International Journal of Machine Learning and Cybernetics
https://doi.org/10.1007/s13042-022-01687-4

ORIGINAL ARTICLE

# Multi-modality deep forest for hand motion recognition via fusing sEMG and acceleration signals

**Yinfeng Fang**[1] · **Huiqiao Lu**[1] · **Han Liu**[2]

Received: 7 May 2022 / Accepted: 10 October 2022

## Abstract

Bio-signal based hand motion recognition plays a critical role in the tasks of human-machine interaction, such as the natural control of multifunctional prostheses. Although a large number of classification technologies have been taken to improve the motion recognition accuracy, it is still a challenge to achieve acceptable performance for multiple modality input. This study proposes a multi-modality deep forest (MMDF) framework to identify hand motions, in which surface electromyographic signals (sEMG) and acceleration signals (ACC) are fused at the input level. The proposed MMDF framework constitutes of three main stages, sEMG and ACC feature extraction, feature dimension reduction, and a cascade structure deep forest for classification. A public database "Ninapro DB7" is used to evaluate the performance of the proposed framework, and the experimental results show that it can achieve a significantly higher accuracy than that of competitors. Besides, our experimental results also show that MMDF outperforms other traditional classifiers with the input of the single modality of sEMG signals. In sum, this study verifies that ACC signals can be an excellent supplementary for sEMG, and MMDF is a plausible solution to fuse mulit-modality bio-signals for human motion recognition.

**Keywords** Deep forest · sEMG signals · Acceleration signals · Hand gesture classification

## 1 Introduction

Amputation is the main cause of disability and prostheses play an important role in assisting amputees to conduct daily activities [1]. Functionality, controllability and aesthetics are three key elements in the design of prosthetic hands [2]. Using surface electromyographic signals(sEMG) to control prostheses owns its natural advantage, since sEMG naturally reflects muscular activities that originally drive the skeleton joints to move [2, 3]. However, for complex hand movements, it is still a challenge to predict the human intentions from multi-channels of sEMG signals. Despite the technological advances, existing prostheses cannot fully meet the actual needs of amputees in terms of dexterity. Therefore, how to naturally interact with a multi-functional and dexterous prosthesis becomes a challenging problem [4].

To make these robotic devices work accurately, the basic problem is how to distinguish users'intentions from the obtained bio-signals. Recent decade witnesses the fast progress of pattern recognition technology, and it has been widely applied to predict human intentions from sEMG signals [5]. However, there exits a major issue regarding technology assessment for a fair comparison, because of the diversity of experimental setups, including the types and the number of gestures, the construction of the dataset, differs in the number of samples obtained, the shape, origin, and type of sensor used [6]. As a result, experimental results are not always reproducible. A possible solution is to establish public benchmark databases and set down rules for the experiments. Ninapro (Non-Invasive Adaptive Hand Prosthetics (http://ninaweb.hevs.ch/) [7–9]) is one of the most famous sEMG databases for the evaluation of machine learning algorithms for upper limb prosthesis control, which somewhat becomes a benchmark database. With different researching targets and experimental setups, Ninapro can

✉ Yinfeng Fang
yinfeng.fang@hdu.edu.cn

Huiqiao Lu
huiqiao.lu@hdu.edu.cn

Han Liu
han.liu@szu.edu.cn

[1] School of Communication Engineering, Hangzhou Dianzi University, Hangzhou 310018, Zhejiang, China

[2] College of Computer Science and Software Engineering, Shenzhen University, Shenzhen 518060, Guangdong, China

Published online: 01 November 2022

be divided into several sub-ones ranged from DB1 to DB9. The number of gestures in Ninapro can be up to 50, which becomes one of its main characteristic.

The main novelties and contributions of this study are as follows: (1) to the best of our knowledge, it is the first attempt to take deep forests in classifying hand motions from sEMG signals; (2) this study modifies the original gcForest [10] via replacing the multi-grained scanning by feature extraction technology, and applies it in the fusion of sEMG and ACC modalities; (3) this study verifies the importance of ACC signals as a complement for hand motion recognition, and the proposed MMDF can be a decent solution for multi-modality data fusion at input level.

## 2 Related works

A pattern recognition system using single modality of sEMG signals would be negatively influenced by the arm position [11–13]. Thus, ACC signals are usually taken as additional modality to counteract it, and they can be easily measured via integrating into the prosthesis receiving cavity without much cost. The combination of sEMG and ACC has been used in motor rehabilitation [14], sign language recognition [15], and prosthesis control [12, 16, 17]. For instance, Fougner et al. [12] demonstrated the average classification error can be reduced from 18% to lower than 6% via fusing ACC signals (collected under different arm positions) with sEMG signals. Khushaba et al. [16] studied the combined effects of forearm orientation and muscle contraction level, and verified that the use of accelerometers is beneficial to the classification performance. Liu et al. [14] take sEMG and ACC as multi-feedback user interface for upper limb motor rehabilitation user training, and the method improves the users' initiative and performance.

In terms of the fusion approach, sEMG and ACC signals can be fused in the feature level or the decision level [18]. Xie et al. [19] utilized EMG and ACC signals to predict hand movements, which EMG signals are used to recognise static hand gesture via a dynamic time warping algorithm and KNN classifier, and ACC signals are used to predict dynamic wrist movements (indicating left, right, up, and down). Eventually, hand gesture and wrist movement are combined at the decision level. Liu et al. [4] fused sEMG and ACC at the feature level to recognise 17 hand movements, where sEMG features and ACC features are conjoined as the input vector for classifiers. Similarly, Krasoulis et al. [20] fused sEMG with various inertial measurements (IM), including accelerometers, gyroscopes and magnetometers, in offline and online hand motion recognition experiments, where several feature combinations (fusion at feature level) from multiple modalities are fed to linear discriminant analysis (LDA) for classification. Regardless fusing data at the feature level or decision level, it is proved that multi-modal solutions have the potential to improve the usability of pattern recognition based upper limb prostheses in practical applications. The current study adopts GcForest algorithm to fuse sEMG and ACC at the feature level.

GcForest is a highly competitive decision tree ensemble method for deep learning [10]. It employs a cascade structure to realize layer-by-layer processing, but its training process does not rely on back-propagation and gradient adjustment. The original framework of GcForest for image classification consists of multi-grained scanning and cascade forest, and some modified models can be found in [21–24]. Compared with deep neural networks, GcForest has fewer hyper-parameters and can achieve excellent performance across various domains by using even the same parameter setting. Daouadi et al. [25] adopted deep forest to classify tweet topics, and find that deep forest does not require a huge amount of labeled data for training, but can achieve better classification performance than that of other state-of-the-art approaches. Ding et al. [26] compared a deep forest model with a deep neural network model in the application of mechanical fault diagnosis, and find that the former is more effective and shows a stronger generalization ability. Sun et al. [24] propose an adaptive feature selection guided deep forest for COVID-19 classification with chest computed tomography (CT), and achieve higher classification accuracy than traditional classifier. Fang et al. [27] propose a multi-feature deep forest (MFDF) model to identify human emotions from EEG signals, and achieve competitive classification performance, which is the first attempt that uses human-crafted features to replace a multi-grained scanning structure in a deep forest framework. The proposed MMDF model of the current study is an extension of MFDF, where multiple modality signals are fused in the feature level.

