

EMG-based Multi-User Hand Gesture Classification via Unsupervised Transfer Learning Using Unknown Calibration Gestures

Haojie Shi¹, Xinyu Jiang¹, Chenyun Dai¹, *Member, IEEE*, and Wei Chen¹, *Senior Member, IEEE*

Abstract—The poor generalization performance and heavy training burden of the gesture classification model contribute as two main barriers that hinder the commercialization of sEMG-based human-machine interaction (HMI) systems. To overcome these challenges, eight unsupervised transfer learning (TL) algorithms developed on the basis of convolutional neural networks (CNNs) were explored and compared on a dataset consisting of 10 gestures from 35 subjects. The highest classification accuracy obtained by CORrelation Alignment (CORAL) reaches more than 90%, which is 10% higher than the methods without using TL. In addition, the proposed model outperforms 4 common traditional classifiers (KNN, LDA, SVM, and Random Forest) using the minimal calibration data (two repeated trials for each gesture). The results also demonstrate the model has a great transfer robustness/flexibility for cross-gesture and cross-day scenarios, with an accuracy of 87.94% achieved using calibration gestures that are different with model training, and an accuracy of 84.26% achieved using calibration data collected on a different day, respectively. As the outcomes confirm, the proposed CNN TL method provides a practical solution for freeing new users from the complicated acquisition paradigm in the calibration process before using sEMG-based HMI systems.

Index Terms—High-density sEMG, hand gesture classification, unsupervised transfer learning, deep learning.

I. INTRODUCTION

HUMAN-MACHINE Interaction (HMI) systems recognize user's intent based on biological signals, and

Manuscript received 30 October 2023; revised 4 February 2024; accepted 27 February 2024. Date of publication 4 March 2024; date of current version 12 March 2024. This work was supported in part by the National Natural Science Foundation of China under Grant 62001122, in part by Shanghai Municipal Science and Technology Project under Grant 20510710500, and in part by the Natural Science Foundation of Shanghai under Grant 20ZR1403400. (Corresponding authors: Chenyun Dai; Wei Chen.)

This work involved human subjects or animals in its research. Approval of all ethical and experimental procedures and protocols was granted by the Ethics Committee of Fudan University under Approval No: BE2035.

The authors are with the Center for Intelligent Medical Electronics, School of Information Science and Technology, Fudan University, Shanghai 200433, China (e-mail: chenrundai@fudan.edu.cn; w_chen@fudan.edu.cn).

Digital Object Identifier 10.1109/TNSRE.2024.3372002

translate them into commands of machine behaviors [1]. Surface Electromyography (sEMG), as one of the widely used biological modalities, has shown great potential in both clinical research and commercial applications [2]. Compared to other applicable signals (e.g., signals from brain or spinal cord), sEMG, which can reflect abundant information of user's motion intention, is non-invasive and easy to be obtained through the electrodes placed on skins. Nowadays, the sEMG-based HMI systems have been widely used in various fields, such as prosthesis control [3], [4], [5], motor rehabilitation [6], [7], [8], biomechanics [9], [10], [11] and ergonomics [12], [13], [14].

For decades, hand gesture recognition has been one of the most intuitive and commonly used techniques for sEMG-based HMIs. A large number of studies have established different models to achieve a precise classification accuracy of gesture recognition. For example, in [15], researchers used a Support Vector Machine (SVM) based on Gaussian radial basis function to classify five gestures and achieved an average accuracy of 89%. In [2], four different traditional classifiers were tested to find the best configuration for identifying 17 diverse hand movements. In recent years, with the development of advanced artificial intelligence technology, hand gesture recognition with deep learning has become a research hotspot in the related fields. Compared with traditional statistical learning algorithms, deep learning has stronger learning ability and better robustness on large datasets, which has proven its superiority in many classification tasks. Wei et al. [16] used a multi-stream divide-and-conquer CNN framework to learn the correlation between individual muscles and specific gestures. Hu et al. [17] proposed a hybrid CNN-RNN network structure as well as a new sEMG image representation for sEMG pattern recognition. Although these relevant studies have achieved higher classification accuracy than traditional classifiers, most of them are subject-specific. Specifically, the HMI systems require labeled data and model calibration procedures for a specific subject before use. Such a time-consuming process largely reduces the convenience of practical use. However, although directly applying models trained from other subjects to a new user can reduce the inconvenience of calibration, it leads to

significant errors. Therefore, developing a subject-generalized model with high recognition accuracy and minimal calibration time is crucial to advance user experience.

Accordingly, many researchers have introduced transfer learning (TL) algorithms in their work [18], [19], [20], [21], [22]. TL can utilize a large number of prior knowledge to train a universal model for new scenarios with a few data, which provides the possibility to solve the problems of model generalization and calibration burden [21]. Vidovic et al. [22] used a small amount of calibration data to learn a linear transformation to project the original and new data into a common subspace, and then trained and tested two Bayesian multi-class classifiers, namely Linear Discriminant Analysis (LDA) and Quadratic Discriminant Analysis (QDA) in that space. With the adaptation, their classification accuracy increased from 75% to 92% for 8 gestures classification task. Chen et al. [21] fine-tuned two networks on target datasets based on a pre-trained model to reduce the training time and guarantee the recognition accuracy of more than 90% for 10 gestures classification task. Ulysse et al. [20] proposed Progressive Neural Networks (PNN) to train a target network with only one cycle data for training. Together with a source network trained on the original datasets, a higher average accuracy of 93.36% can be achieved compared to that of 86.77% for non-TL algorithm for 7 gestures classification task. Although these relevant studies improve the generalization ability of gesture classification models with minimal calibration data or reduce model training time, most studies use supervised TL methods, which still require the labeled data from target users. Moreover, for the deep learning-based studies, researchers usually feed raw sEMG data into their network instead of hand-crafted features. Considerable heuristic knowledge contained in these classic features are ignored, resulting in the requirement of additional data to train the network.

In this paper, we used classic EMG features with unsupervised TL algorithms embedded in a CNN-based deep learning model to implement a cross-subject gesture classification model. Several TL algorithms were compared with the highest accuracy obtained from Correlation Alignment for Deep Domain Adaptation (Deep CORAL). The proposed method can achieve $> 10\%$ accuracy improvement over the baseline without TL technique. Code and data are available on https://github.com/Knight99812/EMG_DeepTL. The main contents and contributions are summarized as follows:

- 1) We validated different TL strategies on a large hybrid dataset (including one public dataset and two private datasets) of 10 gestures from 35 subjects. In this dataset, data from 34 subjects were used as training set and source domain, and data from the remaining 1 subject (considered as target user) were divided into validation set and target domain. By transferring knowledge learned from source domain to target domain, we achieved 3%-10% accuracy improvement over the baseline. The highest accuracy reaches more than 90% which is state-of-the-art to our knowledge for unsupervised gesture classification of cross-subject tasks.

- 2) We investigated the relationship between the number of trials per gesture used for transfer and the model performance.

The results indicate our scheme can reduce the calibration time for new users. Moreover, the possibility of using new gestures that were completely different from the training set for model transfer was verified. The results show that new users do not need to follow a fixed calibration process to obtain a well-performed classification model.

The remainder of this paper is organized as follows. Section II describes the information of database and experimental protocol. The signal preprocessing methods, classification algorithms and validation protocols are introduced in Section III. Moreover, the results are presented in Section IV. Section V discusses and concludes the proposed paper.

II. MATERIALS

In this study, three hand gesture datasets (named V1, V2, and V3) were used. They all follow a similar data acquisition paradigm except for the types of hand gestures involved, and they are collected under the supervision of the same principal investigator. The V1 is one session (pattern recognition session) of an open-sourced dataset containing 34 hand gestures from 20 subjects collected on two separate days [23]. The V2 and V3 are two private datasets with 10 and 11 subjects respectively. However, the V2 and V3 only contain 10 commonly used hand gestures which are also involved in V1. Therefore, we combined the three datasets as one large hybrid dataset to verify the generalization ability of our proposed model. Overall, the final dataset consisted of 10 common gestures (see Fig. 2) from 41 (20+10+11) subjects collected on 2 separate days. However, 6 subjects mistakenly performed at least one specific type of gesture during data collection. To avoid possible bias in TL algorithm due to gesture type, we used the remaining 35 subjects (22 male, 13 female; aged 21-34 years, all right-handed) for further analysis.

A. Subjects

All the subjects involved were informed of the detailed research purpose and experimental procedures in advance, and provided the informed consent. The study was supervised and approved by the ethics committee of Fudan University (approval number: BE2035).

B. Data Acquisition

The data acquisition process of the three datasets is briefly described here. A more detailed description can be found in the previous study [23].

The 256-channel high-density sEMG (HD-sEMG) array was applied in our work because of its high spatiotemporal resolution. Specifically, four 8×8 electrode arrays were mounted on both extensor and flexor (two arrays for each) of the forearm to obtain HD-sEMG signals, as shown in Fig. 1. The right leg drive and reference electrodes were placed on the head of the ulna and the elbow, respectively. The signals were acquired by the Quattrocento system (OT Bioelettronica, Torino, Italy), with a passband filtering of 10-500 Hz, an amplifier gain of 150, a sampling rate of 2048 Hz and a resolution of 16 bits.

