



Dual layer transfer learning for sEMG-based user-independent gesture recognition

Yingwei Zhang^{1,2} · Yiqiang Chen^{1,2,3} · Hanchao Yu¹ · Xiaodong Yang^{1,2} · Wang Lu^{1,2}

Received: 5 September 2019 / Accepted: 10 March 2020 / Published online: 5 April 2020
© Springer-Verlag London Ltd., part of Springer Nature 2020

Abstract

During the last few years, significant attention has been paid to surface electromyographic (sEMG) signal-based gesture recognition. Nevertheless, sEMG signal is sensitive to various user-dependent factors, like skin impedance and muscle strength, which causes the existing gesture recognition models not suitable for new users and huge precision dropping. Therefore, we propose a dual layer transfer learning framework, named dualTL, to realize user-independent gesture recognition based on sEMG signal. DualTL is composed of two layers. The first layer of dualTL leverages the correlations of sEMG signal among different users to label partial gestures with high confidence from new users. Then, according to the consistencies of sEMG signal from the same users, the rest gestures are labeled in the second layer. We compare our method with three universal machine learning methods, seven representative transfer learning methods, and two deep learning-based sEMG gesture recognition methods. Experimental results show that the average recognition accuracy of dualTL is 80.17%. Comparing with SMO, KNN, RF, PCA, TCA, STL, and CWT, the performance improves 24.26% approximately.

Keywords Gesture recognition · Surface electro-myography · User-independent · Transfer learning · Dual layer

1 Introduction

With the development of surface electromyographic (sEMG) signal sensing and analysis technology, it is widely used in many applications [1–6], such as rehabilitation, entertainment, robotics, wheelchairs control, and pedestrian position-

ing. Hand gesture recognition is one of the representative applications of sEMG signal. Compared with other gesture recognition methods, such as Wi-Fi [7], computer vision [8], inertial measurement unit [9], ultrasound [10], electromagnetic wave [11], ultrasound imaging [12], sEMG signal-based methods provide us with significant opportunity to realize natural Human Computer Interaction (HCI) by directly sensing and decoding human muscular activities [13, 14]. sEMG signal-based gesture recognition is not only capable of distinguishing subtle finger configurations, hand shapes, and wrist movements, but also insensitive to environmental light and sound noise. Recently, sEMG signal-based technique attracts more and more attention from researchers. Many sEMG-based gesture recognition methods are proposed [15, 16]. In addition, many commercial gesture recognition productions are available, such as Myo, Econ, and shimmer.

However, sEMG signal has user-dependent property [17], which is the main factor that causes distribution diversities of sEMG signal among different users. Even signals that are acquired at the same position when performing the same gesture are different. The distribution differences are due to the fact that sEMG signal depends on many physical and environmental factors, such as the quantity of subcutaneous fat, skin impedance, muscle strength, the pattern of muscle

✉ Yiqiang Chen
yqchen@ict.ac.cn

Yingwei Zhang
zhangyingwei@ict.ac.cn

Hanchao Yu
yuhanchao@ict.ac.cn

Xiaodong Yang
yangxiaodong@ict.ac.cn

Wang Lu
luwang@ict.ac.cn

¹ Beijing Key Laboratory of Mobile Computing and Pervasive Device, Institute of Computing Technology, Chinese Academic of Sciences, Beijing, China

² University of Chinese Academy of Sciences, Beijing, China

³ Pengcheng Laboratory, Shenzhen, China

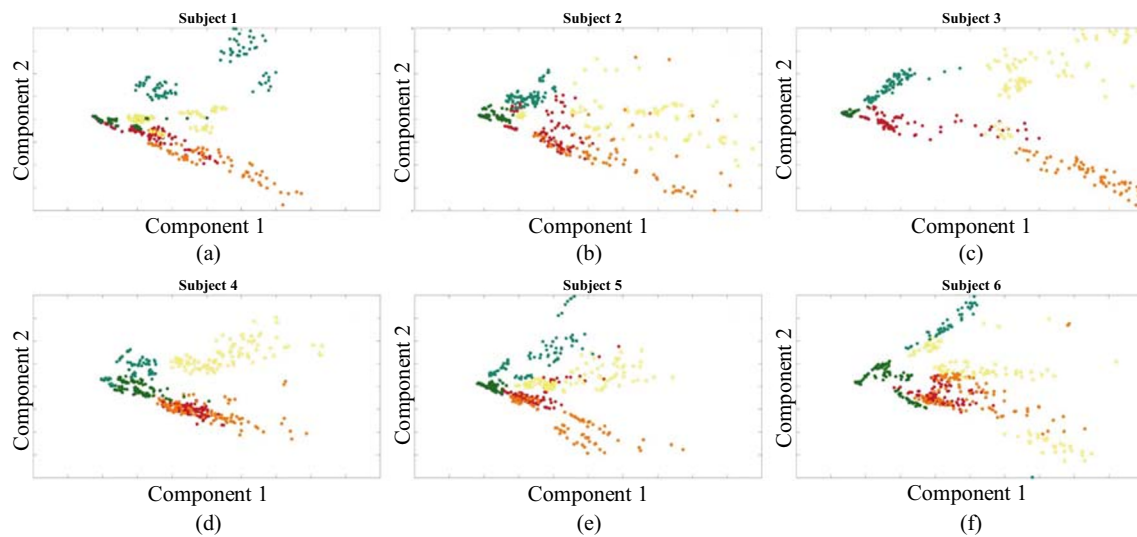


Fig. 1 sEMG signals distribution. The signals come from six different users, performing five hand gestures

synergies, muscle geometry and tone, specific motor unit sizes, length/size of the innervating nerves, and muscle innervation locations [18].

Figure 1 shows the distribution difference between sEMG signal over five hand gestures of six subjects. The data showed in Fig. 1 is the two-dimensional projections of sEMG signal by principal component analysis (PCA). From the analysis of Fig. 1, we observe that signal distribution varies from subject to subject, even when they perform the same hand gesture. Fortunately, there is still some prior knowledge that we can take advantage of. The data of the same hand gesture from one user gather together, for example, the red circles (data of gesture 1) in Fig. 1a. This proves that the same hand gesture of the same user is highly consistent. In addition, the distribution of data from the same gesture of different subject is relational, for example, the red circle in Fig. 1 a and b. This proves that the same hand gesture of different users is weakly correlating.

Previous studies try to construct classifiers for each individual user [19], which means each user must perform quite a long-time gesture and collect enough training data. To eliminate the inconvenience of retraining classifier and data annotation, we propose a novel sEMG signal-based gesture recognition method to realize an efficient and convenient recognition system in this paper. Generally, we design dual layer transfer learning framework, namely dualTL. DualTL is designed based on the prior knowledge attained from Fig. 1. Besides, dualTL is composed of two layers. In the first layer, we use the weak correlation of the same gesture from different users to realize preliminary recognition for part of novel user's gestures. In the second layer, the strong consistency of the same hand gesture from one user is used to realize ultimate recognition.

The structure of this paper is organized as follows: Section 2 reviews the related works about sEMG signal-based gesture recognition, especially some attempts to realize user-independent recognition. Section 3 introduces the proposed dualTL method in detail. Section 4 presents the experiments, including data collection, preprocessing, and recognition performance evaluation. Finally, Section 5 presents our conclusions and future works.

2 Related work

In this section, we briefly discuss the existing research on sEMG-based gesture recognition, user-independent gesture recognition, and transfer learning.