The organization of this study is as follows. Section 3 introduces the proposed methodology for hand motion classification, including the feature extraction method and the MMDF. Sections 4 and 5 demonstrate the experimental results with discussions. Section 6 concludes the study.

## 3 Materials and methods

This section introduces the overall framework for hand motion recognition, including sEMG and ACC signals processing and feature extraction, the algorithm of multi-modality deep forest, and the evaluation procedure.

### 3.1 Hand motion recognition framework

This study proposes a hand motion recognition framework with sEMG and ACC signals as the input, as shown in Fig. 1. It follows the guidelines of Englehart and Hudgins' general

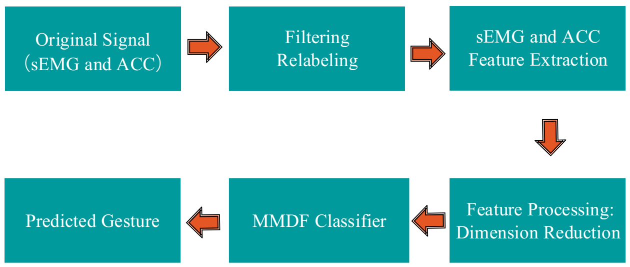

**Fig. 1** The diagram of hand motion recognition framework, where signals reprocessing, feature extraction, feature dimension reduction, and MMDF classifiers are included

framework [28]. Firstly, the original data are filtered to remove unexpected noises, and relabeled for classifier training. Then, features are extracted from sEMG and ACC signals, and the dimensionality of the ACC feature is reduced by PCA(Principal Component Analysis). Thirdly, MMDF classifies the feature vector (the combination of sEMG and ACC features) into one of the hand gestures.

### 3.1.1 Filtering

Following Krasoulis'suggestion [20], the ACC data were low-pass filtered at a cutoff frequency of 5 Hz by a zero-phase second order Butterworth filter to remove high frequency noise components. For sEMG signals, a Hampel filter is applied to remove 50 Hz powerline noise and its harmonics power-line interference.

### 3.1.2 Relabeling

It is unrealistic to make the subjects perfectly mimic the kinematics of the video stimulus due to human reaction delay. The authors of the database had applied an offline generalized likelihood ratio algorithm to correct the fault labels, which realigned the movement boundaries by maximizing the likelihood of a rest-movement-rest sequence [3].

### 3.1.3 Feature extraction

A sliding window is taken in this study for extracting sEMG and ACC features. The choice of the window size trades the balance between the prediction delays and the classification accuracy [8, 28]. This study follows Krasoulis' suggestion and selects the sliding window of 265 ms with the increment of 50 ms to divide the sEMG and ACC signals [20], which meets the suggested maximum allowable delay of 300 ms [16].

Feature selection is an essential stage in sEMG classification, and a large amount of features have been evaluated in myoelectric control design [9]. Although frequency domain features are much more complex than time domain (TD) features, it has not been found that frequency domain features outperform TD features [29]. To guarantee the real-time performance of the recognition framework, time domain (TD) sEMG features are taken in this study. For ACC features, the mean value (MEAN) are calculated from AAC signals, as suggested by Fougner et al. [30].

Five classic TD sEMG features [29, 31], including Mean Absolute Value (MAV), Waveform Length (WL), Zero Crossing(ZC), Slope Sign Changes(SSC) and Autoregressive coefficients 4 (AR4), are used in the current study. The definition of these sEMG features is provided as follows, where $N$ is the sliding window size, and $x_i$ an instant sEMG value at the time point $i$. MAV is the average of absolute value in a window, which somewhat demonstrates the envelope of sEMG signals, and can also be used as muscle activation [32]. It is defined in Eq. (1).

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

WL is the cumulative length of the waveform of EMG signals. It is defined in Eq. (2).

$$\mathrm{WL} = \sum_{i=1}^{N-1}|x_{i+1} - x_i|. \tag{2}$$

ZC is the number of times that the signals crosses zero, which is somewhat associated with the frequency of EMG signals. It is defined in Eq. (3).

$$\mathrm{ZC} = \sum_{i=1}^{N-1}\mathrm{sgn}\left(-x_i x_{i+1}\right), \tag{3}$$

where

$$\mathrm{sgn}(x) = \begin{cases} 1, & x > \varepsilon \\ 0, & x \le \varepsilon \end{cases}, \tag{4}$$

and $\varepsilon$ is the threshold to avoid low-level noises.

SSC provides another measure of the frequency content measuring the number of times the slope of the waveform changes the sign. It is defined in Eq. (5).

$$\mathrm{SSC} = \sum_{i=2}^{N-1} f\left(x_{i-1}, x_i, x_{i+1}\right), \tag{5}$$

where

$$f\left(x_{i-1}, x_i, x_{i+1}\right) = \begin{cases} 1, & \left(x_{i+1} - x_i\right)\left(x_{i-1} - x_i\right) > 0 \quad \text{AND} \\ & \left(|x_{i+1} - x_i| > \varepsilon \text{ OR } |x_{i-1} - x_i| > \varepsilon\right) \\ 0 & \text{else} \end{cases}. \tag{6}$$

In the current study, $\varepsilon$ is set to 2 for both ZC and SSC.

An auto-regressive (AR) model specifies that the output variable depends linearly on its own previous values and a stochastic term, as shown in Eq. (7).

$$x_{i,k} = \sum_{j=1}^{P} \rho_j x_{i,k-j} + \mathcal{E}_t, \tag{7}$$

where $P$ is the order of the model. $\rho_j$ is the $j$th coefficient of the model. $\mathcal{E}_t$ is the residual white noise, and k is the $k$th sampling point in the sEMG sequence.

In addition, the MEAN value is used as the features of ACC signals, which is defined in Eq. (8).

$$MEAN = \frac{1}{N}\sum_{i=1}^{N} x_i, \tag{8}$$

where $N$ is the number of sampling point to calculate the MEAN value.

### 3.2 Multi-modality deep forest

The proposed MMDF can be divided into two parts: the input part and the cascade forest.

The input part is responsible for extracting features from 12 channels of sEMG and 12 channels of ACC, and then combining the features into a feature vector. The schematic diagram of feature fusion is shown in Fig. 2.

After feature extraction, the feature vector of an EMG sample can be defined in Eq. (9).

$$S_{EMG} = \left[f_1, f_2, \dots, f_{12}\right], \tag{9}$$

where

$$f_i = \left[MAV_i, WL_i, AR1_i, AR2_i, AR3_i, AR4_i, SSC_i, ZC_i\right], \tag{10}$$

where $i$ ranges from 1 to 12, indicating the channel number. Therefore, the length of the EMG feature vector is $12\times 8 = 96$ , where 8 is the number of types of features and 12 is the number of channels. For the ACC feature vector, the original length is 36, including 12 channels and 3 sensory directions for each channel. Following Liu's advice [4], to reduce the redundant information in ACC signals, PCA based dimension reduction is applied to obtain the first 18 main components, which can be defined in Eq. (11).

$$S_{ACC} = \left[MEAN_1, MEAN_2, \dots MEAN_{18}\right], \tag{11}$$

where $MEAN_i$ indicates the $i$th main components after PCA. Therefore, the final fusion of $S_{EMG}$ and $S_{ACC}$ is

