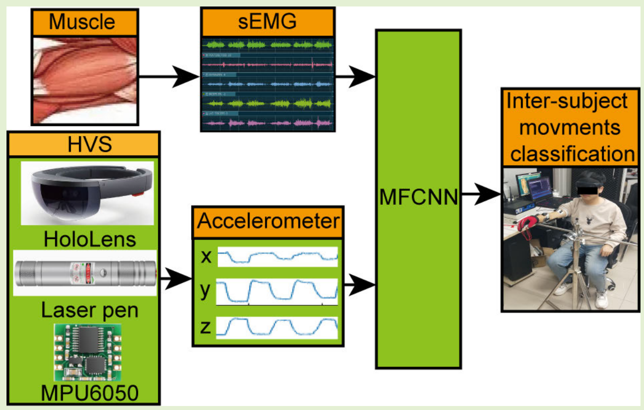
IEEE Sensors Council

Multimodal Fusion Convolutional Neural Network Based on sEMG and Accelerometer Signals for Intersubject Upper Limb Movement Classification

Anyuan Zhang, Qi Li, Zhenlan Li, and Jiming Li



***Abstract*—The variation in the distributions of recorded data between individuals leads to low classification accuracy. To address this issue, we introduce a multimodal fusion convolutional neural network (MFCNN). This network extracts common information from surface electromyography (sEMG) and accelerometer signals of different subjects using a two-stream convolutional neural network (CNN). To enhance the classification accuracy of a particular subject, a fine-tuning approach was implemented. The performance of the proposed method was assessed in four different scenarios, which include intersubject classification, intersubject classification when training data from multiple subjects, fine-tuned intersubject classification, and fine-tuned intersubject classification when training data from multiple subjects. The results demonstrate that in the intersubject scenario, when multiple subjects are available for training, the MFCNN achieves higher classification accuracy ($p < 0.05$) than other neural networks and support vector machines (SVMs) that use sEMG signals [neural network (NN) and SVM], accelerometer signals (accNN and accSVM), sEMG and accelerometer signals [multimodal fusion nerual network (MFNN) and multimodal fusion support vector machine (MFSVM)] as inputs, as well as a CNN that uses sEMG signals as input after fine-tuning. Furthermore, compared with an MFCNN model trained with data from a single subject and an accCNN model trained with data from a single subject or multiple subjects, an MFCNN trained with multiple subjects demonstrated better performance on new subjects after fine-tuning ($p < 0.05$). This method can learn common features among different subjects and improve the performance of classification among subjects. Our proposed method demonstrates the innovation of using a multimodal fusion approach and two-stream CNN to improve intersubject classification accuracy in upper limb movements.**

***Index Terms*—Classification, convolutional neural network (CNN), multimodal, surface electromyography (sEMG), upper limb movement.**

Manuscript received 19 February 2023; revised 5 April 2023; accepted 9 April 2023. Date of publication 18 April 2023; date of current version 31 May 2023. This work was supported by the Department of Science and Technology of Jilin Province under Grant 20200801035GH. The associate editor coordinating the review of this article and approving it for publication was Prof. Yulong Huang. *(Corresponding authors: Qi Li; Zhenlan Li.)*

This work involved human subjects or animals in its research. Approval of all ethical and experimental procedures and protocols was granted by the Changchun University of Science and Technology (CUST) under Application No. 20210013.

Anyuan Zhang and Jiming Li are with the School of Computer Science and Technology, Changchun University of Science and Technology, Changchun 130022, China.

Qi Li is with the School of Computer Science and Technology, Changchun University of Science and Technology, Changchun 130022, China, and also with the Zhongshan Institute of Changchun University of Science and Technology, Zhongshan 528437, China (e-mail: liqi@cust.edu.cn).

Zhenlan Li is with the Department of Rehabilitation Medicine, The First Hospital of Jilin University, Changchun 130012, China (e-mail: zhenlan@jlu.edu.cn).

Digital Object Identifier 10.1109/JSEN.2023.3266872

I. Introduction

THE classification of human movements using surface electromyography (sEMG) is extensively employed in the fields of prosthetics and rehabilitation robotics [1], [2], [3]. Several techniques that employ machine and deep learning have been used for human movement classification, including neural network (NN), support vector machine (SVM), and convolutional neural network (CNN) models. These methods have shown promising results in previous studies [4], [5], [6], [7]. Because sEMG data comprise a nonstationary biomedical signal, human movement classification models based on sEMG must, however, be typically trained separately for each individual user, making intersubject movement classification a challenging task; thus far, these methods have typically exhibited relatively low classification accuracy [8].

The accuracy of movement classification models trained on sEMG signals decreased significantly when used to classify the movements of different individuals, likely due to the differences in data distribution of sEMG signals across subjects. Several studies have attempted to address these issues. Extracting invariant sEMG features of specific influencing factors can strengthen the robustness of human movement classification based on sEMG signals [9], [10], [11]; however, these methods focus only on a few specific interference factors. Intersubject classification involves multiple combined interference factors, such as the position of the electrodes, degree of muscle contraction, and subcutaneous fat. Employing postprocessing strategies can enhance the reliability of real-time methods and reduce the occurrence of misclassifications in the output [12], [13], [14]; however, the postprocessing strategy only adjusts the current outputs of the network and does not change the distribution of the data or model parameters. Therefore, the postprocessing strategy generally does not suffice to address the intersubject classification problem. Multiple techniques have been devised to enhance the robustness of the upper limb movement classification through data augmentation [15], [16], [17], [18], [19]. The inclusion of data possessing different distributions into the training dataset can considerably enhance the model's ability to withstand fluctuations in data distribution. These methods, however, have major limitations. Data collection from multiple sources necessitates a significant time investment for data augmentation techniques. Differences in the data distribution among different subjects may negatively affect the model performance.

Adaptive learning with feature engineering can enhance the robustness of sEMG-based movement classification to diverse data distributions but is time-consuming [20], [21]. CNN models have gained widespread adoption and exhibited satisfactory efficacy [22], [23], [24]. In contrast to manual feature engineering [25], [26], [27], CNN models can automatically learn features contained in an sEMG signal. Furthermore, the fine-tuning strategy can conveniently execute adaptive learning through a CNN model, and only a meager quantity of data is needed for the fine-tuning process [28], [29], [30]. These methods, however, focus only on sEMG signal information. When the data distribution changes frequently, relying solely on sEMG signals to achieve robust classification becomes challenging.

To improve the robustness of human movement classification, multiple sensor modalities are necessary, and multimodal sensing is commonly applied to improve the robustness of human movement classification [31], [32], [33], [34]. Multimodal sensing data combines information from different sensors, and different modalities can compensate for each other to yield better performance [35]. Recently, multistream CNN models have demonstrated good performance in decoding human movements based on sEMG signals [36], [37]. A multistream CNN has the capability to extract sEMG signal information from diverse scales based on varying kernel sizes, thereby augmenting the efficacy of pattern recognition and decoding using sEMG signals.

The current research introduces a multimodal fusion convolutional neural network (MFCNN) that amalgamates multistream CNN and multimodal sensing techniques to enhance the efficacy of a fine-tuning strategy across multiple subjects exhibiting different data distributions. The MFCNN was implemented to acquire common features among various subjects for enhanced classification accuracy of the model in relation to a target subject. To build multimodal sensing data and curtail the gap in data distribution across diverse subjects, sEMG and accelerometer signals were used, which, in turn, further augmented the efficacy of the fine-tuning strategy for classification among subjects.

We evaluated the effectiveness of the MFCNN model proposed in this study by comparing it with several other models, such as NN using sEMG signals (NN), NN using accelerometer signals (accNN), multimodal fusion nerual network (MFNN) using both sEMG and accelerometer signals, SVM using sEMG signals (SVM), SVM using accelerometer signals (accSVM), multimodal fusion support vector machine (MFSVM) using both sEMG and accelerometer signals, CNN using sEMG signals (CNN), and CNN using accelerometer signals (accCNN). This comparison aimed to evaluate the MFCNN's ability to extract common features and improve the classification accuracy of the fine-tuning strategy for intersubject classification. We evaluated these models in four intersubject scenarios: intersubject classification, intersubject classification when training data from multiple subjects, fine-tuned intersubject classification, and fine-tuned intersubject classification when training data from multiple subjects. To showcase the ability of the proposed model to identify similarities among multiple subjects, we assess its performance in scenarios where data from several subjects are available for training. We analyze and discuss the classification accuracy and feature space of nine models across four scenarios to verify the effectiveness of the proposed approach.