2.1 sEMG-based gesture recognition

Recently, due to the advantages showed by sEMG signal-based gesture recognition, such as the ability to recognize subtle gestures, insensitiveness to environmental light and sound noise, and non-intrusion, there emerge numerous works about sEMG signal-based gesture input and control methods in HCI area. Therefore, we review some related works about sEMG-based gesture recognition in this subsection.

Amma et al. [20] used sEMG sensor arrays with 192 electrodes to record high-density sEMG signal of the upper forearm muscles for finger gesture recognition. A baseline system was built to discriminate 27 gestures on their dataset with naive Bayes classifier. Finally, the averaged accuracy was 90% for the within-session scenario and 75% for the cross-session scenario. David et al. [21] designed a PC mouse commanded by sEMG signal from two muscles

of the forearm, palmar longus, and extensor digitorum. The experimental result showed that the classification accuracy was 87% on the predefined hand movement set: rest, flexion, extension, and closure. Saponas et al. [22] researched sEMG signal-based real-time gesture recognition method, and the experimental result demonstrated that the proposed real-time method acquired recognition accuracy of 79%, 85%, and 88% in pinching, holding a travel mug, and carrying a weighted bag gesture, respectively. Further, they showed the generalizability of their method across different arm postures and explored the trade-off of providing real-time visual feedback. McIntosh et al. [15] acquired four channels of sEMG signal and four channels of Force Sensitive Resistor signal by wearable equipment placed on the wrist. Then, they constructed a high-accuracy hand gesture recognition system named EMPress.

2.2 User-independent gesture recognition

Though all works aforementioned reached acceptable recognition accuracy, they did not consider the user-independent challenges. Fortunately, there was already some research trying to realize user-independent gesture recognition. In this subsection, we will review some sEMG signal-based user-independent gesture recognition methods in detail.

Khushaba et al. [23] proposed a framework for multiuser myoelectric interfaces by using canonical correlation analysis, where the data of different users were projected onto a unified-style space. The proposed method was able to overcome the individual differences with an acceptable cross-user accuracy 83%. Nevertheless, their method can not be used to recognize gestures of the new user. Matsubara et al. [24] made use of the bilinear model to construct a multiuser myoelectric interface, where the original sEMG signal was decomposed into motion dependent part and user-dependent part. However, as this paper mentioned, the user-dependent factors were not precise enough and the electrode placement problem was still open. What is more, the dimensions of the style and the content variables were experimentally selected by trial-and-error. In addition, it was reported that the positioning of electrodes, the type of features extracted, and their dimensionality could significantly impact the model's performance. Orabona et al. [25] applied an adaption model by constraining a new model that is mostly closed to multiple pre-trained models stored in the memory at each step. The adaptation process attempted to modify the best matched model to fit a new subject. Nevertheless, this process was executed in a high-dimensional parameter space, which required a large amount of data to make the adaptation complete. Chattopadhyay et al. [26] also presented, using sEMG signal, a user-independent computational feature selection framework to monitor muscle fatigue. A search mechanism toward the vicinity of the best

feature subset was guided by an objective function based on the ratio of between-user to within-user variance for the specific features, and this identified movements across multiple users. However, the main limitations of this method included the time taken to find the best feature subset and the large variance of sEMG signal, which limited the applicability of this feature selection algorithm.

2.3 Transfer learning

Transfer learning aims to relax the assumption in traditional machine learning that the training data and testing data should have an identical probability distribution [27]. It has achieved great success in many areas, such as Wi-Fi localization [28], natural language processing [29], face recognition [30], and human activity recognition [31]. The enlightening works of [32, 33] indicate that many factors (e.g., user habit, wearing position, and equipment fault) tend to influence the distribution of data in behavior and gesture recognition. To overcome these kinds of distribution evolution challenges in gesture recognition, some researchers have made significant explorations.

Goussies et al. proposed a novel algorithm to transfer knowledge from multiple other sources to computer vision-based gesture recognition tasks [34]. Comparative experiments showed transfer learning outperformed other baseline methods and achieved the best results. Costante et al. focused on the view-dependent problem in computer vision-based gesture recognition area and proposed a domain adaptation framework that worked on robust view-invariant self similarity matrix descriptors [35]. To realize rapid construction of gesture recognition model, some studies take advantage of transfer learning to fine-tune the existing convolutional neural network model [36–38]. Among them, Ozcan et al. combined AlexNet model and transfer learning together and verified it on computer vision-based gesture recognition datasets [36]. Cote-Allard aimed to alleviate the data acquisition burdens in sEMG-based gesture recognition by leveraging the data from other users [37]. Bu et al. proposed a Wi-Fi-based gesture recognition method by transforming the amplitude of channel state information into image matrix [38].

However, most of the studies above concentrate on computer vision-based gesture recognition. The performance of transfer learning on sEMG-based gesture recognition and unsupervised cross-user tasks are still unclear.

3 Dual layer transfer learning

In this section, we introduce the proposed dual layer transfer learning (dualTL) framework. Firstly, we present problem definition in Section 3.1. Then, we will detail cross-user and

within-user recognition in Sections 3.2 and 3.4. Candidate optimization methods are presented in Sections 3.3 and 3.5 is the overall procedure of dualTL.

3.1 Problem definition

User-independent gesture recognition system usually contains two kinds of data, the data of existing users $\mathcal{D}_e = \{(x_i, y_i)\}_{i=1}^{n_e}$ and the data of new users $\mathcal{D}_n = \{x_j\}_{j=1}^{n_n}$. \mathcal{D}_e and \mathcal{D}_n have the same dimensionality and label spaces, i.e., $x_i, x_j \in \mathbb{R}^d$, where d is the dimensionality of features, and $c_i \in \mathcal{Y}_e = \mathcal{Y}_n$ is label space. In addition, n_e is the size of data of existing users and n_n is the size of data of new users.

Figure 2 illustrates the main idea of dualTL. DualTL includes three main steps. Initially, dualTL selects candidates for data of new users through cross-user transfer and generates pseudo labels for the candidates. Then, it performs candidate optimization to optimize the selected subset of data. Finally, a cross-user transfer step is performed on the final candidates and the residuals.

3.2 Cross-user transfer

Cross-user transfer is the first layer of dualTL. This layer selects part of data of new users and generates pseudo labels for these selected data. The data that are selected are called candidates and the others are called residuals. The selected operation is based on defined confidence index.

The candidate selection and pseudo label generation are based on similarity comparison. We define the similarity measurement metric as the following:

$$\begin{aligned}
 dist_{ed}(x_i, x_j) &= \|x_i - x_j\| \\
 &= \sqrt{\sum_{k=1}^d |x_{ik} - x_{jk}|^2}
 \end{aligned} \tag{1}$$

This euclidean distance metric measures the similarity of different instances. In this layer, the data of existing users \mathcal{D}_e are used as source data, and the data of new users \mathcal{D}_n are used as target data. Based on metric defined in (1), we find the nearest \mathcal{K}_1 instances in \mathcal{D}_e for every instance in \mathcal{D}_n . Then, information of the \mathcal{K}_1 nearest neighbors are used to generate pseudo labels for instances in \mathcal{D}_n .

