



## An end-to-end lower limb activity recognition framework based on sEMG data augmentation and enhanced CapsNet

Changhe Zhang<sup>a</sup>, Yangan Li<sup>b</sup>, Zidong Yu<sup>a</sup>, Xiaolin Huang<sup>b</sup>, Jiang Xu<sup>b,\*</sup>, Chao Deng<sup>a,\*</sup>

<sup>a</sup> School of Mechanical Science and Engineering, Huazhong University of Science and Technology, Wuhan 430074, China

<sup>b</sup> Department of Rehabilitation Medicine, Tongji Hospital, Tongji Medical College, Huazhong University of Science and Technology, Wuhan 430030, China



### ARTICLE INFO

#### Keywords:

Lower limb activity recognition  
Biomedical signal analysis  
sEMG denoising  
Class-imbalanced problem  
Capsule network

### ABSTRACT

Recently, lower limb activity recognition (LLAR) based on surface electromyography (sEMG) signal has attracted increasing attention, mainly due to its applications in the control of robots and prosthetics, medical rehabilitation, etc. Traditional machine learning-based LLAR methods rely on expert experience for feature extraction. In addition, the noise interference and class-imbalanced problem can also affect the recognition effect. Aiming at these problems, a LLAR framework based on sEMG data augmentation (DA) and enhanced capsule network (ECN) is proposed in this paper. Firstly, a hybrid denoising technique combining variational mode decomposition and non-local means estimation is designed to effectively filter out noise components mixed in the sEMG. Then, K-Means synthetic minority oversampling technique is utilized to synthesize new samples for minority classes, thereby overcoming the influence of class imbalance on recognition reliability. Finally, an ECN model is constructed to implement end-to-end LLAR, in which an efficient channel attention module is embedded to mine sensitive features, thus further improving the feature learning ability of the classifier. Experimental results indicate that the proposed framework is applicable to multiple types of individuals, including healthy subjects, patients with knee abnormalities, and patients with stroke, providing more satisfactory recognition performance and robustness than state-of-the-art methods..

### 1. Introduction

In recent decades, lower limb activity recognition (LLAR) technology has played an increasingly crucial role in many disciplines and fields (Qiu et al., 2022; Vijayvargiya, Singh, Kumar, & Tavares, 2022). For the elderly people, the disabled people and patients with stroke, it is beneficial to secure or ameliorate their activities of daily living. In the rehabilitation field, it can help physicians assess the health status of patients, thus optimizing the treatment regimen. Also, it is widely applied in robot and exoskeleton control, intelligent prosthesis and other fields (Vijayvargiya, Kumar, et al., 2021).

At present, scholars have explored lots of LLAR researches based on electromyography (EMG) signals (Naik et al., 2018; Vijayvargiya, Singh, Kumar, & Dey, 2022; Wang, Dai, et al., 2021). Compared with intramuscular EMG, surface EMG (sEMG) acquisition is non-invasive and more convenient. Since muscles in lower limbs are buried deep under the skin and are more intricate, the recognition of lower limb activities is more difficult than that of upper limbs (Wang, Wu, Dey, et al., 2020).

Recently, scholars have explored many methods that combine sEMG and machine learning (ML) algorithms (Naik et al., 2018; Vijayvargiya et al., 2022; Wang, Dai, et al., 2021; Wei et al., 2022; Zhang, Zhang, & Elsabbagh, 2022). Although certain progress has been made, these methods rely on expert experience to manually extract features of sEMG and are difficult to apply to large datasets (Xiong, Zhang, Zhao, & Zhao, 2021). Nowadays, deep learning (DL) has achieved better performance in many fields due to its excellent feature learning and decision-making ability (Liu, Zhang, & Jiang, 2022). As one of the most representative DL models, convolutional neural network (CNN) can automatically obtain valuable features from sEMG and predict activity types through convolution (Conv) and pooling operations, which has aroused strong interest of scholars (Gautam, Panwar, Biswas, & Acharya, 2020; Issa & Khaled, 2022; Si, Dai, & Wang, 2022). However, there are still some research gaps that need to be considered and addressed, as described below:

1) In most studies, the types and numbers of subjects were limited, and the applicability to individuals with lower limb motor dysfunction,

\* Corresponding authors.

E-mail addresses: [changhe\\_zhang@hust.edu.cn](mailto:changhe_zhang@hust.edu.cn) (C. Zhang), [liyanan@hust.edu.cn](mailto:liyanan@hust.edu.cn) (Y. Li), [zidong\\_yu@hust.edu.cn](mailto:zidong_yu@hust.edu.cn) (Z. Yu), [xiaolin2006@tjh.tjmu.edu.cn](mailto:xiaolin2006@tjh.tjmu.edu.cn) (X. Huang), [xujiang@hust.edu.cn](mailto:xujiang@hust.edu.cn) (J. Xu), [dengchao@hust.edu.cn](mailto:dengchao@hust.edu.cn) (C. Deng).

such as patients with stroke, is still unclear.

2) Due to individual differences and ambient impacts, raw sEMG signals usually contain a large amount of noise (Vijayvargiya, Kumar, Dey, & Tavares, 2020), such as motion artifacts, white Gaussian noise (WGN), etc., posing a challenge to the accuracy and anti-noise ability of LLAR methods.

3) Typically, classifiers are designed for balanced class distribution and require large amounts of training data to improve model performance. However, it is a challenge to collect data of different individuals for a long time. Moreover, for some patients, it is difficult to collect data under specific activities, such as going up/down stairs. The class-imbalanced problem will greatly reduce the reliability of classification results (Vijayvargiya, Prakash, et al., 2021).

4) CNNs have some inherent flaws that may affect their performance, e.g., Conv operation has translation invariance, leading to poor ability to learn affine transformation, and pooling operation may lead to loss of feature information (Liu et al., 2022). Some variants of CNN, such as Inception networks (Szegedy, Ioffe, Vanhoucke, & Alemi, 2017), enhance model performance by increasing network width, but this inevitably leads to a significant increase in parameters and computations.

Based on the above motivations, this paper aims to propose a sEMG-based LLAR framework with high accuracy and strong anti-noise ability, which can be applied to multiple types of individuals, thus providing a research basis for the development of neural-controlled lower limb rehabilitation robot. The main contributions can be summarized as follows:

- 1) An end-to-end LLAR framework based on sEMG data augmentation (DA) and enhanced capsule network (ECN) is proposed, which performs well in LLAR tasks for healthy subjects, patients with knee abnormalities, and patients with stroke.

- 2) A hybrid denoising technique combining variational mode decomposition (VMD), permutation entropy (PE) and non-local means estimation (NLM) is designed to effectively filter out the noise components mixed in the raw sEMG signals, thus improving the anti-noise ability of the proposed framework.

- 3) The K-Means synthetic minority oversampling (K-Means SMOTE) algorithm is applied to synthesize high-quality samples for minority classes, which can greatly eliminate the adverse effects of class-imbalanced on the classification performance of DL models.

- 4) The ECN model exhibits better satisfactory performance than existing literature, in which the embedded efficient channel attention (ECA) module can improve the classifier's feature learning performance by improving its ability to identify sensitive features.

The rest of this paper is organized as follows. Section 2 introduces related work. Section 3 describes the proposed framework in detail. Section 4 contains experimental verification and discussion of two case studies. Section 5 presents the conclusion and future work.

## 2. Related work

### 2.1. Signal sources in LLAR research

At present, the LLAR implementations mainly include: computer vision-based methods (Sheng, Wang, Xiong, Hou, & Zhang, 2019), biomechanical signals-based methods (Feng, Wang, Vladareanu, Chen, & Jin, 2019), bioelectrical signals-based methods (Gu, Jiang, Han, Gan, & Wang, 2023; Lin, Jiang, Chen, Huang, & Liao, 2021) and multi-source data fusion-based methods (Al-Quraishi et al., 2021; Iqbal et al., 2021). The first method is limited by deployment space and privacy. With the rapid development of sensors, more and more wearable devices for collecting human activity data have emerged (Wang, Cang, & Yu, 2019). The second method uses accelerometers, etc., to measure physical changes in human activity, but inherent delay causes a lag in prediction. The third method includes electroencephalogram (EEG) (Gu et al., 2023), electrooculogram (EOG) (Lin et al., 2021), EMG, etc. However,

EEG can only identify limited activity types, and EOG is susceptible to interference from user's unconscious blinking. For the multi-source data fusion methods, system cost will rise rapidly with the increase of sensors. Compared to mechanical signals, EMG has the advantages of directly reflecting muscle contraction state and forming before the onset of motion (Al-Quraishi et al., 2021).

### 2.2. LLAR research based on traditional shallow ML algorithms

Traditional ML-based LLAR mainly includes two steps. Firstly, signal analysis methods are used to extract features of sEMG from the time domain or frequency domain. Then, some ML algorithms, e.g., support vector machine (SVM) (Naik et al., 2018), k-nearest neighbor (KNN) (Wei et al., 2022), random forest (RF) (Vijayvargiya et al., 2020), linear discriminant analysis (LDA) (Vijayvargiya, Gupta, et al., 2021), etc., are used to obtain recognition results. For example, Wei et al. proposed a feature extraction method combining VMD and entropy (Wei et al., 2022). In addition, further feature engineering, such as feature selection or fusion, can be used to improve the feature quality. Naik et al. combined Fisher score and profile likelihood maximization to reduce redundancy features (Naik et al., 2018). Zhang et al. constructed a multi-dimensional feature set of sEMG and used differential evolution for feature fusion (Zhang et al., 2022). However, these methods still have some limitations: 1) The features are extracted manually, and the feature quality determines recognition accuracy; 2) Due to the limited nonlinear mapping ability of ML, it is difficult to handle large datasets.

### 2.3. LLAR research based on DL algorithms

DL is a pattern recognition technology developed in recent years, with powerful nonlinear mapping and automatic feature extraction abilities, which can effectively overcome the defects of traditional ML algorithms. Vijayvargiya et al. proposed a LLAR method based on one-dimensional CNN (1D-CNN) (Vijayvargiya, Kumar, et al., 2021). Gautam et al. constructed MyoNet, which can simultaneously realize LLAR and knee angle prediction (Gautam et al., 2020). Issa et al. preprocessed sEMG by short-time Fourier transform (STFT) and realized LLAR with two-dimensional CNN (2D-CNN) (Issa & Khaled, 2022). However, some inherent flaws of CNN, such as translation invariance and local feature loss, may reduce its classification performance. Capsule network (CapsNet) can overcome these defects by replacing traditional neuron scalar with "capsule vector" and constructing a deeper capsule layer with "dynamic routing" algorithm, which has a promising application prospect in LLAR (Sabour, Frosst, & Hinton, 2017).

In order to further improve the feature learning ability and classification performance of DL models, there are typically two ways: 1) By increasing the network width, i.e., multi-scale feature extraction, to mine richer feature information, but this inevitably leads to a large increase in model parameters and computation. 2) By fusing attention mechanism modules to enhance model's attention to sensitive features and suppress irrelevant features, thereby improving its performance with only a few additional parameters. Therefore, to further improve the classification performance of CapsNet, this paper deeply integrates it with a local cross-channel interaction module, i.e., the efficient channel attention (ECA) (Wang, Wu, Zhu, et al., 2020).

### 2.4. Research on denoising of sEMG signal

Typically, most studies simply filter out noise outside the spectrum of sEMG by band-pass filter (Al-Quraishi et al., 2021; Wei et al., 2022). For noises in the range of sEMG spectrum, such as WGN, wavelet denoising (WD) is one of the commonly used methods (Thukral, Kumar, & Arora, 2019). Also, some mode decomposition techniques, such as empirical mode decomposition (EMD) (Chen, Xu, Liu, McKeown, & Wang, 2017), ensemble EMD (EEMD) (Vijayvargiya, Gupta, et al., 2021), etc., are able to handle time-varying and non-stationary signals without changing

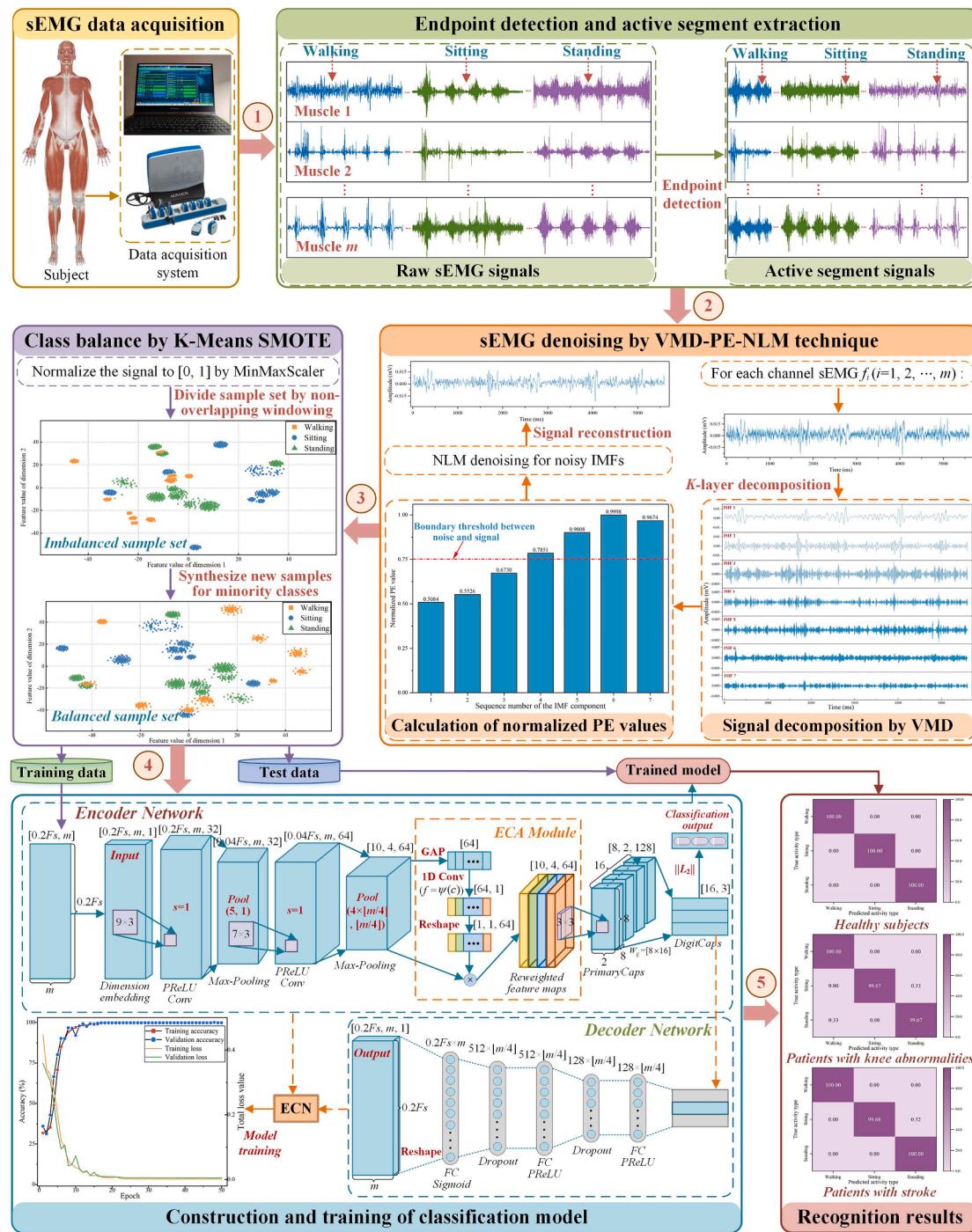


Fig. 1. Flowchart of the proposed LLAR framework.

their characteristics, showing great potential in bioelectrical signal denoising. However, the wavelet basis function and threshold in WD need to be set empirically, and EMD and its variants have inherent drawbacks such as mode aliasing and end effect (Ma, Lv, Lin, Sheng, & Zhu, 2020).

