International Conference on Communication Systems and Network Technologies (Csnt2015)*, pp. 1094-1099, 2015.