Then, we denote these \mathcal{K}_1 instances as $N_{\mathcal{K}_1}(x_j)$. Based on the label of these neighbors, category $F_1(x_j)$ of x_j is determined by the majority voting strategy showed in (2). The classification confidence $C_1(x_j)$ is determined by the probability of voting showed in (3), which represents the degree of confidence that sets the label of x_j as $F_1(x_j)$.

$$F_1(x_j) = \arg \max_{c_i} \frac{\sum_{\{x_{i'}, y_{i'}\} \in N_{\mathcal{K}_1}(x_j)} sgn(y_{i'}, c_i)}{\mathcal{K}_1} \tag{2}$$

$$C_1(x_j) = \sum_{\{x_{i'}, y_{i'}\} \in N_{\mathcal{K}_1}(x_j)} \frac{sgn(y_{i'}, F_1(x_j))}{\mathcal{K}_1} \tag{3}$$

where $\{x_{i'}, y_{i'}\}$ represents an instance in set $N_{\mathcal{K}_1}(x_j)$, $x_{i'}$ is the feature of this instance, and $y_{i'}$ is the label of this instance. In addition, $sgn(y_{i'}, c_i)$ is sign function. The value of this function is 1 when $y_{i'}$ is equal to c_i and the value of this function is 0 when $y_{i'}$ is not equal to c_i .

Due to the distribution difference of sEMG signal among different users, it is arduous to realize high-accuracy recognition just by majority voting. Thus, a filtering strategy is needed to realize recognition with high recall rate. Specifically, we select part of gestures \mathcal{D}'_n with high recognition confidence and keep their classification results:

$$y_j = \begin{cases} F_1(x_j), & C_1(x_j) > \mu \\ -1, & \text{otherwise} \end{cases} \tag{4}$$

The instances with confidence higher than μ are selected as candidates and the others are residuals. After the first layer of dualTL, the data of new users are transformed to $\mathcal{D}'_n = \{\mathcal{D}'_n, \mathcal{D}''_n\}$, $\mathcal{D}'_n = \{(x'_j, F_1(x'_j))\}_{j=1}^m$, $\mathcal{D}''_n = \{x''_j\}_{j=m+1}^{n_n}$. \mathcal{D}'_n is the set of candidates. \mathcal{D}''_n is the set of residuals. m is the number of instances that are selected with high confidence.

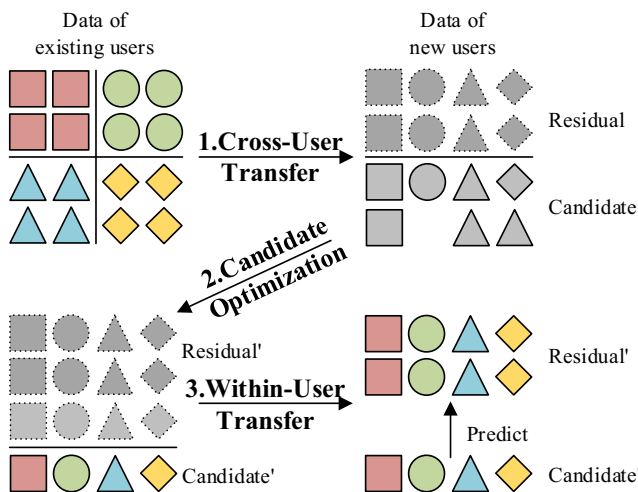


Fig. 2 Framework of dual layer transfer learning, including three main steps: (1) candidates generating via cross-user transfer; (2) candidate optimization through further selection; (3) final label decision through within-user transfer

3.3 Candidate optimization

The purpose of candidate optimization is to select a subset of candidates \mathcal{D}'_n and this operation has two constraints. Firstly, the classification confidence of selected instances should be as high as possible. Secondly, the distribution of selected instances should be as decentralized as possible. The first objective is easy to understand, and the second objective is to avoid all selected instances distributing too centrally so that they can not cover all sample spaces. Consequently, the optimization function are formulated as following:

$$\arg \max_{\mathcal{D}'_n} \sum_{j=1}^{|\mathcal{D}'_n|} C_1(x'_j) + \lambda Distr(\mathcal{D}'_n) \tag{5}$$

where, \mathcal{D}'_n is the selected subset of \mathcal{D}^l_n , λ is coefficient and $Distr(\mathcal{D}'_n)$ is the divergence of set \mathcal{D}'_n .

In addition, we build divergence model according to the feature of instance \mathcal{D}'_n and the procedure of this model is demonstrated in Algorithm 1. Based on the idea of PCA, we project the raw feature to one-dimensional space. And then, we use variance of subset \mathcal{D}'_n measures the divergence of selected data set.

Algorithm 1 Divergence model.

Input: Subset \mathcal{D}'_n of \mathcal{D}^l_n .

Output: Divergence $Distr(\mathcal{D}'_n)$.

- 1: Centralization all samples: $x'_j \leftarrow x'_j - \frac{1}{|\mathcal{D}'_n|} \sum_{j=1}^{|\mathcal{D}'_n|} x'_j$;
 - 2: Calculate covariance matrix $\mathcal{D}'_n \mathcal{D}'_n{}^T$;
 - 3: Decompose eigen of covariance matrix $\mathcal{D}'_n \mathcal{D}'_n{}^T$;
 - 4: Get the corresponding feature vector \vec{w} of maximum feature value;
 - 5: Calculate the projection result: $\bar{\mathcal{D}}'_n = \mathcal{D}'_n \vec{w}$;
 - 6: Calculate the variance: $Distr(\mathcal{D}'_n) = \frac{(\bar{\mathcal{D}}'_n - \bar{\mathcal{D}}'_n)^2}{T}$, where $\bar{\mathcal{D}}'_n = \frac{1}{|\mathcal{D}'_n|} \sum_{j=1}^{|\mathcal{D}'_n|} x'_j$.
-

There are two complete solutions that can find the optimal value of (5). One solution is to enumerate all possible subsets of \mathcal{D}'_n , but this method need to iterate C_{c_i} (calculated in (6)) subsets and it is time consuming.

$$C_{c_i} = \begin{pmatrix} Q_{c_i} \\ P_{c_i} \end{pmatrix} \tag{6}$$

$$Q_{c_i} = \sum_{j=1}^{|\mathcal{D}'_n|} sgn(F_1(x'_j), c_i) \tag{7}$$

$$P_{c_i} = \omega \cdot \sum_{j=1}^{\mathcal{D}'_n} sgn(F_1(x'_j), c_i) \tag{8}$$

where Q_{c_i} is the number of gestures predicted as the c_i gesture in set \mathcal{D}'_n and P_{c_i} is the number of gesture we will select. Another complete solution is to use the idea of dynamic programming. If $f[w_1, w_2, \beta, \sigma]$ is the optimal value when we select w_2^{th} data among the first w_1^{th} data under the restrictions that confidence sum is β and divergence is σ , then the dynamic programming function is $f[w_1, w_2, \beta, \sigma] = \max(f[w_1, w_2 - 1, \beta, \sigma], f[w_1 - 1, w_2 - 1, \beta - C_{c_1}(x'_j), \sigma - \mathcal{D}'_n] + (\mathcal{D}'_n)^2)$. The time complexity of this solution is also high and it requires that the confidence and divergence values are discrete.