II. Methods

A. Subjects

This experiment involved ten male participants with ages ranging from 26 to 30 years who were free of any neuromuscular diseases. Prior to the experiment, the participants were informed about the study's purpose and procedures and gave their consent. On July 3, 2021, the study received approval (approval number: 20210013) from the Ethics Committee of Changchun University of Science and Technology (CUST).

B. Collection of sEMG Data

In this study, the positions of the seven sEMG sensors were carefully chosen, as shown in Fig. 1. Seven different muscles, including the biceps, brachioradialis, flexor carpi radialis, flexor carpi ulnaris, triceps, extensor carpi ulnaris, and extensor digitorum communis were fit with sensors as per the guidelines of Noraxon. To improve the conduction of electrical signals between the skin and electrodes, We use 75% alcohol to wipe the skin before placing the electrodes. The signals were sampled at 1500 Hz, and a bandpass filter with a frequency range of 20 to 500 Hz was applied to the signals for filtering.

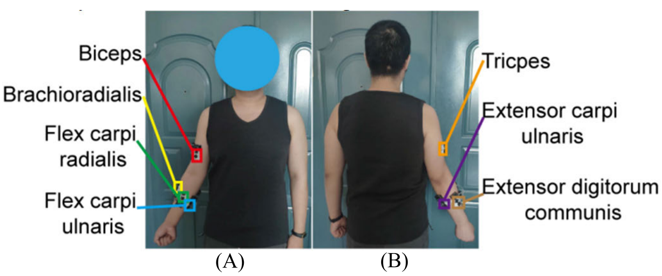


Fig. 1. Placement of seven electrodes on the subjects' muscles. (A) First set of four electrodes were placed on the biceps, brachioradialis, flexor carpi radialis, and flexor carpi ulnaris muscles. (B) Remaining three electrodes were placed on the triceps, extensor carpi ulnaris, and extensor digitorum communis muscles.

C. Collection of Kinematic Data

To construct multimodal sensing data, three-axis Euler angle data of an accelerometer were collected using an MPU6050 sensor fixed to the back of the subjects' hands. The sampling rate for the Euler angles was 100 Hz. To align the sEMG signal with the accelerometer signal, serial communication was used to realize the synchronous acquisition of the accelerometer sensor (MPU6050) and sEMG signal sensor (Noraxon). After synchronization, every 100 accelerometer sampling points correspond to 1500 sEMG signal sampling points. After the acquisition, we normalized the collected Euler angle data and sEMG signals, as shown as follows:

$$x_i = \frac{x_i - u}{\sigma} \tag{1}$$

where x_i refers to the ith sample point of the signal, u denotes the signal's mean value, and σ represents the variance of the signal.

D. Experimental Protocol

A HoloLens virtual system (HVS) was developed to prompt subjects regarding the movements to be performed. The HVS is shown in Fig. 2.

In this study, a red cross was projected into a real environment using a program developed with the Unity development platform and run on a HoloLens device. The participants were equipped with a green laser pen that projected a green cross into the same environment. The subjects were asked to follow the movement of the red cross using the green cross. To carry out this task, all the subjects rested their arms on a metal frame that was shaped like a cross, and their elbows were supported with the help of a bracket. The participants were informed that they had the right to withdraw from the experiment at any time if they experienced any discomfort. All subjects performed nine upper limb movements according to the instructions of the HoloLens. The nine upper limb movements are described in Table I.

All participants were encouraged to exceed the range of each movement. Each movement consisted of five trials, each lasting for 7 s. We chose the data that corresponded to the 5-s period in the middle of the trial. Prior to the experiment, all subjects underwent a familiarization process with the HVS.

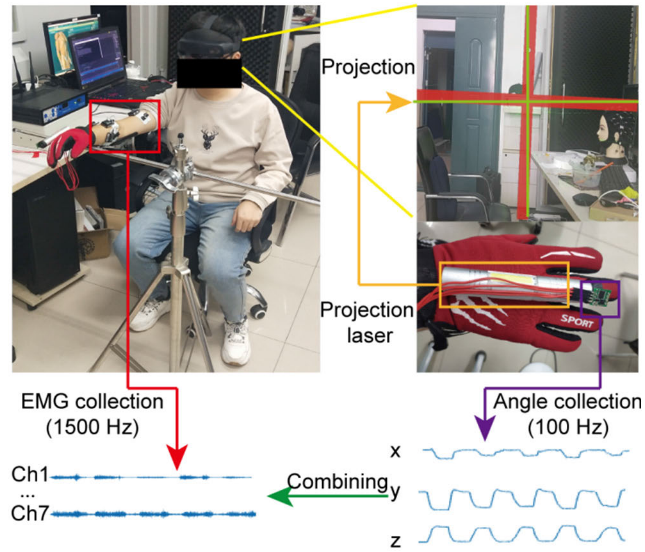


Fig. 2. Prompted subjects' movements were performed using a human–computer interface system (HVS). The movement involved tracking a red cross using a green cross, which was controlled by the subject. During the movement, both sEMG and accelerometer signals were simultaneously collected using the HVS.

TABLE I
DETAILED DESCRIPTION OF NINE UPPER LIMB MOVEMENTS

Protocol	Description	Active DoF
M1	Elbow flexion	Elbow flexion-extension(E-F-E)
M2	Wrist flexion	Wrist flexion-extension(W-F-E)
M3	Wrist extension	Wrist flexion-extension(W-F-E)
M4	Wrist pronation	Wrist pronation-supination(W-P-S)
M5	Wrist supination	Wrist pronation-supination(W-P-S)
M6	Wrist flexion and pronation	W-F-E+W-P-S
M7	Wrist flexion and supination	W-F-E+W-P-S
M8	Wrist extension and pronation	W-F-E+W-P-S
M9	Wrist extension and supination	W-F-E+W-P-S

E. Data Segmentation

Each trial included 5 s sEMG and accelerometer signals. A 250 ms sliding window with an overlap of 80% was used to segment the sEMG and accelerometer signals. The sEMG signal of each sliding window included 375 sampling points and the accelerometer signal of each window included 25 sampling points. Each trial consisted of 96 windows. The data for each subject contained 4320 windows.

F. Feature Extraction for Machine Learning

Five time-domain (TD) features were selected to construct a feature vector. There were five TD features that we considered, namely root mean square (rms), wavelength (WL), zero crossings (ZCs), slope sign change (SSC), and mean average value (MAV). Five TD features were proposed in [38], and satisfactory performance was achieved. The details of these five features are as follows.

1) *Root Mean Square:*

$$\mathrm{RMS} = \sqrt{\frac{1}{N}\sum\nolimits_{i=1}^{N} x_i^2} \tag{2}$$

where x_i denotes the ith sample in the sliding window, N indicates the length of the sliding window, and $N = 375$.

2) *Wavelength:*

$$\mathrm{WL} = \sum_{i=1}^{N} |x_{i+1} - x_i| \tag{3}$$

where x_i denotes the ith sample in the sliding window, N represents the length of the sliding window, and $N = 375$.

3) *Zero Crossing:*

$$\mathrm{ZC} = \sum_{i=1}^{N-1} \left[\mathrm{sgn}\,(x_i \times x_{i+1}) \cap |x_i - x_{i+1}| \geq \text{threshold}\right]$$
$$\mathrm{sgn}\,(x) = \begin{cases} 1, & \text{if } x < 0 \\ 0, & \text{otherwise} \end{cases} \tag{4}$$

where x_i denotes the ith sample in the sliding window, N represents the length of the sliding window, and $N = 375$. The threshold was equal to 0.25 based on the sEMG signals.

4) *Slope Sign Change:*

$$\mathrm{SSC} = \sum_{i=1}^{N-1} \left[f\left[(x_i - x_{i-1}) \times (x_i - x_{i+1})\right]\right],$$
$$f\,(x) = \begin{cases} 1, & \text{if } x < 0 \\ 0, & \text{otherwise} \end{cases} \tag{5}$$

where x_i denotes the ith sample in the sliding window, N refers to the length of the sliding window, and $N = 375$. The threshold was set to 0.10 based on the sEMG signals.