As an adaptive non-recursive decomposition method, VMD can effectively suppress the defects in EMD and EEMD (Dragomiretskiy & Zosso, 2013). After using VMD to decompose sEMG into multiple intrinsic mode functions (IMFs), the following problems still need to be considered: 1) How to identify the boundary between noiseless and noisy IMFs; 2) Some studies used only noiseless IMFs to reconstruct the signal, which may result in the loss of useful information. In this paper,

PE is introduced to address the first issue, mainly due to its advantage in measuring the complexity of time series (Bandt & Pompe, 2002), then 1D NLM algorithm is adopted to denoise the noisy IMFs (Singh, Shah-nawazuddin, & Pradhan, 2018). On this basis, the processed and unprocessed IMFs are combined to obtain the denoised sEMG signal.

## 2.5. Research on classification under the class-imbalanced condition

Classification of class-imbalanced data remains a common and challenging problem. Currently, most studies provided solutions from two aspects: classifier and data. The former is mainly to improve existing classifiers or construct new classifiers, which is difficult to deal with

extreme imbalance. The latter, which generates or synthesizes new samples before classification, is more general and can be divided into the following two types of methods:

1) DA methods represented by generative adversarial networks (GANs). Chen et al. converted sEMG signals into gray images and proposed a deep Conv GAN-based DA method (Chen, Qian, Wang, & Fang, 2022). However, GANs often suffer from the gradient vanishing and mode collapse problems, and the training process usually consumes a lot of computing resources.

2) DA methods represented by SMOTE (Chawla, Bowyer, Hall, & Kegelmeyer, 2002). As a feature space-based oversampling method, SMOTE provided a new solution to this problem. For example, it was applied to balance the class distribution in the sEMG classification of upper limbs (de Freitas et al., 2022). However, the main drawbacks of SMOTE are the possibilities of synthesizing noisy samples and blurring boundaries between different classes.

Aiming at these, K-Means SMOTE was further proposed by combining SMOTE with clustering algorithm (Douzas, Bacao, & Last, 2018). In this paper, to eliminate the adverse effects of class-imbalanced problem on the classification performance of ECN, the K-Means SMOTE is applied to synthesize samples for minority classes, thus balancing the class distribution.

### 3. Materials and methods

#### 3.1. Overview of the proposed LLAR framework

An overview of the proposed end-to-end LLAR framework based on sEMG data DA and ECN model is shown in Fig. 1. Specific steps are described as follows:

**Step 1:** The raw sEMG signals of the relevant muscles of different subjects under different types of lower limb activities are collected through the data acquisition system.

**Step 2:** The active segments of the raw sEMG signals are extracted by the double-threshold method for preliminary preprocessing, and then the noise components in the preprocessed sEMG signals are effectively filtered out by the VMD-PE-NLM technique.

**Step 3:** The denoised sEMG signals are divided into samples of the same size by the non-overlapping windowing technique, and new samples are synthesized by the K-Means SMOTE algorithm for the minority classes in the sample set, thus balancing the class distribution.

**Step 4:** The balanced sample set is divided into two parts according to a certain proportion: training sample set and test sample set. After that, the training set is input into the constructed ECN model for training.

**Step 5:** The test set is input into the trained ECN model to obtain the final LLAR results.

#### 3.2. sEMG denoising based on the VMD-PE-NLM technique

##### 3.2.1. Variational mode decomposition

VMD was an adaptive time-frequency analysis method, which can decompose the signal into a series of sparse IMFs by searching for the optimal solution of constrained variational model. For input signal  $f(t)$ , the constrained variational problem is constructed as Eq. (1), where  $K$  is the number of decomposed modes, and  $k = 1, 2, \dots, K$ ;  $\partial_t$  denotes the partial derivative;  $\delta(t)$  is the unit impulse function;  $j$  is the imaginary unit;  $*$  denotes the Conv operation;  $\mu_k$  and  $\omega_k$  represent the  $k$ th IMF and the corresponding center frequency, respectively.

$$\left\{ \begin{array}{l} \min_{\{\mu_k\}, \{\omega_k\}} \left\{ \sum_k \left\| \partial_t \left[ \left( \delta(t) + \frac{j}{\pi t} \right) * \mu_k(t) \right] e^{-j\omega_k t} \right\|_2^2 \right\} \\ \text{s.t. } \sum_k \mu_k = f \end{array} \right. \quad (1)$$

The quadratic penalty factor  $\alpha$  and Lagrange multiplier  $\lambda(t)$  are introduced to transform Eq. (1) into an augmented Lagrange function:

$$L(\{\mu_k\}, \{\omega_k\}, \lambda) = \alpha \sum_k \left\| \partial_t \left[ \left( \delta(t) + \frac{j}{\pi t} \right) * \mu_k(t) \right] e^{-j\omega_k t} \right\|_2^2 + \left\| f(t) - \sum_k \mu_k(t) \right\|_2^2 + \left\langle \lambda(t), f(t) - \sum_k \mu_k(t) \right\rangle \quad (2)$$

The alternate direction method of multipliers algorithm is used to search the saddle points of Eq. (2) iteratively, i.e., the optimal solution of Eq. (1). The solutions of mode component  $\mu_k$  and center frequency  $\omega_k$  can be obtained according to Eq. (3) and Eq. (4), respectively:

$$\hat{\mu}_k^{N+1}(\omega) = \frac{\hat{f}(\omega) - \sum_{i \neq k} \hat{\mu}_i(\omega) + \hat{\lambda}(\omega)/2}{1 + 2\alpha(\omega - \omega_k)^2} \quad (3)$$

$$\omega_k^{N+1} = \int_0^\infty \omega |\hat{\mu}_k(\omega)|^2 d\omega / \int_0^\infty |\hat{\mu}_k(\omega)|^2 d\omega \quad (4)$$

where  $\hat{\cdot}$  represents the Fourier transform, and  $N$  is the number of iterations. By using Fourier transform, the specific process of VMD to decompose the signal is described as follows:

1) Initialize  $\{\hat{\mu}_k^1\}$ ,  $\{\omega_k^1\}$ ,  $\hat{\lambda}_k^1$  and  $N$  to 0, and set  $K$  to a suitable positive integer.

2) Update  $\mu_k$  and  $\omega_k$  according to Eq. (3) and Eq. (4), respectively.

3) Define  $\tau$  as the time step, then  $\lambda$  can be updated according to the following equation:

$$\hat{\lambda}^{n+1}(\omega) \leftarrow \hat{\lambda}^n(\omega) + \tau \left[ \hat{f}(\omega) - \sum_k \hat{\mu}_k^{n+1}(\omega) \right] \quad (5)$$

4) Define  $\varepsilon$  as the acceptable criterion tolerance. Set  $N = N + 1$  and repeat steps 2 and 3 until the condition  $\sum_k \|\hat{\mu}_k^{n+1} - \hat{\mu}_k^n\|_2^2 / \|\hat{\mu}_k^n\|_2^2 < \varepsilon$  is met. Stop the iteration and obtain  $K$  IMFs.

##### 3.2.2. Permutation entropy

Compared with other entropy variants, PE has higher efficiency and better robustness. Theoretically, the larger the PE value, the more complex the IMF and the more noise it contains.

Given the embedded dimension  $d$  and time delay  $\tau$ , PE firstly reconstructs the time series into a  $d$ -dimensional phase space matrix, and then map the matrix to  $d!$  symbol sequences. Calculate the probability of each sequence occurring ( $P_1, P_2, \dots, P_d$ ), and then PE is defined as:

$$H_P(d) = - \sum_{j=1}^d P_j \ln P_j \quad (6)$$

For convenience,  $H_P(d)$  is normalized to:

$$0 \leq H_P = H_P(d) / \ln(d!) \leq 1 \quad (7)$$

##### 3.2.3. Non-local means estimation

Initially, NLM algorithm was proposed for image denoising (Buades, Coll, & Morel, 2005). On this basis, 1D NLM was proposed for the denoising of ECG signals, as described below.

Suppose the noisy signal is  $z(i)$ , and the desired real signal is  $\hat{z}(i)$  and is given by:

$$\hat{z}(i) = \frac{1}{C(i)} \sum_{j \in \Omega_i} z(j) \omega(i, j), \quad C(i) = \sum_{j \in \Omega_i} \omega(i, j) \quad (8)$$

where  $C(i)$  is the normalization constant;  $\Omega_i$  is the predefined search neighborhood;  $\omega(i, j)$  denotes the weight between  $i$ th searched sample and  $j$ th desired sample in  $\Omega_i$ :

$$0 < \omega(i, j) = \exp \left\{ - \frac{\sum_{\delta \in B} [v(i + \delta) - v(j + \delta)]}{2L\theta^2} \right\} < 1, \quad \sum_j \omega(i, j) = 1 \quad (9)$$

where  $\theta$  is the filter bandwidth parameter;  $B$  is the similar region

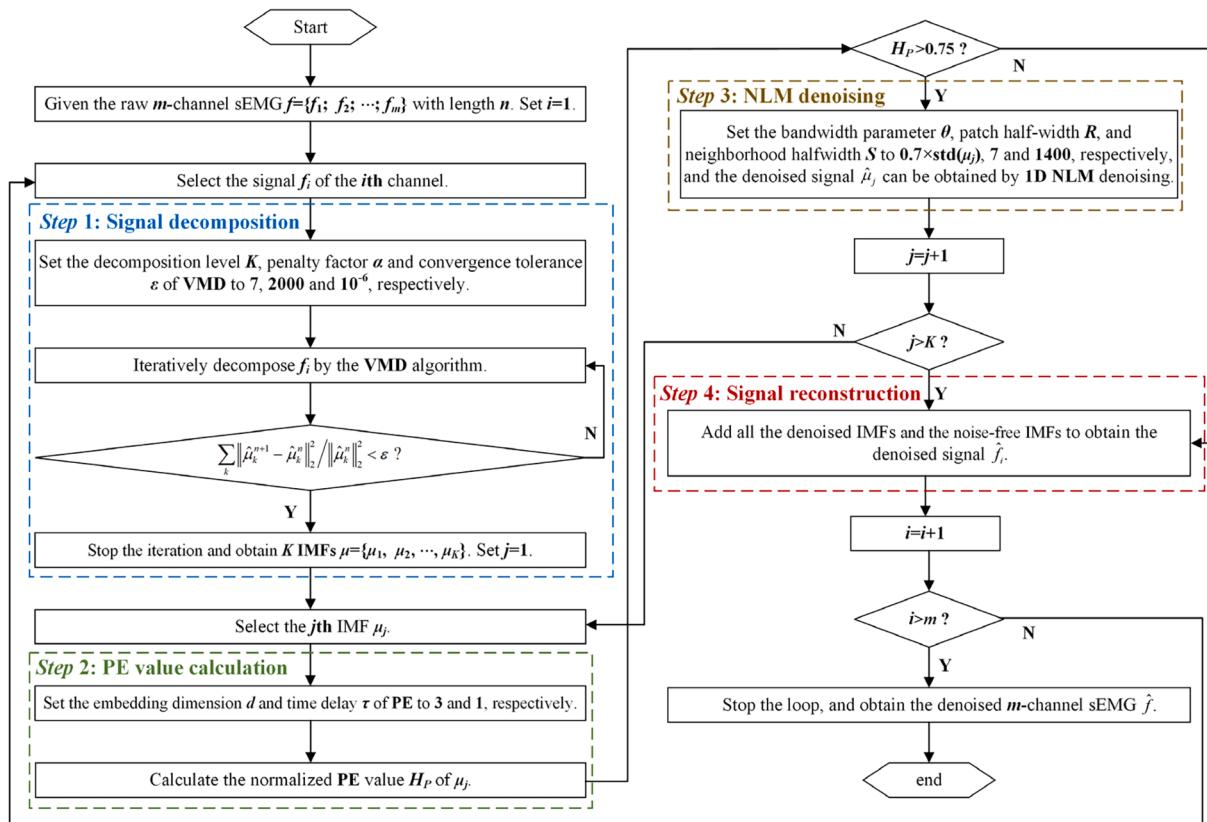


Fig. 2. Specific process of sEMG denoising based on VMD-PE-NLM technique.

**Table 1**  
Parameter settings of the VMD-PE-NLM technique.

Method	Parameter	Description	Value
VMD	K	Decomposition level	7
	alpha	Penalty factor	2000
PE	epsilon	Convergence tolerance	10^-6
	d	Embedding dimension	3
NLM	tau	Time delay	1
	theta	Filter bandwidth	0.7 * STD(x)*
	R	Patch half-width	7
	S	Neighborhood halfwidth	1400

\*Where STD(x) denotes the standard deviation of input signal x.

centered on the  $i$ th sample point containing;  $L = 2R + 1$  is the number of sample points contained in domain  $B$ .

### 3.2.4. The proposed sEMG denoising technique

As shown in Fig. 2, the proposed sEMG denoising technique mainly consists of the following steps: 1) The preprocessed sEMG data is input channel by channel to VMD to be decomposed into  $K$  IMFs; 2) Calculate the PE value of each IMF, and IMFs whose PE value is greater than the set threshold are regarded as noisy IMFs and denoised by 1D NLM; 3) The denoised IMFs and noise-free IMFs are added to obtain the denoised sEMG signal.

The related parameter settings are described in Table 1.

### 3.3. sEMG data augmentation based on the K-Means SMOTE algorithm

In K-Means SMOTE, the synthesis of noisy samples was avoided by oversampling only in safe areas, and both sample imbalance between and within classes were considered. It mainly consists of the following steps:

Step 1: The K-Means algorithm is used to cluster the input samples

into  $K_m$  clusters, where  $K_m$  is adaptively determined by the Mini Batch K-Means algorithm (Feizollah, Anuar, Salleh, & Amalina, 2014).

Step 2: Given a cluster  $c_i$ , its imbalance ratio  $irt(c_i)$  is defined as:

$$irt(c_i) = \frac{\text{majorityCount}(c_i) + 1}{\text{minorityCount}(c_i) + 1} \quad (10)$$

where  $\text{majorityCount}(c_i)$  and  $\text{minorityCount}(c_i)$  represent the sample size of majority and minority classes in cluster  $c_i$ , respectively. Subsequently, the clusters that with a high proportion of samples of minority classes ( $irt(c_i) \leq 1$ ) are selected for oversampling.

Step 3: For each filtered cluster  $f_b$ , calculate the corresponding sampling weight  $SW(f_b)$ :

$$0 \leq SW(f_b) = \text{sparsity}(f_b) / \sum_i \text{sparsity}(f_i) \leq 1 \quad (11)$$

where  $\text{sparsity}(f_b)$  denotes the sparsity factor of  $f_b$ , which is the reciprocal of density factor:

$$\text{density}(f_i) = \text{mincount}(f_i) / \text{averdistance}(f_i)^{m_f} \quad (12)$$

where  $\text{mincount}(f_i)$  and  $\text{averdistance}(f_i)^{m_f}$  represent the Euclidean distance matrix and the average distance of minority samples in  $f_i$ , respectively. Then, for  $f_i$ , the number of synthetic samples is  $||SW(f_i) \times s_n||$ , where  $s_n$  is the total number of samples that need to be synthesized.