Here, we use an approximate solution in our scenario. We sort the confidence value $C_1(x'_j)$ firstly and choose top κ percentage of data with highest confidence to find the optimal value. So, we only need to enumerate C'_{c_i} (showed in (9)) subsets.

$$C'_{c_i} = \begin{pmatrix} \kappa \cdot Q_{c_i} \\ P_{c_i} \end{pmatrix} \tag{9}$$

After candidate optimization, the data of new users are transformed to $\mathcal{D}''_n = \{\mathcal{D}'_n, \mathcal{D}''_n\}$, $\mathcal{D}'_n = \{(x'_j, F_1(x'_j))\}_{j=1}^{m'}$, $\mathcal{D}''_n = \{x''_j\}_{j=m'+1}^{n_n}$. \mathcal{D}'_n is the new set of candidates (i.e., candidate' in Fig. 2), \mathcal{D}''_n is the new set of residual (i.e., residual' in Fig. 2), and m' is the number of instances that are selected.

3.4 Within-user transfer

Following this, we build the concluding transfer (10) with $\mathcal{D}'_n = \{(x'_j, F_1(x'_j))\}_{j=1}^{m'}$ as source data, $\mathcal{D}''_n = \{x''_j\}_{j=m'+1}^{n_n}$ as target data.

$$F_2(x''_j) = \arg \max_{c_i} \frac{\sum_{(x'_j, F_1(x'_j)) \in \mathcal{N}_{\mathcal{K}_2}(x''_j)} sgn(F_1(x'_j), c_i)}{\mathcal{K}_2} \tag{10}$$

The decision strategy is also majority voting, same to the method used in the first layer of dualTL. In this layer, we use the data from new users to recognize his own gestures. This can avoid the distribution drift of different users. After the analysis above, all gestures are recognized accurately. Equation (11) is the distance metric used in the second layer of dualTL:

$$\begin{aligned} dist_{ed}(x_j', x_j'') &= 1 - \frac{\sum_{k=1}^d (x'_{jk} - \bar{x}'_k)(x''_{jk} - \bar{x}''_k)}{\sqrt{\sum_{k=1}^d (x'_{jk} - \bar{x}'_k)^2} \sqrt{\sum_{k=1}^d (x''_{jk} - \bar{x}''_k)^2}} \end{aligned} \tag{11}$$

$$\bar{x}^{l'} = \frac{1}{|\mathcal{D}_n^{l'}|} \sum_{j=1}^{|\mathcal{D}_n^{l'}|} x_j^{l'} \quad (12)$$

$$\bar{x}^{u'} = \frac{1}{|\mathcal{D}_n^{u'}|} \sum_{j=1}^{|\mathcal{D}_n^{u'}|} x_j^{u'} \quad (13)$$

where $\bar{x}^{l'}$ and $\bar{x}^{u'}$ are mean of $x_j^{l'}$ and $x_j^{u'}$, respectively.

3.5 Overall procedure

The overall process of dualTL is described in Algorithm 2. DualTL is a general framework for user-independent gesture recognition based on sEMG signal. On the basis of small data set, we provide the feasible implementation of dualTL. It can also be implemented in different ways according to the specific applications.

Algorithm 2 Overall procedure of dualTL algorithm.

Input: Data of existing users: $\mathcal{D}_e = \{(x_i, y_i)\}_{i=1}^{n_e}$, data of new users: $\mathcal{D}_n = \{x_j\}_{j=1}^{n_e}$, parameters: $\Theta = \{\mathcal{K}_1, \mathcal{K}_2, \lambda, \mu\}$.

Output: The labels for the gestures of new users: $\{y_j\}$.

- 1: Cross-user data transfer with \mathcal{D}_e as source data and \mathcal{D}_n as target data using (2) and (3) to get $F_1(x_j)$ and $C_1(x_j)$;
- 2: High confidence instances selection according to (4): $\{\mathcal{D}_n' = \{\mathcal{D}_n^l, \mathcal{D}_n^u\}\} \leftarrow \mathcal{D}_n$;
- 3: Candidates optimization according to (5): $\{\mathcal{D}_n'' = \{\mathcal{D}_n^{l'}, \mathcal{D}_n^{u'}\}\} \leftarrow \{\mathcal{D}_n' = \{\mathcal{D}_n^l, \mathcal{D}_n^u\}\}$;
- 4: Within-user data transfer with $\mathcal{D}_n^{l'}$ as source data and $\mathcal{D}_n^{u'}$ as target data according to (10);
- 5: Labels of $x_j^{l'} \in \mathcal{D}_n^{l'}$ is $F_1(x_j^{l'})$, labels of $x_j^{u'} \in \mathcal{D}_n^{u'}$ is $F_2(x_j^{u'})$.

4 Experimental evaluation

In this section, we conduct extensive experiments to validate the performance of the proposed dualTL. Except for data acquisition, all experiments are conducted on a Lenovo ThinkCentre M8600t-D065 (Intel Core i7-6700 / 16GB DDR3) desktop computer with Matlab R2016a.

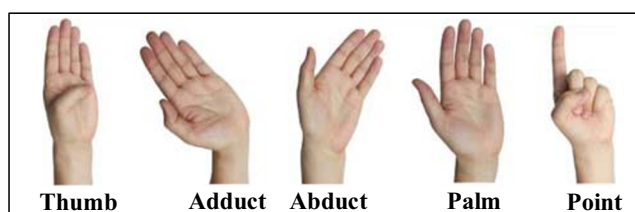


Fig. 3 The details of hand gesture set

Table 1 Physiological information of all subjects

Subject	Age	Sex	Height	Weight	Circum	Hand
s ₁	21	Male	175 cm	70 kg	24 cm	Right
s ₂	23	Female	160 cm	50 kg	22 cm	Right
s ₃	24	Male	176 cm	65 kg	33 cm	Right
s ₄	22	Male	172 cm	60 kg	24 cm	Right
s ₅	26	Male	180 cm	70 kg	26 cm	Right
s ₆	24	Female	164 cm	45 kg	18 cm	Right

4.1 Data acquisition

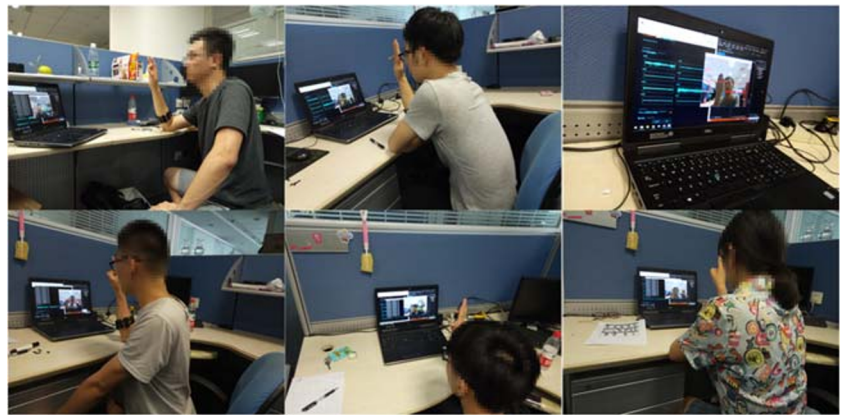
We design a gesture set with five static hand gestures: thumb, adduct, abduct, palm, and point. The details of gesture set are demonstrated in Fig. 3. We recruit a total of six participants (four males and two females) for the experiment. Table 1 details the physiological information of all subjects. All shown in Table 1 are the age, height, weight, and circumference of upper forearm range from 18 to 26, 160 to 180 cm, 45 to 70 kg, and 18 to 34 cm, respectively. All participants are healthy and right-handed.