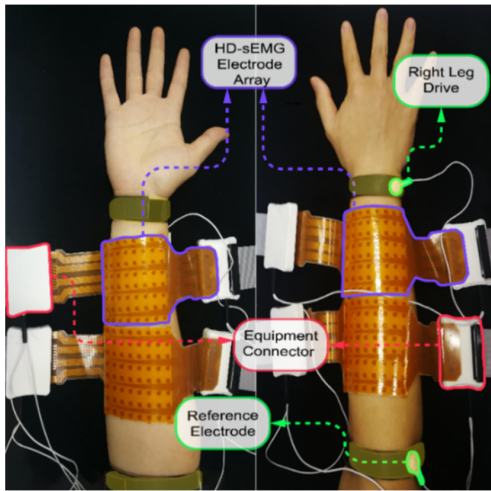


Fig. 1. Electrode placement on the volar and dorsal side of the forearm.

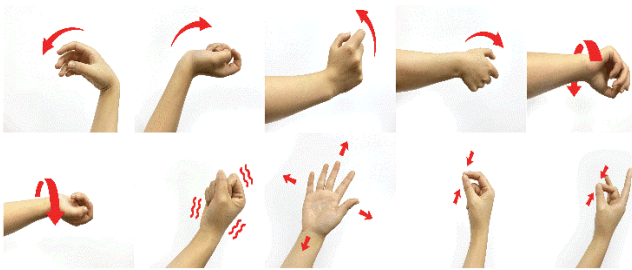


Fig. 2. Involved gestures. The gestures were: 1) Wrist Flexion (WF), 2) Wrist Extension (WE), 3) Wrist Radial (WR), 4) Wrist Ulnar (WU), 5) Wrist Pronation (WP), 6) Wrist Supination (WS), 7) Hand Close (HC), 8) Hand Open (HO), 9) Thumb + Index finger pinch (TI), 10) Thumb + Middle finger pinch (TM).

The 10 hand gestures included in this study are shown in Fig. 2. They are all common gestures in daily life involving the combination of various states of finger and wrist joints. The entire acquisition process was guided by a self-built Graphic User Interface (GUI), and supervised by at least one experiment assistant. During the experiment, the subject was required to perform two repeated trials for each single gesture before they continued to the next one. Each trial included three 1-s dynamic tasks (from the resting state to a designated gesture, and finally back to the initial state) and a 4-s gesture-and-hold task (from the resting state to the end of the ongoing gesture). Only dynamic tasks were used in our experiment. A 2-second inter-trial and a 5-second inter-gesture resting period were provided to avoid the impact of muscle fatigue. If the subject performed a wrong gesture or missed one certain trial, they were asked to inform the experiment assistant. Then, these tasks were removed from the final dataset. On average, 0.43 ± 0.86 trials for each subject were excluded. Each subject performed the experiment on two different days, with an interval from 1 to 22 days. The data from two days were noted as Day 1 and Day 2, respectively.

III. METHODS

The cross-subject hand gesture classification framework based on the transfer learning strategy is shown in Fig. 3. We describe the main step as follows.

A. Data Preprocessing

The acquired HD-sEMG signals were segmented into corresponding tasks with a window length of 1 second. For each task, the first 0.25s reaction time was removed to avoid introducing interference. 4170 tasks were obtained from the hybrid dataset after data segmentation. Then they were filtered with a 10-500 Hz band pass Butterworth filter. A zero-phase filter processing was used to solve the non-linear phase issues, bidirectionally with 8-order for each direction. A notch filter was also used to attenuate power line interference at 50 Hz and its harmonic components up to 400 Hz.

B. Feature Extraction

40 classic features were selected to extract the EMG features based on the previous study [24]. For each feature, a 256-dimensional feature vector was extracted with each dimension representing a specific channel (note: Auto-Regressive Coefficient (AR) feature has 1024 dimensions because it had 4 values per channel). Then we concatenated these feature vectors to obtain a constituent matrix with 11008 dimensions ($39 \times 256 + 1024$) representing a specific trial of a gesture. Therefore, the size of the feature matrix should be 60 ($10 \text{ gestures} \times 6 \text{ repetitions}$) $\times 11008$ for one-day data of each subject.

C. Outlier Recovery

Despite the high spatiotemporal resolution of HD-sEMG, it commonly has a proportion of channels with poor signal quality. The outliers may greatly degenerate classification performance and thus need to be handled before model training. Since an 8×8 feature map can be re-arranged based on channel location of each electrode array, we detected the outliers within each feature if any value in feature map was more than three standard deviations away from the mean value of the ensemble. Then, these outliers were smoothed through replacing the original value by the average of values in their neighboring channels.

D. Feature Selection

We compared the cross-subject classification performance of 40 classic features to obtain the optimum feature combination for our gesture recognition task. Since it is impossible to investigate all combinations, a heuristic search method, Sequential Forward Selection (SFS), was applied. Specifically, the searching process of SFS started with an initial feature, and then was gradually extended by adding one feature whose inclusion resulted in the highest classification accuracy at each forward step. The stepping finally terminated when the accuracy was no longer improved. Since SFS was a greedy algorithm which may fall into a local optimum solution, feature combinations with different initial features were examined.

E. Classifiers

Several traditional classifiers and a deep classifier were used in this study to explore the performance of both conventional machine learning and deep learning methods on our hybrid dataset. Moreover, multiple transfer learning modules were embedded in these classifiers to compare their improvement.

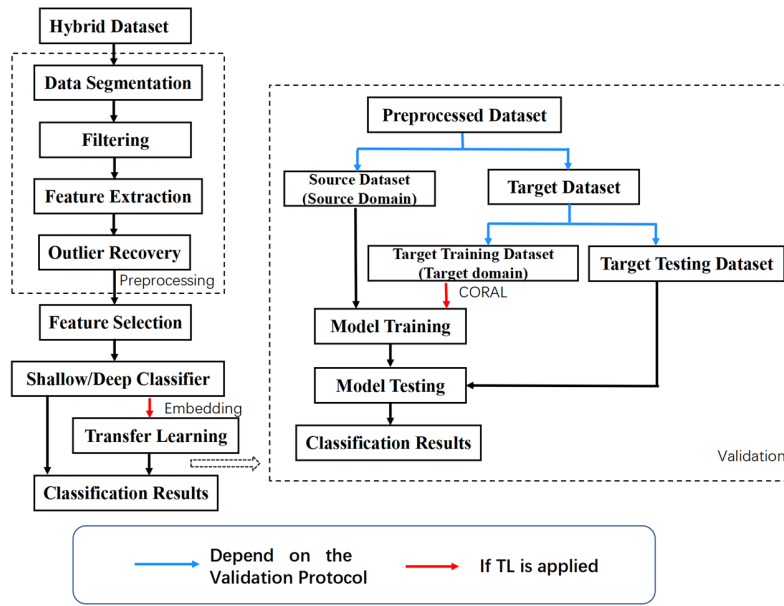


Fig. 3. Block diagram of the proposed cross-subject hand gesture classification framework. The left side was the complete flow chart, including the preprocessing part and the subsequent validation part. The right side was the detailed explanation of the validation part.

1) *Traditional Classifiers*: 4 common traditional classifiers, namely K-Nearest Neighbors (KNN), LDA, SVM, and Random Forest (RF), were applied to the gesture classification task. Once the optimal feature combination was determined, we extracted the corresponding feature matrix with Z-score normalization step within each feature. Then, the processed feature matrix was fed into these classifiers for training. In addition, the hyperparameters were selected as follows: 1) for KNN, the number of neighbors was set to 8; 2) for SVM, LibSVM was used to automatically search for the optimal hyperparameters; 3) for RF, the number of random trees was set to 300.

2) *Deep Classifier*: This study used CNN as the underlying framework, given its great success in sEMG pattern recognition validated by previous studies [16], [18], [21]. We reshaped the 256-dimensional vector for each feature selected from feature selection step into a 16×16 feature map according to the actual 2-D location of the electrodes in Fig. 1. Then, different features can be regarded as different channels in CNN, similar to the RGB channels of pictures in computer vision. These operations aimed to take the advantages of spatial information obtained from HD-sEMG, and facilitate CNN to extract this spatial information. The size of processed input for the CNN had four dimensions: number of tasks (6 repetitions \times 10 gestures = 60), number of features selected (based on the results of SFS feature selection step), width and height (16×16 correspond to actual electrode placement).

The structure of our CNN was based on many well-known architectures developed for image classification. However, the structures were greatly simplified to avoid overfitting since our task with a small amount of training samples was not complex. In detail, the proposed CNN comprised 15 layers including two convolutional layers (with 16 and 32 filters respectively) and three fully-connected layers, each followed by a ReLU layer, two max-pooling layers, two dropout layers, and finally

TABLE I
SUMMARY OF USED HYPERPARAMETERS IN DEEP NEURAL NETWORK

PARAMETER	VALUE
TRAINING EPOCHS	60
BATCH SIZE	20
CONVOLUTION KERNEL SIZE	3×3
MAX POOL SIZE	2×2
LEARNING RATE	10^{-3}
DROPOUT PROBABILITY	0.5
MOMENTUM IN SGD	0.9
TOTAL PARAMETERS	129,914

a classification layer (see Fig. 4). Specifically, the ReLU layer was adopted to avoid vanishing gradient problem. The max-pooling layer was to reduce the dimensions of features and refine the information. The dropout layer was to prevent the overfitting problem. The hyperparameters and other details are described in Table I.