Step 4: The SMOTE algorithm randomly selects a sample  $x_i$  of minority classes in  $f_i$  and selects a sample  $\hat{x}_i$  from its KNN, then synthesizes a new sample  $x_{new}$  by random interpolation:

$$x_{new} = x_i + \text{rand}(0, 1) \times (\hat{x}_i - x_i) \quad (13)$$

where  $\text{rand}(0, 1)$  represents the random number between 0 and 1.

**Table 2**

Details of the network structure and number of parameters of ECN.

Module	Input size	Network layer or operation	Output size	Number of parameters
Encoder	200 × 4 × 1	Conv2D ( $f = 32, k = (9, 3), s = (1, 1)$ , padding = 'same')	200 × 4 × 32	896
	200 × 4 × 32	PReLU	200 × 4 × 32	25,600
	200 × 4 × 32	Max-Pooling2D ( $k = (5, 1), s = (5, 1)$ , padding = 'valid')	40 × 4 × 32	0
	40 × 4 × 32	Conv2D ( $f = 64, k = (7, 3), s = (1, 1)$ , padding = 'same')	40 × 4 × 64	43,072
	40 × 4 × 64	PReLU	40 × 4 × 64	10,240
	40 × 4 × 64	Max-Pooling2D ( $k = (4, 1), s = (4, 1)$ , padding = 'valid')	10 × 4 × 64	0
	10 × 4 × 64	GAP	64	0
	64	Reshape	64 × 1	0
	64 × 1	Conv1D ( $f = 1, k = 3$ , padding = 'same')	64 × 1	4
	64 × 1	Sigmoid	64 × 1	0
	64 × 1	Reshape	1 × 1 × 64	0
	10 × 4 × 64	Multiply	10 × 4 × 64	0
	/1 × 1 × 64			
	10 × 4 × 64	Conv2D ( $f = 64, k = (7, 3), s = (1, 1)$ , padding = 'same')	8 × 2 × 128	73,856
	8 × 2 × 128	Reshape	128 × 16	0
	128 × 16	Squash in the Primarycaps layer	128 × 16	0
	128 × 16	DigitCaps layer with routing algorithm	3 × 16	98,304
Classifier	3 × 16	Mask	48	0
	48	Calculate the length of capsule vectors	3	0
Decoder	48	FC layer	128	6,272
	128	PReLU	128	128
	128	Dropout(0.4)	128	0
	128	FC layer	512	66,048
	512	PReLU	512	512
	512	Dropout(0.4)	512	0
	512	FC (activation = 'sigmoid')	800	410,400
	800	Reshape	200 × 4 × 1	0

### 3.4. Activity recognition based on the ECN model

#### 3.4.1. Overall structure and number of parameters of ECN

The overall structure of the ECN model is shown in Fig. 1, which is mainly composed of encoder network, decoder network and classifier. The details are described as follows:

1) Input module: Considering the processing latency, the window length is set to 200 ms (Al-Quraishi et al., 2021). Given the sampling frequency  $F_s$ , and sort the  $m$ -channel sEMG in parallel, then the input of ECN with the size of  $0.2F_s \times m \times 1$  is obtained by dimension embedding.

2) Basic feature extraction module: The Conv and Max-pooling layers are used to extract the features of input data and transform them into  $10 \times 4 \times 64$  size feature maps. The parametric ReLU (PReLU) activation function (K. He, Zhang, Ren, & Sun, 2015) is adopted to replace ReLU in original CapsNet, which can improve model performance with only a few additional parameters.

3) Dynamic weighting module: The embedded ECA module dynamically weights and reconstructs the input features at the channel level, which mainly includes global average pooling (GAP) layer, 1D Conv layer and reweighting process, and determines the coverage of local cross-channel interaction by adaptively selecting the kernel size  $k_{ECA}$ , namely:

$$k_{ECA} = \psi(ch) = \left| \frac{\log_2(ch)}{\gamma} + \frac{b}{\gamma} \right|_{odd} \quad (14)$$

where  $ch$  is the number of channels;  $\gamma$  and  $b$  are set to 2 and 1 according to the related literature (Wang, Wu, Zhu, et al., 2020), respectively.

4) Deeper feature extraction module: Primary capsule (*PrimaryCaps*) layer transforms the reweighted feature maps into 16 shallow capsule vectors of  $8 \times 2 \times 8$  size. Then, the digital capsule (*DigitCaps*) layer further transforms the output of *Primarycaps* into  $n_c$  deep capsules of length 16, where routing algorithm loops three times and  $n_c$  is the number of activity types.

5) Classification module: The class capsule layer outputs the classification result by calculating the  $L_2$  norm of each capsule vector, i.e., the length of capsule vector  $v_j$  corresponds to the probability value  $p_j$  of being judged as the  $j$ th class:  $p_j = \|v_j\|$ .

6) Decoder module: In the decoder network, the input is the output vectors of *DigitCaps* after mask operation, which passes through several fully connected (FC) layers, PReLU layers, Dropout layers to reduce the risk of overfitting and a sigmoid activation layer, and is finally reshaped into a matrix with the same size as the input. It should be emphasized that only the correctly classified capsule vector is used for training, while the others are masked with 0.

Taking Case I in Section 4 as an example, the number of parameters of ECN is analyzed, in which  $F_s$ ,  $m$  and  $n_c$  are 1000 Hz, 4 and 3, respectively. Let  $f$  denote the number of filters,  $k = (k_w, k_h)$  denote the width and height of kernel in Conv layer or Max-pooling layer, and  $s = (s_1, s_2)$  denote the stride size, then the network structure and the number of parameters of ECN are shown in Table 2. Specially, the ECA module adds only 4 training parameters due to the 1D Conv operation. In addition, compared to encoder network, the decoder network used to reconstruct input feature maps has more network parameters due to the use of FC layers.

#### 3.4.2. Loss function and learning rate

The margin loss is adopted as the classification loss, which is defined as:

$$L_c = T_c \cdot (\max(0, m^+ - \|v_c\|)^2 + \zeta \cdot (1 - T_c) \cdot \max(0, \|v_c\| - m^-)^2) \quad (15)$$

where  $T_c$  is the indicator function of class  $c$  ( $T_c = 1$  if  $c$  exists, 0 otherwise);  $m^+$  and  $m^-$  represent the upper bound to penalizing false positives and lower bound on penalizing false negatives, respectively;  $\zeta$  is the coefficient. In CapsNet, the values of  $m^+$ ,  $m^-$  and  $\zeta$  are set to 0.9, 0.1 and 0.5, respectively, and the reconstruction loss used to train the decoder is defined as the sum of squares of the difference between the corresponding pixels between the input feature map  $y$  of the encoder and the output feature map  $\hat{y}$  of the decoder, i.e.:  $L_r = \sum_i (y_i - \hat{y}_i)^2$ .

Define  $\beta$  as the scale factor of  $L_r$ , then the total loss function can be expressed as:

$$Loss = L_c + \beta \times L_r \quad (16)$$

However, in this paper, mean square error (MSE) is used to replace  $L_r$ , namely:

$$L_r = MSE = \frac{1}{n} \sum_i^n (y_i - \hat{y}_i)^2 \quad (17)$$

where  $n$  represents the number of pixels in the feature map  $y$ . In CapsNet,  $\beta$  was set to 0.0005 to ensure that margin loss dominates model training process, while in this work, it should be equal to  $0.0005 \times n$ . For example, in Case I, the size of input feature map is  $200 \times 4 \times 1$ , which means a total of 800 pixels. Therefore,  $\beta$  should be set to 0.4, i.e.,  $0.4 = 0.0005 \times 800$ . In Case II, the input size is  $400 \times 8 \times 1$ , and then  $\beta$  should be set 1.6, i.e.,  $1.6 = 0.0005 \times 3200$ . The same setting can also be seen in another literature (Mazzia, Salvetti, & Chiaberge, 2021). In addition, Adam algorithm is adopted as the optimizer, and the learning rate is set

**Table 3**

Descriptions of performance evaluation metrics for multi-classification.

Metric (%)	Equation	Calculation method
Accuracy	$(\sum_{i=1}^{n_c} TP_i) / Total$	/
Precision (PPV)	$(\sum_{i=1}^{n_c} Precision_i) / n_c$	Macro-averaged
Sensitivity (Recall)	$(\sum_{i=1}^{n_c} Sensitivity_i) / n_c$	
Specificity	$(\sum_{i=1}^{n_c} Specificity_i) / n_c$	
NPV	$(\sum_{i=1}^{n_c} NPV_i) / n_c$	
F1-Score	$(\sum_{i=1}^{n_c} F1 - score_i) / n_c$	

to a piecewise decay function:

$$\epsilon = \begin{cases} 0.001, & 0 < e \leq 0.4N_e \\ 0.001/(1 + d_e \times iter \times (e - 0.4N_e)), & 0.4N_e < e \leq N_e \end{cases} \quad (18)$$

where  $N_e$  and  $e$  are the total and current training epoch, respectively;  $d_e$  is a decay factor that is set to  $(0.001/N_e) \times 500$ ;  $iter$  is the number of iterations per epoch. The training epoch and batch size are set according to the number of samples in the datasets to ensure model convergence.

### 3.5. Performance evaluation metrics

The classification performance of ECN is evaluated by some statistical metrics such as accuracy, precision (also known as positive predictive value, i.e., PPV), sensitivity (also known as recall), specificity, negative predictive value (NPV), F1-score. In binary classification, TP, TN, FP, and FN are defined to represent the true positive, true negative, false positive and false negative indicators of classification results, respectively. For multi-classification, some metrics can be calculated in

two ways: micro (total average) and macro (category average). In this paper, the latter is chosen. Let  $n_c$  be the number of classes, subscript  $i$  represents which class it belongs to, and then the total number of samples is: Total =  $\sum_{i=1}^{n_c} (TP_i + FN_i) = \sum_{i=1}^{n_c} (TP_i + FP_i)$ . The calculation of these above metrics are shown in Table 3.

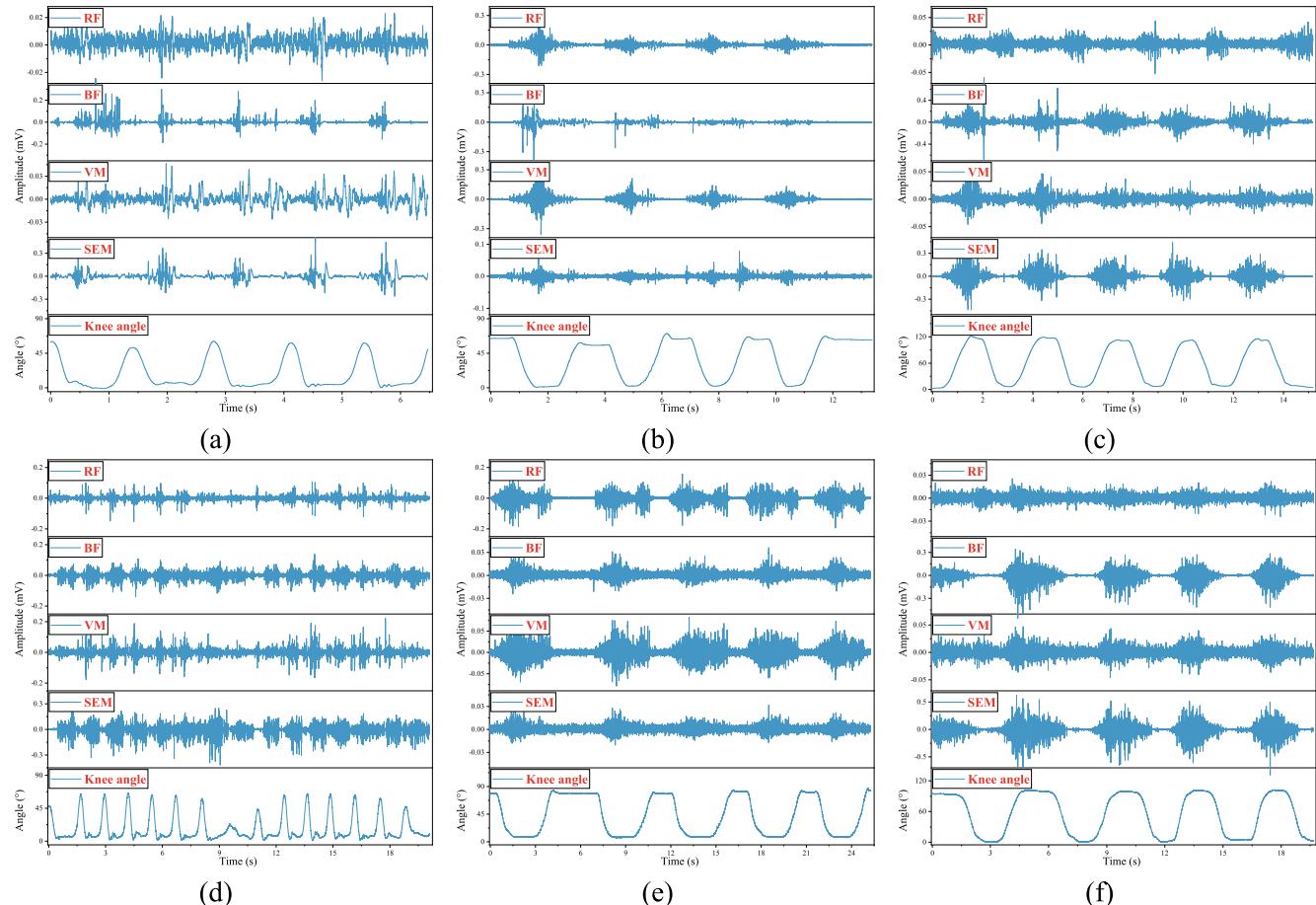
## 4. Experimental verification and discussion

In this paper, two sEMG datasets of lower limb activities were utilized to verify the effectiveness of the proposed LLAR framework for different individuals. Python 3.7.3 and Keras 2.2.4 DL library were used for algorithm design. Experimental platform was configured with AMD Ryzen 7 5800H CPU@ 3.20 GHz, GeForce RTX 3060 GPU and 16G RAM.