Data acquisition is conducted on a Dell Precision 7510 (Intel Core i7-6820HQ / 16 GB DDR3) laptop computer with Visual Studio (VS) 2017 Integrated Development Environment (IDE), OpenCV 2.4.11, and Myo armband. Myo is a wearable myoelectric armband Myo from Thalmic Labs. It has eight evenly distributed electrical chips, which are used to collect sEMG signal and the sampling rate is 200 Hz. In the process of data acquisition, Myo is worn on the subjects' upper forearm, like Fig. 4. Before the beginning of data acquisition of each gesture, the subject has 5 s interval and meanwhile, the standard pose is demonstrated by the guider in order to regularize the motion of subject. The data acquisition time lasts 15 s for each gesture. We perform the data acquisition for all gestures orderly and repeat it eight times. Simultaneously, we also record the motion of subject by the camera to make sure whether the gesture is performed correctly or not. The real scenario of data collection is showed in Fig. 5.



Fig. 4 Data acquisition position on upper forearm of subject

Fig. 5 The real scenario of data acquisition



4.2 Data preprocessing and feature extraction

To reduce the noise of sEMG signal, we will do some preprocessing operation. To begin with, we apply a fourth order butter-worth bandpass filter with pass-band of 30 – 70 Hz to remove the attenuate dcoffset, motion artifacts, and low-frequency and high-frequency noise. Then, a fourth order butter-worth low-pass filter with pass-low of 60 Hz is also applied to capture the “envelope” of sEMG signal. Raw sEMG signal and filtered sEMG signal from the first subject are showed in Fig. 6. The five columns are signal of five hand gestures. The first row and the second row are raw sEMG signal and the filtered sEMG signal, respectively.

Besides, all gestures used in this experiment are static gesture; we use sliding window to segment the data. The length of each window is 1 s and the overlay of adjacent two windows are 50%. Since the sampling rate of Myo is 200 Hz and the scale of electrical chip is 8, there are 200×8 points in each window.

Generally, most of the attempts extracting features from sEMG signal can be classified into three categories, including time domain, frequency domain, and time–frequency domain [39, 40]. In our setting, we only consider the first two categories for computational simplicity [41]. In feature extraction process, we separately extract seven

time-domain features and three frequency domain features from raw sEMG signal, which are described in Table 2, where x_i represents the raw sEMG signal and N is the length of x_i . PSD_i means power spectrum density and M means the length of PSD_i . A_i and f_i indicate magnitude spectrum and frequency respectively.

As we all know, the amplitude of sEMG signal differs greatly among different subjects. To eliminate the influence of distribution diversity of sEMG signal, we calibrate all subjects’ feature by dividing the mean of all features of this subject before using dualTL.

4.3 Comparison methods

We compare dualTL with 14 different methods, including four universal machine learning methods, seven transfer learning methods, two deep learning–based sEMG gesture recognition methods, and one variation of dualTL:

- SMO: sequential minimal optimization [42];
- KNN: K-nearest neighbor [43];
- RF: random forest [44];
- PCA: principal component analysis [45];
- TCA: transfer component analysis [46];
- JDA: joint distribution adaptation [47];
- BDA: balanced distribution adaptation [48];

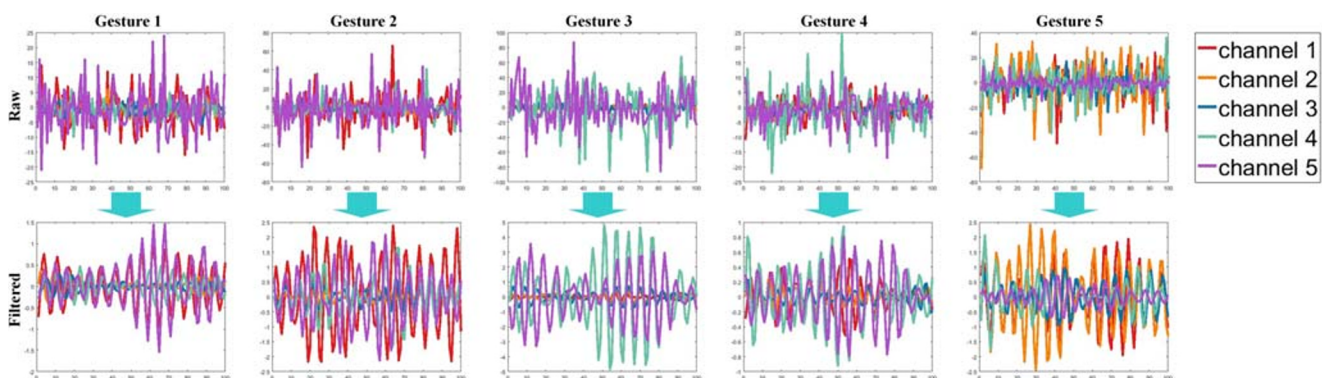


Fig. 6 The comparison of raw sEMG signal and filtered sEMG signal

Table 2 Feature extraction

Extracted feature	Description
Mean absolute value	$MAV_k = \frac{1}{N} \sum_{i=1}^N x_i $
Modified mean	$MMAV1_k = \frac{1}{N} \sum_{i=1}^N w_i x_i $
Absolute value 1	$w(i) = \begin{cases} 1, & 0.25N \leq i \leq 0.75N \\ 0.5, & \text{otherwise} \end{cases}$
Modified mean	$MMAV2_k = \frac{1}{N} \sum_{i=1}^N w_i x_i $
Absolute value 2	$w(i) = \begin{cases} 1, & 0.25N \leq i \leq 0.75N \\ \frac{4i}{N}, & 0.25N > i \\ \frac{4(i-N)}{N}, & 0.75N < i \end{cases}$
Mean absolute value slope	$MAVS_k = MAV_{k+1} - MAV_k$
Root mean square	$RMS_k = \sqrt{\frac{1}{N} \sum_{i=1}^N x_i^2}$
Variance	$VAR_k = \frac{1}{N} \sum_{i=1}^N x_i - \hat{x}^2$
Waveform length	$WL_k = \sum_{i=1}^M x_{i+1} - x_i $
Frequency median	$FMD = \frac{1}{2} \sum_{i=1}^M PSD_i$
Frequency mean	$FMN = \frac{\sum_{i=1}^M f_i PSD_i}{\sum_{i=1}^M PSD_i}$
Modified	$MFMN = \frac{\sum_{j=1}^M f_j A_j}{\sum_{j=1}^M A_j}$
Frequency mean	

- GFK: geodesic flow kernel [49];
- CLGA_s: coupled local–global adaptation with single-source [50];
- CLGA_m: coupled local–global adaptation with multi-source [50];
- STL: stratified transfer learning [51];
- Spectrograms: deep learning–based sEMG gesture recognition method with spectrograms as input [52];
- CWT: deep learning–based sEMG gesture recognition method with continuous wavelet transform (CWT) as input [52];