F. Transfer Learning Methods

Due to the large difference in the distribution of sEMG signals across individuals, the performance of a well-trained gesture classification model for a specific group of subjects may degenerate significantly when applied to a new user. This huge difference can be derived from many factors, for example, individual neuromuscular anatomy, signal quality, position of electrode placement, etc. Therefore, to establish a gesture classification model with a powerful generalization ability, TL is an indispensable technique attracting considerable research interests.

In this paper, we introduced several completely unsupervised TL approaches, also termed Domain Adaptation (DA), to learn domain-invariant features by directly reducing the discrepancy among different distributions. The working mechanism of these methods was similar and can be described in our scenarios as follows: Let $\mathcal{X} = \{x_1, x_2, \dots, x_n\}$ be the input

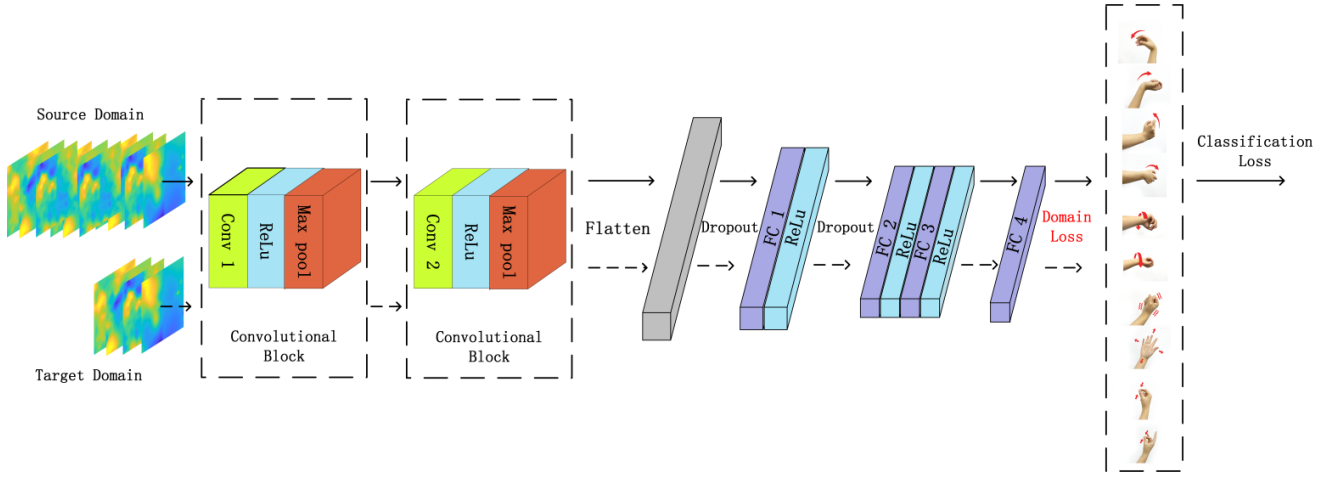


Fig. 4. The architecture of the deep classifier CNN with TL algorithms.

feature space (e.g., sEMG features) and $\mathcal{Y} = \{y_1, y_2, \dots, y_n\}$ the output space (e.g., gesture categories) of the classification task. The source domain with n labeled samples represents sEMG dataset acquired from a large group of subjects for model training, denoted as $\mathcal{S} = \{(x_1^s, y_1^s), \dots, (x_n^s, y_n^s)\}$. Besides, the target domain with m ($m \ll n$) unlabeled samples represents sEMG dataset acquired from a new subject for transfer and model testing, denoted as $\mathcal{T} = \{x_1^t, \dots, x_m^t\}$. The goal is to estimate a strong predictor from \mathcal{S} and \mathcal{T} to classify new gestures from the target user. These DA methods achieve it by adding an inter-domain difference loss to the loss function of model. Eight widely used algorithms (DaNN, DDC, RTN, DAN, JAN, B-JMMD, AssociativeDA, CORAL) based on five different loss measures were compared in this paper.

1) Maximum Mean Discrepancy (MMD): The MMD is one of the most commonly used measures to measure the difference between two probability distributions from their samples. It is an effective criterion that compares distributions without initially estimating their density functions [25], which is defined as

$$\mathcal{MMD}[\mathcal{F}, p, q] = \sup_{f \in \mathcal{F}} (\mathbb{E}_p[f(x^s)] - \mathbb{E}_q[f(x^t)]) \quad (1)$$

where p and q are two probability distributions of source and target domain; \mathcal{F} is a class of functions $f: \mathcal{X} \rightarrow \mathbb{R}$ defined as the unit ball in a universal Reproducing Kernel Hilbert Space (RKHS); $f(x)$ represents the dot product of f and $\varphi(x)$, where $\varphi(x)$ maps the variable to RKHS through kernel function.

Based on (1) and the kernel trick, we can rewrite the kernelized empirical estimate of MMD for our work as

$$\begin{aligned} \mathcal{MMD}[\mathcal{F}, \mathcal{S}, \mathcal{T}] = & \left(\frac{1}{n(n-1)} \sum_{i \neq j}^n k(x_i^s, x_j^s) \right. \\ & + \frac{1}{m(m-1)} \sum_{i \neq j}^m k(x_i^t, x_j^t) \\ & \left. - \frac{2}{mn} \sum_{i=1}^n \sum_{j=1}^m k(x_i^s, x_j^t) \right)^{\frac{1}{2}} \quad (2) \end{aligned}$$

where $k(\cdot, \cdot)$ is a kernel function, e.g., a Gaussian kernel was used in this work.

By measuring MMD at different stages in the model architecture, different MMD-based methods were generated.

We selected DaNN [25], DDC [26] and RTN [27] as representatives to test. Their difference is that DaNN consists of only two fully-connected layers; DDC deepens the network and introduces an adaptive layer; and RTN measures MMD of two layers at the same time and introduces an additional entropy loss.

2) Multiple Kernel variant MMD (MK-MMD): Intuitively, MMD is the upper bound of the difference between the expectations of two distributions after being mapped. The way of mapping, or the choice of kernel function, has a direct impact on the results of MMD. However, so far we have no theoretical support for which kernel function should be selected, thus the MK-MMD was introduced. It was formalized to jointly maximize the two-sample test power and minimize the failure of rejecting a false null hypothesis [28]. The characteristic kernel associated with the feature map is defined as the convex combination of v Positive Semi-Definite (PSD) kernels $\{k_u\}$

$$\mathcal{K} \triangleq \left\{ k = \sum_{u=1}^v \beta_u k_u : \sum_{u=1}^v \beta_u = 1, \beta_u \geq 0, \forall u \right\} \quad (3)$$

where the constraints on coefficients β_u are imposed to guarantee that the derived multiple kernel k is characteristic [28]. The MK-MMD frees us from manually selecting a special kernel function and enhances the model performance by leveraging several different kernels. The deep TL algorithm based on MK-MMD used in this work is DAN [28].

3) Joint MMD (JMMD): Due to the characteristic of MMD, most existing methods apply it to measure the discrepancy between marginal distributions of source and target domains. However, in classification tasks, conditional distributions are equally important. By taking both marginal and conditional distributions into consideration, we can get its measurement of joint distribution called JMMD, which is defined as

$$\begin{aligned} \mathcal{JMMD}[\mathcal{F}, \mathcal{S}, \mathcal{T}] = & \left(\frac{1}{n^2} \sum_{i,j=1}^n \prod_{l \in \mathcal{L}} k^l(z_i^{sl}, z_j^{sl}) \right. \\ & + \frac{1}{m^2} \sum_{i,j=1}^m \prod_{l \in \mathcal{L}} k^l(z_i^{tl}, z_j^{tl}) \\ & \left. - \frac{2}{mn} \sum_{i,j=1}^{n,m} \prod_{l \in \mathcal{L}} k^l(z_i^{sl}, z_j^{tl}) \right)^{\frac{1}{2}} \quad (4) \end{aligned}$$

where \mathcal{L} stands for the part of the network that needs to measure the domain discrepancy; z^l represents the output of l_{th} layer.

JMMD applies non-uniform weights on the kernel function to reflect the influence of other variables in other layers $l \notin \mathcal{L}$. This captures the full interactions between different variables in the joint distributions, which is crucial for DA [29]. The deep TL algorithm JAN [29] and its improved version B-JMMD [30] which adds a balancing coefficient to the JMMD loss are employed to investigate the effect of JMMD in our scenario.