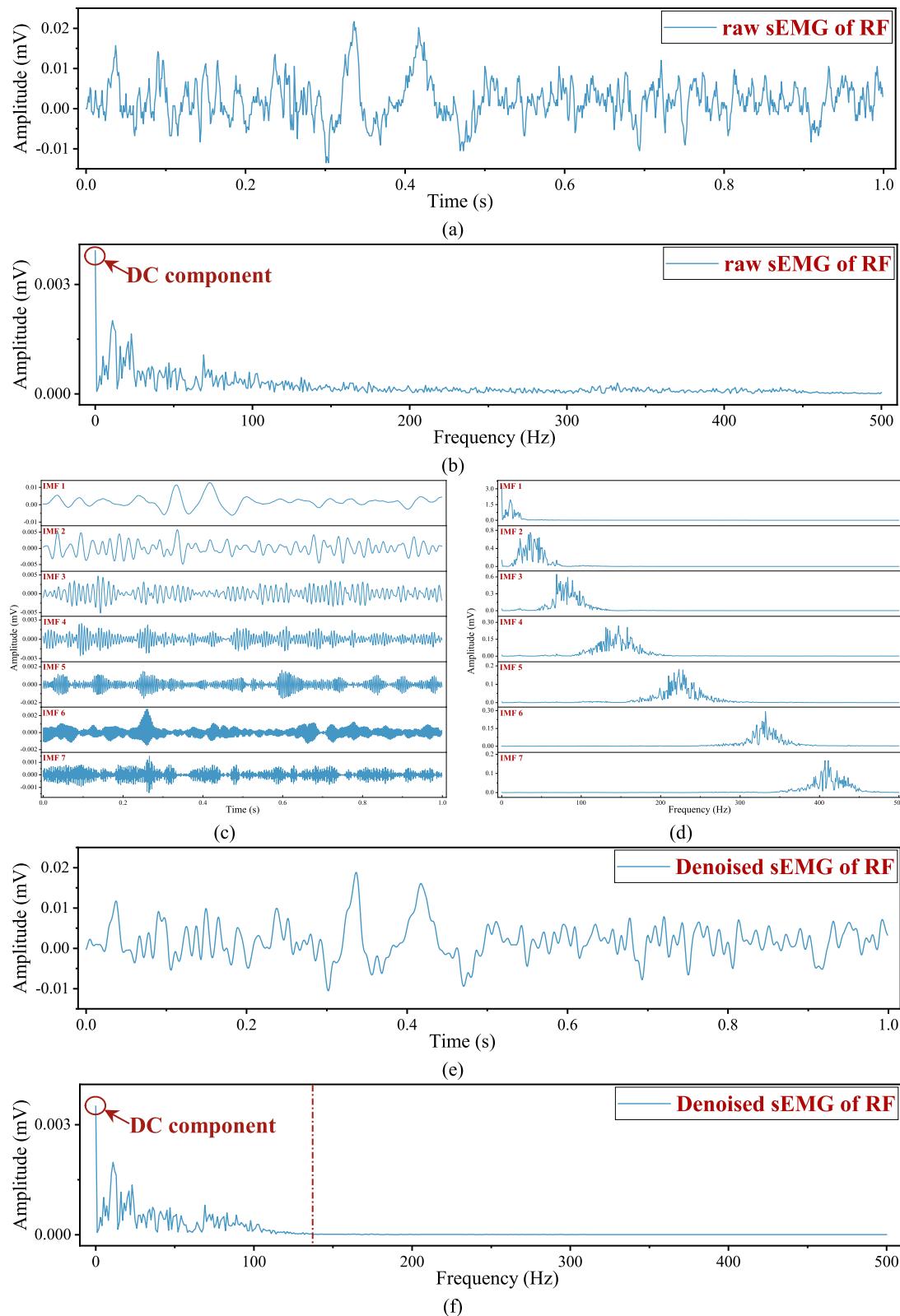
### 4.1. Case I: Experimental verification with the publicly available UCI dataset

#### 4.1.1. Experimental setup and data acquisition

In this study, a publicly available dataset provided by the University of California, Irvine (UCI) Machine Learning Repository was selected for experimental verification, which can be found on the website: <http://archive.ics.uci.edu/ml/datasets/emg> + dataset + in + lower + limb (Sanchez, Sotelo, Gonzales, & Hernandez, 2014). It consisted of 22 male subjects over the age of 18, of whom 11 were healthy (denoted as H1-H11) and the remaining had previously diagnosed different knee abnormalities (denoted as P1-P11): one with sciatic nerve injury, four suffering from meniscus injuries and six had anterior cruciate ligament injuries. The left leg of healthy subjects and affected limb of patients with knee abnormalities were used for data acquisition. All subjects



**Fig. 3.** Raw sEMG and knee angle data: (a)-(c) are the data of subject H5 under Gait, LESit and SitTStand types, respectively; (d)-(f) are the data of subject P5 under Gait, LESit and SitTStand types, respectively.



**Fig. 4.** sEMG denoising process of RF muscle of subject H5: (a), (c) and (e) are the time domain waveforms of raw sEMG, the 7 IMFs decomposed by VMD, the sEMG after denoising by NLM, respectively; (b), (d) and (f) are frequency domain waveforms corresponding to (a), (c) and (e), respectively.

**Table 4**

Normalized PE values corresponding to different IMFs.

IMFs	IMF 1	IMF 2	IMF 3	IMF 4	IMF 5	IMF 6	IMF 7
Normalized PE values	0.492	0.592	0.700	<b>0.842</b>	<b>0.944</b>	<b>0.998</b>	<b>0.964</b>

**Table 5**

Class distribution on the datasets of different types of subjects in Case I.

Subjects	Original dataset	Denoised dataset	Activity type	Class	Number of samples	Balanced dataset
H1-H11	A1	D1	Gait	0	336	BD1
			LESit	1	582	
			SitStand	2	678	
P1-P11	A2	D2	Gait	0	1504	BD2
			LESit	1	908	
			SitStand	2	865	

were required to perform three different movements: walking on level ground (Gait), leg extension from a sitting position (LESit), and flexion of the knee stand up (SitStand) to analyze the behavior of knee muscles. Four channels of sEMG sensors were placed on the surfaces of rectus femoris (RF), biceps femoris (BF), vastus medialis (VM), and semitendinosus (SEM), where the distance between the bipolar electrodes was 20 mm and the input impedance of amplifier was  $>100\text{M}\Omega$ , meaning that sampling could be performed without the need of conducting gels. The SG150 goniometer was kept to the external side of knee joint, with a resolution of  $+0.1^\circ$  in a range of  $180^\circ$ . The data acquisition device was Datalog MWX8 from Biometrics Ltd., which had eight digital channels and four analog channels, one for goniometer and four for sEMG

sampling. Data were acquired at a sampling frequency of 1000 Hz with 14-bit resolution, then passed through a bandpass filter with cutoff frequencies of  $20 \sim 460$  Hz, connected to a MWX8 internal storage computer with a microSD card, and transmitted in real-time via a Bluetooth adapter. In this study, the angle data were ignored.

#### 4.1.2. Signal analysis and data preprocessing

Fig. 3 shows the raw sEMG and knee angle data of subjects H5 and P5 under these three activity types. The data preprocessing mainly includes steps such as active segment extraction, signal denoising, normalization and sample synthesis, as described below.

Firstly, this study designs an adaptive double-threshold method (one of the commonly used methods in voice endpoint detection) to extract the active segments of raw sEMG signals, so as to reduce the impact of data redundancy on recognition accuracy and computational efficiency. The two thresholds are respectively set as the average short-term energy and variance of the 4-channel sEMG signals, and the threshold coefficients are set according to the signal characteristics. Since it is not the focus of this study, it will not be repeated here.

Then, VMD-PE-NLM is used for sEMG denoising. Specific parameter settings are described in Section 3.2. Taking the RF muscle of subject H5 as an example, Fig. 4(a) and (b) are the time domain and frequency domain waveforms of raw sEMG of active segment, respectively. Fig. 4(c) and (d) are the time domain and frequency domain waveforms of the

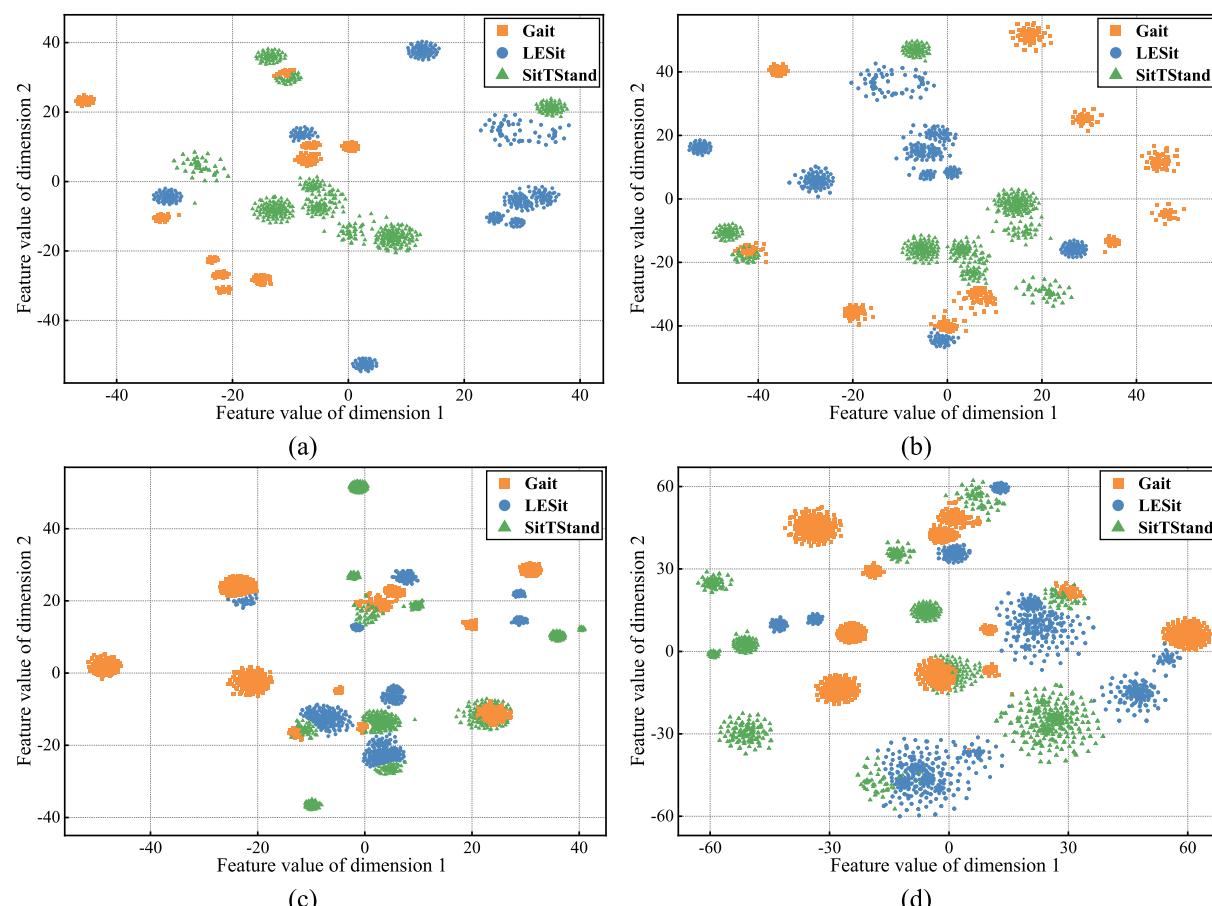


Fig. 5. Using t-SNE to visualize the sample distribution of dataset: (a) D1; (b) D2; (c) BD1; (d) BD2.

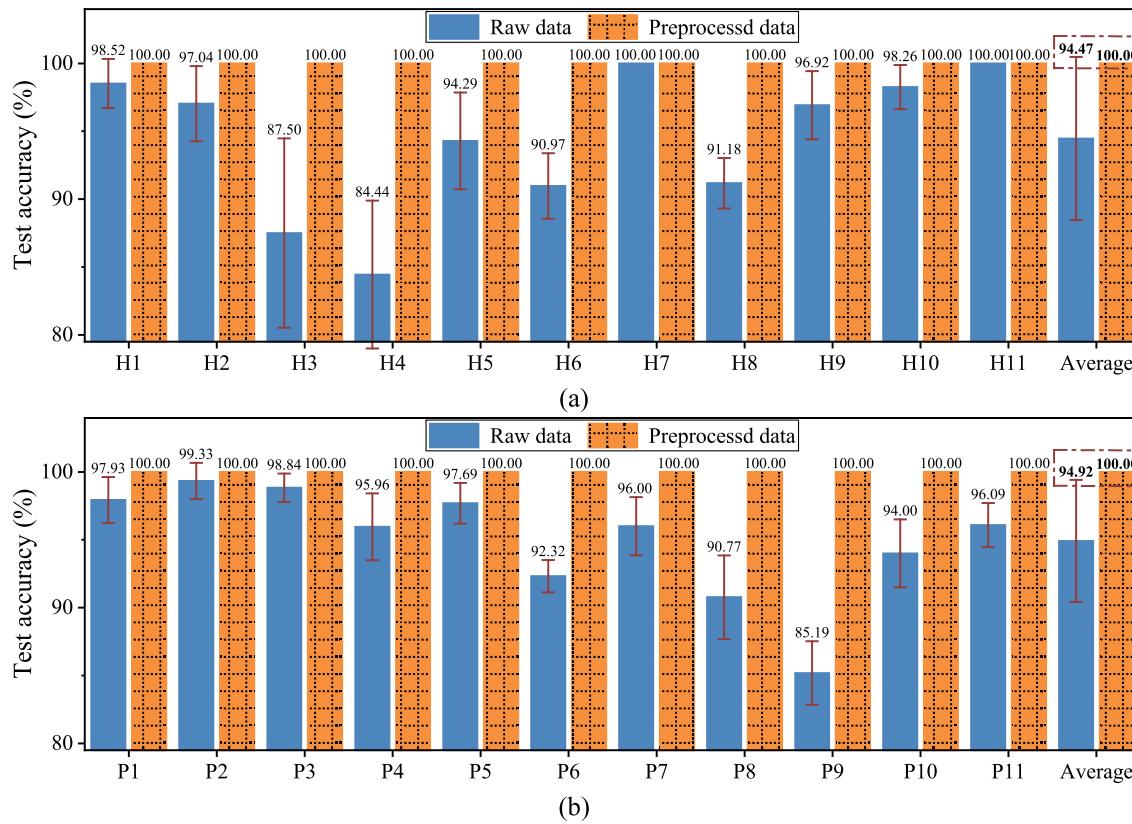


Fig. 6. Recognition accuracy using data from: (a) healthy subjects; (b) patients with knee abnormalities.

7 IMFs decomposed by VMD, respectively. It can be seen that different IMFs separately extract information of sEMG in different frequency bands. Table 4 shows the normalized PE values of different IMFs. According to the characteristics of sEMG,  $H_p = 0.75$  is set as the boundary between noisy and noiseless IMFs, i.e., when  $H_p > 0.75$ , it is considered that the associated IMF (IMFs 4 ~ 7) contains more noise. It can be seen that the noise components in sEMG are mostly concentrated in the frequency band above 150 Hz. Fig. 4(e) and (f) show the time-domain and frequency-domain waveforms of the sEMG after denoising by NLM, respectively. By comparing Fig. 4(a) and (e), it can be seen that VMD-PE-NLM has strong signal fidelity ability, which can effectively eliminate the noise interference while minimizing the loss of effective signal. Generally, the effective frequency band of sEMG is 0 ~ 500 Hz, while the main energy is concentrated in the band of 20 ~ 150 Hz. By comparing Fig. 4(b) and (f), it can be seen that VMD-PE-NLM mainly removes the noise of raw sEMG in the frequency band above 140 Hz, while preserving useful information in the low frequency band.

Next, the *MinMaxScaler* is utilized to normalize the denoised data to [0, 1], which is beneficial to speed up model convergence. On this basis, the normalized sEMG are divided into sample sets of  $0.2F_s \times m$  size by non-overlapping windowing, where  $F_s$  and  $m$  are 1000 and 4, respectively. Table 5 describes the class distribution on the datasets of different types of subjects. It can be seen that there is the class-imbalanced problem in both datasets: for D1, classes 0 and 1 are the minority classes, while for D2, classes 1 and 2 are the minority classes.

