Surface Electromyography Feature Extraction Via Convolutional Neural Network

Hongfeng Chen · Yue Zhang · Gongfa Li · Yinfeng Fang* · Honghai Liu

Received: date / Accepted: date

Abstract Although a large number of surface electromyography (sEMG) features have been proposed to improve hand gesture recognition accuracy, it is still hard to achieve acceptable performance in inter-session and inter-subject tests. To promote the application of sEMG-based human machine interaction, a convolutional neural network based feature extraction approach (CNNFeat) is proposed to improve hand gesture recognition accuracy. A sEMG database is recorded from eight subjects while performing ten hand gestures. Three classic classifiers, including Linear Discriminant Analysis (LDA), Support Vector Machine (SVM) and K Nearest Neighbor (KNN), are employed to compare the CNNFeat with 25 traditional features. This work concentrates on the analysis of CNNFeat through accuracy, safety index and repeatability index. The experimental results show that CNNFeat outperforms all the tested traditional features in inter-subject test and is listed as the best three features in inter-session test. Besides, it is also found that combining CNNFeat with traditional features can further improve the accuracy by 4.35%, 3.62% and 4.7% for SVM,

Hongfeng Chen · Yue Zhang
College of Computer Science and Technology, Zhejiang University of Technology, Hangzhou 310023, China
E-mail: {chenhf1026, zhangyuemessi}@163.com

Gongfa Li
Key Laboratory of Metallurgical Equipment and Control Technology of Ministry of Education, Institute of Precision Manufacturing, 947 Heping Avenue, Wuhan, 430081, China
E-mail: (ligongfa@wust.edu.cn)

Yinfeng Fang
Telecommunication College, Hangzhou Dianzi University, Hangzhou, 310018, China
E-mail: yinfeng.fang@hdu.edu.cn

Honghai Liu
Intelligent Systems and Biomedical Robotics Group, School of Computing, University of Portsmouth, Winston Churchill Avenue, Portsmouth, Hampshire, PO1 2UP
E-mail: honghai.liu@icloud.com

LDA and KNN, respectively. Additionally, this work also demonstrates that CNNFeat can be potentially enhanced with more data for model training.

Keywords Surface EMG · CNN · Traditional Classifiers · Feature Combination · Hand Motion

1 Introduction

Surface electromyography (sEMG) is a bioelectrical signal that characterizes motor unit impulses of skeletal muscle fibers during excitation and contraction [1]. It is closely related to the state of muscle activity and has been widely used in auxiliary diagnostic research [2] and human-machine interaction [3,4]. The sEMG signals captured from the forearm contain sufficient information about hand movements and the decoding of sEMG signals can be applied to the control of external devices. However, sEMG is sensitive to a variety of interference due to its inherent characteristics. Although sEMG-based hand motion recognition accuracy reaches up to more than 90% in well controlled experimental setup, it is hard to achieve the similar result in the practical environment, where electrode shift occurs during different sessions (*i.e.* inter-session scenario) and muscular structure varies among different subjects (*i.e.* inter-subject scenario).

To fill the gap, machine learning technology has been widely applied in myoelectric control system [5]. A typical machine learning system includes three basic components: data preprocessing, feature extraction and classification. Among them feature extraction plays a critical role to achieve high classification accuracy. Till now, a large number of sEMG features have been investigated for hand gesture classification. They can generally be divided into four types: time domain (TD) feature, spectral domain or frequency domain (FD) feature, time-scale or time-frequency domain (TFD) feature, and parametric model analysis based feature. Phinyomark *et al.* [6] compared 37 types of sEMG features for hand gesture classification. Among them sample entropy (SampEn) achieved the best performance. In addition, the experimental result also demonstrated that the best three sEMG features could receive recognition accuracy over 80%, while the accuracy of the system employing the worst three features was below 20% [7]. Moreover, it was also found that most of the TD features were redundant, and FD features could not achieve good performance in their study.

In recent years, with the rapid development of computer science, Convolutional Neural Network (CNN) has outperformed most traditional methods in many fields, such as object detection [8], speech recognition and natural language processing [9], *etc.* Several studies have also demonstrated the advantage of CNN-based hand gesture recognition. Manfredo *et al.* [10] discovered that a simple architecture CNN (four convolutional layers, one fully connected layer and a softmax function) could reach the accuracy of 66.59%±6.4% on dataset one of the NinaPro database, which was higher than the average result (62.06%±6.07%) by the classical classification methods. Park *et al.* [11]

proposed a EMG signal decoding approach for motion intent prediction using deep feature learning method. It enhanced system robustness, receiving the inter-subject classification accuracy over 90%. Geng *et al.* [12] applied a deep CNN to high-density EMG (128 channels) based gesture recognition, and achieved an accuracy of 89.3% with only one single frame as input and 99.0% with over 40 frames as input.

In these works, they all use deep learning methods in the myoelectric, but under different experimental setup, such as different data sets (even for the same one, for NinaPro dataset, different parts are chosen), different sizes of network input, different standards for training and testing. In this paper, we focus on analyzing the performance of EMG feature extracted by CNN. Traditionally, the entire CNN is designed as a classifier, in which a fully connected layer is applied at the end to generate the class label by using a softmax function. This end-to-end CNN model is used to solve a global optimization problem. In this paper, CNN is constructed for sEMG features extraction only, while traditional classifiers are further applied to classify hand gestures. Related works from Niu *et al.* and Bluche *et al.* have demonstrated the advantage of combining CNN with traditional method in hand-written recognition [13, 14].

The rest of the paper is organised as follows: Section 2 presents the materials for experimental setup, data collection protocol, and the designed CNN. Section 3 demonstrates the experimental results with further discussion in Section 4. Section 5 concludes the paper.

2 MATERIAL AND METHODS

2.1 Subjects

The experiment is carried out in accordance with the Declaration of Helsinki. Eight subjects (7 males, 1 female, age: 25±5, height: 175±10cm, weight: 65±10kg) volunteered in the experiment. They are all right-handed and have no previous history of neuropathies or traumas to the upper limbs. Besides, the acquisition of sEMG data is under the relevant guidelines during the entire experiment.

2.2 Apparatus

A multi-channel EMG acquisition system (Elonxi Ltd, UK) designed by our research group is used to record sEMG signals. It mainly includes an EMG acquisition device and an electrode cuff. As demonstrated in Fig. 1, a customised sEMG electrode cuff is fitted with 18 electrodes in Zig configuration [15]. Each electrode is about 12mm diameters. Besides, the vertical and horizontal distances between two electrodes are 25mm and 30mm, respectively. Sixteen channels are recorded synchronously. Besides, one reference electrode is marked as 17 and one bias electrode is marked as 18 in Fig. 1.