1) *Association Loss (ASSOC)*: ASSOC loss can be divided into walker loss and visit loss. Denote the embedding vectors derived from the labeled source domain data and the unlabeled target domain data by network as \mathcal{A} and \mathcal{B} . Then imagine a walker going from \mathcal{A} to \mathcal{B} according to the mutual similarities, and back [31]. The walk is correct once he ended up at the same class as he started from. The walk loss penalizes incorrect walks and encourages walks to a uniform probability distribution of the correct class, which is defined as the cross-entropy \mathcal{H} between the target distribution of correct round-trips \mathcal{RT} and the round-trip probabilities \mathcal{P}^{aba} ,

$$\mathcal{L}_{walker} = \mathcal{H}(\mathcal{RT}, \mathcal{P}^{aba}) \quad (5)$$

with the target distribution

$$\mathcal{RT}_{ij} = \begin{cases} 1/n_i & \text{class}(A_i) = \text{class}(A_j) \\ 0 & \text{else} \end{cases} \quad (6)$$

where n_i is the number of samples of class A_i , and the two-step round-trip probability

$$\mathcal{P}^{aba} = (\mathcal{P}^{ab} \mathcal{P}^{ba})_{ij} \quad (7)$$

where \mathcal{P}^{ab} is the transition probability from A_i to B_j . The visit loss is a regularizer to make each target sample be visited with equal probability, which is defined as the cross-entropy \mathcal{H} between the uniform distribution over target samples and the probability of visiting some target sample start in any source sample,

$$\mathcal{L}_{visit} = \mathcal{H}(\mathcal{V}, \mathcal{P}^{visit}) \quad (8)$$

where

$$\mathcal{P}_j^{visit} = \sum_{x_i \in \mathcal{S}} \mathcal{P}_{ij}^{ab}, \quad \mathcal{V}_j = \frac{1}{|\mathcal{B}|} \quad (9)$$

2) *CORrelation Alignment (CORAL)*: Besides MMD and its variants, a different distance metric CORAL, which reduces the discrepancy between domains by aligning their second-order statistics [32], was also performed. It is defined as

$$\mathcal{COHAL} = \frac{1}{4d^2} \|C_S - C_T\|_F^2 \quad (10)$$

where d denotes the dimensions of the input feature vector; $\|\cdot\|_F^2$ denotes the squared matrix Frobenius norm; C_S and C_T are the feature covariance matrices of the source and target domains, respectively. They can be given by

$$C_S = \frac{1}{n-1} \left(D_S^T D_S - \frac{1}{n} (\mathbf{1}^T D_S)^T (\mathbf{1}^T D_S) \right) \quad (11)$$

$$C_T = \frac{1}{m-1} \left(D_T^T D_T - \frac{1}{m} (\mathbf{1}^T D_T)^T (\mathbf{1}^T D_T) \right) \quad (12)$$

where n and m are the same as defined in previous section; D_S and D_T is the input feature matrix; and $\mathbf{1}$ is a column vector with all elements equal to 1.

Additionally, when combined with traditional classifiers, CORAL is a subspace-based TL method that transfers knowledge in the feature aspect [32]. The optimization function is:

$$\min_A \|A^T C_S A - C_T\|_F^2 \quad (13)$$

where A is a second-order feature transformation matrix; C_S , C_T and $\|\cdot\|_F^2$ are defined the same as in deep version. The features in source domain were transferred into the same space as those in target domain through the mapping matrix A . Then, different machine learning classifiers were trained in the projected space.

G. Evaluation

Three different validation protocols were used to evaluate model performance.

1) *Protocol 1*: to maximize the use of data, “leave-one-subject-out” cross-validation protocol was implemented. Each subject was treated as the target user in turn, whose model was trained by the data acquired from the remaining 34 subjects as source domain. Then, a portion of unlabeled data from the target user (e.g., 2 trials for each gesture) were used to reduce the domain difference between source domain and target domain if TL was applied. The classification model was tested on the remaining data (except the trials used for transfer) from Day 1 or Day 2 of the target user. Thus, two accuracy values (Day 1 and Day 2) were obtained for each subject, and their average of two days was set as the evaluation results. This protocol was mainly to evaluate the generalization ability of our gesture classification models when applied to new users, and to compare the performance between different models both with and without TL.

2) *Protocol 2*: to simulate real-world scenarios, we also accessed the cross-day robustness of our model with TL methods. Specifically, for target user, unlabeled data in Day 1 was employed as the target domain to implement TL, and only data in Day 2 constituted the testing set whose accuracy was reported. This protocol was used to demonstrate that our model with TL also have a cross-day robustness for new users even in the absence of current-day data, significantly reducing the calibration time for practical use.

3) *Protocol 3*: to further validate the robustness of the proposed model, we investigated the influence of the types of gesture in target domain if data in TL step were different from the training data in source domain or testing data in target domain. In this protocol, only V1 dataset, which has 24 additional gestures (V1 has 34 gestures, and 10 gestures were used in *Protocol 1* and 2), was used for validation. In detail, the training set and testing set from 10 gestures were the same as that in *Protocol 1* and 2. The data of 24 new gestures from the target user in a specific Day were considered as a random transfer pool. We randomly selected a portion of data from this pool to carry out the calibration for target

domain. This protocol can verify that our model with TL has a strong robustness against the types of gestures for new users even if they do not necessarily perform the same gestures as training or testing stage during calibration.

It is worth mentioning that we compared the performance of several models both with and without TL in all 3 validation protocols. To get a fairer comparison result, the training set composition and testing set composition was the same for TL and non-TL scenarios. The difference between them was that when using TL, the target domain was employed to map features to a common subspace by traditional model or to calculate a domain difference loss by deep model. When TL was not used, the target domain was simply ignored.

The main evaluation results are organized and presented in next chapter as follows:

First, we employed SFS to find the optimal feature combination for cross-subject gesture classification task in the classic sEMG feature set. Second, based on the optimal feature combination, we compared the classification performance of different classifiers, including four common traditional classifiers and a deep CNN-based classifier. Third, we compared the performance variation of deep classifier embedded with several TL (DA) methods. The best TL algorithm termed CORAL was obtained. Moreover, the comparison between traditional classifiers and deep classifier with CORAL were also performed. Fourth, different *Protocols* (1, 2 and 3) were further performed for deep CORAL to test the robustness of the proposed method.

All the traditional classifiers were implemented in Matlab 2021b and the deep classifier in Pytorch (version:1.10.0). All these evaluations were conducted on the hybrid dataset and all the results were reported as average over at least three random training/testing runs.

H. Statistical Analysis

Because the results followed tests of normality, the performance differences were tested using multivariate repeated measures analysis of variance (RANOVA). When degree of sphericity (ϵ) was < 0.75 , degrees of freedom was adjusted by Greenhouse-Geisser; and when $0.75 < \epsilon < 1$, it was adjusted by Huynh-Feldt. Post hoc pair-wise comparisons were conducted using paired t-tests with Bonferroni correction for multiple comparisons. The differences were considered significant for $p < 0.05$.

IV. RESULTS

A. Comparison of Different Feature Combination

As mentioned above, we employed an intuitive but efficient approach, SFS, to find the optimal feature combination for cross-subject gesture classification task in *Protocol 1*. Since this was an iterative process, LDA was chosen as the classifier considering its low time-consuming and relatively high accuracy. We started the algorithm with 5 different initial features and recorded the top 5 features with highest accuracy in each step. The top 10 feature combinations were shown in [Table II](#).

For all the acquired combinations, the ZC, SSC and SKEW had the highest selection frequency (almost appear in all top

TABLE II
SUGGESTED FEATURE COMBINATIONS AND THE CORRESPONDING FEATURE DIMENSIONS AND CLASSIFICATION ACCURACY

No.	FEATURE COMBINATION	DIMENSIONS	ACCURACY(%)
1	LTKEO+ZC+SSC+SKEW	1024	70.08
2	LDASDV+ZC+SSC+SKEW	1024	69.77
3	MFL+LDASDV+ZC+SSC+SKEW	1280	69.68
4	LTKEO+ZC+SSC+SKEW+ASM	1280	69.59
5	LTKEO+ZC+SSC+SKEW+ASS	1280	69.56
6	LTKEO+ZC+SSC+SKEW+EMAV	1280	69.40
7	LDASDV+ZC+SSC+SKEW+ASS	1280	69.40
8	LDASDV+ZC+SSC+SKEW+ASM	1280	69.38
9	LDAMV+ZC+SSC+SKEW+ASM	1280	69.24
10	MFL+ZC+SSC+LDASDV	1024	69.14

10 combinations), indicating their superiority in cross-subject gesture classification task. The optimal feature combination was LTKEO + ZC + SSC + SKEW, with a 70.08% classification accuracy achieved. Thus, this combination was selected as the model input for the further analysis.

B. Results of Protocol 1

When TL was not applied, the deep classifier CNN obtained the highest classification accuracy (79.95%). The traditional classifiers, KNN, LDA, SVM and RF achieved 68.42%, 69.84%, 79.26% and 77.94% classification accuracy, respectively (see [Fig. 6](#)). The results of RANOVA showed that CNN, SVM and RF achieved higher classification accuracy than KNN and LDA ($p < 0.05$).

We embedded eight TL algorithms with the best-performing CNN as the underlying framework. These methods were distinguished by different domain difference metrics and different transfer stages. However, for a fair comparison, their network structure and hyperparameter settings were the same as the baseline version. The results are shown in [Table III](#). CORAL achieved higher classification accuracy than other TL algorithms in *Protocol 1* and *Protocol 3* ($p < 0.05$).

Since the CORAL achieved more than 90% classification accuracy (90.19%, 94.33% for V1), the rest analysis are mostly based on this TL algorithm. As shown in [Fig. 5](#) and [Fig. 9](#), CORAL significantly improved the classification accuracy on almost all subjects and all gestures. To demonstrate its effect more intuitively, we selected a representative subject to visualize its training process both with and without CORAL. If the loss function included a CORAL loss, the feature difference between source domain and target domain learned by the network was limited within a small range. However, if only the classification loss was considered, the network focused on learning inherent characteristics of the source domain, resulting in an increasing distance between two domains as the iteration (see [Fig. 10](#)). Moreover, the t-SNE embeddings of gestures from the testing set were plotted to demonstrate the separability of the features using CORAL (see [Fig. 11](#)).