5) *Mean Average Value:*

$$\mathrm{MAV} = \frac{1}{N}\sum_{i=1}^{N} |x_i| \tag{6}$$

where x_i denotes the ith sample in the sliding window, N indicates the length of the sliding window, and $N = 375$.

G. Multimodal Fusion Convolutional Neural Network (MFCNN)

The purpose of this study is to introduce an MFCNN as a solution to the challenge of reducing accuracy in human movement classification based on sEMG signals. This issue arises due to the variations in data distribution across different subjects. The structure of the MFCNN is shown in Fig. 3. The proposed model for this study is a two-stream CNN consisting of a first stream and a second stream. To extract meaningful information from the sEMG signals, we created the first stream of a CNN, which comprised three convolutional modules. To accomplish this, seven channels of raw sEMG signals were used as inputs for the first-stream network. This resulted in a matrix of size 375 × 7 (250 ms window length × 7 channels) being used as input for the first-stream CNN. All convolution layers use the same padding method to extract features from the signal.

The three convolution layers use 64, 96, and 128 filters, respectively, with kernel sizes of 23, 12, and 11. The filter was then moved at a stride of one. The first and second CNN modules of the first-stream network comprise five layers: convolutional, batch normalization, activation, pooling, and dropout. The batch normalization layers were assigned an epsilon of 0.001, and the activation function used was a rectified linear unit (ReLU). We configured the first pooling layer to have a size and stride of 15, while the size and stride of the second pooling were set to 5. For the dropout layer, we set the dropout rate to 0.5. The third module of the CNN contained only convolution, batch normalization, and activation layers. The second-stream network was primarily used to extract information from the accelerometer signal. The three-axis Euler angle was used as the input for the second-stream network. As for the second-stream CNN in this study, a matrix of size 25 × 3 (250 ms window length × 3 axes) was used as input. The second-stream network included only a single convolutional module of layers, including convolutional, batch normalization, activation, pooling, and dropout layers. The convolutional, batch normalization, activation, and dropout layers used the same configuration as that of the first-stream CNN. In this study, the pooling layer of the second-stream CNN had a pooling size and stride set to five. After the features were extracted by the two-stream CNN, the outputs of the two convolutional modules and two-stream CNNs were concatenated. Once we combined the outputs of the two-stream CNN, we supplied the resulting matrix to a fully connected layer, which was followed by a softmax layer. The softmax layer is responsible for generating the probability distribution of each class. The model's prediction was determined by selecting the class with the highest probability among all possible classes.

To train the model, we employed a stochastic gradient descent (SGD) optimizer with momentum, and we executed the training process on a personal computer equipped with an NVIDIA Quadro P620 GPU. The momentum and learning rate values of the optimizer were set to 0.8 and 0.01, respectively. These values were chosen based on the outcomes of various experiments.

H. Methods of Comparison

To assess the efficacy of the proposed method in classifying upper limb movements, we employed several alternative methods for comparison. These included NN, aaccNN, MFNN, SVM, accSVM, MFSVM, CNN, and accCNN. Additionally, several features such as rms, WL, SSC, ZC, and MAV were extracted from the sEMG and accelerometer signals [38]. For the NN and SVM, these features were extracted from the sEMG data of seven channels. Thus, each sample was a 35-dimensional vector. For accNN and accSVM, we extracted these features from the accelerometer signals of the three channels. Each sample was a 15-dimensional vector. For the MFNN and MFSVM, we extracted these features from the

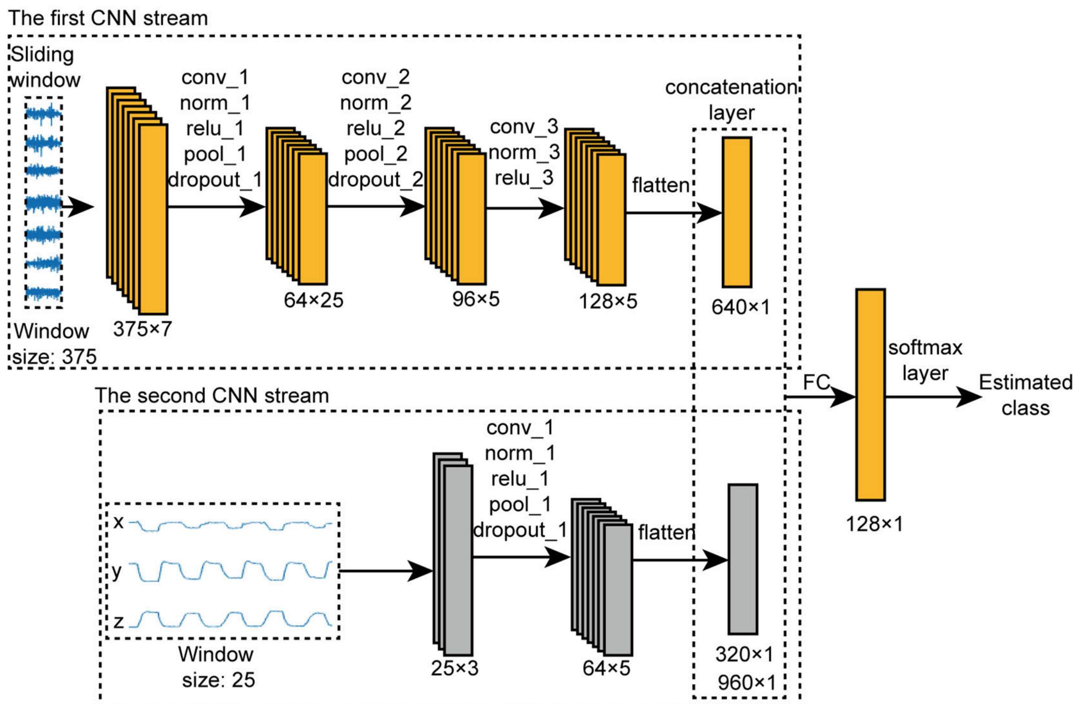


Fig. 3. MFCNN architecture comprises two streams, each with a specific purpose. The first stream is designed to extract information from the sEMG signal, while the second stream is focused on extracting information from the accelerometer. The MFCNN is capable of simultaneously extracting information from both sources, allowing for more comprehensive analysis.

sEMG and accelerometer signals, respectively. The extracted features were combined into a 50-dimensional vector. The NN, accNN, and MFNN comprise a single input layer, two hidden layers, one dropout layer, a softmax layer, and an output layer; 1024 hidden units exist in the first and second hidden layers. We set the dropout rate to 0.5 in the dropout layer. A softmax layer was used to calculate the probability of each class. We set the learning rate to 0.0001. For SVM, accSVM, and MFSVM, we employed the radial basis function (RBF) as the kernel function. The parameter values for c and g in SVM were 1.0 and 0.11, respectively. For the CNN, we used the same configuration as that of the first-stream CNN of the MFCNN. For accCNN, we used the same configuration as the second-stream CNN of the MFCNN. The last flattened layers of both CNN and accCNN were linked directly to a fully connected layer with 128 dimensions. After that, a softmax layer was employed to generate the predicted data.

I. Model Adaptability Analysis

We compared the results of upper-limb movement classification obtained using the MFCNN model with those of other methods, including NN, accNN, MFNN, SVM, accSVM, MFSVM, CNN, and accCNN, to evaluate the effectiveness of our proposed approach. The nine models were tested under four scenarios: intersubject classification, intersubject classification using training data from multiple subjects, fine-tuned intersubject classification, and fine-tuned intersubject classification using training data from multiple subjects. The details of the four scenarios are described below.

Scenario 1: For intersubject classification, we compiled a training set by selecting data from a single subject, which consisted of 4320 samples (96 samples per movement $\times$ 5 trials per movement $\times$ 9 upper-limb movements). The testing set for each subject contained 4320 samples (96 samples $\times$ 5 trials $\times$ 9 upper-limb movements). We evaluated the performance of the models using the testing set comprising data from the remaining subjects, and we measured the classification accuracy on this set. To ensure the reliability of the results, we conducted twofold cross-validation to analyze the results of each pair of subjects.