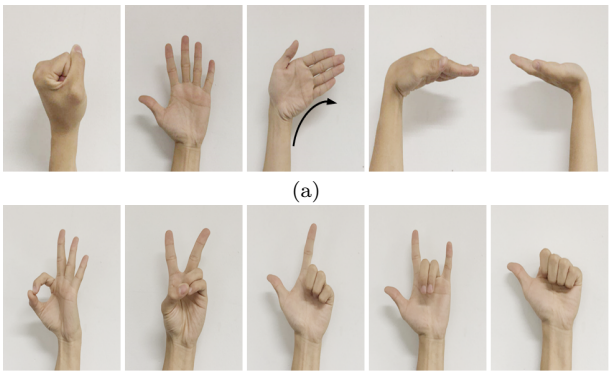
Fig. 1: The electrode cuff for sEMG signal recording. Sixteen monopolar channels record the differential voltage between electrodes 1-16 and electrode 17. Electrode 18 provides the bias voltage to eliminate DC offset.

The sEMG signals are sampled at 1kHz sampling frequency. Two Sallen-Key filters are utilized to make up a band pass filter, which contributes to remove low frequency motion artefact as well as the high frequecy white noise out of the band of sEMG signal (20-500Hz). The entire device is powered by a 3.3V rechargeable lithium battery. It can somewhat reduce the influence of the 50 Hz powerline noise from the cable, but it is not able to remove the noise caused by capacitive coupling. Therefore, we use a notch filter with center frequency at 50Hz to suppress power line noise that is permeated into sEMG signals through capacitive coupling. The noise of sEMG data in each channel is less than 1μV. sEMG data obtained by the acquisition module is packaged and transmitted to the computer through two Bluetooth modules. A software is designed to display and record multi-channel sEMG data. Besides, off-line analysis is provided for quick screening of the recorded data.

2.3 Acquisition Protocol

Before the experiment, every subject is familiar with the entire experimental procedure. They wear the electrode cuff and practice the intended hand gestures. The selection of hand gestures in this experiment is based on the NinaPro database [16] and the CSL-HDEMG database [17]. Two groups of gestures are involved. The first group: (1) hand close (hc), (2) hand open (ho), (3) wrist radial deviation (wrd), (4) wrist extension (we) and (5) wrist flexion (wf), as shown in Fig. 2 (a). The second group includes five finger-related hand gestures: (1) tip pinch, (2) flexion of ring and little finger, thumb flexed over ring and little finger, (3) flexion of middle, (4) ring and little finger, middle and ring flexion, and (5) thump up, as shown in Fig. 2 (b).

During data collection, each subject wears the electrode cuff on the left forearm and sits comfortably on the chair, as depicted in Fig. 3. The gestures are performed in sequence, and each gesture maintains for at least 12s. In order to avoid muscle fatigue, subjects are asked to rest for at least 10s between two gestures. The first session is implemented to collect sEMG signals while



(a)

(b)

Fig. 2: Two groups of gestures are used in this experiment. (a) The first five gestures. (b) The second group includes five finger-related hand gestures.

performing ten hand gestures within a short period. Large body movements are not required during this period. Half an hour later, the second session is held in the same way. During the period of half an hour, subjects can move around the office to have a rest. The third session is conducted by the same method after half an hour later of the succeeding session. After a three-day interval, another three sessions are conducted. It is worth pointing out that no predefined contraction force or elbow angle are applied in the experiment to mimic the real application of a sEMG-based human machine interface (HMI), although it is well known that muscular contract force and arm position can influence the robustness of pattern recognition system [18].

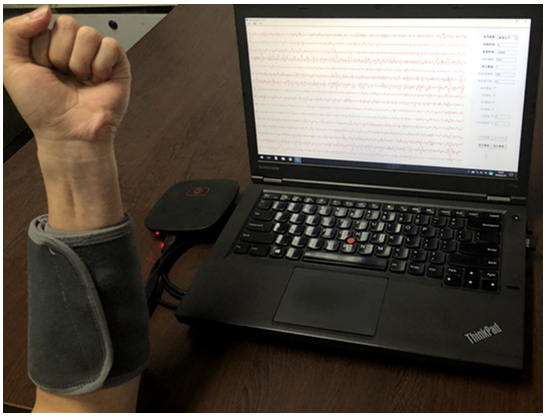


Fig. 3: The scenes for sEMG data recording, in which the electrode sleeve was worn on the left forearm.

2.4 Data Preprocessing

In order to exclude the transient state between two gestures and avoid diverse response times, only the latter 10s signals (each gesture remains for about 12s) are labeled for classification. Consequently, ten seconds steady sEMG signals containing 10000 frames are extracted for further analysis.

The whole database (Elonxi DB) is published in the link[1], and it is organised in mat format with the architecture as described in Table 1. Besides, there are two formats for these files: *aaa-ccc*.mat and *aaa-bbb-ccc*.mat. They indicate the raw and preprocessed signals, respectively, where *aaa* is the subject ID, *bbb* is the gesture ID, and *ccc* is the session ID. For example, 001-002.mat contains the raw sEMG data captured from subject 1 in the second session, and 003-004-005.mat contains the steady-state sEMG data of gesture 4, captured from subject 3 in the fifth session. In addition, although intra-session test (same subject, same data distribution) is out of the scope of the present study, a basic test is carried out by SVM, LDA[19] and KNN under traditional EMG feature set (modied mean absolute value (MAV), waveform length (WL) and zero crossing (ZC)), and receives the accuracy at 98.51%, 99.12% and 99.64%, respectively.

Table 1: The format of EMG database.

aaa-ccc.mat

Name	Type	Description
subject	scalar	The subject ID
group	scalar	Gesture group ID
data	110000×16 matrix	Raw sEMG signals of five gestures

aaa-bbb-ccc.mat

Name	Type	Description
subject	scalar	The subject ID
gesture	scalar	The gesture ID
trial	scalar	The trail ID
data	10000×16 matrix	sEMG data for one gesture

2.5 Convolutional Neural Network

Convolutional Neural Network has become one of the research hotspots in various scientific fields, especially in the field of pattern classification. It is capable of processing original signal without the input of human-crafted features. In the current study, CNN is employed to extract the EMG characteristics from

[1] https://github.com/taowucheng1026/CNN-LDA-SVM-KNN-for-EMG

raw EMG signals. In general, the basic CNN structure mainly consists of two layers. One is the feature extraction layer. It extracts the local feature, and each input neuron is connected to the local accepted field of the previous layer. Once the features are extracted, the positional relationship with other features is also determined. The second is the feature mapping layer. Each computing layer of the network consists of several feature maps, and each feature is mapped to a plane, in which all neurons have equal weights (it greatly reduces the parameter training and the risk of overfitting). In the feature mapping structure, the sigmoid function is generally used as the activation function of convolution network, which makes the feature mapping have a good advantage: shifting invariance. In general, each feature extraction layer in CNN is closely followed by a computational layer (feature mapping) for local average and secondary extraction. This structure makes the network have high distortion tolerance to input samples in pattern recognition.