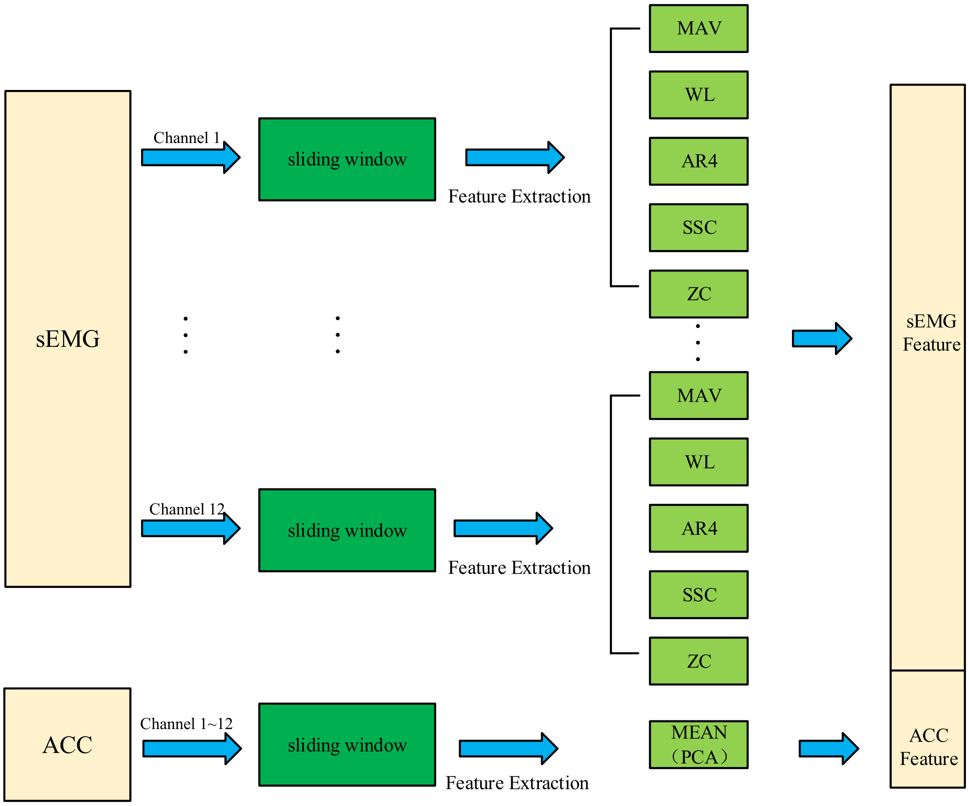

**Fig. 2** The structure of feature extraction; The feature extraction process of the sEMG and the ACC has a lot in common, and it is necessary to perform window segmentation on the data of each channel separately, and then extract the features. The difference is that five features of MAV, WL, AR4, SSC, and ZC are extracted for the sEMG, and only the average value is extracted for the ACC. Finally, the two data types are merged into one feature vector

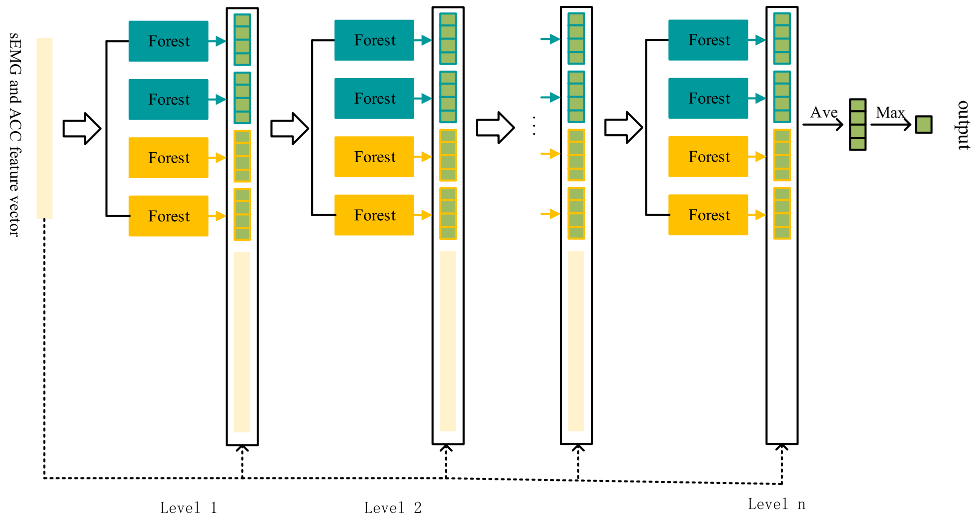

**Fig. 3** The structure of cascade forest; each layer of the cascade consists of two types of random forests with different color blocks, and the number of layers is optimised during the training stage

$$S = \left[S_{EMG}, S_{ACC}, y\right], \tag{12}$$

where $y$ indicates the label in 40 types of gestures.

The architecture of the cascade forest can be found in Fig. 3, where the input is in the form of a feature vector, which is processed by several layers of the forest group ordered from level 1 to level $n$, leading to the classification output. Four forests compose a forest group, and each group contains two types of forests: completely-random tree forest and random forest [10, 33, 34].

Given an instance, each forest can produce an estimation of the class probability distribution, by counting the percentage of each class of training examples at the leaf node in which the concerned instance falls, and then averaging the class probability distributions estimated by various trees in the same forest. After expanding a new level, the performance of the whole cascade can be estimated on a validation set, and the training procedure will terminate if there is no significant performance gain. Thus, the number of cascade levels can be automatically determined.

The training procedure of a deep forest can be formalized as follows. Considering the supervised learning problem of learning a mapping from the feature space $\mathcal{X}$

to the label space $\mathcal{Y}$ where $\mathcal{Y} = \{1, 2, \ldots, C\}$, $\mathcal{Z} = [0, 1]^C$ and training set $S = \left(\left(\mathbf{x}_1, y_1\right), \ldots, \left(\mathbf{x}_m, y_m\right)\right)$ can be drawn from distribution $\mathcal{D}$. A deep forest model is defined by a triplet $(\mathbf{h}, \mathbf{f}, \mathbf{l})$ where

- $\mathbf{h} = \left(h_1, \ldots, h_T\right)$, where $h_t$ is the ensemble of forests at level $t$, and $h_t$ is a member of hypothesis class $H_t$.
- $\mathbf{f} = \left(f_1, \ldots, f_T\right)$, where $f_t$ is the cascade of ensembles of forests up to level $t$.
- $\mathbf{l} = \left(l_1, \ldots, l_T\right)$, where $l_t$ is the validation error at level $t$.

At level $t \in \{1, \ldots, T\}$, $f_t : \mathcal{X} \to \mathcal{Z}$ is defined as follows:

$$f_t(\mathbf{x}) = \begin{cases} h_1(\mathbf{x}) & t = 1, \\ h_t\left(\left[\mathbf{x}, f_{t-1}(\mathbf{x})\right]\right) & t > 1. \end{cases} \tag{13}$$

At every level $t$, $h_t(\cdot)$ and $f_t(\cdot)$ output a class vector $\left[p_1^t, \ldots, p_C^t\right]$, where $p_i$ is the prediction of class $i$. The input of $h_t$ is $\left[\mathbf{x}, f_{t-1}(\mathbf{x})\right]$ except at level $t$=1, where its input is $\mathbf{x}$.