Table 3 Comparisons of accuracy over all subjects, the average accuracy, and standard deviation of recognition accuracy

	s ₁	s ₂	s ₃	s ₄	s ₅	s ₆	av.	std
SMO	47.83%	45.97%	37.74%	38.12%	41.09%	38.23%	41.50%	± 4.39%
KNN	45.87%	43.62%	49.80%	34.38%	35.43%	37.22%	41.05%	± 6.28%
RF	51.30%	45.32%	45.32%	45.62%	40.88%	38.13%	44.43%	± 4.53%
PCA	49.13%	23.83%	33.77%	36.88%	41.30%	26.20%	35.18%	± 9.45%
TCA	46.30%	20.00%	32.67%	31.25%	29.77%	23.55%	30.59%	± 9.10%
JDA	40.65%	32.13%	27.37%	30.42%	31.87%	22.63%	30.84%	± 5.98%
BDA	39.35%	31.91%	31.57%	28.96%	28.30%	22.32%	30.40%	± 5.58%
GFK	44.35%	23.62%	30.02%	33.96%	33.12%	27.42%	32.08%	± 7.11%
GLGA_s	46.52%	25.74%	31.79%	35.83%	38.57%	24.87%	33.89%	± 8.21%
GLGA_m	45.87%	29.36%	37.53%	34.17%	43.40%	25.18%	35.92%	± 7.99%
STL	45.43%	24.89%	30.36%	36.67%	33.96%	27.83%	33.19%	± 7.32%
Spectrograms	62.13%	54.62%	58.61%	54.33%	55.91%	49.85%	55.91%	± 4.17%
CWT	60.60%	52.32%	58.56%	53.69%	54.04%	48.19%	54.57%	± 4.45%
dualTL_wo	70.30%	81.67%	79.34%	72.83%	82.34%	54.77%	73.54%	± 10.40%
dualTL	79.39%	86.38%	86.53%	79.54%	86.15%	63.05%	80.17%	± 9.04%

The bold entries are used to highlight the best results achieved within every subject and emphasized the method that achieves the best result

- dualTL_wo: variation of dualTL, and the candidates are not optimized with the second step, i.e., candidate optimization.

where SMO, KNN, RF, and PCA are four universal machine learning methods. TCA, JDA, BDA, GFK, CLGA_s, CLGA_m, and STL are seven representative transfer learning methods. Spectrograms and CWT are deep learning–based sEMG gesture recognition methods and these two methods are initially supervised transfer learning methods to recognize gestures of new user. However, this setting is different from dualTL. Thus, we remove the fine-tune process of spectrograms and CWT. DualTL_wo is a variation of dualTL, which remove the candidate optimization process in the second step.

4.4 Experimental setting

In experimental process, the parameters $\Theta = \{\mathcal{K}_1, \mathcal{K}_2, \lambda, \mu\}$ of dualTL are set to $\mathcal{K}_1 = 5, \mathcal{K}_2 = 1, \lambda = 0.5, \mu = 0.4$, respectively. These four parameters are calculated by grid search. For SMO, the kernel function is radial basis function and the punishment factor is 100. For KNN, the number of neighbor is 5. For RF, the number of tree is 30. All other eight methods require dimensionality reduction. Therefore, we set them the same dimension 30.

4.5 Recognition performance

We evaluate the performance of dualTL by recognition accuracy of novel subject using the leave-one-out validation. In this process, the sEMG signals of one subject are used as testing data, and the remaining signals are used as training data to construct the recognition model. We repeat the

performance evaluation process until all subjects’ data are once used as testing data.

4.5.1 Recognition accuracy

The recognition accuracy over all subjects and the average accuracy are showed in Table 3. From Table 3, we can know that the average accuracies of four universal methods are 41.50%, 41.05%, 44.43%, and 35.18%. The average accuracies of seven traditional transfer learning methods are 30.59%, 30.84%, 30.40%, 32.08%, 33.89%, 35.92%, and 33.19%, respectively. These results are not very good that it is hard to realize natural HCI using hand gestures. Moreover, common transfer learning methods can not achieve better recognition results compared with SMO, KNN, RF, and PCA (Fig. 7). The average accuracy of two deep learning–based sEMG gesture recognition methods, i.e., spectrograms and CWT, are 55.91% and 54.57%, respectively. Compared with universal four learning methods and seven transfer learning methods, spectrograms and CWT achieve better results (Fig. 7). DualTL achieves the best performance among all 15 methods. The accuracy of dualTL is 80.17% and 24.26% better than the first 13 methods, including four traditional machine learning method, seven transfer learning methods, and two deep learning methods that are designed for sEMG gesture recognition. Also, dualTL is 6.63% better than dualTL_wo, proving the effectiveness of the second step (i.e., candidate optimization).

4.5.2 Confusion matrix

Besides, we also analyze the confusion matrix among all users. Here we only present the average confusion matrix

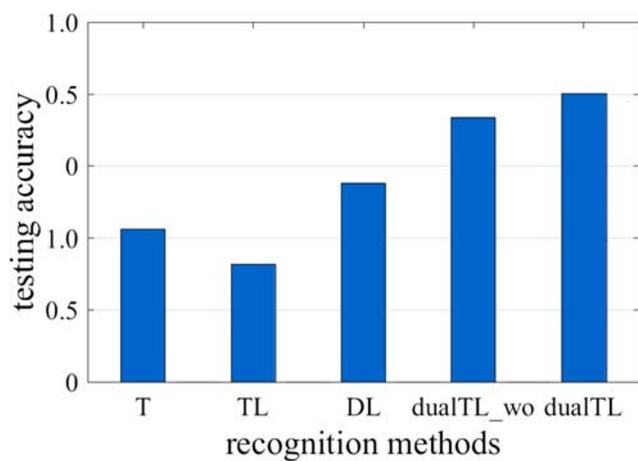


Fig. 7 Average accuracy comparison of five types of methods. T represents average accuracy of four universal machine learning methods. TL represents average accuracy of seven transfer learning machines. DL represents average accuracy of two deep learning–based sEMG gesture recognition methods

of all subjects, which is showed in Fig. 8a, and the confusion matrix of the fourth subject, which is showed in Fig. 8b. From the analysis of averaged confusion matrix, we can know that the fifth hand gesture “point” reaches the highest recognition accuracy 91%. But the accuracy of the second hand gesture is only 58%, which is the lowest among all gestures. Compared with the averaged confusion matrix, there are some differences in confusion matrix of the fourth subject, which reaches the highest recognition accuracy 91% in the third hand gesture “abduct” and the fourth hand gesture “palm.” The gesture with the lowest recognition accuracy is “adduct,” which is consistent with the averaged confusion matrix. By comparing these two confusion matrices, we can know that the performance is good among the third, the fourth, and the fifth hand gestures while the performance is poor on the second gesture. The lesson we can learn from the analysis aforementioned is that excellent gesture set design is essential to constructing high-accuracy hand gesture recognition system.