A CNN framework is designed for the extraction of CNNFeat in this paper. A total of 300 continuous frames are used as the network input, and each frame refers to an one-dimensional array (1×16). In other words, CNN input is a 300×16 matrix. The number 16 indicates 16 channels of sEMG signal and 300 indicates 300 frames. This study takes 300 frames as the input because of the following two reasons. On the one hand, if the number of frames is too small, it may be unable to express the content in a short period of time. On the other hand, it corresponds to the traditional feature extraction method, which makes the comparison more convincing (in the traditional method, a sliding window with 300ms length is applied), as shown in Fig. 4.

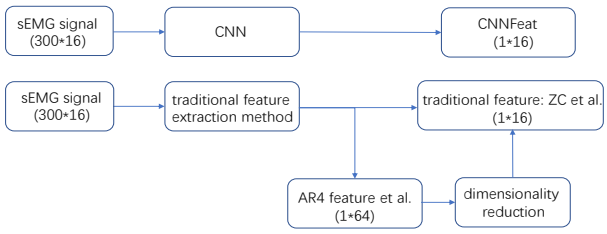


Fig. 4: The feature extraction procedure by CNN and traditional methods.

The entire network contains 11 layers, as shown in Fig. 5. The first two layers are convolutional layers, and 64 convolution kernels are applied for each layer. To choose proper kernel size, several combinations are tested. As illustrated in Fig. 6, the results indicate that the combination of 5×5 and 3×3 can achieve better results in two scenarios. Therefore, two kernels with the size of 5×5 and 3×3, and the stride of 1 and the padding of 1 are taken in the first two convolutional layers. Closely following, two local connection layers with 64 non-overlapping convolution kernels (1×1) are used. These layers can provide wealth of nonlinear capabilities. The next six hidden layers are fully connected and consists of 512, 256, 128, 64, 32 and 16 units, respectively. The network

ends with a softmax function and a 5-way output layer. In order to keep the same dimension for CNNFeat and traditional features, the output of the last fully connected layer is used as the CNNFeat for evaluation. Besides, Dropout with a probability of 0.5 [20] is applied after the third and fourth connection layers, and the first full connection layer to prevent the model from overfitting. In addition, batch normalization [21] and ReLU non-linearity [22] operations are also added in the input layer and behind each hidden layer.

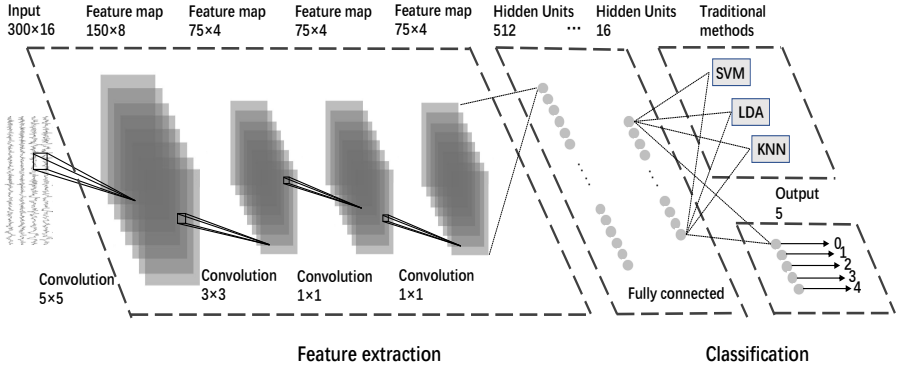


Fig. 5: The architecture of the convolutional neural network, and its modified version via replacing the last layer by traditional classification algorithms.

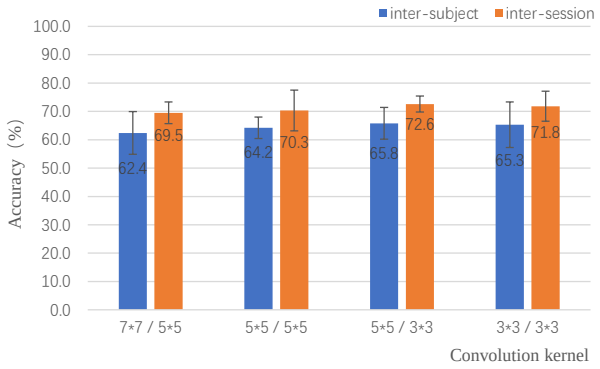


Fig. 6: The effect of different convolution kernel size on gesture recognition.

The last layer of the current trainable classifier is a fully connected layer with a softmax function, which is the probability estimation for the input raw sEMG signal. It is hard to interpret the features extracted by each inter-layer from a human's point of view, but it is practically possible to extract these

features as the input for other non neural network based classifiers. Therefore, we use SVM, LDA and KNN classifiers to take the place of the last fully connected layer to obtain a better classification effect. As shown in Fig. 5, the original sEMG image is sent to the input layer. After four convolution operations, we get a lot of local features. Then, after marking the second-to-last layer, sixteen units are obtained by passing through the six fully connected layers, and the result is sent to the output layer. The neural network is trained until the system is convergent, and then the previously marked 16 features are extracted and sent to three classifiers. Once the classifiers have been trained well, identification task will be performed on the testing set. It will make new decisions with these automatically learned features.

This work implemented the deep learning framework based on Tensorflow [23]. It is an open source framework with good flexibility and scalability, which supports distributed computing of heterogeneous devices. In the training procedure, stochastic gradient descent (SGD) method is applied [24]. The batch size is set to 100 with the 40000 epoch (if the network has converged before this value, it can stop ahead of schedule). In addition, for the multi-class SVM classifier, the penalty coefficient c is set to 0.9 (this value is verified by the experiment).

2.6 CNNFeat Evaluation and Qualitative Analysis

To compare with CNN-based feature extraction, 25 traditional features are tested in this work, including TD features, FD features and parametric model analysis based features. All of these features have been used in the analysis of sEMG signals. Considering the proper comparison with CNNFeat obtained from each 300×16 input, a sliding window with 300ms length and 50ms increment is applied to calculate the sEMG features.

In order to demonstrate the impact of CNNFeat for our dataset, we show how to separate these clusters from the original high-dimensional space. The safety index is chosen to measure the separateness between two clusters, which is defined as

$$d_{ij} = \frac{max\{\sigma_i\}}{\|C_i - C_j\|} \tag{1}$$

where C_i and C_j denote two clusters, and σ_i is the standard deviation of C_i. Therefore, d_{ij} is a ratio of the maximal standard deviation of cluster C_i and the Euclidean distance between two clusters. And obviously, a small value of d_{ij} indicates that most elements in cluster C_i are far away from cluster C_j.