At level $t$, if the difference $(l_{t-1} - l_t)$ between the validation error $l_{t-1}$ of an input at level $t - 1$ and the validation error $l_t$ at level $t$ is less than or equal to the threshold $\eta$, it needs to go through the next level. Until the $T$th layer is reached, the difference between the validation error and the T − 1 layer is greater than the threshold $\eta$ , and the final layer number T of the deep forest is determined. Each triplet $(\mathbf{h}, \mathbf{f}, \mathbf{l})$ defines a deep forest model $g : \mathcal{X} \to \mathcal{Y}$ as follows:

$$g(\mathbf{x}) = \underset{c \in \{1, \ldots, C\}}{\arg\max} \left[f_{t'}(\mathbf{x})\right]_c, \tag{14}$$

where $t' = \arg_{t \in \{1, \ldots, T\}} \left(l_{t-1} - l_t\right) \leqq \eta$.

Algorithm 1 summarizes the training algorithm for a deep forest.

**Algorithm 1** The training algorithm for a deep forest.

```
Input: Training set $S$, validation set $S_v$, learning algorithm $\mathcal{A}$ and the
    maximal number of cascade levels $T$.
Output: The deep forest model $g$.
initial $S_1$=$S$, $\ell_0$=1 and $t$=1
while t $\leq$ T do
    $h_t$=$\mathcal{A}\,(S_t)$
    Get $f_t$ according to Eq. (13)
    Get $g_t$ according to Eq. (14)
    Compute the validation error $\ell_t$=$L_{S_v}\,(g_t)$
    if ( $\ell_{t-1} - \ell_t$ )> $\eta$ then
        return $g$.
    end if
    $g$=$g_t$.
    $S_{t+1}$= $S_t$+$S$.
    t=t+1.
end while
return $g$.
```

In the current study, each forest in a layer can generate 40 probability values for each class (i.e. gestures). Therefore, each layer would generate 160 probability values as enhanced features produced by four random forests. Except the input of the first layer, all the following layers are fed by the contact of 160 enhanced features and the original 114 features from EMG and ACC.

### 3.3 Model evaluation

The database utilized in this study is the seventh version (DB7) of the publicly available Non Invasive Adaptive Prosthetics(NinaPro) database [35], which is designed to promote the state of sEMG controlled hand prosthetics. Data were collected offline from 20 able-bodied and 2 amputee

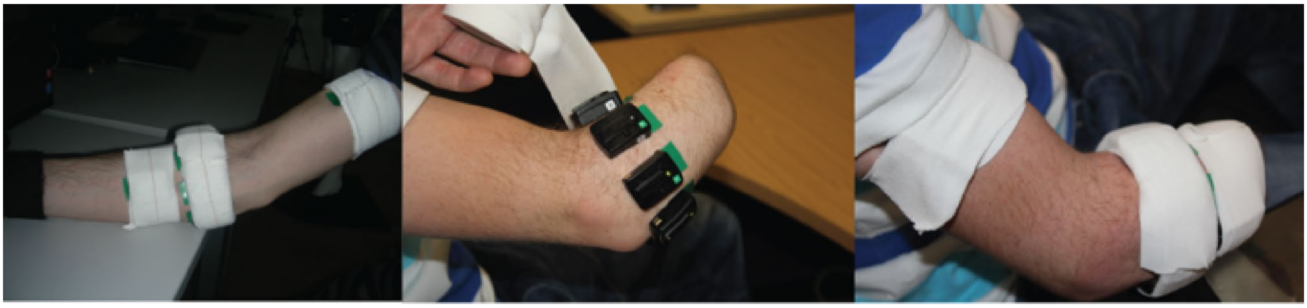

**Fig. 4** Sensor placement on able-bodied (left) and amputee subjects (centre, right) [20]. Eight EMG-IM sensors are equally spaced around the participants'forearm (3 cm below the elbow). Two are place on EDC and FDS muscles, and the rest two on biceps and triceps muscles. Elastic bandage is used to fix sensors

**Table 1** The medical records of two amputee subjects

| Gender | Age | Type of amputation | Cause of amputation | Years | Missing limb | Hand dominance (prior to amputation) | Prosthesis use |
|---|---|---|---|---|---|---|---|
| Male | 28 | Transradial | Car accident | 6 | Right | Right | Split hook |
| Male | 54 | Transradial | Car accident (epitheliod sarcoma) | 18 | Right | Right | Split hook |

subjects by adopting the NinaPro protocol [7, 8]. Subjects were asked to reproduce a series of 40 motions, including various individuated-finger, hand, wrist, grasping and functional movements. Each movement was repeated six times with 5 s resting interim between each trail. Two amputee volunteers were instructed to perform bilateral imaginary mirrored movements.

In NinaPro DB7, EMG and IM data were collected by using 12 Trigno Wireless electrodes (Delsys, Inc, www.delsys.com). The sampling frequency was set to 2 kHz for myoelectric signals and 128 Hz for ACC data. This database followed the NinaPro protocol to place 12 sensors [7]. Figure 4 demonstrates the electrode placement for a healthy subjects and an amputee. Two amputee's medical records is also provided in Table 1.

As shown in Fig. 3, each layer in the cascade forest structure contains four random forests. The number of trees in a random forest affects its performance. Training more trees is very likely to bring in higher classification accuracy, but increase the computational burden. In [10], the number of trees was set to 500 for both completely random forest or an ordinary random forest, as suggested in [36]. This study explores the most appropriate number of trees to determine the number of trees in a forest, and the data of one subject (s1) are used to evaluate how the number of trees influences the classification accuracy. Figure 5 shows the accuracy change along the increase of the number of trees. It can be found that with the increase of the number of trees, the accuracy shows a clear improvement. When the number of trees increases to 125, accuracy enhancement becomes stable. In order to balance computing resources and classification accuracy, this study sets the number of trees in each random forest to 125.

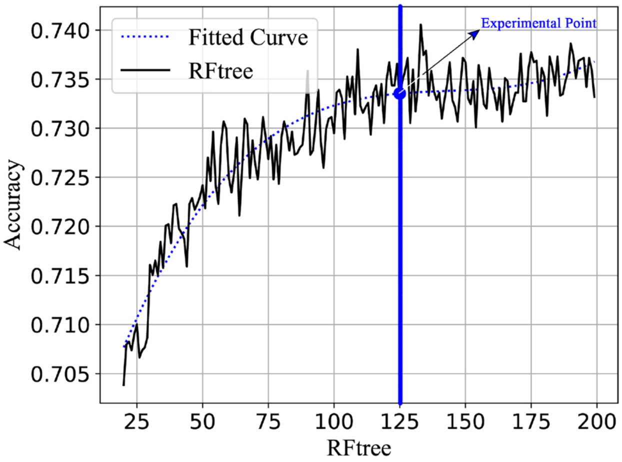

**Fig. 5** The change of classification accuracy when the use of different number of trees in each forest

Following the same protocol proposed in [20], 6-fold cross-validation is applied to obtain the average classification accuracy, where five repetitions are used to train the MMDF, and the remaining one for testing. This study evaluates MMDF in six aspects.