Scenario 2: Intersubject classification using training data from multiple subjects. We tested the proposed model's ability to learn common features from different subjects by training it with data from multiple subjects and evaluating its performance. To examine how well the model can learn from multiple subjects, the data from nine subjects were selected to construct the training set, comprising a total of 38 880 samples (96 samples per movement $\times$ 5 trials per movement $\times$ 9 upper-limb movements $\times$ 9 subjects). We then used data from the remaining subjects as the testing set, which contained 4320 samples (96 samples $\times$ 5 trials $\times$ 9 upper-limb movements $\times$ 1 subject). The model's classification accuracy was recorded for each subject after being trained on data from multiple subjects. We conducted ten-fold cross-validation on ten subjects to ensure the reliability of the results.

Scenario 3: Fine-tuned intersubject classification. For our experiment, we chose to use data from a single subject as the training set, which comprised 4320 samples (96 samples $\times$ 5 trials $\times$ 9 movements). The data from the other subjects were used as the calibration and testing sets. Specifically, we selected one trial from each of the remaining subjects to form the calibration set, while the remaining four trials of each subject were combined to form the testing set. Because the fine-tuning strategy only applies to deep learning methods, we used calibration sets to train new models for the NN, accNN, MFNN, SVM, accSVM, and MFSVM methods. The retrained models are tested using a testing set. To improve the performance of the CNN, accCNN, and MFCNN models, we fine-tuned them using a calibration set after training them on the previous subject's data. The performance of the fine-tuned models was then assessed on the testing set. To evaluate the performance of the NN, accNN, MFNN, SVM, accSVM, MFSVM, CNN, accCNN, and MFCNN models, we conducted fivefold cross-validation on the results obtained from each of them. We also carried out twofold cross-validation for every pair of participants to guarantee the accuracy and dependability of the results.

Scenario 4: Fine-tuned intersubject classification using training data from multiple subjects. Data from nine subjects were selected as training data. There were 38 880 samples in the training set (96 samples $\times$ 5 trials $\times$ 9 movements $\times$ 9 subjects). We chose the data of the remaining participants as the calibration and testing data. We evaluated the performance of various models using data from multiple subjects. We divided the data into a calibration set and a testing set from the remaining subjects, where one trial was selected as the calibration set, and the remaining four trials were used as the testing set. For the NN, accNN, MFNN, SVM, accSVM, and MFSVM models, we retrained them using the calibration set and tested them on the testing set of the remaining subjects. The CNN, accCNN, and MFCNN models were fine-tuned using the pretrained model from multiple subjects with the calibration set of the remaining subject and then tested on the testing set of the remaining subject. We used fivefold cross-validation to analyze the results of all the models. Moreover, to ensure the reliability of the findings, we also conducted ten-fold cross-validation on the results of all ten subjects.

J. Feature Visualization Method

Principal component analysis (PCA) was employed to visualize the feature space in this study. To conveniently describe the feature space, we reduced the high-dimensional feature vector to three dimensions using PCA and used a scatter plot to display the feature space. For NN and SVM, we applied PCA to the 35-dimensional feature vector to reduce it to three dimensions. For accNN and accSVM, the 15-dimensional feature vector was reduced to three dimensions using PCA. For MFNN and MFSVM, PCA was used to reduce the 50-dimensional feature vector to three dimensions. For CNN and accCNN, we decreased the dimensionality of the flattening layer after feature extraction using convolution modules to three dimensions. For the MFCNN model, we reduced the dimensionality of the concatenation layer extracted by the two-stream CNN to three dimensions.

K. Statistical Analysis

We performed statistical analysis using IBM[1] SPSS Statistics 22 software, setting the significance level at $p < 0.05$. A repeated measures analysis of variance (ANOVA) was used, followed by a post hoc analysis and Bonferroni correction to investigate the significance of the results.

[1]Registered trademark.

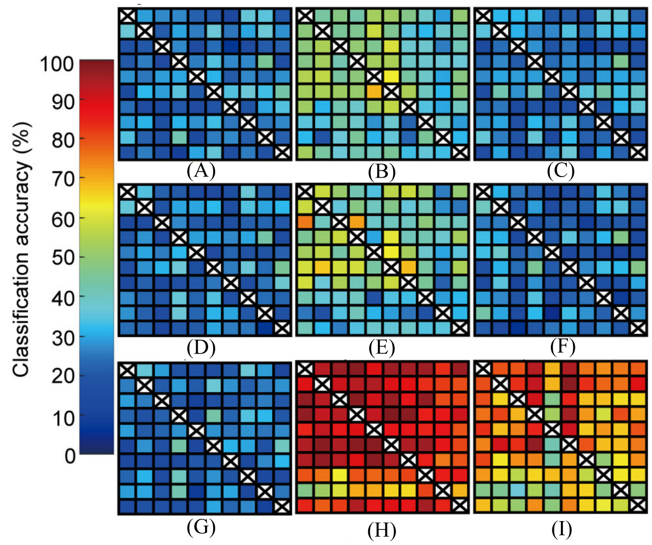


Fig. 4. Confusion matrices of different classification models under the intersubject classification scenario. (A) NN. (B) accNN. (C) MFNN. (D) SVM. (E) accSVM. (F) MFSVM. (G) CNN. (H) accCNN. (I) MFCNN.

III. Results

A. Intersubject Classification

Fig. 4 illustrates the confusion matrices of the nine models for intersubject classification.

Each small square represents the classification accuracy obtained from different subjects through training (rows) and testing (columns). To examine the impact of the classification models on classification accuracy, an ANOVA was conducted. The findings revealed a significant influence of the classification model on classification accuracy ($F_{8,37} = 451.932$, $p < 10^{-3}$). The NN, MFNN, SVM, and CNN showed comparable levels of classification accuracy, with no statistically significant differences observed among them ($p > 0.05$). Despite the fact that the neural network's classification accuracy was notably lower than both accNN and accSVM ($p < 0.05$), it still exhibited a significantly higher accuracy rate than MFNN, SVM, MFSVM, and CNN ($p < 0.05$). There were no significant differences in classification accuracy between accNN and accSVM ($p > 0.05$) and also between MFNN, SVM, and MFSVM ($p > 0.05$). It was observed that the MFNN had a significantly lower classification accuracy than accSVM ($p < 0.05$). On the other hand, the MFNN showed a significantly higher classification accuracy than CNN ($p < 0.05$). Furthermore, it was found that SVM had a lower classification accuracy compared to accSVM ($p < 0.05$). There were no significant differences in classification accuracy observed among SVM, MFSVM, and CNN ($p > 0.05$). In addition, the results showed that accSVM achieved significantly higher classification accuracy than both MFSVM and CNN ($p < 0.05$). Moreover, there were no significant differences in classification accuracy between MFSVM and CNN ($p > 0.05$). Finally, the accCNN and

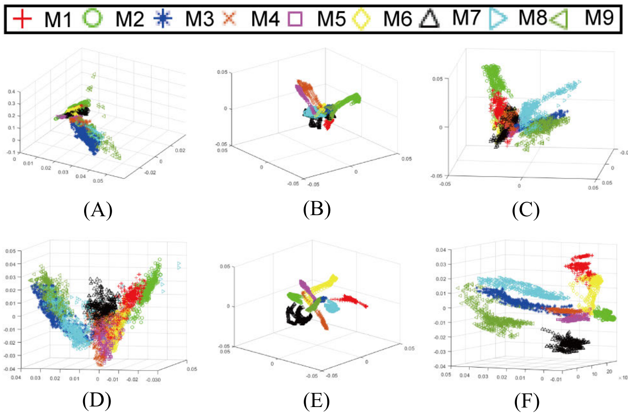


Fig. 5. Feature spaces of nine models under intersubject classification scenario. (A) Feature space of NN and SVM. (B) Feature space of accNN and accSVM. (C) Feature space of MFNN and MFSVM. (D) Feature space extracted by the CNN model. (E) Feature space extracted by accCNN. (F) Feature space extracted by MFCNN.