Besides, repeatability index (RI) is also considered to measure the repeatability of model among different trials. The RI is calculated as one-half the average Mahalanobis distance [25] between the feature vector centroid for a training ($\mu_{T_{rj}}$) and testing ($\mu_{T_{eij}}$) trial. It can be obtained from

$$RI = \frac{1}{M}\sum_{j=1}^{M}\left(\frac{1}{N}\sum_{i=1}^{N}\frac{1}{2}\sqrt{\left(\mu_{T_{rj}} - \mu_{T_{eij}}\right)^T C_{T_{rj}}^{-1}\left(\mu_{T_{rj}} - \mu_{T_{eij}}\right)}\right) \tag{2}$$

where N denotes testing trial, M denotes class, $C_{T_{rj}}$ is the covariance of the training data for class j.

This work carries out four types of evaluation strategy to demonstrate the advantages of CNNFest, as listed below. Moreover, NinaPro database is also taken to verify the effectiveness of our method.

- *Inter-subject evaluation* refers to testing a classifier by a single subject's data, and the remaining (seven subjects' data) as the training set [17]. Depending on the data of each subject, eight groups of experiments are conducted. Averaged hand gesture classification accuracy is provided for comparison, where 26 types of single sEMG feature, including CNNFeat, are evaluated.
- *Inter-session evaluation* refers to using the data of the previous three sessions for training and the data that collected three days later for testing. Although the training and testing data are collected from the same subject, it can exist an enormous difference due to electrode shift, *etc.* Averaged hand gesture classification accuracy across eight subjects is provided for comparison, where 26 sEMG features are evaluated.
- *Feature combination evaluation* refers to evaluating the combinations of traditional features (including feature combinations) with CNNFeat in inter-session scenario, and only the first group of gestures is considered in this test. The best traditional features are selected according to the achieved classification accuracy in the mentioned inter-session and inter-subject evaluation.
- *Evaluation of training dataset for CNNFeat extraction* refers to choosing one subject's data as testing set, and enlarging the training set from the remaining one subject to seven subjects steadily. It is designed to check if the performance of CNNFeat can be further enhanced with more sEMG data.

3 Results

3.1 Inter-subject evaluation

The experimental results demonstrate that the CNNFeat outperforms 25 traditional features in inter-subject hand gesture recognition, as listed in Table 2. It obtains an accuracy of 65.8%±5.6% and 38.1%±10.3% by using CNN for basic and finger gesture classifications. In addition, for classifying five basic gestures, SVM, LDA and KNN achieve the accuracy of 68.7%±4.7%, 67.4%±5.6%, 68.5%±4.1%, respectively. These accuracy results are 5.3%, 11.3% and 10.2% higher than the second successful traditional features. They are also 2.9%, 1.6% and 2.7% higher than the accuracy of CNN for classification. While for five finger gestures, three classifiers obtain the accuracy of 40.3%±10.3%, 39.9%±10.1%, 41.0%±9.9%, respectively, which are 1.3%, 9.2% and 0.1% higher than the second successful traditional features. Similarly, they are 2.2%,

1.8% and 1.9% higher than the result of CNN. Besides, it is also found that CNNFeat is less sensitive to the use of classifiers. The maximum deviation of the accuracy among different classifiers is less than 2%, for both basic and finger gesture classifications.

The best three features in each test are highlighted in Table 2, which will be further evaluated in feature combination evaluation. In total, six traditional features are selected: AAC, LOG, WL, AR4, DASDV and MNP.

3.2 Inter-session evaluation

As shown in Table 3, similar to the result of inter-subject test, the classification effect of CNNFeat is better than most of the traditional features, although it is not always the best. For example, in the classification of the first group of gestures, SVM classifier with AR4 feature achieves the accuracy of 74.4%±10.8%, which is 0.6% higher than the result of CNNFeat. However, it still belongs to the best three features in each test. Besides, it is also noted that CNNFeat can provide more stable accuracies across different classifiers. The maximum accuracy deviation is less than 2.1%. In contrast, the accuracy of AR4 reaches to 74.4% for SVM, but it falls to 62.2% for KNN. Moreover, six selected features (AAC, LOG, WL, AR4, DASDV and MNP) obtained in inter-subject evaluation can still attain satisfactory results.

Fig. 7: The safety matrix of CNNFeat which is calculated from data of subject 1. d_{ij} is the ratio of the maximal standard deviation of cluster C_i and the Euclidean distance between the two clusters.

In addition, each pair of clusters (the first five gestures, as shown in Fig. 2 (a)) that used CNNFeat is computed, and a safety matrix $D = \{d_{ij}\}$ is obtained, as illustrated in Fig. 7. The results indicate that the highest value

Table 2: A comparison of 25 traditional sEMG features with CNN-based feature in inter-subject evaluation.