- Healthy subjects vs amputation subjects. The signals acquisition protocol for healthy subjects and amputees are different, which leads to the divergence of two groups of data. This study evaluates MMDF on both data sets from twenty healthy subjects and two amputees to evaluate its robustness.
- Feature fusion vs single modality. To verify the integration of the ACC signals with EMG can improve the classification accuracy, this study compares the achieved accuracy by MMDF in three input modalities, including EMG + ACC, EMG, and ACC.
- ACC feature reduction. MMDF's performance are compared before and after ACC feature dimension reduction, which aims to prove why the dimension of ACC feature need to be reduced.
- Comparison with other traditional classifiers on accuracy and speed. This study also compares MMDF with SVM, KNN, RF and the original GcForest. Some key parameters are adjusted empirically purchasing a higher classification accuracy. For KNN, parameter 'k' is set to 5. For SVM, RBF-kernel is selected. For RF, the number of the decision tree is set to 100. For the evaluation of the original GcForest, the original data are processed by multi-grained scanning before feeding to the deep forest.
- Confusion matrix. To observe how does the misclassification happen in MMDF, the confusion matrices of s1 (healthy subject) and s21 (amputee subject) are demonstrated.
- Comparison with literature. Relevant studies that take EMG and ACC as the input for hand motion classification are surveyed for comparison. Since the difference on the number of channels, classes, data sets, etc. It is hard to make a precise comparison on the performance of different classifier. Therefore, this study lists as much details as possible to demonstrate the merit of MMDF. The hardware environment and software environment used in this article are: 3.00GHz CPU 32G memory and python 3.8.6.

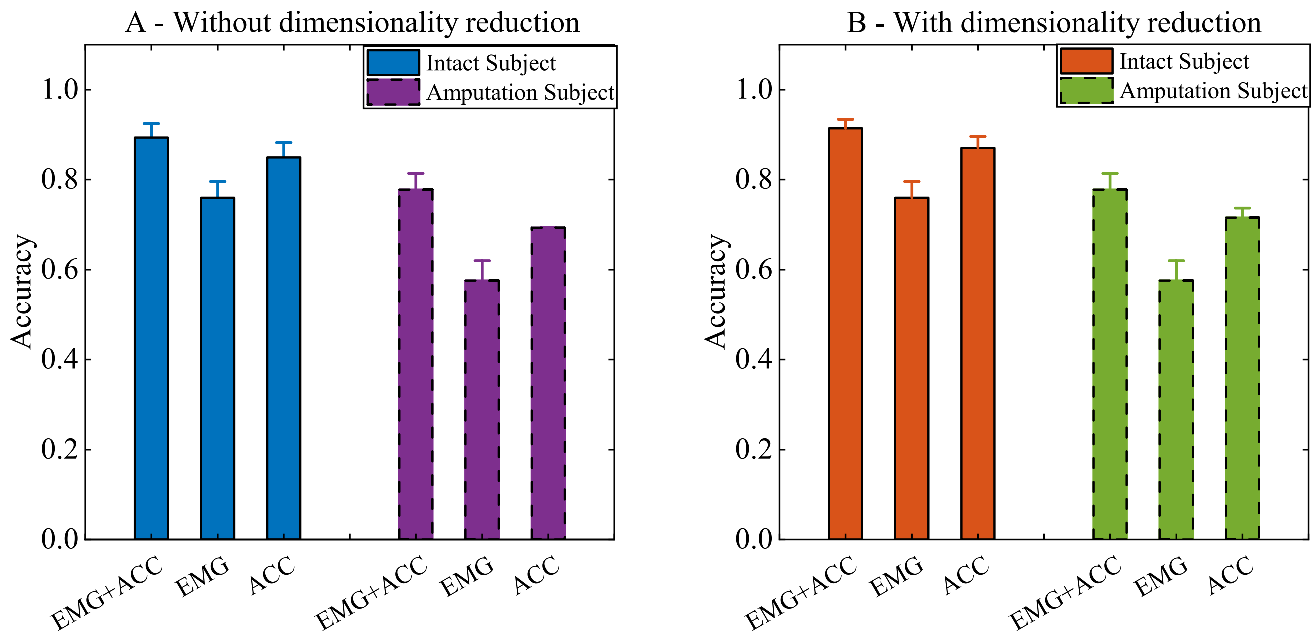

**Fig. 6** The above figure respectively shows the classification accuracy of the intact subject and the amputee subject for three different data input EMG+ACC, EMG, ACC, where **A** is the experimental result before dimensionality reduction, **B** is the result of dimensionality reduction processing on the ACC features using PCA. For intact subjects, each bar shows the average classification accuracy across 20 complete subjects for 40 types of gestures. For amputation subjects, each bar shows the average accuracy of two amputees, where 38 motions are classified

**Table 2** The hand motion recognition accuracy by GcForest under different conditions

| Data input | Subject type | EMG + ACC | EMG | ACC | EMG + ACC (dimensionality reduction) | ACC (dimensionality reduction) |
|---|---|---|---|---|---|---|
| Accuracy (%) | Healthy subjects | 89.33 ± 3.14 | 75.97 ± 3.61 | 84.92 ± 3.30 | 91.40 ± 2.02 | 87.04 ± 2.57 |
| | Amputation subjects | 75.23 ± 1.94 | 57.54 ± 4.44 | 69.30 ± 0.06 | 77.80 ± 9.61 | 71.53 ± 2.10 |

## 4 Results

Figure 6 shows the classification accuracy achieved by MMDF, in which the performance scores are compared between the settings with and without dimensionality reduction, and between intact subjects and amputees. It is found that the obtained accuracy after fusing EMG and ACC features is much higher than that of single EMG or ACC for all situations. For intact subject, the classification accuracy is 89.33 ± 3.14% after fusion, which is 13.35% and 4.41% higher than using single modality of sEMG or ACC signals when the feature dimensionality of ACC is not reduced (Fig. 6A). It also can be seen that using single modality of ACC signals can obtain higher classification accuracy than using only sEMG signals. Consistent results can be also found in amputee subjects, regardless whether dimensionality reduction is applied or not. In the comparison between intact subjects and amputee subjects, the accuracy improvement (13.35% vs 20.26%) obtained after fusing EMG and ACC features is more significant for intact subjects.

Figure 6 also reflects that dimensionality reduction on the ACC signals can enhance the performance of MMDF. In particular, the accuracy is 91.40 ± 2.02% for EMG+ACC with dimensionality reduction in intact subjects, which is 2.07% higher than the one obtained without dimensionality reduction. Also, the accuracy is 77.80 ± 9.61% for EMG + ACC with dimensionality reduction in amputee subjects, which is 2.57% higher than the one obtained without dimensionality reduction. These results are also reflected in Table 2.