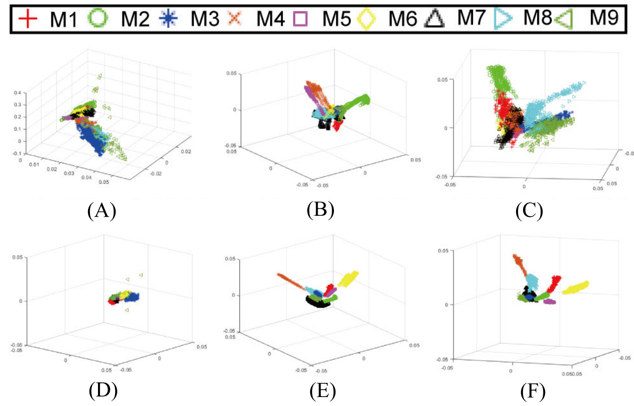


Fig. 7. Feature spaces of nine models in the scenario of intersubject classification when training data from multiple subjects. (A) Feature space of NN and SVM. (B) Feature space of accNN and MFSVM. (C) Feature space of MFNN and MFSVM. (D) Feature space extracted by CNN. (E) Feature space extracted by accCNN. (F) Feature space extracted by MFCNN.

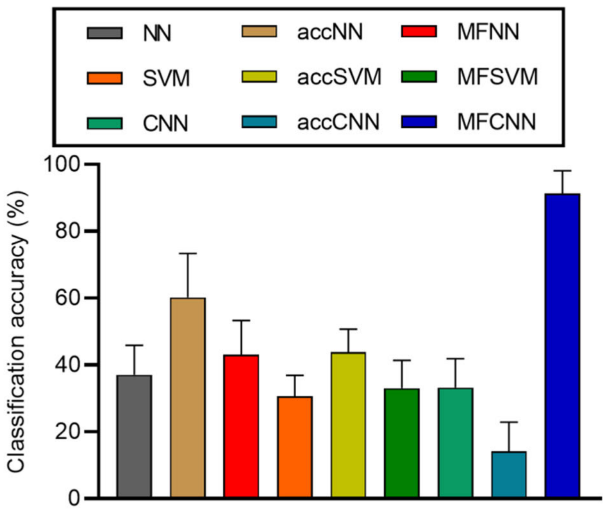


Fig. 6. Classification accuracy of models trained on data from nine subjects when applied to data from a target subject.

MFCNN had higher classification accuracy than the NN, accNN, MFNN, SVM, accSVM, MFSVM, and CNN ($p < 10^{-3}$) models. Nevertheless, it was observed that the classification accuracy of accCNN was significantly greater than that of MFCNN ($p < 10^{-3}$).

Fig. 5 shows the feature spaces of subject 2 extracted using different methods under the intersubject classification scenario. We found that the boundary between classes in the NN, SVM, and CNN feature spaces was considerably fuzzy, whereas the boundaries between classes in the feature spaces of accNN, accSVM, MFNN, MFSVM, accCNN, and MFCNN were relatively clearer.

B. Intersubject Classification Using Training Data From Multiple Subjects

The classification accuracy of the models, which were trained using data from nine subjects and tested on the data of the target subject, is presented in Fig. 6. To explore the possibility of applying the proposed model to new subjects without the need for individualized training, we investigated the feasibility of using data from multiple subjects for training. Specifically, we examined whether the model could extract common features from the data of multiple subjects. The results of the ANOVA revealed that the classification accuracy was significantly affected by the classification model employed ($F_{2,8} = 1298.015$, $p = 10^{-3}$). There were no significant differences in the NN, accNN, MFNN, SVM, accSVM, MFSVM, or CNN ($p > 0.05$). Nonetheless, it was observed that the classification accuracy of accNN was significantly superior to that of SVM, MFSVM, and CNN ($p < 0.05$). There were no significant differences in classification accuracy observed between accNN, MFNN, and accSVM ($p > 0.05$). It was found that the classification accuracy of MFNN was significantly greater than that of CNN ($p = 0.013$). There were no significant differences in classification accuracy observed among MFNN, SVM, accSVM, and MFSVM ($p > 0.05$). Similarly, there were no significant differences in classification accuracy found among SVM, accSVM, MFSVM, and CNN ($p > 0.05$). Among the nine models, the accCNN exhibited the lowest classification accuracy ($p < 0.05$). On the contrary, among the nine models, the MFCNN exhibited the highest classification accuracy ($p < 10^{-3}$).

The feature spaces of the different models for intersubject classification with training data from multiple subjects are shown in Fig. 7. As the number of subjects increased, the feature space boundaries of the NN, ACCNN, MFNN, SVM, ACCSVM, and MFSVM remained blurred and indistinct. The distances between classes in the feature spaces of the CNN, accCNN, and MFCNN decreased. Compared to CNN, the accCNN and MFCNN models exhibited clearer boundaries between classes in their feature space.

C. Performance Comparison of MFCNN Trained by Multiple Subjects and accCNN Trained by a Single Subject

To evaluate the effectiveness of the training strategy across different subjects, we compared the performance of

TABLE II
COMPARISON OF CLASSIFICATION ACCURACY (%) BETWEEN accCNN TRAINED FROM A SINGLE SUBJECT AND MFCNN TRAINED FROM MULTIPLE SUBJECTS UNDER A FINE-TUNING STRATEGY

Subject	accCNN (single subject)	MFCNN (multiple subjects)
1	85.00	97.00
2	85.67	96.00
3	84.22	94.00
4	89.78	96.00
5	88.22	96.00
6	88.22	95.00
7	86.89	89.00
8	89.11	87.00
9	83.22	75.00
10	82.44	88.00
Average	86.28	91.30
Std	2.55	6.83

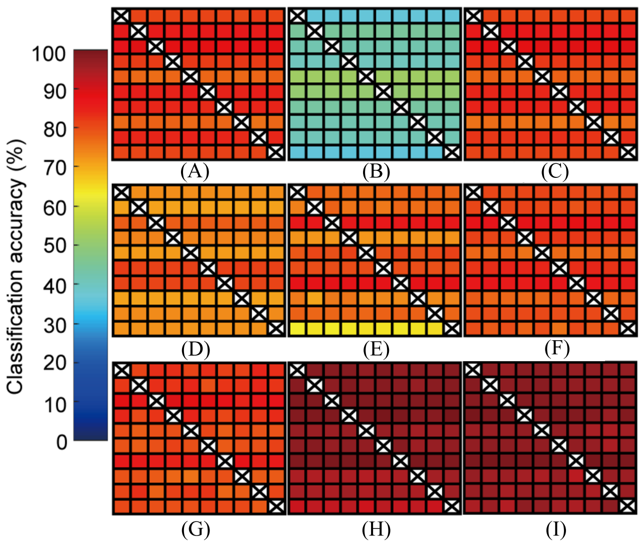


Fig. 8. Confusion matrices of different classification models under the fine-tuned intersubject classification scenario. (A) NN. (B) accNN. (C) MFNN. (D) SVM. (E) accSVM. (F) MFSVM. (G) CNN. (H) accCNN. (I) MFCNN.

models trained on single-subject data with those trained on multiple-subject data. The comparison of classification accuracy between an accCNN model trained on data from a single subject and an MFCNN model trained on data from multiple subjects is presented in Table II. We conducted a t-test to determine the statistical significance of the results. The results showed that the MFCNN model, which was trained using data from multiple subjects, achieved a significantly higher classification accuracy compared to the accCNN model, which was trained using data from a single subject ($p = 0.031$).

D. Fine-Tuned Intersubject Classification

To enhance the classification accuracy for a particular subject, fine-tuning strategies are frequently employed. We also evaluated the effectiveness of these fine-tuning strategies in our models. The corresponding confusion matrices for this scenario are presented in Fig. 8. Each small square in the figure represents the classification accuracy achieved through training (rows) and testing (columns) on different subjects. ANOVA demonstrated that the classification model significantly influenced the classification accuracy in this scenario. According to the results of the ANOVA, the classification model had a significant impact on the classification accuracy ($F_{8,37} = 2979.108$, $p < 10^{-3}$). Compared to accNN, SVM, accSVM, and MFSVM, the NN attained a significantly higher classification accuracy ($p < 0.05$). There were no significant differences in classification accuracy among the NN, MFNN, and CNN models, according to our analysis ($p > 0.05$). Out of the nine models, the accNN achieved the lowest classification accuracy ($p < 10^{-3}$). On the other hand, our findings indicated that the MFNN model performed better than the SVM, accSVM, and MFSVM models in terms of classification accuracy ($p < 10^{-3}$). The classification accuracy of the MFNN and CNN models did not differ significantly according to our statistical analysis ($p > 0.05$). The classification accuracy of both accCNN and MFCNN models was significantly higher than the remaining seven models ($p < 0.05$). In contrast, the SVM model demonstrated a significantly lower classification accuracy than the accSVM, MFSVM, and CNN models ($p < 10^{-3}$), while the accSVM exhibited a significantly lower classification accuracy compared to the MFSVM and CNN models ($p < 0.05$). Our analysis did not find any statistically significant differences in classification accuracy between the MFSVM and CNN models ($p > 0.05$). Moreover, both the accCNN and MFCNN models demonstrated similar classification accuracies, and there were no significant differences between them ($p > 0.05$).