Feature[1]	Basic gestures			Finger gestures		
	SVM	LDA	KNN	SVM	LDA	KNN
MAV	59.8±8.8	52.3±6.9	48.3±6.0	34.9±10.6	28.6±7.4	40.8±9.0
ZC	54.7±13.3	46.5±14.5	34.8±6.5	24.8±3.5	25.0±4.9	24.4±3.9
SSC	55.6±11.2	45.9±10.1	32.8±5.9	25.3±4.1	24.6±5.5	24.3±4.7
AAC	57.1±7.0	55.3±3.3	51.4±4.1	38.1±13.2	**30.7±11.6**	**41.0±7.1**
IAV	57.3±6.1	52.3±6.9	48.2±6.0	35.5±10.9	28.6±7.4	40.7±9.0
IEMG	57.2±6.4	52.3±6.9	48.2±6.0	34.6±10.1	28.6±7.4	40.7±9.0
LOG	58.3±6.9	51.7±7.6	49.7±5.7	36.2±11.6	**30.6±6.4**	39.8±7.5
RMS	55.3±5.1	52.1±6.5	49.1±7.2	34.1±9.4	28.2±8.0	40.7±9.1
RPcoes	53.7±3.7	50.9±4.4	50.5±5.5	32.4±8.8	29.8±6.3	33.6±9.7
TM3	27.1±2.9	40.9±6.3	40.5±8.6	28.1±5.0	22.1±6.6	30.1±6.0
VAR	56.0±4.0	52.3±6.9	48.2±6.0	35.5±8.1	28.6±7.4	40.7±9.0
WL	**63.4±7.7**	55.7±5.2	**58.4±6.4**	**39.0±13.6**	30.2±12.3	40.0±8.4
AR4	53.9±13.9	**55.9±11.6**	34.4±3.7	25.7±3.2	26.0±4.2	23.5±2.2
DASDV	**60.7±8.4**	**56.1±4.3**	**52.3±3.4**	35.0±9.9	30.4±11.7	**40.8±7.1**
MYOP	37.8±6.9	37.9±7.6	34.9±1.9	31.3±1.4	32.3±4.4	32.3±3.2
SSI	43.6±10.0	50.4±6.2	42.3±10.4	29.8±5.2	24.1±9.9	40.5±8.2
V_order	57.9±5.6	51.9±6.4	49.6±8.2	37.7±9.5	27.4±8.5	40.6±9.4
WAMP	55.8±2.7	54.5±3.6	50.2±6.1	29.7±6.5	28.1±8.6	35.3±9.8
SampEn	53.2±4.0	50.0±2.8	49.0±6.8	26.9±2.9	26.5±2.3	27.5±3.5
MNF	27.4±1.5	27.5±2.6	26.4±1.7	21.2±0.5	20.6±1.1	21.7±1.1
MDF	57.0±4.3	54.6±4.6	49.6±6.6	36.2±8.6	30.0±7.9	38.7±10.7
PKF	55.1±2.5	47.9±7.7	47.9±4.9	34.2±4.8	26.5±9.6	35.4±9.1
MNP	58.9±3.5	54.6±4.6	49.6±6.5	**38.6±9.8**	30.0±7.9	38.7±10.7
TTP	58.5±4.5	54.6±4.6	49.6±6.5	38.3±9.1	30.0±8.0	38.5±10.4
FR	23.3±1.9	27.3±2.9	21.1±0.4	22.2±1.6	22.3±1.7	20.3±0.4
CNNFeat	**68.7±4.7**	**67.4±5.6**	**68.6±4.1**	**40.3±10.3**	**39.9±10.1**	**41.1±9.9**

Abbreviations: Modified mean absolute value(MAV), Zero crossing(ZC), Slope sign change(SSC), Average amplitude change(AAC), Integral absolute value(IAV), Integrated EMG(IEMG), Log detector(LOG), Root mean square(RMS), Absolute temporal moment(TM3), Variance(VAR), Waveform length(WL), Auto-regressive coefficients(AR4), Difference absolute standard deviation value(DASDV), Myopulse percentage rate(MYOP), Simple square integral(SSI), Willison amplitude(WAMP), Sample entropy(SampEn), Mean frequency(MNF), Median frequency(MDF), Peak frequency(PKF), Mean power(MNP), Total power(TTP), Frequency ratio(FR), Convolutional neural network(CNN).

Table 3: A comparison of 25 traditional sEMG features with CNN-based feature in inter-session evaluation.

Feature	Basic gestures			Finger gestures		
	SVM	LDA	KNN	SVM	LDA	KNN
MAV	66.6±7.6	62.1±13.9	65.2±8.5	35.7±7.4	45.3±8.9	35.2±7.2
ZC	66.7±9.9	67.0±7.0	51.6±5.9	31.5±5.4	39.7±9.8	34.3±8.7
SSC	65.6±8.2	69.6±4.4	50.4±9.2	36.4±4.2	46.6±1.4	36.3±5.8
AAC	68.0±8.7	65.9±8.6	68.3±5.8	41.2±9.0	37.5±2.9	**47.9±8.8**
IAV	47.9±5.4	41.7±5.3	45.2±10.9	22.4±3.6	21.0±8.4	22.2±9.1
IEMG	66.8±7.9	62.1±10.9	65.2±8.5	36.6±4.8	45.3±8.8	35.2±7.7
LOG	65.4±7.8	64.2±8.9	64.2±8.8	41.6±8.9	**50.0±6.2**	42.1±6.4
RMS	67.5±5.3	66.6±9.9	65.5±7.8	46.2±3.0	46.6±6.6	28.0±3.3
RPcoes	67.5±8.0	61.0±12.8	59.9±7.5	37.1±9.7	39.0±8.4	32.6±4.5
TM3	57.3±7.8	53.2±5.9	58.3±6.9	35.9±5.3	34.2±6.8	32.8±9.5
VAR	68.7±8.8	62.0±8.0	65.2±8.5	34.7±5.9	45.3±6.3	35.2±4.8
WL	65.7±3.1	**72.2±4.0**	**71.8±6.1**	40.1±3.9	33.6±9.9	43.0±5.1
AR4	**74.4±10.8**	**74.1±2.1**	62.2±8.2	**48.4±7.8**	**48.8±10.7**	38.8±2.3
DASDV	**71.0±3.2**	69.1±5.8	**68.9±8.9**	42.5±1.4	29.1±8.0	**47.2±12.3**
MYOP	47.0±2.4	40.7±10.7	34.2±8.6	37.1±3.9	35.0±7.2	30.7±7.2
SSI	46.1±6.1	68.8±8.5	64.2±9.7	35.2±9.8	33.7±6.9	23.4±6.7
V_order	67.1±5.1	69.0±8.4	65.6±7.2	43.6±0.9	46.3±5.0	27.8±10.3
WAMP	57.0±4.6	56.4±7.5	56.1±6.3	39.2±3.4	38.5±5.1	39.4±9.2
SampEn	56.6±7.7	51.8±7.1	44.4±7.2	40.0±6.8	41.3±10.7	39.2±3.1
MNF	30.5±2.8	30.4±3.5	30.3±3.8	22.2±8.0	21.0±8.0	23.1±8.6
MDF	70.1±6.0	69.5±12.0	55.0±12.3	34.0±6.0	35.6±10.3	35.0±10.3
PKF	65.9±10.9	63.6±7.7	59.7±6.3	47.3±8.6	46.9±6.7	37.8±7.0
MNP	69.0±5.2	65.8±9.9	66.9±7.5	**47.7±1.9**	35.5±8.9	33.9±6.1
TTP	68.9±5.2	65.8±9.9	66.9±7.5	47.6±6.4	35.4±9.1	33.9±4.6
FR	20.8±0.4	32.3±3.1	22.9±1.1	20.4±3.8	25.1±4.7	22.5±9.5
CNNFeat	**73.8±8.2**	**72.9±6.8**	**74.1±5.7**	**49.9±3.9**	**47.8±4.2**	**48.5±4.8**

in the matrix is 0.409 (wf and wrd), which is lower than the same result of features: ACC (2.144), WL (2.217), LOG (2.047), AR (3.570), DASDV (2.105) and MNF (2.330) (for the same operation, we calculate the distance between each cluster of these six features, and finally extract the maximum value). Thus, the EMG data are extremely well separated in the input space and can be effectively classified with CNNFeat. Besides, according to the formula (2), the RI of CNNFeat is calculated: 0.359, which is higher than the results of

ACC (0.112), WL (0.011), DASDV (0.080) and MNF (0.019), but it is lower than the results of LOG (0.464) and AR (0.900) features.