Figure 7 and Table 3 show the comparison of MMDF with SVM, KNN and RF and the original GcForest. For healthy subjects, the average classification accuracy obtained by MMDF is up to 91.40 ± 2.02%, while they are only 68.88 ± 4.23%, 72.43 ± 5.49%, 72.91 ± 2.73%, and

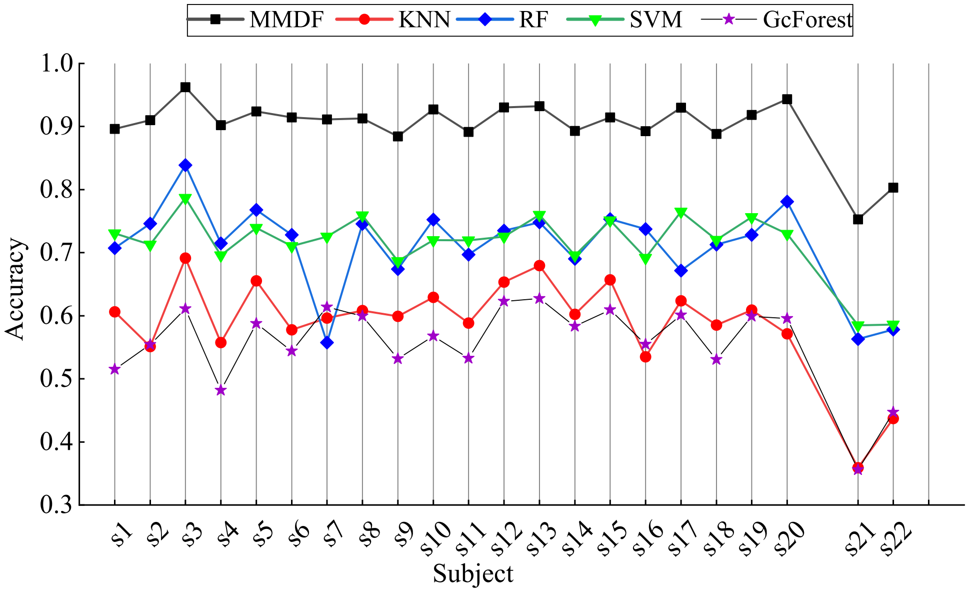

**Fig. 7** The classification accuracy comparison among MMDF (this study), RF, SVM, KNN and the original GcForest across all different subjects, including two amputees s21 and s22

**Table 3** The averaged classification accuracy comparison among MMDF (this study), RF, SVM, KNN and the original GcForest for healthy subjects

| Classifier | KNN | RF | SVM | GcForest | MMDF |
|---|---|---|---|---|---|
| Accuracy (%) | 68.88 ± 4.23 | 72.43 ± 5.49 | 72.91 ± 2.73 | 57.31 ± 4.06 | 91.40 ± 2.02 |

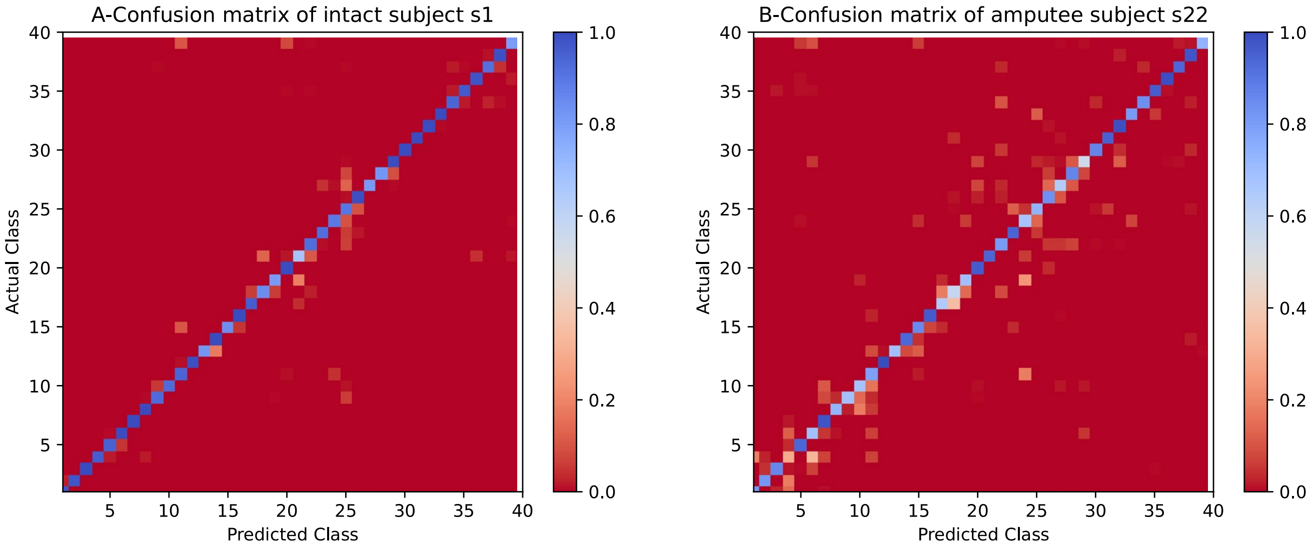

**Fig. 8** The Confusion matrices obtained from subject s1 (intact) and subject s21 (amputee), and are displayed in the left panel and the right panel, respectively. The x-axis of the matrices is the class label of the predicted results by the proposed methods MMDF, while the y-axis shows the actual class label. The change of color shows the accuracy

57.31 ± 4.06% for KNN, RF, SVM, and the original GcForest, respectively. For the amputees (i.e. s21 and s22), MMDF achieves the average accuracy at 77.80 ± 9.61%, which is 38.00%, 20.75%, 19.26%, and 20.47% higher than that of KNN, RF, SVM, and the original GcForest, respectively. Besides, it can be found that the accuracy obtained from amputees are much lower than that of healthy subjects.

**Table 4** Related studies on the fusion of sEMG and ACC signals to recognise hand motions on healthy subjects

| References | Electrodes | Classes | Features | Database | Classifier | Accuracy (%) | Number of subject |
|---|---|---|---|---|---|---|---|
| [20] | 12 | 40 | sEMG (MAV/ML/AR4/LogVar) + ACC (MEAN) | Ninapro DB7 | LDA | About 77 | 20 (intact) |
| [19] | 8 | 12 | sEMG (MAV) + ACC (MEAN) | Self-test | KNN | 96.11 | 5 (intact) |
| [37] | 8 | 8 | sEMG) +) ACC | Self-test | LSTM | 96.87 | 5 (intact) |
| [4] | 12 | 18 | sEMG (MAV) + ACC (MEAN) | Ninapro DB5 | SVM | 88.7 ± 2.6 | 5 (amputee) |
| This study | 12 | 40 | sEMG (MAV/ML/AR4/SSC) + ACC (MEAN) | Ninapro DB7 | Deep Forest | 91.40 | 20 (intact) |

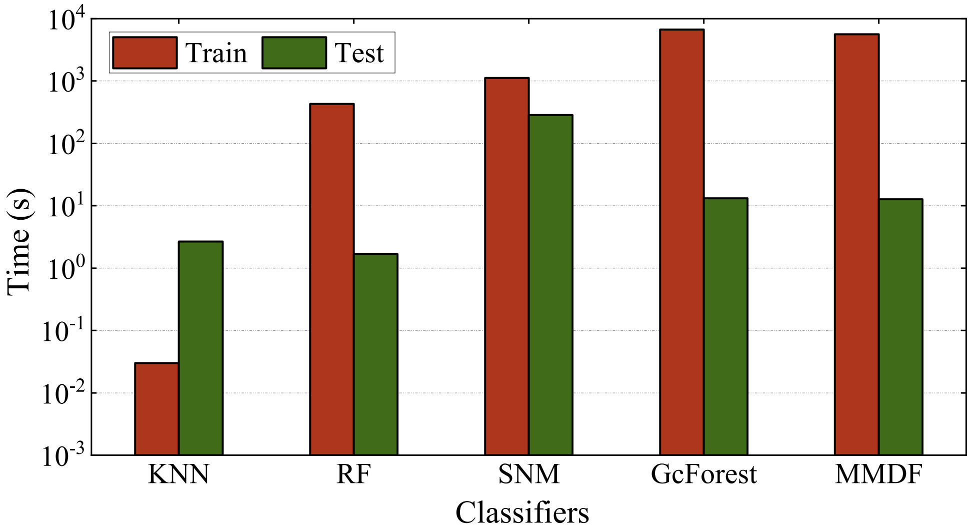

**Fig. 9** The comparison of time cost on different classifiers

Figure 8 shows the confusion matrices of the intact subject s1 and the amputee subject s22 using the proposed MMDF method. The overall accuracy for s1 and s22 are around 90% and 80%, respectively. It can be seen from both matrices that most of the classes are correctly predicted, and MMDF performs better on the intact subjects than on amputee subjects. Besides, it can be also found that most misclassifications happen among neighboring classes because of their similarity.