4.5.3 Pseudo label analysis

DualTL is a kind of unsupervised gesture recognition method to recognize the unlabeled data of the new user. To realize high-accuracy gesture recognition, dualTL firstly labels part of gestures from the new user with high confidence. Then, all data of the new user are classified with the help of these pseudo labels. The reliability of the pseudo label is important for final gesture recognition results. Thus, we analyze the recognition accuracy of candidates after cross-user transfer in the first step, the accuracy of new

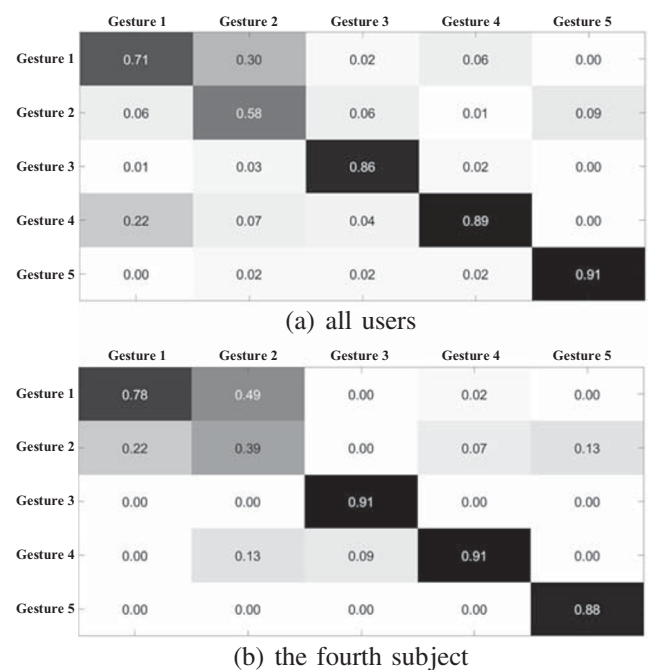


Fig. 8 Confusion matrix

Table 4 Gesture recognition accuracy for candidates after cross-user transfer and candidate optimization, and final recognition accuracy for all instances

	Cross-user transfer (step 1)	Candidate optimization (step 2)	dualTL (step 3)
s_1	80.23%	84.12%	79.39%
s_2	84.89%	92.90%	86.38%
s_3	87.12%	91.65%	86.53%
s_4	83.61%	85.10%	79.54%
s_5	86.68%	90.45%	86.15%
s_6	70.70%	72.13%	63.05%
<i>av.</i>	82.21%	86.06%	80.17%

candidates after candidate optimization in the second step, and final gesture recognition accuracy in the third step. Table 4 presents the analysis results. As Table 4 shows, the recognition accuracy of new candidates in the second step is highest, proving the effectiveness of candidate optimization. Besides, the recognition results in the first and the second step are not 100% correct. The average recognition accuracy in the first, the second, and the third step are 82.21%, 86.06%, and 80.17%, respectively. Compared with the results in the second step, the results in the third step have some degrees of decline. Fortunately, these declines are not very serious, proving the reliability of dualTL.

5 Conclusions and future works

5.1 Conclusions

In this work, we propose dualTL, a dual layer transfer learning method to realize user-independent hand gesture recognition. The weak correlation of the same hand gesture from different users and the strong consistency of the same hand gesture from one user are both used in this method. To evaluate the effectiveness of the proposed approach, a verification experiment is designed. From the analysis of experiment result, the recognition accuracy of the proposed method is 80.17%, which improves about 24.26% compared with conventional machine learning algorithm, such as SMO, KNN, and RF, and even state-of-the-art transfer learning and other methods specifically designed for sEMG gesture recognition.

5.2 Future works

However, there are still some limits in our approach. Firstly, the gesture set is small and only static gesture is taken into consideration. We will apply our method on other gesture set in the future. Secondly, we will explore how to combine the dual layer recognition framework with other conventional machine learning algorithm to realize more accurate and robust user-independent gesture recognition.

Funding information This work is financially supported by the National Key Research and Development Plan of China (2017YFB1002801); Natural Science Foundation of China under Grant No. 61502456 and No. 61972383; R & D Plan in Key Field of Guangdong Province (No. 2019B010109001); and by Alibaba Group through Alibaba Innovative Research (AIR) Program.

References

- Zhang Y, Chen Y, Yu H, Yang X, Lu W, Liu H (2018) Wearing-independent hand gesture recognition method based on emg armband. *Personal and Ubiquitous Computing* 22(3):511–524
- Moseley JB, JR FW, Pink M, Perry J, Tibone J (1992) Jobe Emg analysis of the scapular muscles during a shoulder rehabilitation program. *Am J Sports Med* 20(2):128–134
- Kawamoto H, Lee S, Kanbe S, Sankai Y (2003) Power assist method for hal-3 using emg-based feedback controller. In: 2003 IEEE international conference on systems, man and cybernetics, vol 2. IEEE, pp 1648–1653
- Sears HH, Shaperman J (1991) Proportional myoelectric hand control: an evaluation. *Am J Phys Med Rehabil* 70(1):20–28
- Neto AF, Celeste WC, Martins VR, Bastos Filho TF, Sarcinelli Filho M (2006) Human-machine interface based on electro-biological signals for mobile vehicles. In: 2006 IEEE International Symposium on Industrial Electronics, vol 4. IEEE, pp 2954–2959
- Gao N, Zhao L (2016) A pedestrian dead reckoning system using semg based on activities recognition. In: 2016 IEEE Chinese guidance, navigation and control conference (CGNCC). IEEE, pp 2361–2365
- Chen J, Li F, Chen H, Yang S, Wang Y (2019) Dynamic gesture recognition using wireless signals with less disturbance. *Pers Ubiquit Comput* 23(1):17–27
- Song J, Sörös G, Pece F, Hilliges O (2015) Real-time hand gesture recognition on unmodified wearable devices. In: IEEE conference on computer vision and pattern recognition
- Ducloux J, Colla P, Petrashin P, Lancioni W, Toledo L (2014) Accelerometer-based hand gesture recognition system for interaction in digital tv. In: 2014 IEEE international instrumentation and measurement technology conference (I2MTC) Proceedings. IEEE, pp 1537–1542
- Nandakumar R, Iyer V, Tan D, Gollakota S (2016) Fingerio: using active sonar for fine-grained finger tracking. In: Proceedings of the 2016 CHI conference on human factors in computing systems. ACM, pp 1515–1525
- Lien J, Gillian N, Karagozler ME, Amihoud P, Schwesig C, Olson E, Raja H, Poupyrev I (2016) Soli: ubiquitous gesture sensing with millimeter wave radar. *ACM Trans Graph (TOG)* 35(4):142
- McIntosh J, Marzo A, Fraser M, Phillips C (2017) Echoflex: hand gesture recognition using ultrasound imaging. In: Proceedings of