3.3 Feature Combination Evaluation

Six traditional features (AAC, LOG, WL, AR4, DASDV and MNP) selected in the previous experiment generally perform better than the rest of the features. In this paper, we combine these features with CNNFeat to check whether additional CNNFeat could enhance hand gesture classification accuracy.

Table 4 shows the experimental results of inter-session test across eight subjects for the classification of the first group of gestures. The results indicate that merging CNNFeat with every single traditional feature can further improve the classification accuracy. In addition, this study combines AR4, WL, DASDV features and then reduces the dimension to 16 by PCA for additinal comparison with CNNFeat. The results indicate that CNNFeat still has a similar effect (it can be a decent complement to traditional combination features). In detail, when combining traditional feature with CNNFeat, the average accuracy shows 4.35%, 3.62% and 4.7% improvement for SVM, LDA, KNN, respectively.

Table 4: Improved gesture recognition accuracy by combining traditional feature with CNN-based feature.

Feature	SVM		LDA		KNN	
	ACC	RE(+CNN)	ACC	RE(+CNN)	ACC	RE(+CNN)
AAC	68.02	74.54	65.90	73.69	68.32	68.71
LOG	65.38	74.71	64.19	73.66	64.16	74.99
WL	65.74	77.81	72.23	74.09	71.75	71.84
AR4	74.42	74.95	74.10	74.24	62.23	75.44
DASDV	71.03	74.72	69.11	73.70	68.91	69.50
MNP	68.96	69.04	65.80	73.38	66.91	66.93
AR4+WL	71.64	73.27	74.01	74.55	64.25	74.55
AR4+DASDV	72.80	74.55	72.68	73.19	65.44	71.21
DASDV+WL	72.03	75.26	72.95	74.88	71.34	72.06
AR4+WL+DASDV	71.14	75.31	72.08	73.95	69.76	74.89
AVERAGE	**70.07**	**74.42**	**70.31**	**73.93**	**67.31**	**72.01**

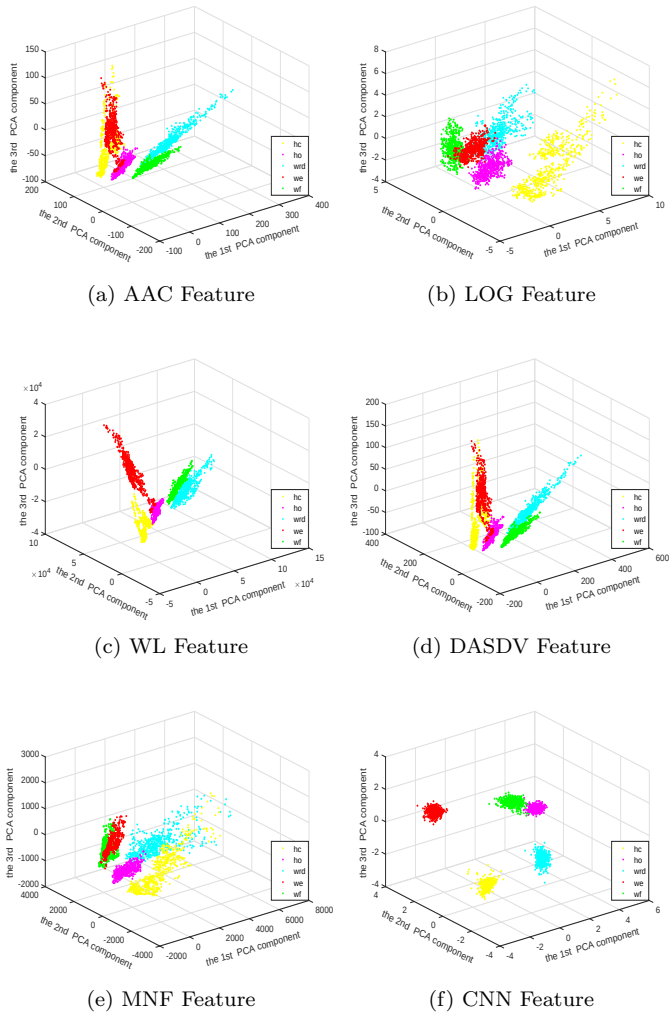
To demonstrate the sEMG samples in different feature space, the extracted 16-dimensional features (from the above-mentioned six traditional features and CNNFeat) are dimensionally reduced to three through principal component analysis (PCA). Among them, the contribution rate of AR4 is 62.42%, which is lower than the predetermined value 85% (general setting, this value depends on the tolerance of the original information). In other words, this

three-dimensional feature can not represent the original feature information effectively (it requires more dimensions). Therefore, for the sake of convenience, three-dimensional visualization operation is not conducted on the AR feature. The contribution rate of other five traditional features and CNNFeat are 91.64%, 88.75%, 92.00%, 91.27%, 90.67% and 88.27%, respectively. Fig.8 demonstrates the samples of different classes in six feature spaces mentioned above. It can be noted that cluster overlapping happens more frequently in traditional feature space than in the CNN-based feature space. In other words, CNNFeat can separate each class better, which is reflected in both aspects of inter-class distance and intra-class similarity.

3.4 Evaluation of Training Dataset for CNNFeat Extraction

As illustrated in Fig. 9, the accuracy of CNNFeat and three classic classifiers is higher than the result of CNN in both basic gesture and finger gesture classification. Besides, classification accuracy rises steadily along with the increasing size of training dataset. Among them, the polylines in Fig. 9 (a) represent the recognition rates of the first group of gestures, while the lines in Fig. 9 (b) indicate the recognition results of five finger gestures. In this figure, the horizontal axis represents the amount and source of data employed in the experiment. Number denotes the label of subjects' data. For example, the abscissa of the starting point of the polyline is 1, which indicates that only the data of subject 1 is used. Similarly, the abscissa of the last point of the polyline means that all EMG data except the data of subject 5 are utilized. Although there is a certain gap in the averaged classification accuracy between two groups of gestures (the recognition of the first group of gestures is obviously more accurate than the finger gestures), both curves show an obvious upward tendency. However, there exists an exception in the experiment. For example, when the training set increases the data of the sixth subject, the recognition accuracy of the corresponding basic gestures have a descending phenomenon. Similarly, for five finger gestures, there is a decline in accuracy when the training set appends the data of the third subject. It seems the obvious data distribution difference between the newly added sEMG data and the testing set results in failing to model the testing data from the training data. In general, with the increasing amount of data, the extracted CNNFeat can be more representative to achieve higher classification accuracy.