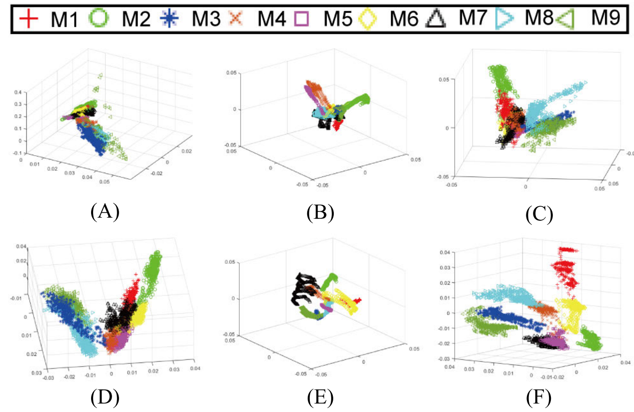


Fig. 9. Feature spaces of NN, SVM, CNN, accCNN, and MFCNN methods in the scenario of fine-tuned intersubject classification. (A) Feature space of NN and SVM. (B) Feature space of accNN and MFSVM. (C) Feature space of MFNN and MFSVM. (D) Feature space extracted by CNN. (E) Feature space extracted by accCNN. (F) Feature space extracted by MFCNN.

Fig. 9 displays the feature spaces of various methods used in the fine-tuned intersubject classification scenario.

When compared to other models such as NN, accNN, MFNN, SVM, accSVM, and MFSVM, the CNN exhibited a clearer boundary between classes; however, the accCNN and MFCNN models showed even clearer boundaries than the CNN model with a fine-tuning strategy.

E. Fine-Tuned Intersubject Classification When Training Data From Multiple Subjects

To assess whether our proposed model is capable of extracting common information from multiple subjects, we conducted

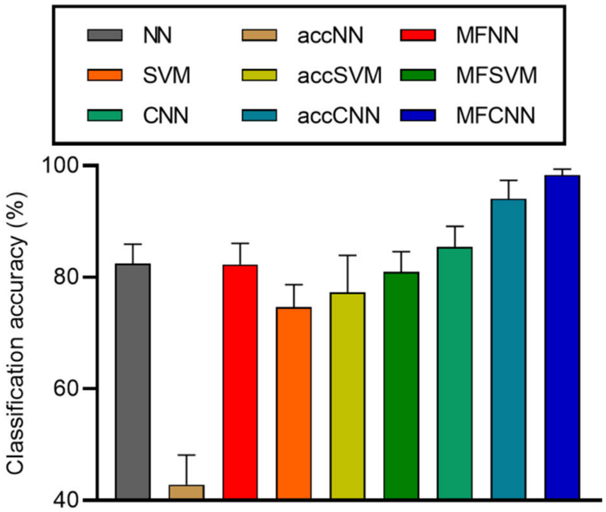


Fig. 10. Classification accuracy of models trained the data of nine subjects on the data of the target subject.

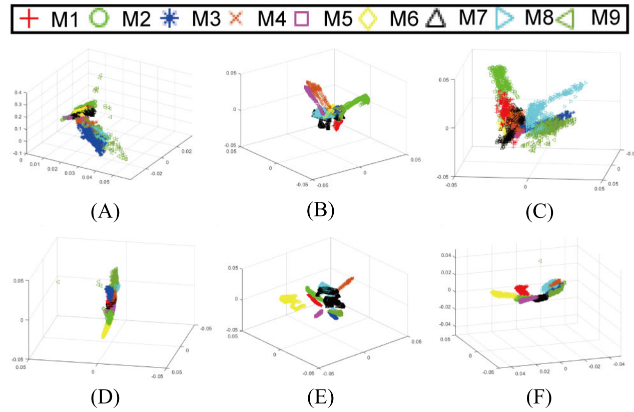


Fig. 11. Feature spaces of NN, SVM, CNN, accCNN, and MFCNN models in the scenario of fine-tuned intersubject classification when training data from multiple subjects. (A) Feature space of NN and SVM. (B) Feature space of accNN and MFSVM. (C) Feature space of MFNN and MFSVM. (D) Feature space extracted by CNN. (E) Feature space extracted by accCNN. (F) Feature space extracted by MFCNN.

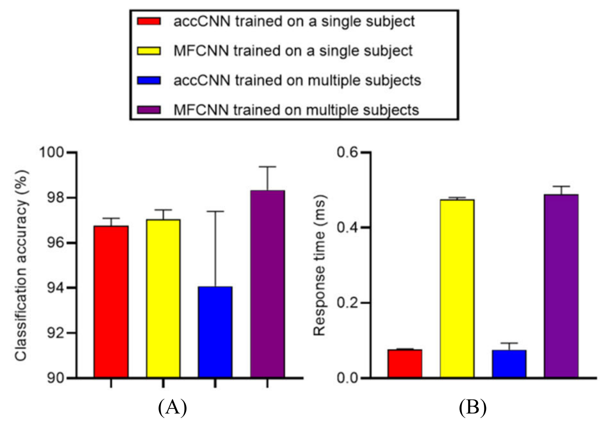


Fig. 12. Classification accuracy and computational time of accCNN trained on a single subject, MFCNN trained on a single subject, accCNN trained on multiple subjects, and MFCNN trained on multiple subjects. (A) Classification accuracy of different models. (B) Computational time of different models.

an intersubject classification scenario with both a fine-tuning strategy and training data from multiple subjects. Fig. 10 illustrates the classification accuracy of the models trained on data from nine subjects and tested on data from the target subject.

ANOVA indicated that the classification model significantly influenced classification accuracy ($F_{2,8} = 198.976$, $p = 0.005$). The classification accuracy of the NN model was significantly higher than that of the accNN and SVM models ($p < 0.05$); however, there was no statistically significant difference in classification accuracy among the NN, MFNN, accSVM, MFSVM, and CNN models ($p > 0.05$). Among the nine models, accNN achieved the lowest classification accuracy ($p < 0.05$), whereas the MFSVM outperformed the SVM in terms of classification accuracy ($p = 0.005$). There was no significant difference in classification accuracy among the MFNN, accSVM, MFSVM, and CNN models ($p > 0.05$). The classification accuracy of MFSVM and CNN was significantly higher than that of SVM ($p < 0.05$). The classification accuracy of accCNN was significantly higher than that of CNN ($p = 0.032$), while there was no significant difference between accSVM and MFSVM ($p > 0.05$). The MFSVM model showed a statistically significant higher classification accuracy than the CNN model ($p = 0.019$). Furthermore, both the accCNN and MFCNN models outperformed the other seven models in terms of classification accuracy ($p < 0.05$), but there were no significant differences in classification accuracy between the accCNN and MFCNN models ($p > 0.05$).

The feature spaces of the different methods in the scenario of fine-tuned intersubject classification when training data from multiple subjects are shown in Fig 11. The feature space boundaries of the NN, accNN, MFNN, SVM, accSVM, and MFSVM were unclear. With an increase in the number of subjects, the CNN, accCNN, and MFCNN showed a decrease in the interclass distance in the feature spaces. Following fine-tuning, the degree of separation between classes in the CNN feature space showed an increase. The boundaries between classes in the accCNN and MFCNN feature spaces are clearer than those in the CNN feature space.