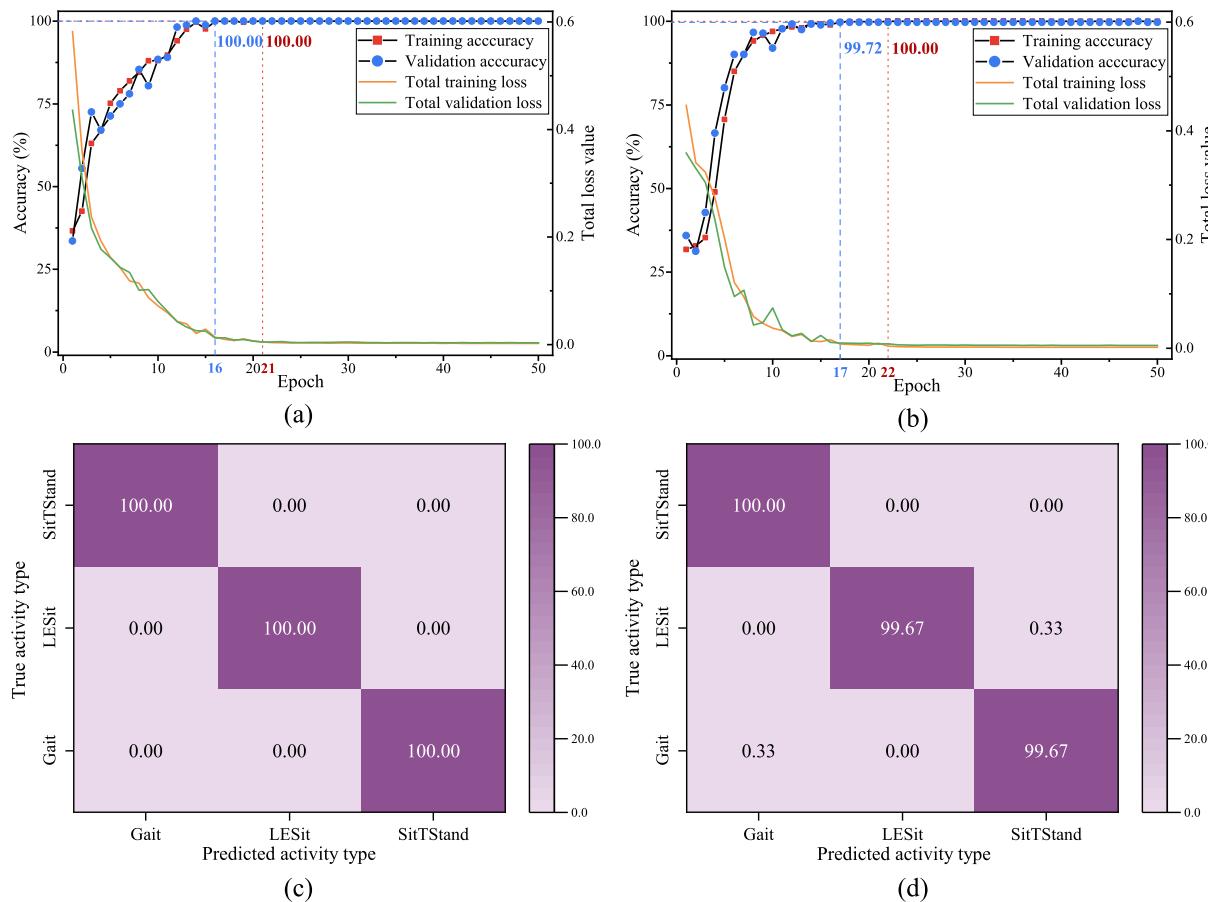
Next, K-Means SMOTE is used to synthesize new samples for minority classes in datasets D1 and D2, and the balanced datasets are denoted as BD1 and BD2, respectively. To more intuitively observe the difference between synthesized and original samples in feature space, t-SNE algorithm (Van der Maaten & Hinton, 2008) is used to visualize the sample distribution in 2D space, as shown in Fig. 5. By comparing Fig. 5 (a) to (d), it can be seen that K-Means SMOTE can synthesize high-quality samples that conform to the feature distribution of original

data, which has the ability to expand imbalanced data. In addition, as a feature space-based oversampling technique, K-Means SMOTE may produce some data that may exist in the case of real-time collection, which is conducive to improving the model generalization ability.

#### 4.1.3. Recognition results

Firstly, the data of each subject are used individually to verify the applicability of the proposed LLAR framework to healthy people and patients with different types of knee abnormalities. Samples of each subject are randomly divided into the training set and test set in a ratio of 4:1, and 10% of the training set is randomly divided as the validation set to prevent model overfitting. In this study, the training epoch and batch size are set to 50 and 32, respectively. Five replicated experiments are conducted for each subject to reduce the influence of accidental factors, and ensure that each experiment will yield different data partitioning results by generating different random number seeds. The average test accuracy and standard deviation (STD) are shown in Fig. 6. It can be seen that: 1) For healthy subjects, the average accuracy using raw data is the lowest of 84.44% for H4 and the highest of 100.00% for H7, while for patients with knee abnormalities, it fluctuates in the range of 85.19%~99.33%; 2) For the two types of subjects, using raw data directly can also achieve a relatively satisfactory accuracy, which is 94.47% and 94.92%, respectively; 3) For each subject, the accuracy reaches 100.00% using the preprocessed data. Experimental results show that the proposed framework can be applied to patients with different types of knee abnormalities, and the proposed preprocessing techniques can greatly improve the classification accuracy.

Then, recognition performance on datasets BD1 and BD2 that mixed with the same type of subjects is investigated separately, and the training process of ECN and classification results are shown in Fig. 7. It can be seen from Fig. 7(a) and (b) that: 1) In terms of accuracy, the training accuracy of BD1 and BD2 reaches 100% when epoch is 21 and 22, respectively, while validation accuracy is stable to 100% and



**Fig. 7.** Experimental results: (a)-(b) model training process on BD1 and BD2, respectively; (c)-(d) normalized confusion matrices of the test results on BD1 and BD2, respectively.

99.72%, respectively; 2) In terms of loss values, model converges quickly and stably on both BD1 and BD2. It can be seen from Fig. 7(c) and (d) that for patients with knee abnormalities, only 0.33% test samples of the LESit and SitTStand types are misclassified, while the test accuracy of healthy subjects reaches 100%.

#### 4.1.4. Comparison with the state-of-the-art methods

Firstly, to illustrate the advantages of K-Means SMOTE, some other data synthesis methods are used for comparison, including random over sampling (ROS) (Menardi & Torelli, 2014), SMOTE, adaptive synthetic (ADASYN) (He, Bai, Garcia, & Li, 2008), and SMOTE-Tomek (Batista, Prati, & Monard, 2004). Related parameter settings are as follows: 1) The sampling strategy is to resample all classes except the majority class; 2) The number of nearest neighbors is set to 2 for K-Means SMOTE and the rest to 5 according to the related literatures. Experiments results on are shown in Fig. 8, in which “None” indicates without the step of data synthesis. It can be seen from Fig. 8 that: 1) For both two types of subjects, better performance can be achieved using data synthesis methods than using class-imbalanced data directly; 2) For D1, there is little difference in using five synthesis methods, while for D2, all the metrics after using K-Means SMOTE are the highest. Experimental results indicate that K-Means SMOTE has the ability to solve the class-imbalanced problem, greatly improving the performance of ECN.

Then, to illustrate the advantages of ECN, several DL and ML models are used for comparison, as detailed in Table 6. For ML classifiers, the extracted sEMG features include: 1) Time domain features such as variance, root mean square, integrated EMG, mean absolute value, waveform length and zero-crossing rate; 2) Frequency domain features such as frequency ratio, mean power, total power, mean frequency, median frequency and peak frequency. Performance of different

classifiers using raw data and preprocessed data are compared, and the mean and STD of five replicated experiments are shown in Table 7 and Table 8.

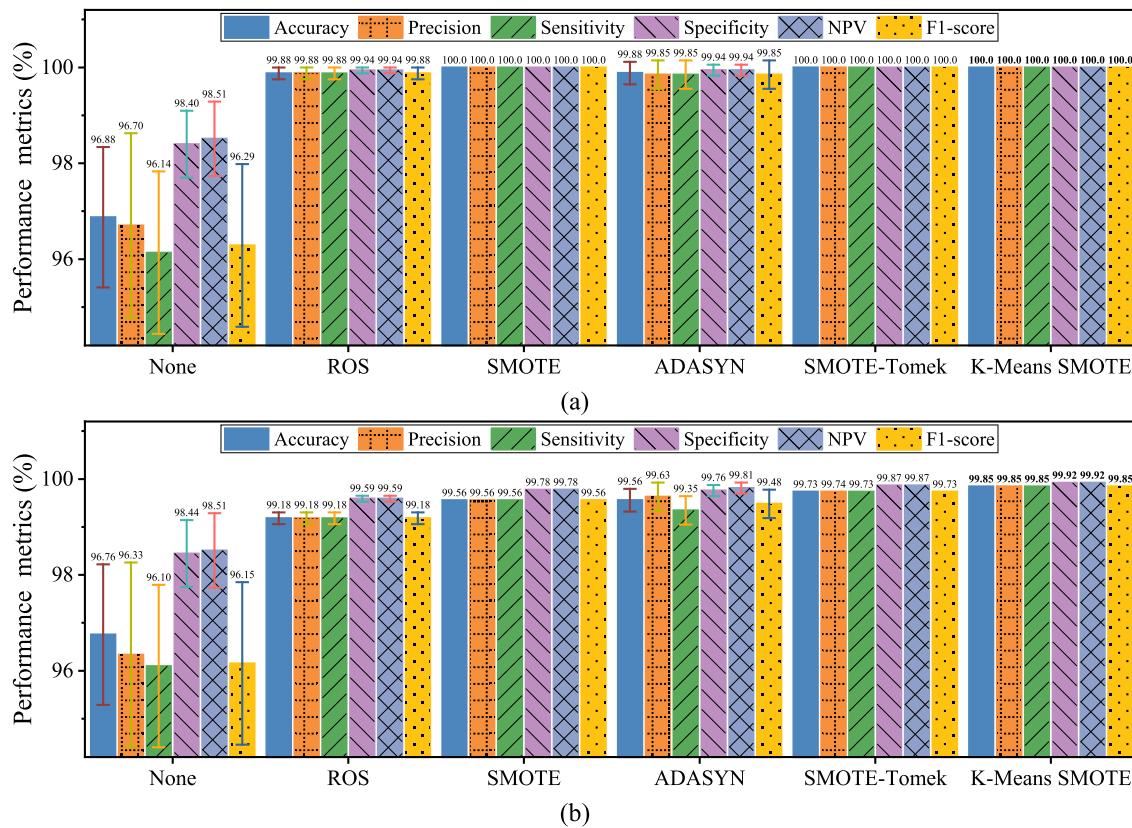
It can be seen from Table 7 and Table 8 that: 1) On the original datasets A1 and A2, the performance of ECN far exceeds that of other classifiers, reaching the average accuracy of 93.94% and 92.35% respectively; 2) When using the preprocessed dataset BD1 and BD2, the performance of CapsNet is closest to that of ECN, which proves its advantages over other DL models; 3) In the ML models, RF has the highest accuracy, mainly due to the extracted high-quality feature set and optimized hyperparameters.

Finally, the proposed framework is compared with some state-of-the-art methods that use the UCI dataset, as shown in Table 9. Experimental results show that compared with existing literature, the proposed framework exhibits more satisfactory performance in terms of recognition accuracy.

#### 4.2. Case II: Experimental verification with the dataset of patients with stroke

##### 4.2.1. Experimental setup and data acquisition

In this study, the lower limb sEMG data of 7 patients with stroke (denoted as SP1-SP7) from the Department of Rehabilitation Medicine, Tongji Hospital, Tongji Medical College, Huazhong University of Science and Technology were collected. For comparison, sEMG data of 7 healthy male subjects (denoted as HS1-HS7) over the age of 18 (height  $178 \pm 5$  cm, weight  $72.57 \pm 7.36$  kg) were also collected. Inclusion criteria for patients with stroke were: 1) male or non-pregnant female aged 18 to 80 years; 2) first-ever subacute stroke (2 weeks to 3 months) and unilateral hemiplegia; 3) able to walk with or without manual



**Fig. 8.** Classification performance using different data synthesis methods on dataset: (a) D1; (b) D2.

**Table 6**  
Parameter settings of classification models used for comparison.

Type	Classifier	Parameter settings
DL	1D-CNN; 2D-CNN; CapsNet; ResNetV1 (He, Zhang, Ren, & Sun, 2016).	1) Training epoch $N_e$ : 50; 2) Batch size: 32; 3) Optimizer: Adam; 4) Learning rate $\epsilon$ : 0.001.
ML	KNN; LDA, SVM; RF.	1) Feature extraction: for each channel, 12 features are extracted; 2) Parameters: optimized by Bayesian optimization (Snoek, Larochelle, & Adams, 2012).

assistance; 4) ability to understand and follow instructions and give informed consent. Exclusion criteria were: 1) excessive spasticity of lower limbs ( $\geq 3$  on modified Ashworth scale); 2) wounds or pressure ulcers on the lower limbs; 3) recent fractures or osteoporosis of lower limbs; 4) pre-existing neurological diseases, such as cerebral palsy, spinal cord injury; 5) serious comorbidities: severe impairment of cardiopulmonary function, malignant tumor, infections, etc. Details of the patients with stroke are shown in Table 10. Approval of all ethical and experimental procedures and protocols was granted by the medical ethics committee of Tongji Medical College of Huazhong University of Science and Technology under Application No. [2020] S296-1.

The trial consists of three tasks suitable for rehabilitation training for patients with early subacute stroke: the Gait and SitTSit tasks described in Case I, and stand-to-sit (StandTSit). Task 1 was to walk for one minute at a comfortable pace under the protection of the staff. Tasks 2 and 3 required subjects to sit still in a height-adjustable chair and then perform 10 repetitions of the corresponding movement without hand or staff assistance, with a 5 ~ 10 s rest between tasks to avoid fatigue interference. Data acquisition system was the wireless Ultimu EMG system (Noraxon USA Inc., Scottsdale, AZ, USA), which can record up to

32 channels simultaneously at a rate of  $\sim 2$  ksp/s, as detailed below.

For each lower limb of each subject, Noraxon Dual EMG electrodes (Noraxon USA Inc., Scottsdale, AZ, USA) were placed on the skin surface of the following eight muscles: BF, gluteus medius (GM), medial gastrocnemius (MG), RF, SEM, soleus (SOL), tibialis anterior (TA) and VM. The electrodes were placed parallel to the direction of the muscle fibers and the skin was cleansed using alcohol wipes to ensure low skin impedance. The bipolar electrodes were attached to wireless Ultium EMG sensors (Noraxon USA Inc., Scottsdale, AZ, USA), where the distance between electrodes was 20 mm, with extremely low baseline noise ( $<1 \mu\text{V RMS}$ ) and extremely high input impedance ( $>100\text{M}\Omega$ ). Each sensor module was integrated with a 9-axis inertial measurement unit. Data were sampled at a frequency of 2000 Hz with 24-bit resolution, and then sent directly from the point of origin to the Ultium EMG receiver connected to a portable laptop running myoMUSCLE™ software (MR3 myoMUSCLE™, Noraxon USA Inc., Scottsdale, AZ, USA) via direct-function wireless technology. Fig. 9 shows the hardware of data acquisition system and the placement of sensors. All operations followed the SENIAM guidelines (Hermens et al., 1999). Besides, the sEMG data of right legs of healthy subjects and affected limbs of patients with stroke were selected for experimental verification.

#### 4.2.2. Signal analysis and data preprocessing

Fig. 10 shows the raw sEMG data of 8 muscles of subject SP1 under three activity types. Firstly, the data during the rest period is eliminated, and the remaining preprocessing process is as described in Case I. It should be noted that since the  $F_s$  and  $m$  in this case are 2000 Hz and 8, respectively, the sample size is  $400 \times 8$ .