TABLE III
PERFORMANCE OF DIFFERENT TRANSFER LEARNING METHODS IN 3 PROTOCOLS

METHOD	DISTANCE METRIC	TRANSFER STAGE	ACCURACY(%)		
			PROTOCOL 1	PROTOCOL 2	PROTOCOL 3
DANN	MMD	HIDDEN LAYER	85.03±10.66	83.41±9.05	86.91±10.25
DDC	MMD	HIDDEN LAYER	83.94±12.58	81.11±11.33	85.58±9.51
CORAL	CORAL	OUTPUT LAYER	90.19±9.99	84.26±9.27	87.7±10.79
DAN	MK-MMD	HIDDEN+OUTPUT LAYER	83.23±11.04	80.88±11.73	85.50±9.54
JAN	JMMD	HIDDEN+OUTPUT LAYER	86.68±11.93	83.08±10.40	86.76±10.33
B-JMMD	BALANCED JMMD	HIDDEN+OUTPUT LAYER	86.70±12.11	83.30±9.95	86.94±9.22
RTN	MMD(WITH AN ENTROPY LOSS)	HIDDEN+OUTPUT LAYER	80.86±13.11	80.99±11.14	85.95±10.02
ASSOCIATEDA	ASSOC	HIDDEN+OUTPUT LAYER	88.22±12.39	84.35±9.69	85.69±8.90

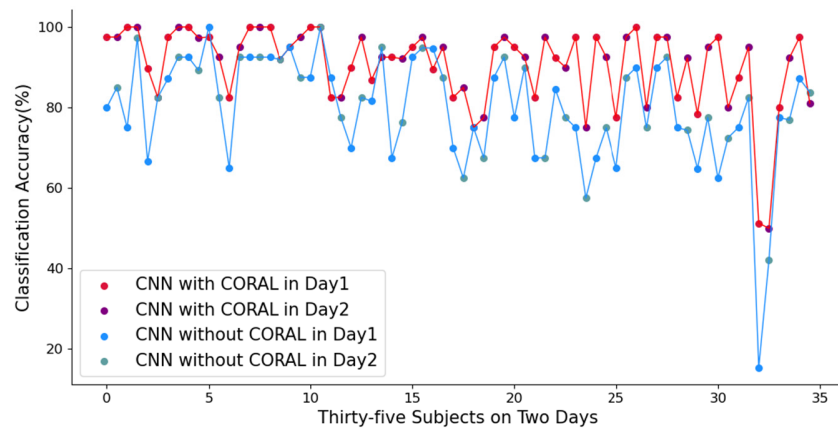


Fig. 5. The classification accuracy of thirty-five subjects on two days with (red line) and without (blue line) CORAL in Protocol 1.

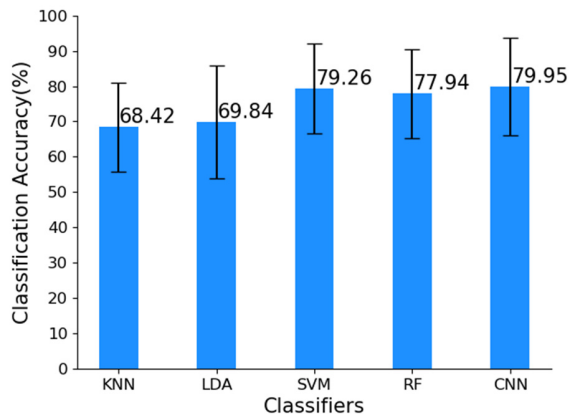


Fig. 6. Performance of five classifiers without CORAL in Protocol 1.

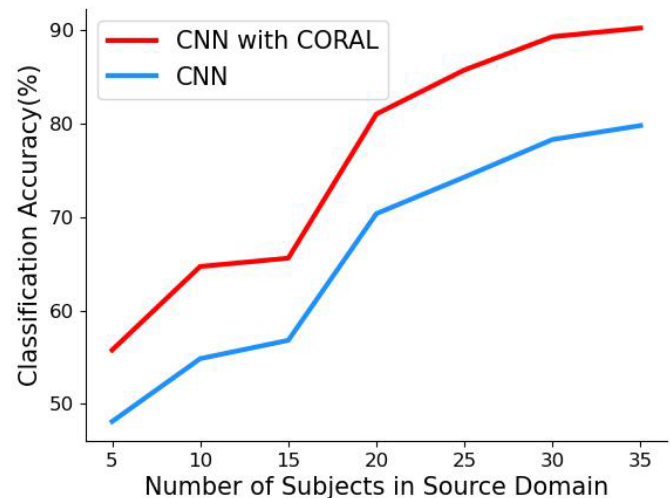


Fig. 7. The relationship between the number of subjects in the source domain and classification accuracy with and without CORAL.

The relationships between the model performance and (1) the number of subjects in the source domain (in the training set) or (2) the number of each gesture used for transfer were investigated, respectively. The results are listed as follows: (1) whether with or without CORAL, the classification accuracy improved significantly with the increase of subjects in source domain (see Fig. 7); (2) two repeated trials for each

gesture were enough to achieve a high level of transfer (see Fig. 8).

We also tried embedding CORAL into traditional classifiers. As shown in Fig. 12, most traditional classifiers underwent

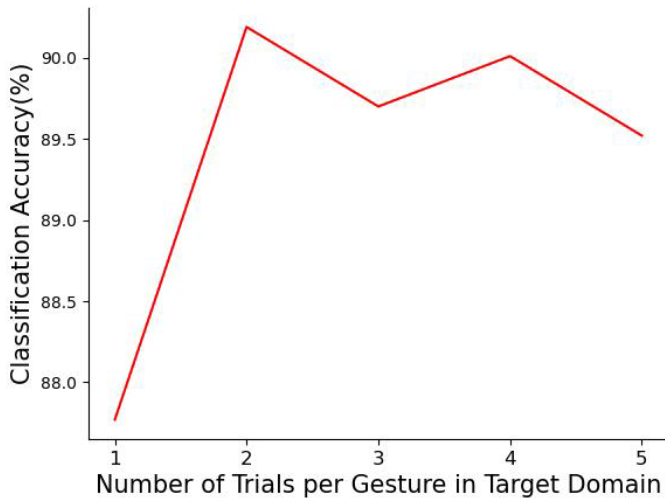


Fig. 8. The relationship between the number of trials per gesture in the target domain and classification accuracy.

negative transfer after combining with CORAL. The KNN, LDA and SVM degenerated significantly by 3.5%, 24.14% and 24.28%, respectively. The RF was almost unchanged. The CNN with CORAL achieved higher classification accuracy than other classifiers both with and without CORAL ($p < 0.05$).

C. Results of Protocol 2

In *Protocol 2*, we evaluated different classifiers performance and different TL algorithms when transferring data from a second-day. When TL was not applied, CNN achieved the highest classification accuracy (80.86%). The traditional classifiers, KNN, LDA, SVM and RF achieved 68.37%, 72.71%, 80.70% and 79.38% classification accuracy, respectively. The results of RANOVA showed that CNN, SVM and RF achieved higher classification accuracy than KNN and LDA ($p < 0.05$). After combining with CORAL, all the traditional classifiers have undergone negative transfer, KNN, LDA, SVM and RF degenerated by 6.45%, 14.6%, 18.08% and 1.18%, as shown in Fig. 13.

Among the 8 different TL methods, the highest classification accuracy was achieved by AssociateDA (84.35%), and CNN with CORAL achieved 84.26% (88.59% for V1). Compared to the non-TL version, they improved by 3.49% and 3.4%, respectively. After performing the statistical analysis, we can find there was no statistical difference between classification performance with these two methods ($p > 0.05$). Although some improvement has been achieved, their accuracy has declined compared with the results of *Protocol 1*.

D. Results of Protocol 3

In *Protocol 3*, we randomly selected data of new gestures from the target user to construct the target domain. When TL was not applied, CNN achieved the highest classification accuracy (85.47%). The traditional classifiers, KNN, LDA, SVM and RF achieved 72.55%, 63.43%, 84.01% and 82.22% classification accuracy, respectively. The results of RANOVA

TABLE IV
REVIEW OF PREVIOUS RESEARCH AND THIS STUDY

PREVIOUS RESEARCHES	NO. SUBJECT	CLASSIFIER	ACCURACY(%)
Meng et al. [33]	41 (20 for V1)	CNN	86.33
Fan et al. [34]	20 (only V1)	KNN	84.80
Li et al. [35]	20 (only V1)	Deep Forest	84.27
Wang et al. [36]	41 (20 for V1)	SVM	75.23
This Study	35 (17 for V1)	CNN	90.60 (94.33 for V1)

showed that CNN and SVM achieved higher classification accuracy than other classifiers ($p < 0.05$). After combining with CORAL, all the traditional classifiers have undergone negative transfer, KNN, LDA, SVM and RF degenerated by 3.6%, 0.86%, 6.66% and 6.94%, as shown in Fig. 14.

The result presented in Table III demonstrated that CORAL achieved the highest classification accuracy (87.7%), surpassing the non-TL version by 1.85%, indicating the model is robust to the selection of gesture categories. By embedding TL module, the network can effectively acquire relevant information from the target domain, thereby expediting model training and extracting features that are better suited for the target user.