F. Comparison of Performance Among MFCNN Trained With Data From Multiple Subjects, accCNN Trained With Data From Multiple Subjects, MFCNN Trained With Data From a Single Subject, and accCNN Trained With Data From a Single Subject Using Fine-Tuning Strategy

Fig. 12 shows the outcomes of a comparison between four models: a fine-tuned accCNN model trained from a single subject, a fine-tuned MFCNN model trained from a single subject, a fine-tuned accCNN model trained from multiple subjects, and a fine-tuned MFCNN trained from multiple subjects. The comparison includes classification accuracy and computational time. A statistical analysis using ANOVA revealed that the MFCNN model, which was trained on data from multiple subjects and fine-tuned, achieved significantly higher classification accuracy than the fine-tuned accCNN model, trained on data from multiple subjects, the fine-tuned accCNN model trained on data from multiple subjects, and the fine-tuned MFCNN model trained on data from a single subject ($p < 0.05$). No significant difference in computational time was found between the MFCNN when training data from a single subject or multiple subjects ($p = 0.468$), or between the accCNN when training data from a single subject or multiple subjects ($p > 0.05$). The computational time of the models based on MFCNN was significantly longer than that of the models based on accCNN ($p < 0.05$).

IV. Discussion

MFCNN improves intersubject upper limb movement classification using sEMG and accelerometer signals. It employs a two-stream CNN to extract features and achieves higher accuracy than existing models. Results show its effectiveness for new subject data.

The accCNN and MFCNN can achieve considerable performance in intersubject classification tasks. Because of the differences in sEMG signals across different individuals, the classification accuracies of the NN, accNN, MFNN, SVM, accSVM, MFSVM, and CNN models were not satisfactory for intersubject classification. The differing distribution of sEMG signal data among subjects had a notable negative impact on the classification accuracy of the seven models. Both accNN and accSVM outperformed NN, MFNN, SVM, and MFSVM in terms of classification accuracy. The visual representation of the accelerometer signal can illustrate the features of upper limb movements. Compared to the features of a single mode, the manual features of the multimode combined with sEMG and accelerometer signals did not improve the classification accuracy between different subjects. Multimodal manual features based on feature extraction methods cannot handle changes in data distribution when the subjects change. The CNN model did not perform well in intersubject classification, which is different from the results of previous studies that focused on intrasubject classification. The boundaries between classes in the CNN feature space are extremely fuzzy. The unimodal CNN model cannot handle the low accuracy problem of intersubject human movement classification based on sEMG signals. In contrast to the NN, accNN, MFNN, SVM, accSVM, MFSVM, and CNN models, the accCNN and MFCNN achieved satisfactory intersubject classification accuracy. Incorporating accelerometer information into the feature set was found to increase the similarity of data distribution across different subjects. Thus, the use of accCNN and MFCNN can mitigate the impact of variations in sEMG signal distributions across different subjects, resulting in improved classification accuracy. The study found that while the MFCNN trained from a single subject achieved a lower classification accuracy in intersubject classification compared to the accCNN trained from a single subject, the accCNN failed to handle the classification task when trained on data from multiple subjects. This led to a significantly lower classification accuracy than the MFCNN model, which was likely due to the different distributions of accelerometer data across different subjects [39]; however, the MFCNN model trained from multiple subjects outperformed both the accCNN trained from multiple subjects and the accCNN trained from a single subject in intersubject classification. Compared to the accCNN model trained on data from a single subject, the MFCNN model can extract shared features across multiple subjects and attain superior classification accuracy in intersubject classification.

To enhance the classification accuracy for a particular subject, a fine-tuning approach was used. The NN, MFNN, SVM, and MFSVM achieved satisfactory classification accuracy in fine-tuned intersubject classification scenarios using only a small amount of data. Particularly, the classification accuracy of the accNN was low. This could be attributed to the fact that an accNN is unable to effectively learn the features with limited data. While the classification accuracy of the NN, MFNN, and MFSVM did not differ significantly, the MFSVM showed a significantly better classification accuracy than the SVM and accSVM. Multimodal manual features can learn the features of upper-limb movements from limited data and achieve a higher classification accuracy than single-modal manual features. The results showed that a CNN model trained on a single EMG modality could achieve a classification accuracy similar to that of NN, MFNN, and MFSVM with limited fine-tuning data. By using the limited raw sEMG signal from new subjects to transfer the pretrained model, a CNN can prevent information loss during artificial feature construction and facilitate end-to-end control. The accCNN and MFCNN trained from a single subject or trained from multiple subjects achieved higher classification accuracy than the other nine models after fine-tuning. To assess the effectiveness of the MFCNN, we conducted a comparison of the classification accuracy and response time between the accCNN and MFCNN models. These models were trained using data from either a single subject or multiple subjects. After fine-tuning, the MFCNN, when training data from multiple subjects, achieved a higher classification accuracy than the other three models. With a greater volume of data from different subjects, the MFCNN model can learn the generalized initial weights from the data of multiple subjects. The MFCNN was fine-tuned using only a limited number of data points, which led to a significant improvement in its classification accuracy. Although the MFCNN has more trainable parameters, which leads to a notably longer response time than accCNN, all models exhibit a response time of under 0.6 ms, indicating their feasibility for real-time control of the model. In future work, it may be beneficial to explore the use of lightweight models to further reduce the response time of MFCNN.

V. Conclusion

This study presents a novel deep learning model, referred to as MFCNN, for classifying intersubject upper-limb movements using sEMG and accelerometer signals. The proposed method demonstrated remarkable performance in scenarios involving intersubject classification. The proposed multimodal model can extract information from the sEMG and accelerometer signals and increase the similarity of the data distribution of different subjects by complementing the information between different modalities. Thus, the model is capable of identifying shared characteristics across multiple subjects, thereby enabling it to achieve high performance even with a small dataset from new subjects, which can be used to recalibrate and improve its performance. In this study, however, we only considered the influence of the multimodal method on the accuracy of the intersubject upper-limb movement classification and overlooked the method used to reduce the variation in the distribution of data between different subjects. As a future direction, we aim to explore domain adaptation methods to decrease the variance in data distribution among different subjects. This could enhance the performance of upper-limb movement classification methods based on sEMG signals in intersubject scenarios.

References

[1] A. Fleming, N. Stafford, S. Huang, X. Hu, D. P. Ferris, and H. Huang, "Myoelectric control of robotic lower limb prostheses: A review of electromyography interfaces, control paradigms, challenges and future directions," *J. Neural Eng.*, vol. 18, no. 4, Aug. 2021, Art. no. 041004.

[2] X. Li et al., "A new strategy based on feature filtering technique for improving the real-time control performance of myoelectric prostheses," *Biomed. Signal Process. Control*, vol. 70, Sep. 2021, Art. no. 102969.

[3] C. Xie, Q. Yang, Y. Huang, S. Su, T. Xu, and R. Song, "A hybrid arm-hand rehabilitation robot with EMG-based admittance controller," *IEEE Trans. Biomed. Circuits Syst.*, vol. 15, no. 6, pp. 1332–1342, Dec. 2021.

[4] S. Guo, M. Pang, B. Gao, H. Hirata, and H. Ishihara, "Comparison of sEMG-based feature extraction and motion classification methods for upper-limb movement," *Sensors*, vol. 15, no. 4, pp. 9022–9038, Apr. 2015.

[5] J. Liu, "Adaptive myoelectric pattern recognition toward improved multifunctional prosthesis control," *Med. Eng. Phys.*, vol. 37, no. 4, pp. 424–430, Apr. 2015.

[6] M. Zanghieri, S. Benatti, A. Burrello, V. Kartsch, F. Conti, and L. Benini, "Robust real-time embedded EMG recognition framework using temporal convolutional networks on a multicore IoT processor," *IEEE Trans. Biomed. Circuits Syst.*, vol. 14, no. 2, pp. 244–256, Apr. 2020.

[7] X. Liu et al., "Real-time control of intelligent prosthetic hand based on the improved TCN," *Appl. Bionics Biomech.*, vol. 2022, May 2022, Art. no. 6488599.

[8] D. Xiong, D. Zhang, X. Zhao, and Y. Zhao, "Deep learning for EMG-based human-machine interaction: A review," *IEEE/CAA J. Autom. Sinica*, vol. 8, no. 3, pp. 512–533, Mar. 2021.

[9] R. Boostani and M. H. Moradi, "Evaluation of the forearm EMG signal features for the control of a prosthetic hand," *Physiol. Meas.*, vol. 24, no. 2, pp. 309–319, May 2003.

[10] D. Tkach, H. Huang, and T. A. Kuiken, "Study of stability of time-domain features for electromyographic pattern recognition," *J. Neuroeng. Rehabil.*, vol. 7, no. 1, p. 21, May 2010.