The same experimental operation is performed on the traditional features, and the results show that the change of the training data has no significant effect on the classification results. In addition, for the mixed features mentioned in Section 3.3 (traditional features plus CNNFeat), when the amount of training data is small, the traditional feature plays a major role, and the recognition result is similar to the result by using this traditional feature alone. As the volume of data increases, the CNNFeat is able to learn more EMG patterns, therefore, the classification result tends to be similar to the result of the mixed features in Table 4.



(a) AAC Feature (b) LOG Feature

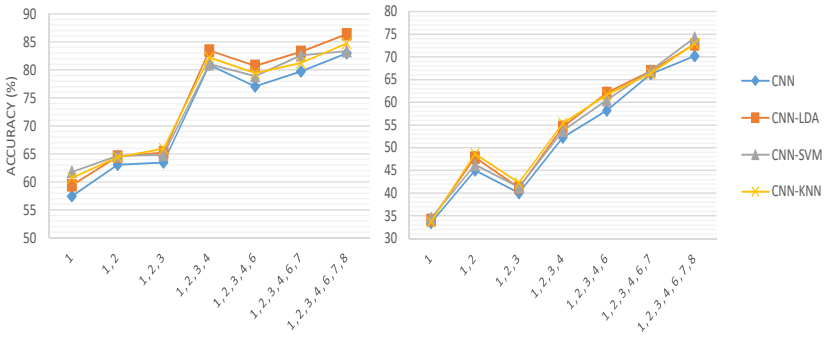
(c) WL Feature (d) DASDV Feature

(e) MNF Feature (f) CNN Feature

Fig. 8: The first three PCA components in six feature spaces.

3.5 Evaluation of Comparative Experiments

This study further takes DB1 of the NinaPor database to evaluate the proposed method. The DB1 database is collected from 27 intacted subjects, including 20 males and 7 females, aging from 22 to 44 with different height and weight. In order to compare with our own dataset, we select the same five actions as the gestures of our first group. The sampling frequency of this device is 100Hz. In order to make the input sEMG signal still represent 300ms information, the size of the network input is set to 30×10. The experimental results demonstrate



(a) Basic gestures (b) Finger gestures

Fig. 9: The classification accuracy with different scale of training set.

that SVM, LDA and KNN achieve the accuracy of 46.73%, 46.78% and 46.04% in inter-subject test, respectively. These accuracy are 1.17%, 1.22% and 0.48% higher than the result by using a softmax function (45.56%). In inter-session test, sEMG data is divided into training and testing sets with the ratio of 2:1. And the accuracy of three classifiers are 84.12%, 83.57%, 84.01%, respectively. Similar to the result in inter-subject test, the accuracy are 1.63%, 1.1%, 1.52% higher than the result of the basic network (82.49%).

We also compare our method with other deep learning methods. The results are shown in Table 5. For NinaPro DB1, the network structure proposed by Park *et al.* [11], Manfredo *et al.* [10] and Geng *et al.* [12] can obtain the accuracy of 60% (only 6 hand movements, non-adapation), 66.59%±6.4% (52 gestures), and 67.4% (52 gestures, input signal over 40 frames, vote to select result), respectively. Our method achieves the average accuracy of 66.9% (52 gestures, vote to select result). Similarly, for our own EMG dataset, the first three methods can obtain the accuracy of 62.02%, 65.72% and 67.25% for basic gestures in the inter-subject scenario, respectively (follow our input and output specifications). In this case, the average result of our method is 68.23%.

Table 5: The experimental results compared with related network algorithms

Database	Park et al.	Manfredo et al.	Geng et al.	**Ours**
NinaPro DB1	60%	66.59%	67.4%	**66.9%**
Elonxi DB	62.02%	65.72%	67.25%	**68.23%**

Besides, we also consider the computational performance of the network. When the model training is completed (about several hours), the time to extract CNNfeat from single EMG image (300*16) is calculated. The result is about 1.1ms, which fully meets the real-time requirements from the point of

view of computational performance. Therefore, the network proposed in this paper is effective for myoelectric applications.

4 Discussion

4.1 Classifiers

As reflected in both inter-subject and inter-session scenarios, the use of traditional methods instead of the last fully connected layer in the network can improve the accuracy by more than 1.5%. This may be benefited from the fact that the final classifier enhances the generalization ability of the whole system.

The convolutional neural network is essentially an input-to-output mapping technology. Each parameter is trained by back-propagation algorithm [26], with an intention to continuously reduce the error of the training set. Besides, CNN is designed as a classifier, in which a fully connected layer is applied at the end to generate the class label by using a softmax function. In contrast, when traditional classifiers are used for recognition, the situation is different. For example, SVM (based on the statistical theory of Vapnik-Chervonenkis Dimension (VC dimension) theory [27] and structural risk minimization [28] theory) attempts to get such a hyperplane that separates two categories and maximize their classification interval. Therefore, when we use SVM with BRF kernel function instead of the output layer in the network, it may exert the generalization capability of SVM to improve the classification effect. From this point of view, the method proposed in this paper is a two-step optimization process. CNN is used to extract features, while traditional classifiers are used for gesture recognition.

4.2 CNN features

CNNFeat is proposed in this paper to achieve better hand motion recognition accuracy for both inter-subject and inter-session scenarios. CNNFeat outperforms all traditional features in inter-subject evaluation, and outperforms most traditional features except AR4 in some cases in inter-session evaluation. It is worth noting that CNN is trained by much more data in inter-subject evaluation than that in inter-session evaluation, which may become one reason why CNNFeat performs better in inter-subject test. It can be partly proved in the evaluation of training dataset for CNNFeat extraction, in which the CNNfeat can perform increasingly better with more data for training. Besides, it may also imply that CNNFeat could function better in complex classification situation (i.e. inter-subject hand gesture classification).

It is also found that CNNFeat can be a decent complement to traditional sEMG features, reflected by the experimental results in Table 4. Integrating CNNFeat improves the accuracy by more than 3% for all tested classifiers. As visualised in Fig. 8, the CNNFeat space can map the raw sEMG signals

into a separable state than the other features. In comparison with traditional features that are obtained through well-designed feature extractors based on statistical theory and extracted statistical values from time-series signals or spectrum information, CNNFeat requires the training data set to learn the mapping rule from the raw sEMG signals to the feature. Therefore, CNNFeats can be invariant to electrode shift and sudden noise, such as waveform spikes. Moreover, CNN takes the relation between channels into account during feature extraction, while traditional sEMG feature extraction methods do not.