Taking the TA muscle of subject SP1 as an example, Fig. 11(a) to (d) show the sEMG denoising process based on VMD-PE-NLM. It can be seen that IMFs 5 to 7 are the modes containing noise, and the proposed technique can well remove noise interference in the raw sEMG, thereby improving the signal-to-noise ratio (SNR) of useful signal.

**Table 7**

Performance of different classification models on dataset of healthy subjects.

Dataset	Model	Accuracy ± STD (%)	Precision ± STD (%)	Sensitivity ± STD (%)	Specificity ± STD (%)	NPV ± STD (%)	F1-score ± STD (%)
A1	1D-CNN	90.00 ± 2.02	88.62 ± 1.96	88.73 ± 2.27	95.20 ± 0.95	95.12 ± 1.00	88.26 ± 2.22
	2D-CNN	90.88 ± 1.50	91.60 ± 0.79	86.84 ± 2.97	95.05 ± 1.00	96.13 ± 0.42	88.02 ± 2.35
	ResNetV1	91.04 ± 2.13	90.85 ± 1.10	88.24 ± 2.92	95.24 ± 1.32	95.89 ± 0.86	89.15 ± 2.39
	CapsNet	93.44 ± 1.01	92.66 ± 1.17	91.90 ± 1.64	96.70 ± 0.53	96.86 ± 0.50	92.11 ± 1.32
	KNN	89.24 ± 1.55	90.13 ± 1.25	84.43 ± 1.66	94.13 ± 0.83	95.47 ± 0.60	85.72 ± 1.77
	LDA	90.03 ± 1.60	89.12 ± 2.08	87.37 ± 2.01	94.87 ± 0.79	95.26 ± 0.79	87.95 ± 2.00
	SVM	90.19 ± 1.58	89.49 ± 1.86	86.81 ± 2.12	94.82 ± 0.86	95.51 ± 0.76	87.65 ± 2.01
	RF	92.74 ± 1.31	92.37 ± 1.42	90.11 ± 1.92	96.18 ± 0.70	96.69 ± 0.58	90.94 ± 1.74
	ECN	<b>93.94 ± 0.42</b>	<b>93.10 ± 0.93</b>	<b>92.47 ± 0.38</b>	<b>96.97 ± 0.17</b>	<b>97.09 ± 0.31</b>	<b>92.69 ± 0.38</b>
BD1	1D-CNN	99.41 ± 0.37	99.42 ± 0.36	99.41 ± 0.37	99.71 ± 0.18	99.71 ± 0.18	99.41 ± 0.37
	2D-CNN	99.95 ± 0.10	99.95 ± 0.10	99.95 ± 0.10	99.98 ± 0.05	99.98 ± 0.05	99.95 ± 0.10
	ResNetV1	99.76 ± 0.38	99.76 ± 0.37	99.76 ± 0.38	99.88 ± 0.19	99.88 ± 0.19	99.76 ± 0.38
	CapsNet	99.95 ± 0.10	99.95 ± 0.10	99.95 ± 0.10	99.98 ± 0.05	99.98 ± 0.05	99.95 ± 0.10
	KNN	99.88 ± 0.14	99.88 ± 0.14	99.88 ± 0.14	99.94 ± 0.07	99.94 ± 0.07	99.88 ± 0.14
	LDA	89.73 ± 1.47	90.38 ± 1.36	89.78 ± 1.47	94.87 ± 0.77	94.97 ± 0.76	89.73 ± 1.42
	SVM	99.33 ± 0.33	99.34 ± 0.32	99.33 ± 0.33	99.67 ± 0.17	99.67 ± 0.16	99.33 ± 0.33
	RF	99.90 ± 0.14	99.90 ± 0.14	99.90 ± 0.14	99.95 ± 0.07	99.95 ± 0.07	99.90 ± 0.14
	ECN	<b>100.00 ± 0.00</b>	<b>100.00 ± 0.00</b>	<b>100.00 ± 0.00</b>	<b>100.00 ± 0.00</b>	<b>100.00 ± 0.00</b>	<b>100.00 ± 0.00</b>

**Table 8**

Performance of different classification models on dataset of patients with knee abnormalities.

Dataset	Model	Accuracy ± STD (%)	Precision ± STD (%)	Sensitivity ± STD (%)	Specificity ± STD (%)	NPV ± STD (%)	F1-score ± STD (%)
A2	1D-CNN	87.04 ± 0.98	86.85 ± 1.17	86.61 ± 1.11	93.41 ± 0.53	93.40 ± 0.50	86.44 ± 1.07
	2D-CNN	90.21 ± 1.18	91.26 ± 1.26	88.87 ± 1.71	94.59 ± 0.75	95.35 ± 0.54	89.60 ± 1.44
	ResNetV1	89.42 ± 1.31	89.28 ± 1.62	88.80 ± 1.75	94.46 ± 0.75	94.65 ± 0.73	88.89 ± 1.43
	CapsNet	90.61 ± 0.76	90.28 ± 0.82	90.42 ± 1.07	95.24 ± 0.49	95.12 ± 0.35	90.24 ± 0.85
	KNN	86.91 ± 1.23	87.05 ± 1.31	85.81 ± 1.46	92.98 ± 0.67	93.38 ± 0.61	86.27 ± 1.37
	LDA	82.63 ± 1.22	82.13 ± 1.46	81.47 ± 1.43	90.91 ± 0.65	91.13 ± 0.65	81.66 ± 1.40
	SVM	87.66 ± 1.20	87.77 ± 1.26	86.68 ± 1.28	93.38 ± 0.66	93.73 ± 0.64	87.13 ± 1.25
	RF	91.17 ± 0.91	91.42 ± 0.97	90.00 ± 0.95	95.22 ± 0.47	95.70 ± 0.47	90.57 ± 0.94
	ECN	<b>92.35 ± 0.95</b>	<b>92.09 ± 1.10</b>	<b>91.79 ± 0.80</b>	<b>96.02 ± 0.46</b>	<b>96.11 ± 0.56</b>	<b>91.92 ± 0.92</b>
BD2	1D-CNN	98.73 ± 1.16	98.77 ± 1.12	98.73 ± 1.16	99.36 ± 0.58	99.38 ± 0.56	98.72 ± 1.17
	2D-CNN	99.29 ± 1.36	99.35 ± 1.24	99.29 ± 1.36	99.65 ± 0.68	99.66 ± 0.66	99.30 ± 1.35
	ResNetV1	99.49 ± 0.33	99.50 ± 0.32	99.49 ± 0.33	99.75 ± 0.16	99.75 ± 0.16	99.49 ± 0.33
	CapsNet	99.65 ± 0.29	99.65 ± 0.28	99.65 ± 0.29	99.82 ± 0.15	99.82 ± 0.14	99.65 ± 0.29
	KNN	98.92 ± 0.29	98.93 ± 0.29	98.92 ± 0.29	99.46 ± 0.14	99.46 ± 0.15	98.92 ± 0.29
	LDA	79.29 ± 1.48	79.53 ± 1.48	79.31 ± 1.49	89.65 ± 0.73	89.63 ± 0.73	79.34 ± 1.49
	SVM	98.58 ± 0.35	98.62 ± 0.33	98.59 ± 0.34	99.29 ± 0.18	99.29 ± 0.18	98.58 ± 0.34
	RF	99.64 ± 0.37	99.64 ± 0.36	99.64 ± 0.37	99.82 ± 0.18	99.82 ± 0.18	99.64 ± 0.37
	ECN	<b>99.85 ± 0.13</b>	<b>99.85 ± 0.13</b>	<b>99.85 ± 0.13</b>	<b>99.92 ± 0.07</b>	<b>99.92 ± 0.07</b>	<b>99.85 ± 0.13</b>

**Table 9**

Comparison with the state-of-the-art methods.

Literature	Processing of input data	Classifier	Average test accuracy (%)	
			Healthy subjects	Patients with knee abnormalities
(Naik et al., 2018)	Time domain features	LDA	96.10%	86.20%
(Vijayvargiya, Gupta, et al., 2021)	WD-EEMD denoising	LDA	90.69%	97.45%
(Vijayvargiya, Kumar, et al., 2021)	Raw sEMG data	1D-CNN	99.35%	97.63%
(Gautam et al., 2020)	Raw sEMG data	MyoNet	98.10%	92.40%
(Issa & Khaled, 2022)	STFT	2D-CNN	92.00%	95.00%
Proposed	Proposed techniques	ECN	<b>100.00%</b>	<b>99.85%</b>

The division and class distribution of different datasets are shown in Table 11. Similarly, the sample distribution of datasets D4 and BD4 is visualized using t-SNE, respectively, as shown in Fig. 12.

#### 4.2.3. Recognition results

First, data of each subject are used individually for experimental verification. Division of training set, validation set and test set is described in Case I, where the epoch and batch size are set to 15 and 32, respectively. Results of five replicated experiments are shown in Fig. 13.

As can be seen from Fig. 13, except that the average accuracy and F1-score of subject SP3 are 99.83%, the average values of the above two metrics of the other subjects all reach 100.00%. Then, the recognition performance on datasets BD3 and BD4 that mixed with the same type of subjects is investigated, and the results are shown in Fig. 14. As can be seen from Fig. 14(b), for patients with stroke, only 0.32% test samples of the SitTStand are misclassified as StandTSit. In Fig. 14(c), the class distribution of different activity types has obvious boundaries. It can be seen from Fig. 14(d) that one sample in the feature interval of SitTSand is incorrectly labeled as StandTSit, which also confirms the classification result in Fig. 14(b). Experimental results show that the proposed method also performs well for patients with stroke.

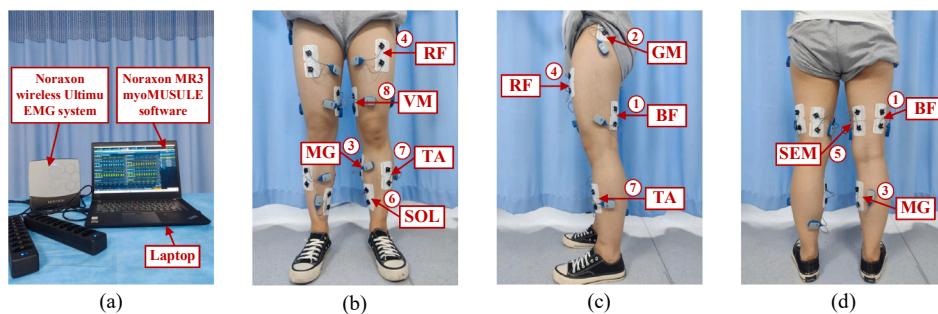
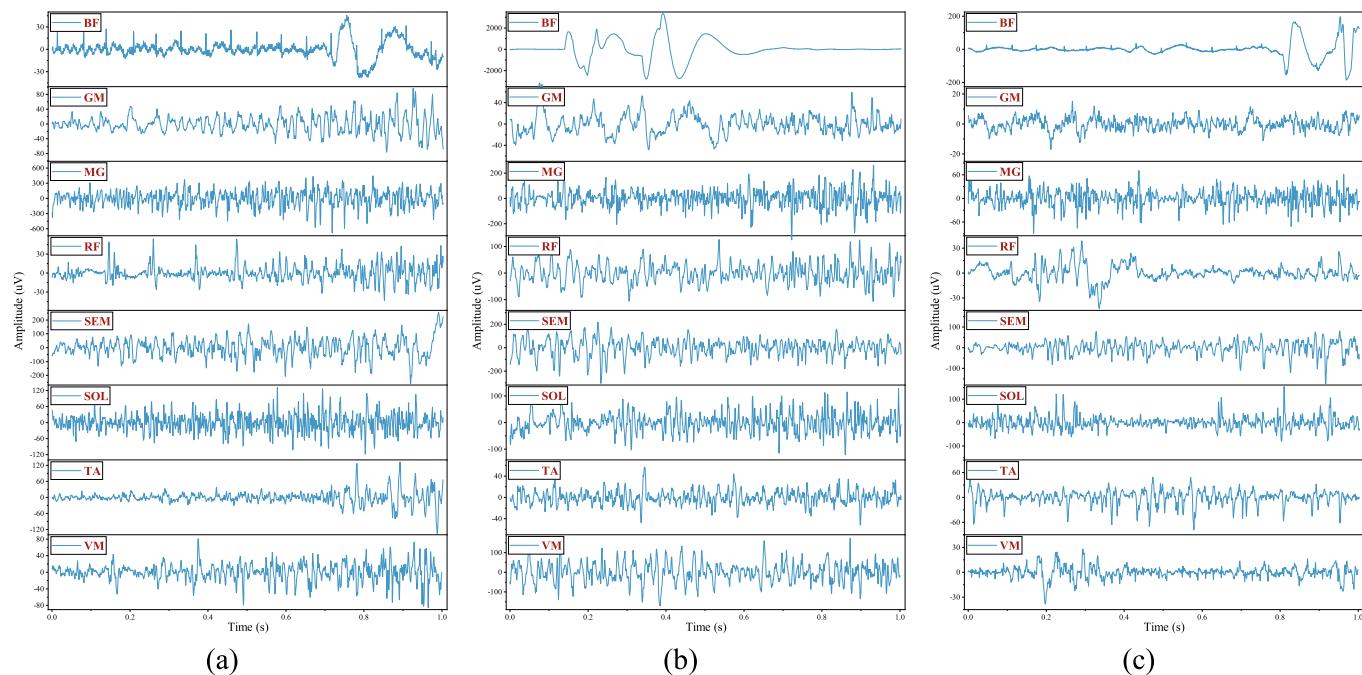
#### 4.2.4. Comparison with the state-of-the-art methods

To further prove the advantages of the proposed framework, ECN is compared with several state-of-the-art DL models: 2D-CNN, ResNetV1, InceptionV4 (Szegedy et al., 2017) and MobileNetV3 (Howard et al., 2019). The mean and STD of five replicated experiments are shown in

**Table 10**

Details of patients with stroke.