V. DISCUSSION

A. Review of Previous Studies on Cross-Subject Hand Gesture Classification

The poor generalization performance and heavy training burden of the gesture classification model are always important factors that hinder the application of sEMG-based HMI systems. To solve this problem, we compared different feature combination and explored different classifiers embedded with eight unsupervised transfer learning algorithms to build a cross-subject gesture classification model. The optimal configuration led to a classification accuracy over 90% on a hybrid dataset including 10 gestures from 35 subjects on two different days. Validated by three different validation protocols, our work is able to significantly reduce the training burden for new users of sEMG-based HMI systems in the calibration process. To clarify our contributions more specifically, we listed the results of previous research using the same public dataset (or partially using it) in Table IV. The superiorities of our study mainly lied in the following aspects:

- 1) we employed a heuristic method SFS to find the optimal feature combination for cross-subject gesture recognition task. Then the selected features were fed into the network instead of raw sEMG data to accelerate the model training.
- 2) we built a simple network architecture with number of trainable parameters below 130K. By embedding an unsupervised transfer learning module in the network, an accuracy improvement of more than 10% could be achieved while only using a small amount of data from target user for training.
- 3) we proposed two new validation protocols to demonstrate the robustness of our model for cross-day and cross-gesture scenarios.

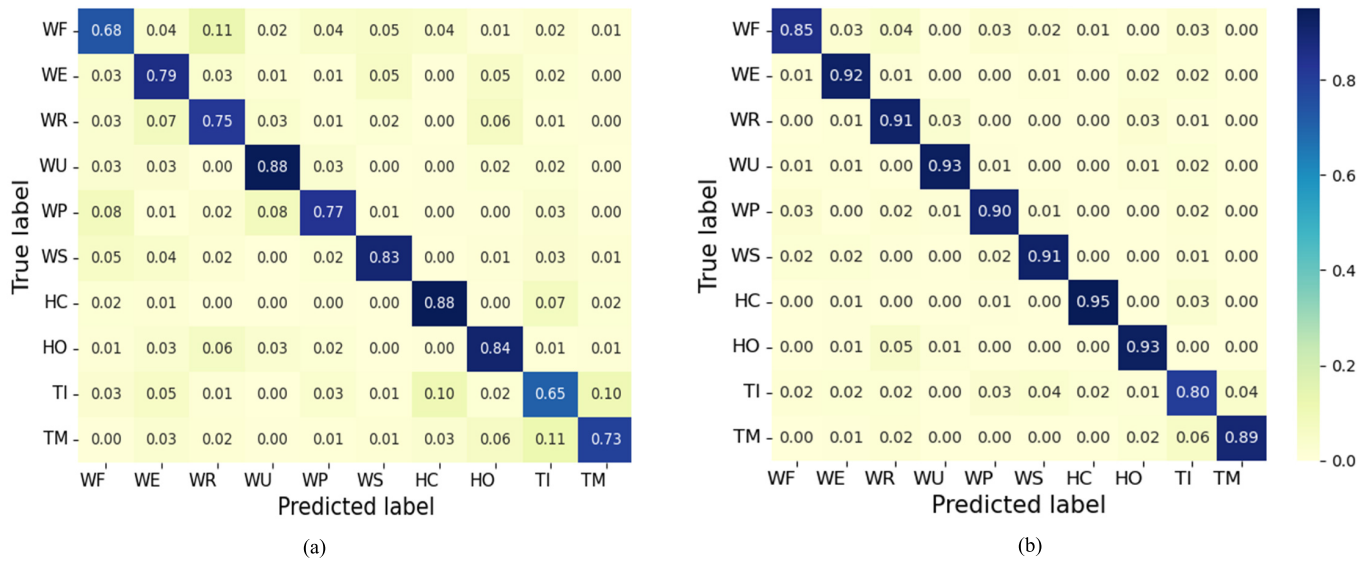


Fig. 9. (a). Confusion matrix for ten-gesture classification accuracy without CORAL (b). Confusion matrix for ten-gesture classification accuracy with CORAL.

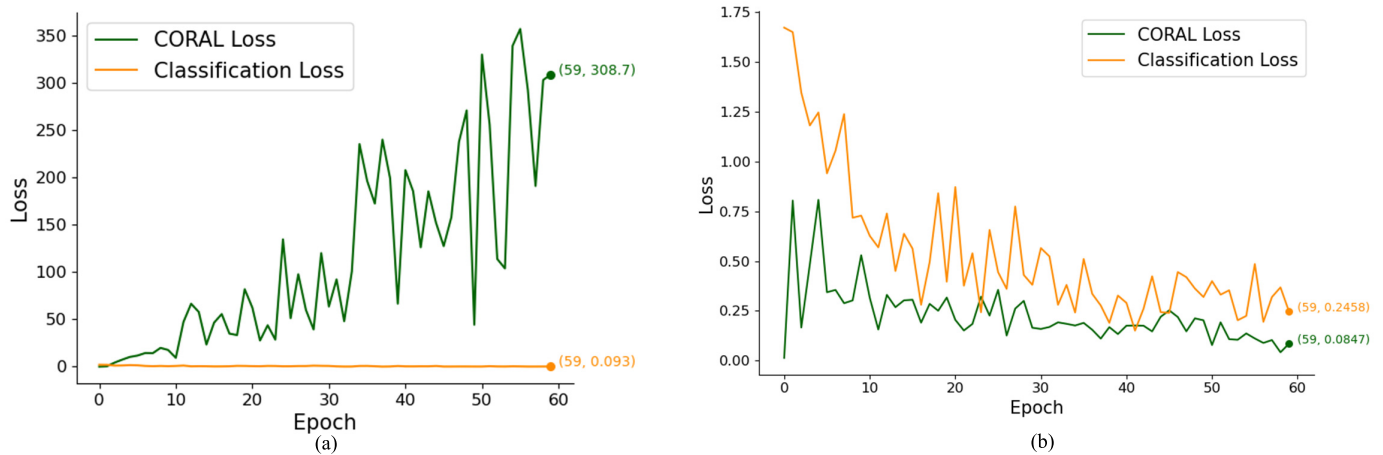


Fig. 10. (a). Changes in the loss of the network without CORAL during training (b). Changes in the loss of the network with CORAL during training.

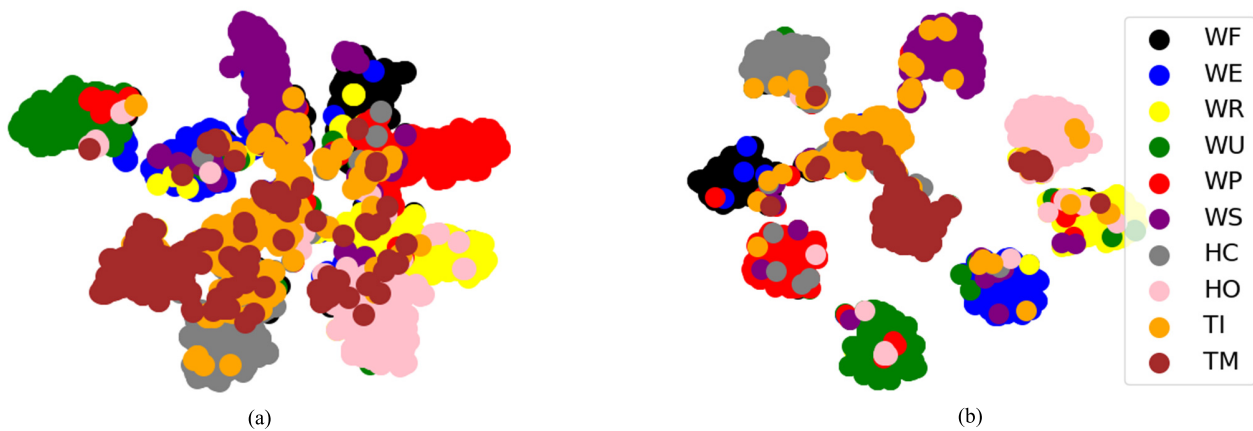


Fig. 11. (a). The t-SNE visualization of network outputs of testing set without CORAL (b). The t-SNE visualization of network outputs of testing set with CORAL.

B. Effects of Feature Selection

In the feature selection stage, we employed SFS to find the optimal combination in a feature set of 40 classic features.

The combination of LTKEO + ZC + SSC + SKEW achieved the highest classification accuracy using LDA classifier. When selecting a combination of features, the m best selected

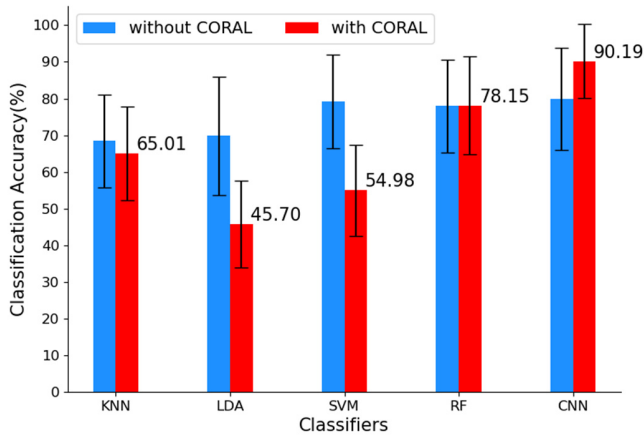


Fig. 12. Performance of five classifiers with and without CORAL in Protocol 1.