5 Conclusion

A sEMG database is constructed in this work to evaluate the performance of CNN-based sEMG feature. A 16-channel EMG acquisition device is used to collect the data from eight volunteers while performing ten hand gestures. Four evaluation methods are designed to compare 25 traditional EMG features with CNNFeat for hand gesture recognition by three classic classifiers (SVM, LDA and KNN). The experimental results show that CNNFeat outperforms all the tested traditional features in inter-subject test, and is listed as the best three features in inter-session test. Besides, with the increasing amount of training data, the high-dimensional features extracted by the trained network are more representative and achieve better classification results. We also find that combining CNNFeat with traditional features (including feature combinations) can further improve the recognition rate. Finally, we test our method on NinaPro DB1. The results show that using CNNfeat and three traditional classifiers can still improve the recognition rate of gestures in different scenarios. In summary, this work preliminarily demonstrates that CNN can be a useful tool to extract sEMG feature for hand gesture recognition, and a well trained CNN can transfer the raw sEMG into a low-dimensional sEMG feature space with better inter-class distinguishability and intra-class similarity.

Acknowledgements The authors would like to acknowledge the support from the EU Seventh Framework Programme (FP7)-ICT under Grant No. 611391, Natural Science Foundation of China under Grant No. 51575338, 51575407, 51475427, and the open fund of the key laboratory for metallurgical equipment and control of ministry of education in wuhan university of science and technologyunder Grant No. 2017B03.

References

1. Atzori M, Gijsberts A, Kuzborskij I, Elsig S, Hager A G, Deriaz O, Castellini C, Muller H, Caputo B (2015) Characterization of a Benchmark Database for Myoelectric Movement Classification. IEEE Trans Neural Syst Rehabil Eng 23(1):73-83
2. Xu Z, Tian Y, Li Y (2015) sEMG Pattern Recognition of Muscle Force of Upper Arm for Intelligent Bionic Limb Control. J Bionic Eng 12(2):316-323
3. Altimemy A H, Bugmann G, Escudero J, Outram N (2013) Classification of finger movements for the dexterous hand prosthesis control with surface electromyography. IEEE J Biomed Health Inform 17(3):608-618

4. Feng Z, Li P, Hou Z, Zeng L, Chen Y, Li Q, Min T (2012) sEMG-based continuous estimation of joint angles of human legs by using BP neural network. Neurocomputing 78(1):139-148
5. White M M, Zhang W, Winslow A T, Zahabi M, Fan Z, He H, Kaber D B (2017) Usability Comparison of Conventional Direct Control Versus Pattern Recognition Control of Transradial Prostheses. IEEE Trans Human-Mach Syst PP(99):1-12
6. Phinyomark A, Phukpattaranont P, Limsakul C (2012) Feature reduction and selection for EMG signal classification. Expert Syst Appl 39(8):7420-7431
7. Phinyomark A, Quaine F, Charbonnier S, Serviere C, Tarpin-Bernard F, Laurillau Y (2013) EMG feature evaluation for improving myoelectric pattern recognition robustness. Expert Syst Appl 40(12):4832-4840
8. Cireşan D C, Meier U, Masci J, Gambardella L M, Schmidhuber J (2011) High-Performance Neural Networks for Visual Object Classification. arXiv preprint arXiv:1102.0183
9. Kim Y, Jernite Y, Sontag D, Rush A M (2016) Character-Aware Neural Language Models. AAAI 2016: 2741-2749
10. Atzori M, Cognolato M, Müller H (2016) Deep learning with convolutional neural networks applied to electromyography data: A resource for the classification of movements for prosthetic hands. Frontiers in Neurorobotics 10: 9
11. Park K H, Lee S W (2016) Movement intention decoding based on deep learning for multiuser myoelectric interfaces. Proc 4th Int Winter Conf Brain-Comput Interface (BCI) pp 1-2
12. Geng W, Du Y, Jin W, Wei W, Hu Y, Li J (2016) Gesture recognition by instantaneous surface EMG images. Sci Rep 6: 36571
13. Niu X X, Suen C Y (2012) A novel hybrid CNN–SVM classifier for recognizing handwritten digits. Pattern Recog 45(4): 1318-1325
14. Bluche T, Ney H, Kermorvant C (2013) Tandem HMM with convolutional neural network for handwritten word recognition. IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP) 2390-2394
15. Fang Y, Liu H (2014) Robust sEMG electrodes configuration for pattern recognition based prosthesis control, Systems Man and Cybernetics (SMC) 2014 IEEE International Conference on 2210-2215
16. Atzori M, Gijsberts A, Castellini C, Caputo B, Hager A M, Elsig S, Giatsidis G, Bassetto F, Müller H (2014) Electromyography data for non-invasive naturally-controlled robotic hand prostheses. Sci Data 1: 140053
17. Amma C, Krings T, Böer J, Schultz T (2015) Advancing muscle-computer interfaces with high-density electromyography. Proc ACM Conference on Human Factors in Computing Systems (CHI'15) 929-938
18. Phinyomark A, Phukpattaranont P, Limsakul C (2012) Feature reduction and selection for EMG signal classification. Expert Syst Appl 39(8): 7420-7431
19. Sharma A, Paliwal K (2008) A Gradient Linear Discriminant Analysis for Small Sample Sized Problem. Neural Processing Letters. 27(1):17-24.
20. Srivastava N, Hinton G, Krizhevsky A, Sutskever I, Salakhutdinov R (2014) Dropout: a simple way to prevent neural networks from overfitting. J Mach Learn Res 15(1): 1929-1958
21. Ioffe S, Szegedy C (2015) Batch normalization: Accelerating deep network training by reducing internal covariate shift. arXiv:1502.03167
22. Krizhevsky A, Sutskever I, Hinton G E (2012) Imagenet classification with deep convolutional neural networks. Proc Neural Information and Processing Systems 1097-1105
23. Abadi M, Agarwal A, Barham P, Brevdo E, Chen Z, Citro C, Corrado G S, Davis A, Dean J, Devin M, et al (2016) TensorFlow: Large-Scale Machine Learning on Heterogeneous Distributed Systems. arXiv preprint arXiv:1603.04467
24. Sutskever I, Martens J, Dahl G E, Hinton G E (2013) On the importance of initialization and momentum in deep learning. J Mach Learn Res 1139-1147
25. Mahalanobis P C (1936) On the generalized distance in statistics. Proc Nat Inst Sci India
26. LeCun Y, Bottou L, Bengio Y, Haffner P, et al (1998) Gradient-based learning applied to document recognition. Proc IEEE 86(11): 2278-2324
27. Blumer A, Ehrenfeucht A, Haussler D, Warmuth M K (1989) Learnability and the Vapnik-Chervonenkis dimension. J Ass Comput Mach 36(4): 929-965

28. Burges C J C (1998) A tutorial on support vector machines for pattern recognition. Data Min Knowl Discovery 2(2): 121-167