Table 4 summaries some related studies regarding the fusion of sEMG and ACC signals to conduct hand motion recognition. To the best of our knowledge, the only study that utilises the same database (i.e. Ninapro DB7) to implement hand gesture recognition is by Krasoulis et al. [20]. Using the combination of sEMG features (i.e. MAV/ML/AR4/LogVar) and ACC feature (i.e.MEAN), they obtained the accuracy about 77% for 20 healthy subjects, which is much lower than 91.4% achieved by the proposed MMDF. Although Xie et al. [19] and Kong et al. [37] achieve a very high classification accuracy ( Both are about 96%) using both KNN and LSTM, the number of gestures is only 12 and 8, respectively. Based on Ninapro DB5, another study obtains the accuracy of 88.7% for the classification of 18 gestures [4]. Although the number of gestures is half of that of the current study, the accuracy is still 2.7% lower.

Figure 9 shows the training time and testing time of each classifier under the same laboratory environment. With the same amount of data, the time-consuming of the classifier with deep forest structure is much greater than that of traditional classifiers. This is due to the complex structure of deep forests. In the comparison with the original GcForest, the time-consuming of MMDF is slightly reduced. In practical applications, time-consuming is a very important factor. In future research, more effort will be devoted to reducing time-consuming.

## 5 Discussion

This section discusses the study in three aspects: (1) why grained scanning is replaced by human-crafted feature extraction; (2) the selected features and the advantage of GcForests; (3) the benefit of ACC signals in sEMG based hand motion recognition.

**Table 5** The accuracy, recall, precision and F-score of each classifier obtained from subject s1 (intact) after dimensionality reduction

| Classifier | KNN | RF | SVM | GcForest | MMDF |
|---|---|---|---|---|---|
| Accuracy (%) | 60.61 | 70.74 | 73.07 | 55.42 | 88.26 |
| Recall (%) | 61.34 | 72.44 | 74.86 | 58.02 | 90.12 |
| Precision (%) | 60.81 | 63.96 | 72.91 | 58.33 | 83.89 |
| F-score (%) | 61.07 | 67.94 | 73.81 | 58.17 | 86.89 |

### 5.1 The change in GcForest

The original GcForest algorithm consists of multi-grained scanning and cascading forests [10]. Multi-grained scanning similar to sliding window technology, which is usually used to process raw data for data argumentation for better training. For instance, Zhai et al. obtain image features through grained scanning to predict the level of facial appearance [38], and Liu et al. introduce a multi-grained scanning method to enrich features for credit scoring [39]. The current study replaces the grained scanning by human-crafted feature extraction, which significantly reduces the dimension of the input for cascading forest, and speeds up model converging. It is probably the reason why the proposed MMDF can achieve higher accuracy than the original GcForest, which is consistent with the results in [27].

### 5.2 Feature selection and the advantage of GcForest

Feature extraction is a key factor in sEMG based pattern recognition, and the quality of feature selection directly affects its performance. This study selects five time domain features for the experiments as mostly suggested in [31, 40]. However, it is worthy to be studied that whether including more features can further improve its performance.

Random forest, as the basic element of GcForest, it gets much lower classification accuracy than that of MMDF in our experiments, as summarised in Table 3. It indicates that the deep layer-by-layer cascade structure plays a critical role in GcForest, where the probability output of each layer is used as the input of the next layer. But the reason why such cascade structure would improve the performance still need further investigation. Whether KNN and SVM can be also constructed in cascade structure can be further studied. Besides, the advantage of MMDF can also be reflected by other classifier metrics, including Recall, Precision and F-score, seen in Table 5.

### 5.3 The benefit of accelerometers

The current study proves that ACC plays a significant impact on the motion classification performance, which is consistent with the ones reported in [4, 19, 20, 37]. For a simple comparison, ACC even outperforms sEMG, as shown in Fig. 7. ACC seems to be a better choice when single modality signals is considered. However, it needs to be noticed that the robustness of ACC have not been widely investigated, such as the impact of arm position, electrode displacement, etc. More importantly, there exists fundamental theory that sEMG is somewhat the intrinsic cause of hand motions including the applied force, but ACC is only the external appearance of the motion from a specific angle of view. ACC cannot replace the function of sEMG on detecting the onset/offset of movement and predicting the muscle force [41]. In sum, ACC can be an important supplement to sEMG for hand motion classification, but could not replace EMG. Since multi-channel ACC signals contain a large amount of redundant information [4], it is necessary to reduce its dimension before classification, which is also pointed out in [19].

## 6 Conclusion

This study proposes a multi-modality deep forest (MMDF) framework to identify hand motions, which fuses the input of sEMG and ACC signals. Instead of using original EMG signals, five EMG features (i.e. MAV, WL, ZC, SSC and AR4) and an ACC feature (i.e. MEAN) are selected as the input of MMDF. Ninapro DB7 is utilised to validate the effectiveness of the proposed framework. The accuracy reaches up to 91.40 ± 2.02% (intact subjects) and 77.80 ± 9.61% (amputation subjects) for the classification of 40 types of hand movements, which is significantly higher than that of the competing classifiers. Besides, the study also shows that ACC is an excellent supplementary modality for myo-control, but it is suggested to reduce the feature dimensionality of ACC signals. Moreover, the experimental result shows that although the training cost of MMDF is expensive, its prediction delay is acceptable.

In the future, deep forest structure will be further optimized to improve the classification accuracy and reduce its computing burden, and its robustness will also be tested towards electrode shift, muscle fatigue, etc.

**Acknowledgements** This work is supported in part by the basic public welfare research program of Zhejiang Province (No. LQ20F020016), the national natural science foundation of China (No. 62106147).