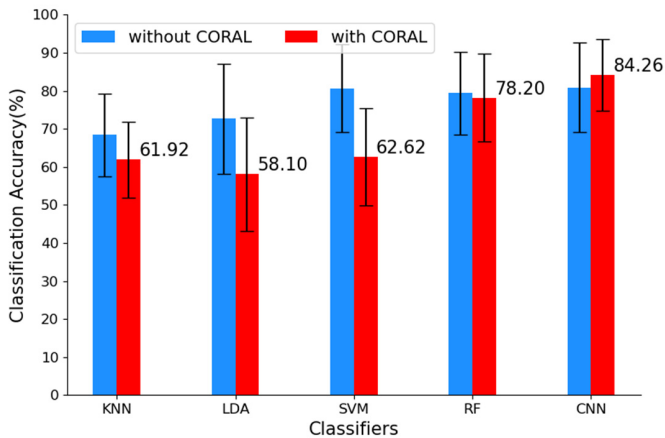


Fig. 13. Performance of five classifiers with and without CORAL in Protocol 2.

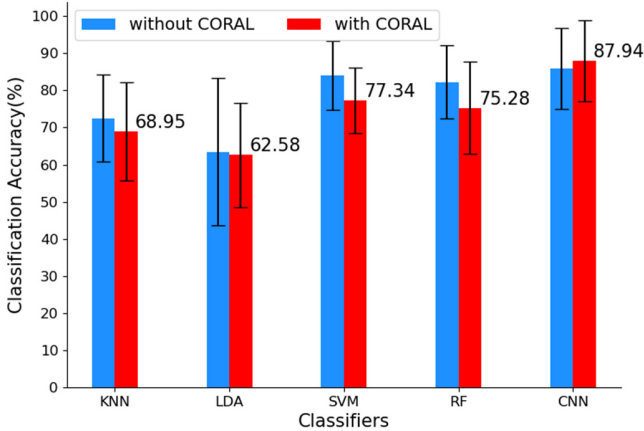


Fig. 14. Performance of five classifiers with and without CORAL in Protocol 3.

features may not necessarily be the best m features globally. This is because besides the relevance of features with ground truth labels, the dependency and redundancy between features may also largely impact classification performance [24]. With different initializations, three features ZC, SSC and SKEW had the highest chance to be selected. One possible reason is

that ZC or SSC counts the number of times where EMG signal crosses zero or peak, and SKEW presents the distribution of sEMG. Therefore, these features can well characterize the signal intensity without measuring signal amplitude which is the most representation between individuals. On the other hand, feeding the selected feature combinations into the network instead of the raw sEMG signals can speed up the convergence of model training. Our model can achieve more than 90% classification accuracy within a mere 60 iterations, which usually requires hundreds of iterations in other researches.

C. Effects of Classifiers

In the classifier stage, CNN and RF showed stronger generalization ability over the other classifiers. This may be due to the fact that they both contain the idea of ensemble learning which often has high classification performance when dealing with high dimensional data. For RF, multiple decision trees are generated, with each one handling a feature subset or a sample subset. The final decision given by the constituent trees is hence less sensitive to characteristics of a specific user. For CNN, the *dropout* layers prevent the model from overfitting the training set. Each decision can be seen as a result of multiple neurons. The impact of differences between different users is thus attenuated.

D. Effects of Transfer Learning Methods

For the TL algorithms, all methods made an improvement when combined with deep classifier. By adding domain difference loss to the loss function, CNN tends to learn domain-invariant features, which contain more gesture information than user information. Compared with other TL algorithms, CORAL achieved higher classification accuracy in most scenarios. This may be because the second-order statistical information is better representative of the domain difference than MMD and its variants for sEMG signals. By comparing the confusion matrix and the t-SNE embeddings, we can see the target categories become more distinguished after transfer. A good performance can be obtained through transferring only two unlabeled trials per gesture, proving the reliability and robustness of this method. In practical use, the subject only needs to perform a simple calibration step (a couple of gestures with no requirement of gesture labels), thereby reducing the training burden for new users of the gesture classification model.

However, CORAL lost its effectiveness when combined with traditional classifiers. This is because CORAL in this scenario is applied by directly projecting source features into an ideal space that aligns the second-order statistics of distributions with target features. This projection solution has a prerequisite that source features should satisfy independent and identically distributed (i.i.d.) assumption, which does not meet the requirement of our case since we have multiple subjects in source domain with each one considered as a distinct distribution. Negative transfer happened on almost all cases, indicating that the traditional classifier with CORAL is not a good option for the cross-subject scenario.

E. Effects of Validation Protocols

During the validation stage, we conducted three different protocols to further validate the robustness of the model. In *Protocol 1*, we used two repeated trials per gesture from the new user for transfer. Those gestures for transfer were from the same day as those used for testing. This protocol was designed to simulate the most common scenarios for cross-subject hand gesture recognition: a new user needs to perform several fixed gestures as instructed before model training. It is not surprising to get the conclusion that this protocol performed the best since the transfer data contained the richest and the most accurate information about the target user.

In *Protocol 2*, the training set composition was the same as in *Protocol 1*, which was data acquired from 34 subjects on two different days. The data of the target user on Day 1 and Day 2 were employed for transfer and testing, respectively. This protocol was designed to simulate a scenario that the target user uses their own trained model without same-day recalibration. The classification accuracy in *Protocol 2* was 5.93% lower than that in *Protocol 1* since the introduction of environmental noises and variation of sEMG characteristics over days.

In *Protocol 3*, only V1 dataset, which has 17 subjects and 24 additional gestures, was used for validation. Twenty new gestures were randomly selected for transfer, whose classes were different from those in the testing set. This protocol was designed to simulate a scenario that the target user perform random gestures during calibration process. Despite the variations in gesture categories, the target domain still encapsulated the individual-specific information of the target user, thereby facilitating the model training.

Overall, the CNN with CORAL achieved a great improvement over their non-TL version in all three protocols, with an accuracy of 90.19%, 84.26% and 87.94%, respectively. These experimental results prove that our cross-subject gesture classification model also has great cross-gesture robustness and good cross-day robustness. Since the data used for domain transfer only require a very small number of unlabeled trials of any gesture, new users are free from the complex acquisition paradigm in the calibration process before using sEMG-based HMI systems. Moreover, our proposed method can provide a theoretical foundation and some alternative technical components for relevant researchers and engineers.

Although our study shows that the use of unsupervised transfer learning to improve the generalization ability of gesture recognition model is very promising, a couple of limitations exist in our current research. First, although HD-sEMG can provide muscle information with a high spatial resolution, it is unsuitable for wearable devices. Future studies need to investigate the trade-off between classification accuracy and convenience of the practical use after performing a channel reduction step. Second, all the experiments in this study were performed offline. Additional works still need to be explored for the real scenarios, such as reducing the size of the sliding window for data segmentation.

VI. CONCLUSION

This work proposes a new model based on an unsupervised transfer learning algorithm CORAL with CNN for cross-subject gesture classification task, which achieves the highest classification accuracy of 90.19% using the optimal feature combination. Three different protocols have been employed to prove that our model has great cross-gesture robustness and good cross-day robustness, which achieved 87.94% and 84.26%, respectively. Our work provides a promising solution to avoid a complex and time-consuming calibration process before using sEMG-based HMI systems.