[11] A. Phinyomark, F. Quaine, S. Charbonnier, C. Serviere, F. Tarpin-Bernard, and Y. Laurillau, "EMG feature evaluation for improving myoelectric pattern recognition robustness," *Expert Syst. Appl.*, vol. 40, no. 12, pp. 4832–4840, Sep. 2013.

[12] K. Englehart and B. Hudgins, "A robust, real-time control scheme for multifunction myoelectric control," *IEEE Trans. Biomed. Eng.*, vol. 50, no. 7, pp. 848–854, Jul. 2003.

[13] S. Amsuss, P. M. Goebel, N. Jiang, B. Graimann, L. Paredes, and D. Farina, "Self-correcting pattern recognition system of surface EMG signals for upper limb prosthesis control," *IEEE Trans. Biomed. Eng.*, vol. 61, no. 4, pp. 1167–1176, Apr. 2014.

[14] X. Zhang, X. Li, O. W. Samuel, Z. Huang, P. Fang, and G. Li, "Improving the robustness of electromyogram-pattern recognition for prosthetic control by a postprocessing strategy," *Frontiers Neurorobotics*, vol. 11, p. 51, Sep. 2017.

[15] D. Yang, Y. Gu, L. Jiang, L. Osborn, and H. Liu, "Dynamic training protocol improves the robustness of PR-based myoelectric control," *Biomed. Signal Process. Control*, vol. 31, pp. 249–256, Jan. 2017.

[16] Y. Geng, O. W. Samuel, Y. Wei, and G. Li, "Improving the robustness of real-time myoelectric pattern recognition against arm position changes in transradial amputees," *Biomed Res. Int.*, vol. 2017, Apr. 2017, Art. no. 5090454.

[17] W. Yang, D. Yang, Y. Liu, and H. Liu, "Decoding simultaneous multi-DOF wrist movements from raw EMG signals using a convolutional neural network," *IEEE Trans. Human-Machine Syst.*, vol. 49, no. 5, pp. 411–420, Oct. 2019.

[18] H. Dantas, D. J. Warren, S. M. Wendelken, T. S. Davis, G. A. Clark, and V. J. Mathews, "Deep learning movement intent decoders trained with dataset aggregation for prosthetic limb control," *IEEE Trans. Biomed. Eng.*, vol. 66, no. 11, pp. 3192–3203, Nov. 2019.

[19] D. Yang and H. Liu, "An EMG-based deep learning approach for multi-DOF wrist movement decoding," *IEEE Trans. Ind. Electron.*, vol. 69, no. 7, pp. 7099–7108, Jul. 2022.

[20] M. Vidovic, H.-J. Hwang, S. Amsuss, J. Hahne, D. Farina, and K.-R. 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.

[21] J. Liu, X. Sheng, D. Zhang, J. He, and X. Zhu, "Reduced daily recalibration of myoelectric prosthesis classifiers based on domain adaptation," *IEEE J. Biomed. Health Informat.*, vol. 20, no. 1, pp. 166–176, Jan. 2016.

[22] M. Atzori, M. Cognolato, and H. Müller, "Deep learning with convolutional neural networks applied to electromyography data: A resource for the classification of movements for prosthetic hands," *Frontiers Neurorobotics*, vol. 10, p. 9, Sep. 2016.

[23] W. Geng, Y. Du, W. Jin, W. Wei, Y. Hu, and J. Li, "Gesture recognition by instantaneous surface EMG images," *Sci. Rep.*, vol. 6, no. 1, p. 36571, 2016.

[24] J. Chen, X. Zhang, Y. Cheng, and N. Xi, "Surface EMG based continuous estimation of human lower limb joint angles by using deep belief networks," *Biomed. Signal Process. Control*, vol. 40, pp. 335–342, Feb. 2018.

[25] A. Phinyomark, F. Quaine, S. Charbonnier, C. Serviere, F. Tarpin-Bernard, and Y. Laurillau, "Feature extraction of the first difference of EMG time series for EMG pattern recognition," *Comput. Methods Program. Biomed.*, vol. 117, no. 2, pp. 247–256, Nov. 2014.

[26] A. Stango, F. Negro, and D. Farina, "Spatial correlation of high density EMG signals provides features robust to electrode number and shift in pattern recognition for myocontrol," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 23, no. 2, pp. 189–198, Mar. 2015.

[27] A.-C. Tsai, J.-J. Luh, and T.-T. Lin, "A novel STFT-ranking feature of multi-channel EMG for motion pattern recognition," *Expert Syst. Appl.*, vol. 42, no. 7, pp. 3327–3341, 2015.

[28] U. Côté-Allard et al., "Deep learning for electromyographic hand gesture signal classification using transfer learning," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 27, no. 4, pp. 760–771, Apr. 2019.

[29] 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, Dec. 2020.

[30] K.-T. Kim, C. Guan, and S.-W. Lee, "A subject-transfer framework based on single-trial EMG analysis using convolutional neural networks," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 28, no. 1, pp. 94–103, Oct. 2020.

[31] C.-W. Yeh, T.-Y. Pan, and M.-C. Hu, "A sensor-based official basketball referee signals recognition system using deep belief networks," in *Proc. Int. Conf. Multimedia Modeling*, 2017, pp. 565–575.

[32] W. Wang, B. Chen, P. Xia, J. Hu, and Y. Peng, "Sensor fusion for myoelectric control based on deep learning with recurrent convolutional neural networks," *Artif. Organs*, vol. 42, no. 9, pp. E272–E282, Sep. 2018.

[33] X. Zhang, Z. Yang, T. Chen, D. Chen, and M. C. Huang, "Cooperative sensing and wearable computing for sequential hand gesture recognition," *IEEE Sensors J.*, vol. 19, no. 14, pp. 5775–5783, Mar. 2019.

[34] Q. Gao, J. Liu, and Z. Ju, "Hand gesture recognition using multimodal data fusion and multiscale parallel convolutional neural network for human–robot interaction," *Expert Syst.*, vol. 38, no. 5, p. e12490, 2021.

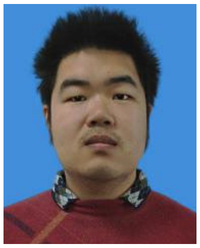
[35] D. Ramachandram and G. W. Taylor, "Deep multimodal learning: A survey on recent advances and trends," *IEEE Signal Process. Mag.*, vol. 34, no. 6, pp. 96–108, Nov. 2017.

[36] X. Fan, Q. Yao, Y. Cai, F. Miao, F. Sun, and Y. Li, "Multiscaled fusion of deep convolutional neural networks for screening atrial fibrillation from single lead short ECG recordings," *IEEE J. Biomed. Health Inform.*, vol. 22, no. 6, pp. 1744–1753, Nov. 2018.

[37] G. Hajian and E. Morin, "Deep multi-scale fusion of convolutional neural networks for EMG-based movement estimation," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 30, pp. 486–495, 2022.

[38] M. Hakonen, H. Piitulainen, and A. Visala, "Current state of digital signal processing in myoelectric interfaces and related applications," *Biomed. Signal Process. Control*, vol. 18, pp. 334–359, Apr. 2015.

[39] C. Sun, A. Shrivastava, S. Singh, and A. Gupta, "Revisiting unreasonable effectiveness of data in deep learning era," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Oct. 2017, pp. 843–852.



Anyuan Zhang is currently pursuing the Ph.D. degree with the Changchun University of Science and Technology, Jilin, China.

His research interests include machine learning, EMG pattern recognition, signal processing, and human–machine interaction.

Qi Li received the Ph.D. degree from Okayama University, Okayama, Japan, in 2010.

He is currently a Professor with the Changchun University of Science and Technology, Jilin, China. He is also with the Zhongshan Institute of Changchun University of Science and Technology, Zhongshan, China. His research interests include human–machine interaction, pattern recognition, and signal processing.

Zhenlan Li received the Ph.D. degree from Jilin University, Changchun, China, in 2008.

She is currently a Professor with The First Hospital of Jilin University, Jilin, China. Her research interests include neuroplasticity of the brain after cerebrovascular disease and neuroimaging of the vegetative state of brain injury.

Jiming Li is currently pursuing the B.Sc. degree with the Changchun University of Science and Technology, Jilin, China.

His research interests include machine learning, EMG pattern recognition, signal processing, and human–machine interaction.