Serial number	Gender	Age	Height (cm)	Weight (kg)	Stroke type	Paretic Side	Time from stroke onset
SP1	Male	45	175	70	Hemorrhagic stroke	Right	71 days
SP2	Female	54	160	60	Ischemia stroke	Right	40 days
SP3	Male	53	176	73	Hemorrhagic stroke	Right	50 days
SP4	Male	49	173	66	Ischemia stroke	Left	44 days
SP5	Male	47	170	64	Hemorrhagic stroke	Right	70 days
SP6	Male	45	168	62	Ischemia stroke	Left	23 days
SP7	Female	33	160	40	Ischemia stroke	Right	90 days

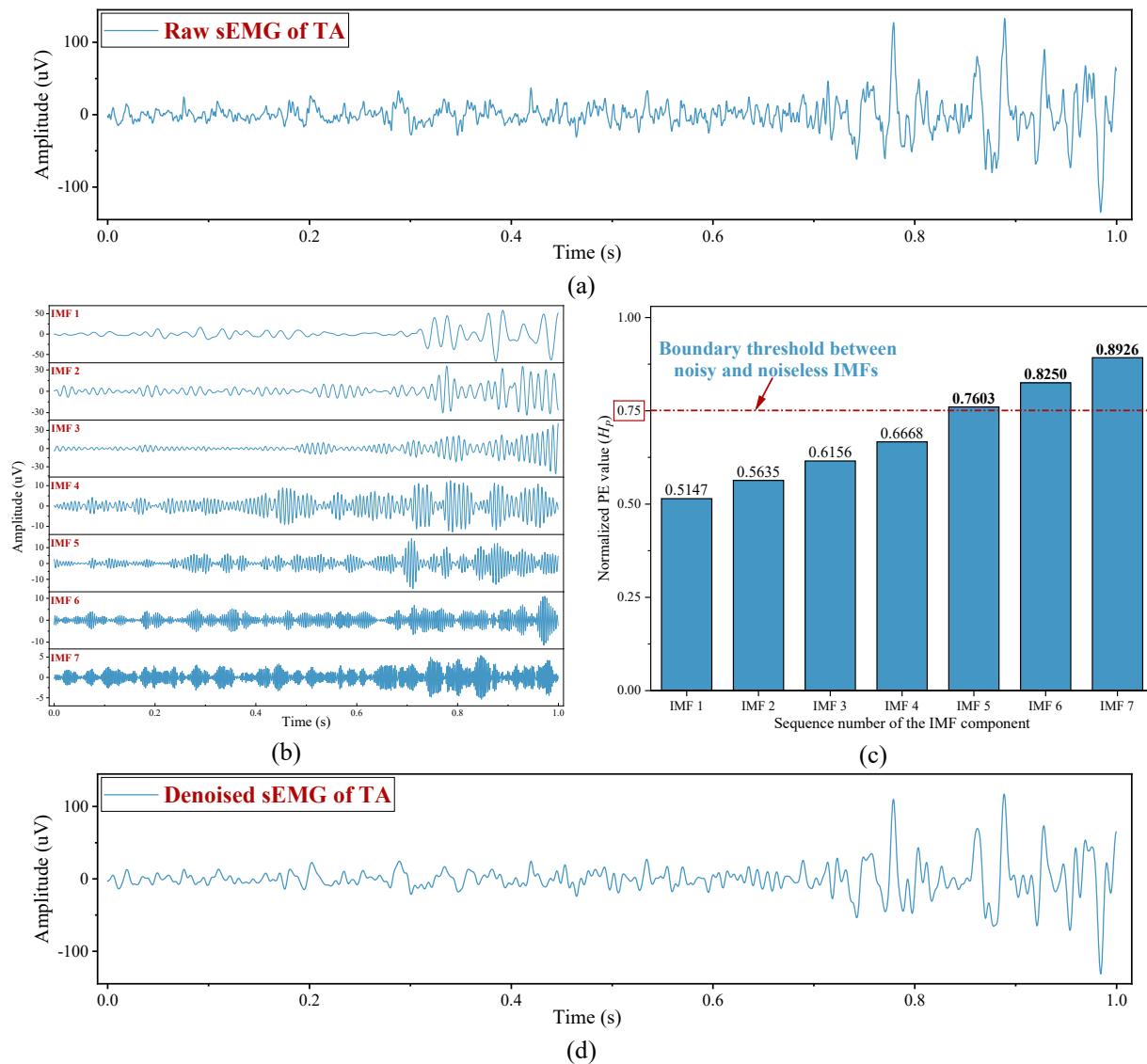
**Fig. 9.** Data acquisition system, position of selected muscles and placement of electrodes: (a) data acquisition system; (b) front side; (c) side face; (d) back side.**Fig. 10.** Raw sEMG of subject SP1 under the activity types of: (a) Gait; (b): SitTStand; (c): StandTSit.

**Fig. 15.** As can be seen: 1) For the two types of subjects, when directly using the original datasets, the performance of InceptionV4 and MobileNetV3 is very close, while ECN outperforms other models, with average accuracy of 87.19% and 84.10% respectively; 2) Performance of all DL models is improved to varying degrees on the preprocessed datasets; 3) When combining the proposed preprocessing techniques and ECN, the accuracy of the two types of subjects reaches the highest, 99.95% and 99.89%, respectively. Experimental results further verify the applicability of the proposed LLAR framework for patients with stroke.

#### 4.3. Discussion

In this work, the proposed LLAR framework based on sEMG DA and ECN model exhibits excellent performance in the LLAR tasks of multiple types of individuals, including healthy subjects, patients with knee abnormalities, and patients with stroke. Since the input to ECN is the multi-channel sEMG without additional feature extraction, end-to-end LLAR is achieved.

Next, the proposed framework will be further discussed from the following aspects:



**Fig. 11.** sEMG denoising process of TA muscle of subject SP1 under the Gait type: (a) raw sEMG signal; (b) 7 IMFs decomposed by VMD; (c) normalized PE values of different IMFs; (d) denoised sEMG signal.

**Table 11**

Class distribution on the datasets of different types of subjects in Case II.

Subjects	Original dataset	Denoised dataset	Activity type	Class	Number of samples	Balanced dataset
HS1-HS7	A3	D3	StandTSit	0	735	BD3
			SitTStand	1	683	
			Gait	2	1408	
SP1-SP7	A4	D4	StandTSit	0	982	BD4
			SitTStand	1	766	
			Gait	2	1539	

#### 4.3.1. Selection of sliding window length

Typically, recognition performance improves as window length  $W_L$  (s) increases. However, it is also crucial to keep the system delay within the range of less than 300 ms (Gautam et al., 2020). Therefore, the effect of  $W_L$ , in other words, the sample length  $S_L$  ( $W_L \times F_s$ ), on recognition accuracy is investigated on datasets BD1 and BD2. Results of five replicated experiments are shown in Fig. 16. As can be seen, when  $S_L = 200$ , BD1 and BD2 both reach the highest accuracy. Therefore, considering the processing latency and accuracy, it is reasonable to set  $W_L$  to

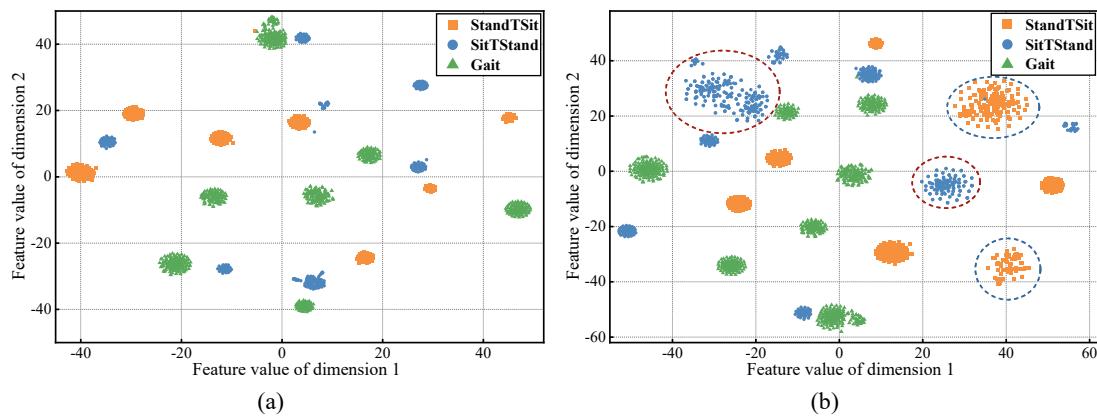
0.2 s.

#### 4.3.2. Method of dividing the dataset

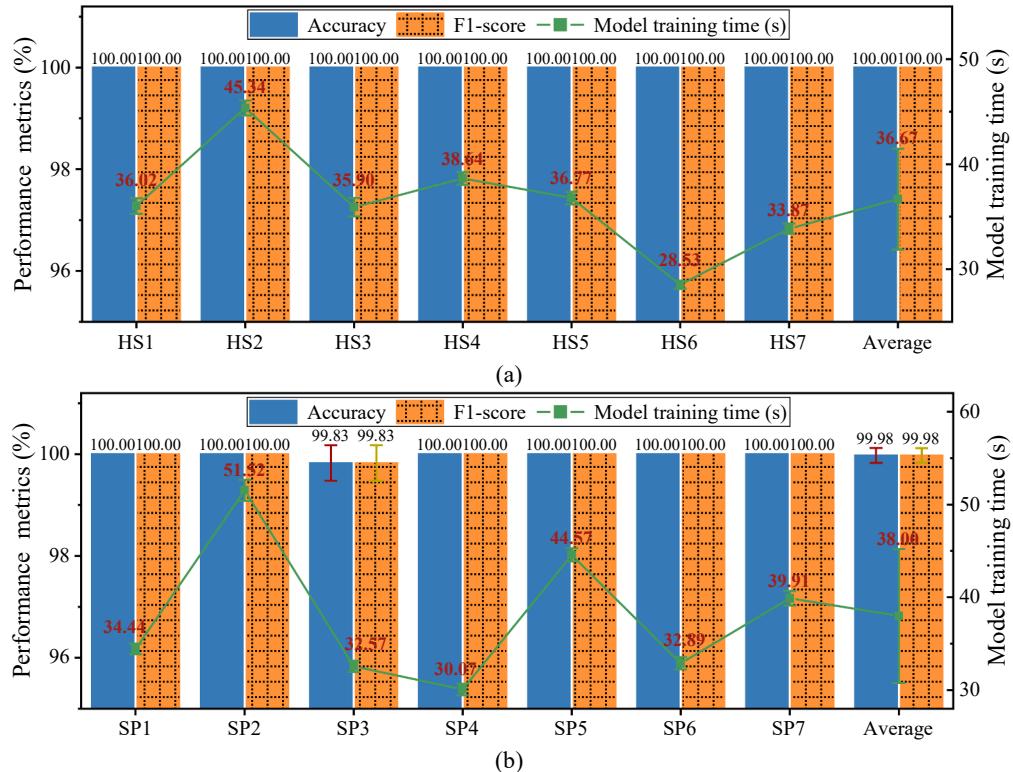
In model training and evaluation, K-fold cross validation and repeated experiments are two effective methods to reduce the interference of accidental factors. In this work, the latter is mainly used to divide the dataset. Therefore, both 5-fold cross validation (denoted as Method 1) and 5 replicate experiments (denoted as Method 2) are conducted on the datasets of the two cases to discuss the effect of different division methods on experimental results, as shown in Table 12. As can be seen, there is little difference between different methods, which further verifies the excellent performance of the proposed LLAR framework.

#### 4.3.3. Complexity of different DL models

Typically, the complexity of DL models is measured from two aspects. One is spatial complexity, i.e., the number of parameters. The other is time complexity, which is usually measured by floating point operations (FLOPs). Moreover, the FLOPs can only be used as an indirect index of computational speed and cannot fully reflect the running time (Wang et al., 2021b). Here, the complexity of ECN and other DL models is discussed from three perspectives: the number of model parameters,



**Fig. 12.** Using t-SNE algorithm to visualize sample distribution in 2D feature space of dataset: (a) D4; (b) BD4.



**Fig. 13.** Recognition performance using data collected from: (a) healthy subjects; (b) patients with stroke.

FLOPs, and model training time, as shown in Table 13. It can be seen that, compared with CapsNet, ECN has some additional parameters, which is mainly due to the added PReLU. In addition, 1D-CNN has the least training time, ECN is slightly higher than CapsNet and very close to 2D-CNN, while ResNetV1 has the longest time.

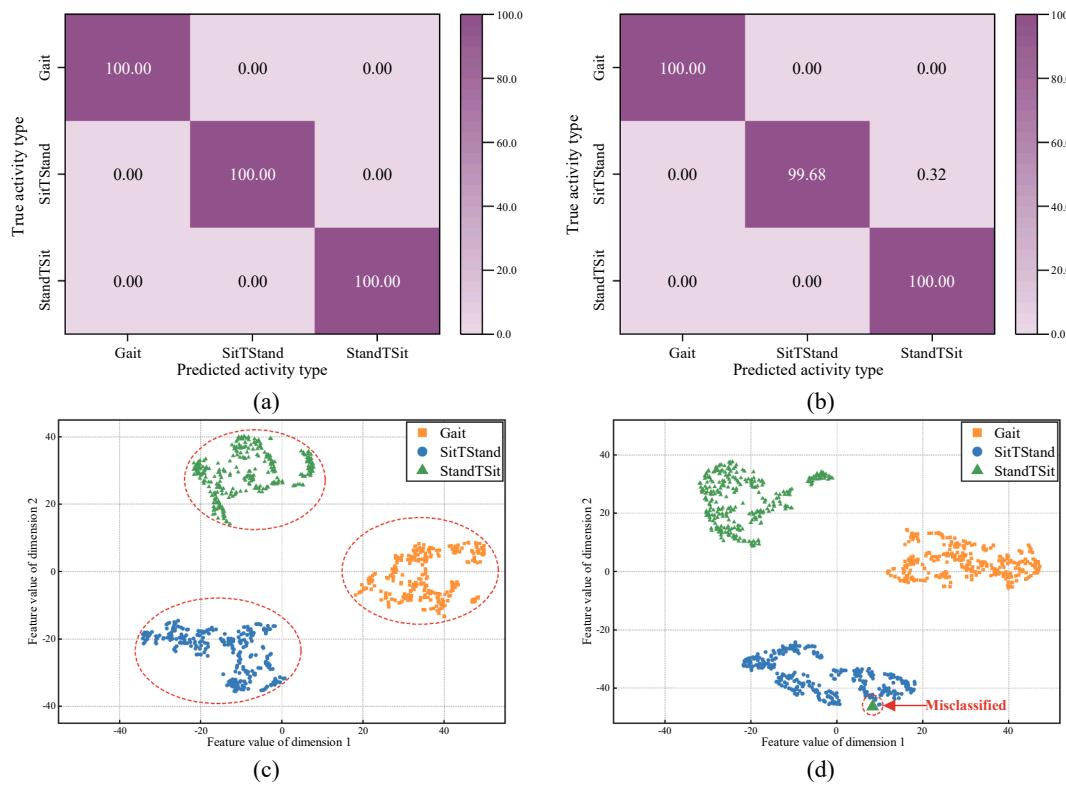
#### 4.3.4. Effect of VMD-PE-NLM on the anti-noise ability

To further verify the anti-noise ability of the proposed framework, the denoising effect of VMD-PE-NLM under different noise backgrounds is analyzed. Taking the RF muscle of subject H5 in Case I as an example, additive WGN signals with different SNRs are added to the raw sEMG shown in Fig. 4(a) to simulate the noise interference that may be encountered in real situations, and then form different levels of noisy signals. These noisy signals are then denoised using VMD-PE-NLM, and the experimental results are shown in Fig. 17. By comparing Fig. 17 with Fig. 4(a), it can be seen that: 1) VMD-PE-NLM is able to restore the real

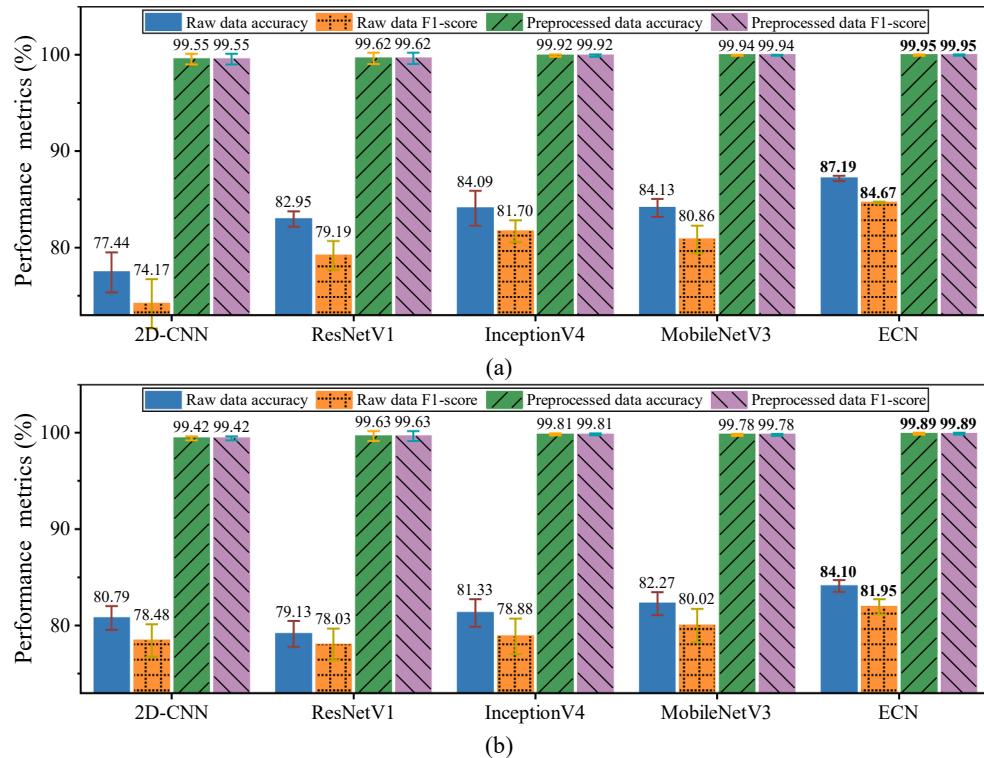
and useful sEMG signal excellently in a weaker noise background ( $\text{SNR} = 5 \text{ dB}$ ); 2) It can also achieve a satisfactory denoising result even in a strong noise background ( $\text{SNR} = -5\text{dB}$ ).