REFERENCES

- [1] D. Farina et al., "Man/machine interface based on the discharge timings of spinal motor neurons after targeted muscle reinnervation," *Nature Biomed. Eng.*, vol. 1, no. 2, p. 25, Feb. 2017, doi: [10.1038/s41551-016-0025](https://doi.org/10.1038/s41551-016-0025).
- [2] J. Wei, Q. Meng, and A. Badii, "Classification of human hand movements using surface EMG for myoelectric control," in *Advances in Computational Intelligence Systems*, vol. 513, P. Angelov, A. Gegov, C. Jayne, and Q. Shen, Eds. Cham, Switzerland: Springer, 2017, pp. 331–339, doi: [10.1007/978-3-319-46562-3_22](https://doi.org/10.1007/978-3-319-46562-3_22).
- [3] D. Blana, T. Kyriacou, J. M. Lambrecht, and E. K. Chadwick, "Feasibility of using combined EMG and kinematic signals for prosthesis control: A simulation study using a virtual reality environment," *J. Electromyogr. Kinesiol.*, vol. 29, pp. 21–27, Aug. 2016, doi: [10.1016/j.jelekin.2015.06.010](https://doi.org/10.1016/j.jelekin.2015.06.010).
- [4] N. Parajuli et al., "Real-time EMG based pattern recognition control for hand prostheses: A review on existing methods, challenges and future implementation," *Sensors*, vol. 19, no. 20, p. 4596, Oct. 2019, doi: [10.3390/s19204596](https://doi.org/10.3390/s19204596).
- [5] L. Resnik, H. Huang, A. Winslow, D. L. Crouch, F. Zhang, and N. Wolk, "Evaluation of EMG pattern recognition for upper limb prosthesis control: A case study in comparison with direct myoelectric control," *J. NeuroEng. Rehabil.*, vol. 15, no. 1, p. 23, Dec. 2018, doi: [10.1186/s12984-018-0361-3](https://doi.org/10.1186/s12984-018-0361-3).
- [6] A. Sarasola-Sanz et al., "A hybrid brain-machine interface based on EEG and EMG activity for the motor rehabilitation of stroke patients," in *Proc. Int. Conf. Rehabil. Robot. (ICORR)*, Jul. 2017, pp. 895–900, doi: [10.1109/ICORR.2017.8009362](https://doi.org/10.1109/ICORR.2017.8009362).
- [7] M. S. P. de Melo, J. G. da Silva Neto, J. M. X. N. Teixeira, A. E. F. Da Gama, and V. Teichrieb, "An EMG-based virtual reality application for motor rehabilitation," in *Proc. 21st Symp. Virtual Augmented Reality (SVR)*, Oct. 2019, pp. 170–177, doi: [10.1109/SVR.2019.00041](https://doi.org/10.1109/SVR.2019.00041).
- [8] N. Irastorza-Landa et al., "EMG discrete classification towards a myoelectric control of a robotic exoskeleton in motor rehabilitation," in *Converging Clinical and Engineering Research on Neurorehabilitation II*, vol. 15, J. Ibáñez, J. González-Vargas, J. M. Azorín, M. Akay, and J. L. Pons, Eds. Cham, Switzerland: Springer, 2017, pp. 159–163, doi: [10.1007/978-3-319-46669-9_29](https://doi.org/10.1007/978-3-319-46669-9_29).
- [9] J. Taborri et al., "Sport biomechanics applications using inertial, force, and EMG sensors: A literature overview," *Appl. Bionics Biomechanics*, vol. 2020, pp. 1–18, Jun. 2020, doi: [10.1155/2020/2041549](https://doi.org/10.1155/2020/2041549).
- [10] A. de Rugy, G. E. Loeb, and T. J. Carroll, "Virtual biomechanics: A new method for online reconstruction of force from EMG recordings," *J. Neurophysiol.*, vol. 108, no. 12, pp. 3333–3341, Dec. 2012, doi: [10.1152/jn.00714.2012](https://doi.org/10.1152/jn.00714.2012).
- [11] G. L. Soderberg and T. M. Cook, "Electromyography in biomechanics," *Phys. Therapy*, vol. 64, no. 12, pp. 1813–1820, Dec. 1984, doi: [10.1093/ptj/64.12.1813](https://doi.org/10.1093/ptj/64.12.1813).
- [12] S. H. Roy, P. Bonato, and M. Knafitz, "EMG assessment of back muscle function during cyclical lifting," *J. Electromyogr. Kinesiol.*, vol. 8, no. 4, pp. 233–245, Aug. 1998, doi: [10.1016/s1050-6411\(98\)00010-8](https://doi.org/10.1016/s1050-6411(98)00010-8).
- [13] C. Pontonnier, "Cutting force and EMG recording for ergonomics assessment of meat cutting tasks: Influence of the workbench height and the cutting direction on muscle activation levels," *Eng. Syst. Des. Anal.*, vol. 44878, pp. 207–216, Jul. 2012.

- [14] S. E. Mathiassen and A. Burdorf, "Statistical power and measurement allocation in ergonomic intervention studies assessing upper trapezius EMG amplitude A case study of assembly work," *J. Electromyogr. Kinesiol.*, vol. 12, no. 1, pp. 45–57, 2002.
- [15] W. Chen and Z. Zhang, "Hand gesture recognition using sEMG signals based on support vector machine," in *Proc. IEEE 8th Joint Int. Inf. Technol. Artif. Intell. Conf. (ITAIC)*, May 2019, pp. 230–234, doi: [10.1109/ITAIC.2019.8785542](https://doi.org/10.1109/ITAIC.2019.8785542).
- [16] W. Wei, Q. Dai, Y. Wong, Y. Hu, M. Kankanhalli, and W. Geng, "Surface-electromyography-based gesture recognition by multi-view deep learning," *IEEE Trans. Biomed. Eng.*, vol. 66, no. 10, pp. 2964–2973, Oct. 2019, doi: [10.1109/TBME.2019.2899222](https://doi.org/10.1109/TBME.2019.2899222).
- [17] Y. Hu, Y. Wong, W. Wei, Y. Du, M. Kankanhalli, and W. Geng, "A novel attention-based hybrid CNN-RNN architecture for sEMG-based gesture recognition," *PLoS ONE*, vol. 13, no. 10, Oct. 2018, Art. no. e0206049, doi: [10.1371/journal.pone.0206049](https://doi.org/10.1371/journal.pone.0206049).
- [18] A. Ameri, M. A. Akhaee, E. Scheme, and K. Englehart, "A deep transfer learning approach to reducing the effect of electrode shift in EMG pattern recognition-based control," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 28, no. 2, pp. 370–379, Feb. 2020, doi: [10.1109/TNSRE.2019.2962189](https://doi.org/10.1109/TNSRE.2019.2962189).
- [19] Y. Zou and L. Cheng, "A transfer learning model for gesture recognition based on the deep features extracted by CNN," *IEEE Trans. Artif. Intell.*, vol. 2, no. 5, pp. 447–458, Oct. 2021, doi: [10.1109/TAI.2021.3098253](https://doi.org/10.1109/TAI.2021.3098253).
- [20] U. Côté-Allard et al., "Deep learning for electromyographic hand gesture signal classification using transfer learning," 2018, *arXiv:1801.07756*.
- [21] X. Chen, Y. Li, R. Hu, X. Zhang, and X. Chen, "Hand gesture recognition based on surface electromyography using convolutional neural network with transfer learning method," *IEEE J. Biomed. Health Informat.*, vol. 25, no. 4, pp. 1292–1304, Apr. 2021, doi: [10.1109/JBHI.2020.3009383](https://doi.org/10.1109/JBHI.2020.3009383).
- [22] M. M.-C. Vidovic, H. Hwang, S. Amsüss, J. M. Hahne, D. Farina, and K. Müller, "Improving the robustness of myoelectric pattern recognition for upper limb prostheses by covariate shift adaptation," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 24, no. 9, pp. 961–970, Sep. 2016, doi: [10.1109/TNSRE.2015.2492619](https://doi.org/10.1109/TNSRE.2015.2492619).
- [23] X. Jiang et al., "Open access dataset, toolbox and benchmark processing results of high-density surface electromyogram recordings," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 29, pp. 1035–1046, 2021, doi: [10.1109/TNSRE.2021.3082551](https://doi.org/10.1109/TNSRE.2021.3082551).
- [24] X. Jiang et al., "Optimization of HD-sEMG-based cross-day hand gesture classification by optimal feature extraction and data augmentation," *IEEE Trans. Human-Mach. Syst.*, vol. 52, no. 6, pp. 1281–1291, Dec. 2022.
- [25] M. Ghifary, W. B. Kleijn, and M. Zhang, "Domain adaptive neural networks for object recognition," in *PRICAI 2014: Trends in Artificial Intelligence*, vol. 8862, D.-N. Pham and S.-B. Park, Eds. Cham, Switzerland: Springer, 2014, pp. 898–904, doi: [10.1007/978-3-319-13560-1_76](https://doi.org/10.1007/978-3-319-13560-1_76).
- [26] E. Tzeng, J. Hoffman, N. Zhang, K. Saenko, and T. Darrell, "Deep domain confusion: Maximizing for domain invariance," 2014, *arXiv:1412.3474*.
- [27] M. Long, H. Zhu, J. Wang, and M. I. Jordan, "Unsupervised domain adaptation with residual transfer networks," 2016, *arXiv:1602.04433*.
- [28] M. Long, Y. Cao, J. Wang, and M. I. Jordan, "Learning transferable features with deep adaptation networks," in *Proc. Int. Conf. Mach. Learn.*, 2015, pp. 97–105.
- [29] M. Long, H. Zhu, J. Wang, and M. I. Jordan, "Deep transfer learning with joint adaptation networks," 2016, *arXiv:1605.06636*.
- [30] C. Meng, C. Xu, Q. Lei, W. Su, and J. Wu, "Balanced joint maximum mean discrepancy for deep transfer learning," *Anal. Appl.*, vol. 19, no. 3, pp. 491–508, May 2021, doi: [10.1142/S0219530520400035](https://doi.org/10.1142/S0219530520400035).
- [31] P. Häusser, A. Mordvintsev, and D. Cremers, "Learning by association—a versatile semi-supervised training method for neural networks," 2017, *arXiv:1706.00909*.
- [32] B. Sun, J. Feng, and K. Saenko, "Correlation alignment for unsupervised domain adaptation," 2016, *arXiv:1612.01939*.
- [33] L. Meng et al., "User-tailored hand gesture recognition system for wearable prosthesis and armband based on surface electromyogram," *IEEE Trans. Instrum. Meas.*, vol. 71, pp. 1–16, 2022.
- [34] J. Fan, X. Jiang, X. Liu, L. Meng, F. Jia, and C. Dai, "Surface EMG feature disentanglement for robust pattern recognition," *Expert Syst. Appl.*, vol. 237, Mar. 2024, Art. no. 121224, doi: [10.1016/j.eswa.2023.121224](https://doi.org/10.1016/j.eswa.2023.121224).
- [35] J. Li, X. Jiang, X. Liu, F. Jia, and C. Dai, "Optimizing the feature set and electrode configuration of high-density electromyogram via interpretable deep forest," *Biomed. Signal Process. Control*, vol. 87, Jan. 2024, Art. no. 105445, doi: [10.1016/j.bspc.2023.105445](https://doi.org/10.1016/j.bspc.2023.105445).
- [36] Z. Wang et al., "Optimization of inter-subject sEMG-based hand gesture recognition tasks using unsupervised domain adaptation techniques," *Biomed. Signal Process. Control*, vol. 92, Jun. 2024, Art. no. 106086, doi: [10.1016/j.bspc.2024.106086](https://doi.org/10.1016/j.bspc.2024.106086).