- the 2017 CHI conference on human factors in computing systems. ACM, pp 1923–1934
13. Zhang X, Chen X, Li Y, Lantz V, Wang K, Yang J (2011) A framework for hand gesture recognition based on accelerometer and emg sensors. *IEEE Trans Syst Man Cybern-Part A Syst Humans* 41(6):1064–1076
 14. Zhang X, Chen X, Wang W-H, Yang J-H, Lantz V, Wang K-Q (2009) Hand gesture recognition and virtual game control based on 3d accelerometer and emg sensors. In: *Proceedings of the 14th international conference on intelligent user interfaces*. ACM, pp 401–406
 15. McIntosh J, McNeill C, Fraser M, Kerber F, Löchtefeld M, Krüger A (2016) Empress: practical hand gesture classification with wrist-mounted emg and pressure sensing. In: *Proceedings of the 2016 CHI conference on human factors in computing systems*. ACM, pp 2332–2342
 16. Benatti S, Casamassima F, Milosevic B, Farella E, Schönle P, Fateh S, Burger T, Huang Q, Benini L (2015) A versatile embedded platform for emg acquisition and gesture recognition. *IEEE Trans Biomed Circ Syst* 9(5):620–630
 17. Matsubara T, Hyon S-H, Morimoto J (2011) Learning and adaptation of a stylistic myoelectric interface: Emg-based robotic control with individual user differences. In: *IEEE International Conference on Robotics and Biomimetics (ROBIO)*, 2011. IEEE, pp 390–395
 18. Merletti R, Parker PA, Parker PJ (2004) *Electromyography: physiology, engineering, and non-invasive applications*, vol 11. Wiley, New York
 19. Khushaba RN, Al-Ani A, Al-Jumaily A (2010) Orthogonal fuzzy neighborhood discriminant analysis for multifunction myoelectric hand control. *IEEE Trans Biomed Eng* 57(6):1410
 20. Amma C, Krings T, Böer J, Schultz T (2015) Advancing muscle-computer interfaces with high-density electromyography. In: *Proceedings of the 33rd annual ACM conference on human factors in computing systems*. ACM, pp 929–938
 21. David RL, Cristian CL, Humberto LC (2015) Design of an electromyographic mouse. In: *2015 20th symposium on signal processing, images and computer vision (STSIVA)*. IEEE, pp 1–8
 22. Saponas TS, Tan DS, Morris D, Balakrishnan R, Turner J, Landay JA (2009) Enabling always-available input with muscle-computer interfaces. In: *Proceedings of the 22nd annual ACM symposium on user interface software and technology*. ACM, pp 167–176
 23. Khushaba RN (2014) Correlation analysis of electromyogram signals for multiuser myoelectric interfaces. *IEEE Trans Neural Syst Rehabil Eng* 22(4):745–755
 24. Matsubara T, Morimoto J (2013) Bilinear modeling of emg signals to extract user-independent features for multiuser myoelectric interface. *IEEE Trans Biomed Eng* 60(8):2205–2213
 25. Orabona F, Castellini C, Caputo B, Fiorilla AE, Sandini G (2009) Model adaptation with least-squares svm for adaptive hand prosthetics. In: *2009 ICRA'09. IEEE international conference on robotics and automation*. IEEE, pp 2897–2903
 26. Chattopadhyay R, Pradhan G, Panchanathan S (2011) Subject independent computational framework for myoelectric signals. In *2011 IEEE instrumentation and measurement technology conference (I2MTC)*. IEEE, pp 1–4
 27. Pan SJ, Yang Q (2010) A survey on transfer learning. *IEEE Trans Knowl Data Eng* 22(10):1345–1359
 28. Pan SJ, Kwok JT, Yang Q (2008) Transfer learning via dimensionality reduction. *AAAI* 8:677–682
 29. Marinelli B, Kang M, Martini M, Zech JR, Titano J, Cho S, Costa AB, Oermann EK (2019) Combination of active transfer learning and natural language processing to improve liver volumetry using surrogate metrics with deep learning. *Radiology: Artificial Intelligence* 1(1):e180019
 30. Yin X, Yu X, Sohn K, Liu X, Chandraker M (2018) Feature transfer learning for deep face recognition with long-tail data. arXiv:1803.09014
 31. Zhao Z, Chen Y, Liu J, Shen Z, Liu M (2011) Cross-people mobile-phone based activity recognition. In *22nd international joint conference on artificial intelligence*
 32. Wang Z, Guo B, Yu Z, Zhou X (2018) Wi-fi csi-based behavior recognition: From signals and actions to activities. *IEEE Commun Mag* 56(5):109–115
 33. Yu Z, Du H, Yi F, Wang Z, Guo B (2019) Ten scientific problems in human behavior understanding. *CCF Trans Pervasive Comput Int* 1(1):3–9
 34. Goussies NA, Ubalde S, Mejail M (2014) Transfer learning decision forests for gesture recognition. *The J Mach Learn Res* 15(1):3667–3690
 35. Costante G, Galieni V, Yan Y, Fravolini ML, Ricci E, Valigi P (2014) Exploiting transfer learning for personalized view invariant gesture recognition. In: *2014 IEEE international conference on acoustics, speech and signal processing (ICASSP)*. IEEE, pp 1250–1254
 36. Ozcan T, Basturk A (2019) Transfer learning-based convolutional neural networks with heuristic optimization for hand gesture recognition. *Neural Comput Appl* 31(12):8955–8970
 37. Cote-Allard U, Fall CL, Campeau-Lecours A, Gosselin C, Laviolette F, Gosselin B (2017) Transfer learning for semg hand gestures recognition using convolutional neural networks. In: *2017 IEEE international conference on systems, man, and cybernetics (SMC)*. IEEE, pp 1663–1668
 38. Bu Q, Yang G, Feng J, Ming X (2018) Wi-fi based gesture recognition using deep transfer learning. In: *2018 IEEE SmartWorld, ubiquitous intelligence & computing, advanced & trusted computing, scalable computing & communications, cloud & big data computing, internet of people and smart city innovation (SmartWorld/SCALCOM/UIC/ATC/CBDCCom/IOP/SCI)*. IEEE, pp 590–595
 39. Oskoei MA, Hu H (2007) Myoelectric control systems—a survey. *Biomed Signal Process Control* 2(4):275–294
 40. Rechy-Ramirez EJ, Hu H (2011) Stages for developing control systems using emg and eeg signals: a survey, School of computer science and electronic engineering, University of Essex, pp 1744–8050
 41. Phinyomark A, Limsakul C, Phukpattaranont P (2009) A novel feature extraction for robust emg pattern recognition. arXiv:0912.3973
 42. Zeng Z-Q, Yu H-B, Xu H-R, Xie Y-Q, Gao J (2008) Fast training support vector machines using parallel sequential minimal optimization. In: *2008 3rd international conference on intelligent system and knowledge engineering*, vol 1. IEEE, pp 997–1001
 43. Cover TM, Hart PE et al (1967) Nearest neighbor pattern classification. *IEEE Trans Inf Theory* 13(1):21–27
 44. Liaw A, Wiener M et al (2002) Classification and regression by randomforest. *R news* 2(3):18–22
 45. Wold S, Esbensen K, Geladi P (1987) Principal component analysis. *Chemometr Intell Laboratory Syst* 2(1-3):37–52
 46. Pan SJ, Tsang IW, Kwok JT, Yang Q (2011) Domain adaptation via transfer component analysis. *IEEE Trans Neural Netw* 22(2):199–210
 47. Long M, Wang J, Ding G, Sun J, Yu PS (2013) Transfer feature learning with joint distribution adaptation. In: *Proceedings of the IEEE international conference on computer vision*, pp 2200–2207
 48. Wang J, Chen Y, Hao S, Feng W, Shen Z (2017) Balanced distribution adaptation for transfer learning. In: *2017 IEEE*

- international conference on data mining (ICDM). IEEE, pp 1129–1134
49. Gong B, Shi Y, Sha F, Grauman K (2012) Geodesic flow kernel for unsupervised domain adaptation. In: 2012 IEEE conference on computer vision and pattern recognition. IEEE, pp 2066–2073
 50. Liu J, Li J, Lu K (2018) Coupled local–global adaptation for multi-source transfer learning. *Neurocomputing* 275:247–254
 51. Wang J, Chen Y, Hu L, Peng X, Philip SY (2018) Stratified transfer learning for cross-domain activity recognition. In: 2018 IEEE international conference on pervasive computing and communications (PerCom). IEEE, pp 1–10
 52. Côté-Allard U, Fall CL, Drouin A, Campeau-Lecours A, Gosselin C, Glette K, Laviolette F, Gosselin B (2019) Deep learning for electromyographic hand gesture signal classification using transfer learning. *IEEE Trans Neural Syst Rehabil Eng* 27(4):760–771

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.