## References

1. Joshi D, Nakamura BH, Hahn ME (2015) High energy spectrogram with integrated prior knowledge for EMG-based locomotion classification. Med Eng Phys 37(5):518–524

2. Matrone GC, Cipriani C, Secco EL, Magenes G, Carrozza MC (2010) Principal components analysis based control of a multi-DoF underactuated prosthetic hand. J Neuroeng Rehabil 7(1):1–13
3. Kuzborskij I, Gijsberts A, Caputo B.(2012) On the challenge of classifying 52 hand movements from surface electromyography. In: 2012 Annual international conference of the IEEE engineering in medicine and biology society, IEEE, p 4931–4937
4. Liu J, Chen W, Li M, Kang X (2016) Continuous recognition of multifunctional finger and wrist movements in amputee subjects based on sEMG and accelerometry. Open Biomed Eng J 10:101
5. Chen H, Zhang Y, Li G, Fang Y, Liu H (2020) Surface electromyography feature extraction via convolutional neural network. Int J Mach Learn Cybern 11(1):185–196
6. Nogales RE, Benalcázar ME (2021) Hand gesture recognition using machine learning and infrared information: a systematic literature review. Int J Mach Learn Cybern 12(10):2859–2886
7. Atzori M, Gijsberts A, Castellini C, Caputo B, Hager A-GM, Elsig S, Giatsidis G, Bassetto F, Müller H (2014) Electromyography data for non-invasive naturally-controlled robotic hand prostheses. Sci Data 1(1):1–13
8. Atzori M, Gijsberts A, Kuzborskij I, Elsig S, Hager A-GM, Deriaz O, Castellini C, Müller H, Caputo B (2014) Characterization of a benchmark database for myoelectric movement classification. IEEE Trans Neural Syst Rehabil Eng 23(1):73–83
9. Gijsberts A, Atzori M, Castellini C, Müller H, Caputo B (2014) Movement error rate for evaluation of machine learning methods for sEMG-based hand movement classification. IEEE Trans Neural Syst Rehabil Eng 22(4):735–744
10. Zhou Z-H, Feng J (2019) deep forest: towards an alternative to deep neural networks. In: Proceedings of the twenty-sixth international joint conference on artificial intelligence (IJCAI-17), p 3553–3559
11. Jiang N, Muceli S, Graimann B, Farina D (2013) Effect of arm position on the prediction of kinematics from EMG in amputees. Med Biol Eng Cmput 51(1):143–151
12. Fougner A, Scheme E, Chan AD, Englehart K, Stavdahl Ø (2011) Resolving the limb position effect in myoelectric pattern recognition. IEEE Trans Neural Syst Rehabil Eng 19(6):644–651
13. Geng Y, Zhou P, Li G (2012) Toward attenuating the impact of arm positions on electromyography pattern-recognition based motion classification in transradial amputees. J Neuroeng Rehabil 9(1):1–11
14. Liu L, Chen X, Lu Z, Cao S, Wu D, Zhang X (2016) Development of an EMG-ACC-based upper limb rehabilitation training system. IEEE Trans Neural Syst Rehabil Eng 25(3):244–253
15. Zhang X, Chen X, Li Y, Lantz V, Wang K, Yang J (2011) A framework for hand gesture recognition based on accelerometer and EMG sensors. IEEE Trans Syst Man Cybern Part A Syst Hum 41(6):1064–1076
16. Khushaba RN, Al-Timemy A, Kodagoda S, Nazarpour K (2016) Combined influence of forearm orientation and muscular contraction on EMG pattern recognition. Expert Syst Appl 61:154–161
17. Young A, Kuiken T, Hargrove L (2014) Analysis of using EMG and mechanical sensors to enhance intent recognition in powered lower limb prostheses. J Neural Eng 11(5):056021
18. Atrey PK, Hossain MA, El Saddik A, Kankanhalli MS (2010) Multimodal fusion for multimedia analysis: a survey. Multimedia Syst 16(6):345–379
19. Xie X, Liu Z (2017) Dynamic gesture recognition method based on EMG and ACC signal. J Comput Appl 37:2700
20. Krasoulis A, Kyranou I, Erden MS, Nazarpour K, Vijayakumar S (2017) Improved prosthetic hand control with concurrent use of myoelectric and inertial measurements. J Neuroeng Rehabil 14(1):1–14
21. Guo Y, Liu S, Li Z, Shang X (2018) Bcdforest: a boosting cascade deep forest model towards the classification of cancer subtypes based on gene expression data. BMC Bioinform 19(5):1–13
22. Liu P, Wang X, Yin L, Liu B (2020) Flat random forest: a new ensemble learning method towards better training efficiency and adaptive model size to deep forest. Int J Mach Learn Cybern 11:2501–2513
23. Zhang Y, Xu T, Chen C, Wang G, Zhang Z, Xiao T (2021) A hierarchical method based on improved deep forest and case-based reasoning for railway turnout fault diagnosis. Eng Fail Anal 127:105446
24. Sun L, Mo Z, Yan F, Xia L, Shan F, Ding Z, Song B, Gao W, Shao W, Shi F (2020) Adaptive feature selection guided deep forest for COVID-19 classification with chest CT. IEEE J Biomed Health Inform 24(10):2798–2805
25. Daouadi KE, Rebaï RZ, Amous I (2021) Optimizing semantic deep forest for tweet topic classification. Inf Syst 101:10–18
26. Ding J, Wu Y, Luo Q, Du Y (2021) A fault diagnosis method of mechanical bearing based on the deep forest. J Vib Shock 40:107–113
27. Fang Y, Yang H, Zhang X, Liu H, Tao B (2020) Multi-feature input deep forest for EEG-based emotion recognition. Front Neurorobotics 14:617531
28. Englehart K, Hudgins B, Parker PA, Stevenson M (1999) Classification of the myoelectric signal using time-frequency based representations. Med Eng Phys 21(6–7):431–438
29. Oskoei MA, Hu H (2008) Support vector machine-based classification scheme for myoelectric control applied to upper limb. IEEE Trans Biomed Eng 55(8):1956–1965
30. Fougner A, Scheme E, Chan AD, Englehart K, Stavdahl Ø (2011) A multi-modal approach for hand motion classification using surface emg and accelerometers. In: 2011 Annual international conference of the IEEE engineering in medicine and biology society, IEEE, p 4247–4250
31. Phinyomark A, Phukpattaranont P, Limsakul C (2012) Feature reduction and selection for EMG signal classification. Expert Syst Appl 39(8):7420–7431
32. Fang Y, Yang J, Zhou D, Ju Z (2021) Modelling EMG driven wrist movements using a bio-inspired neural network. Neurocomputing 470:89–98
33. Cheng J, Chen M, Li C, Liu Y, Song R, Liu A, Chen X (2020) Emotion recognition from multi-channel EEG via deep forest. IEEE J Biomed Health Inf 25(2):453–464
34. Yao H, He H, Wang S, Xie Z 2019) EEG-based emotion recognition using multi-scale window deep forest. In: 2019 IEEE symposium series on computational intelligence (SSCI), IEEE, p 381–386 (
35. Sebelius FC, Rosen BN, Lundborg GN (2005) Refined myoelectric control in below-elbow amputees using artificial neural networks and a data glove. J Hand Surg 30(4):780–789
36. Liu FT, Ting KM, Yu Y, Zhou Z-H (2008) Spectrum of variable-random trees. J Artif Intell Res 32:355–384
37. Kong D, Zhu J (2019) Gesture recognition based on fusion of surface electromyography and acceleration information. Electron Meas Technol

38. Zhai Y, Lv P, Deng W, Xie X, Yu C, Gan J, Zeng J, Ying Z, Labati RD, Piuri V (2020) Facial beauty prediction via deep cascaded forest. Int J High Perform Syst Archit 9(2–3):97–106
39. Liu W, Fan H, Xia M (2021) Step-wise multi-grained augmented gradient boosting decision trees for credit scoring. Eng Appl Artif Intell 97:104036
40. Fang Y, Hettiarachchi N, Zhou D, Liu H (2015) Multi-modal sensing techniques for interfacing hand prostheses: a review. IEEE Sens J 15(11):6065–6076
41. Castellini C, Van Der Smagt P (2009) Surface EMG in advanced hand prosthetics. Biol Cybern 100(1):35–47

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