## 5. Conclusion and future work

In this work, an end-to-end LLAR framework based on sEMG DA and ECN is proposed. In the LLAR process, a hybrid VMD-PE-NLM technique is designed to effectively filter out the noise interference in raw sEMG signals, thus improving the anti-noise ability and accuracy of the proposed method. By implementing the K-Means SMOTE strategy, more high-quality sample data can be synthesized to balance the class distribution, thus effectively solving the class-imbalanced problem in the DL-based LLAR methods. The constructed ECN model can effectively improve the recognition performance by enhancing the feature learning ability of the classifier. Verification experimental results on the UCI



**Fig. 14.** Experimental results: (a)-(b) normalized confusion matrices of the test results on BD3 and BD4, respectively; (c)-(d) feature distribution of test samples at class capsule layer on BD3 and BD4, respectively.



**Fig. 15.** Performance comparison of different DL models: (a) healthy subjects; (b) patients with stroke.

dataset and dataset of patients with stroke show that it performs well in recognizing lower limb activities of multiple types of individuals, providing more satisfactory recognition performance than state-of-the-

art methods.

Notwithstanding the promising outcomes in the pilot studies, there are inevitably some limitations that warrant further investigations, as

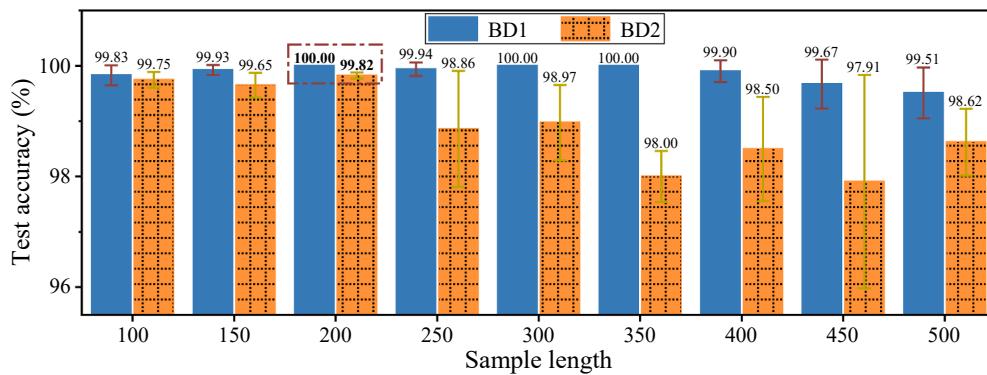


Fig. 16. Relationship between the window length and recognition accuracy.

**Table 12**  
Comparison of different data partitioning methods.

Dataset	Method	Accuracy ± STD (%)	Precision ± STD (%)	Sensitivity ± STD (%)	Specificity ± STD (%)	NPV ± STD (%)	F1-score ± STD (%)
BD1	1	99.88 ± 0.15	99.86 ± 0.19	99.86 ± 0.20	99.94 ± 0.08	99.94 ± 0.08	99.86 ± 0.18
	2	100.00 ± 0.00	100.00 ± 0.00	100.00 ± 0.00	100.00 ± 0.00	100.00 ± 0.00	100.00 ± 0.00
BD2	1	99.51 ± 0.15	99.50 ± 0.16	99.46 ± 0.19	99.75 ± 0.07	99.76 ± 0.07	99.48 ± 0.18
	2	99.85 ± 0.13	99.85 ± 0.13	99.85 ± 0.13	99.92 ± 0.07	99.92 ± 0.07	99.85 ± 0.13
BD3	1	99.86 ± 0.13	99.85 ± 0.12	99.82 ± 0.17	99.90 ± 0.10	99.92 ± 0.07	99.83 ± 0.15
	2	99.95 ± 0.10	99.95 ± 0.09	99.95 ± 0.10	99.98 ± 0.05	99.98 ± 0.05	99.95 ± 0.09
BD4	1	99.79 ± 0.21	99.78 ± 0.25	99.78 ± 0.23	99.89 ± 0.10	99.89 ± 0.11	99.78 ± 0.24
	2	99.89 ± 0.10	99.89 ± 0.10	99.89 ± 0.10	99.95 ± 0.05	99.95 ± 0.05	99.89 ± 0.10

**Table 13**  
Complexity comparison of different DL models.

Model	Number of parameters	FLOPs	Model training time (s)	
			BD1	BD2
1D-CNN	74,211	66,496	27.57 ± 1.47	47.48 ± 2.04
2D-CNN	184,515	206,016	83.70 ± 1.83	174.59 ± 1.44
ResNetV1	279,267	826,304	121.91 ± 2.98	277.62 ± 6.30
CapsNet	698,848	1,783,808	79.87 ± 1.94	147.93 ± 2.45
ECN	735,332	1,786,560	90.25 ± 1.69	182.67 ± 5.45

described below:

1) Although the current study discussed the applicability of the proposed framework to different types of individuals, only data from fewer subjects were used. Especially for stroke patients, factors such as

physiological structure and motor function vary greatly from patient to patient, which may affect the recognition accuracy. Therefore, it is necessary to collect sEMG data from more subjects to form a more complete dataset and conduct further evaluation.

2) The current study lacked consideration of other common lower limb activities in daily life, e.g., stepping over obstacles, ascending/descending stairs, etc. Recognizing this, more types of lower limb activities need to be added in future work. Meanwhile, it is considered to integrate incremental learning into the proposed framework to construct an ECN model with incremental classification capability.

3) The proposed framework has only been evaluated offline, and online testing may face some unknown challenges. Therefore, future research needs to apply it to the real-time control and evaluation of neural-controlled lower limb rehabilitation robot. In addition, light-weight improvements of ECN can also be investigated to facilitate its application in embedded systems.

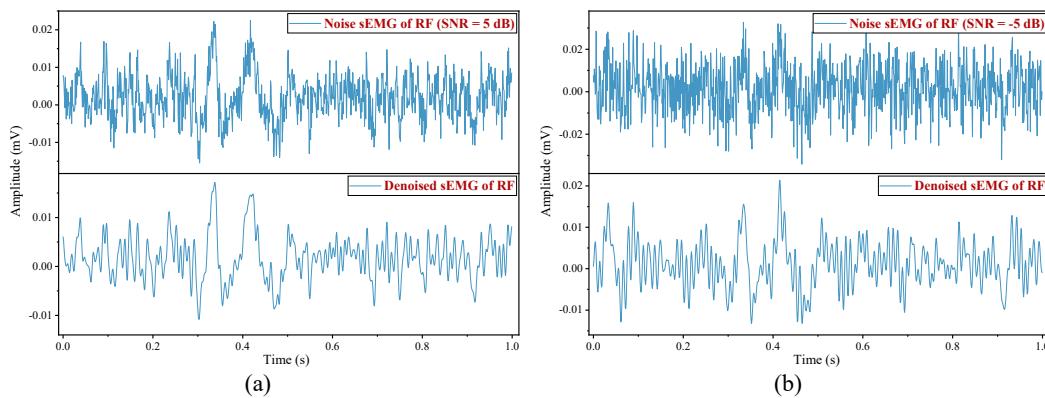


Fig. 17. Denoising effect of VMD-PE-NLM under different noise backgrounds: (a) SNR = 5 dB; (b) SNR = -5 dB.

## CRediT authorship contribution statement

**Changhe Zhang:** Conceptualization, Methodology, Writing – original draft. **Yangan Li:** Data curation, Formal analysis. **Zidong Yu:** Software, Visualization. **Xiaolin Huang:** Funding acquisition. **Jiang Xu:** Investigation, Validation. **Chao Deng:** Supervision, Writing – review & editing.

## Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

## Data availability

Data will be made available on request.

## Acknowledgement

The work here is supported by the National Natural Science Foundation of China under Grant U1913601. The authors would also like to thank Dr. ZeJian Chen for insightful discussion and YiLin Liu for proofreading this paper.

## References

- Al-Quraishi, M. S., Elamvazuthi, I., Tang, T. B., Al-Qurishi, M., Parasuraman, S., & Borboni, A. (2021). Multimodal fusion approach based on EEG and EMG signals for lower limb movement recognition. *IEEE Sensors Journal*, 21(24), 27640–27650.
- Bandt, C., & Pompe, B. (2002). Permutation entropy: A natural complexity measure for time series. *Physical Review Letters*, 88(17), Article 174102.
- Batista, G. E., Prati, R. C., & Monard, M. C. (2004). A study of the behavior of several methods for balancing machine learning training data. *ACM SIGKDD Explorations Newsletter*, 6(1), 20–29.
- Buades, A., Coll, B., & Morel, J.-M. (2005). A review of image denoising algorithms, with a new one. *Multiscale Modeling & Simulation*, 4(2), 490–530.
- Chawla, N. V., Bowyer, K. W., Hall, L. O., & Kegelmeyer, W. P. (2002). SMOTE: Synthetic minority over-sampling technique. *Journal of Artificial Intelligence Research*, 16, 321–357.
- Chen, X., Xu, X., Liu, A., McKeown, M. J., & Wang, Z. J. (2017). The use of multivariate EMD and CCA for denoising muscle artifacts from few-channel EEG recordings. *IEEE Transactions on Instrumentation and Measurement*, 67(2), 359–370.
- Chen, Z., Qian, Y., Wang, Y., & Fang, Y. (2022). Deep convolutional generative adversarial network-based EMG data enhancement for hand motion classification. *Frontiers in Bioengineering and Biotechnology*, 10.
- de Freitas, R. C., Naik, G. R., Valença, M. J. S., Bezerra, B. L. D., de Souza, R. E., & dos Santos, W. P. (2022). Surface electromyography classification using extreme learning machines and echo state networks. *Research on Biomedical Engineering*, 38(2), 477–498.
- Douzas, G., Bacao, F., & Last, F. (2018). Improving imbalanced learning through a heuristic oversampling method based on k-means and SMOTE. *Information Sciences*, 465, 1–20.
- Dragomiretskiy, K., & Zosso, D. (2013). Variational mode decomposition. *IEEE Transactions on Signal Processing*, 62(3), 531–544.
- Feizollahi, A., Anuar, N. B., Salleh, R., & Amalina, F. (2014). Comparative study of k-means and mini batch k-means clustering algorithms in android malware detection using network traffic analysis. *International Symposium on Biometrics and Security Technologies (ISBAST)*, 2014, 193–197.
- Feng, Y., Wang, H., Vladareanu, L., Chen, Z., & Jin, D. (2019). New motion intention acquisition method of lower limb rehabilitation robot based on static torque sensors. *Sensors*, 19(15), 3439.
- Gautam, A., Panwar, M., Biswas, D., & Acharyya, A. (2020). MyoNet: A transfer-learning-based LRCN for lower limb movement recognition and knee joint angle prediction for remote monitoring of rehabilitation progress from sEMG. *IEEE Journal of Translational Engineering in Health and Medicine*, 8, 1–10.
- Gu, L., Jiang, J., Han, H., Gan, J. Q., & Wang, H. (2023). Recognition of unilateral lower limb movement based on EEG signals with ERP-PCA analysis. *Neuroscience Letters*, 800, Article 137133.
- He, H., Bai, Y., Garcia, E. A., & Li, S. (2008). ADASYN: Adaptive synthetic sampling approach for imbalanced learning. *IEEE International Joint Conference on Neural Networks (IEEE World Congress on Computational Intelligence)*, 2008, 1322–1328.
- He, K., Zhang, X., Ren, S., & Sun, J. (2015). Delving deep into rectifiers: Surpassing human-level performance on imagenet classification. *Proceedings of the IEEE International Conference on Computer Vision*, 1026–1034.
- He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (pp. 770–778).
- Hermens, H. J., Freriks, B., Merletti, R., Stegeman, D., Blok, J., Rau, G., ... Hägg, G. (1999). European recommendations for surface electromyography. *Roessingh Research and Development*, 8(2), 13–54.
- Howard, A., Sandler, M., Chu, G., Chen, L.-C., Chen, B., Tan, M., ... Vasudevan, V. (2019). Searching for mobilenetv3. *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 1314–1324.
- Iqbal, N., Khan, T., Khan, M., Hussain, T., Hameed, T., & Bukhari, S. A. C. (2021). Neuromechanical signal-based parallel and scalable model for lower limb movement recognition. *IEEE Sensors Journal*, 21(14), 16213–16221.
- Issa, S., & Khaled, A. R. (2022). Lower limb movement recognition using EMG signals. *Intelligent Systems Design and Applications: 21st International Conference on Intelligent Systems Design and Applications (ISDA 2021) Held During December 13–15, 2021*, 336–345.
- Lin, C.-T., Jiang, W.-L., Chen, S.-F., Huang, K.-C., & Liao, L.-D. (2021). Design of a wearable eye-movement detection system based on electrooculography signals and its experimental validation. *Biosensors*, 11(9), Article 343.
- Liu, J., Zhang, C., & Jiang, X. (2022). Imbalanced fault diagnosis of rolling bearing using improved MsR-GAN and feature enhancement-driven CapsNet. *Mechanical Systems and Signal Processing*, 168, Article 108664.
- Ma, S., Lv, B., Lin, C., Sheng, X., & Zhu, X. (2020). EMG signal filtering based on variational mode decomposition and sub-band thresholding. *IEEE Journal of Biomedical and Health Informatics*, 25(1), 47–58.
- Mazzia, V., Salvetti, F., & Chiaberge, M. (2021). Efficient-capsnet: Capsule network with self-attention routing. *Scientific Reports*, 11(1), Article 14634.
- Menardi, G., & Torelli, N. (2014). Training and assessing classification rules with imbalanced data. *Data Mining and Knowledge Discovery*, 28, 92–122.
- Naik, G. R., Selvan, S. E., Arjunan, S. P., Acharyya, A., Kumar, D. K., Ramanujam, A., & Nguyen, H. T. (2018). An ICA-EBM-based sEMG classifier for recognizing lower limb movements in individuals with and without knee pathology. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 26(3), 675–686.
- Qiu, S., Zhao, H., Jiang, N., Wang, Z., Liu, L., An, Y., ... Fortino, G. (2022). Multi-sensor information fusion based on machine learning for real applications in human activity recognition: State-of-the-art and research challenges. *Information Fusion*, 80, 241–265.
- Sabour, S., Frosst, N., & Hinton, G. E. (2017). Dynamic routing between capsules. *Advances in Neural Information Processing Systems*, 30.
- Sanchez, O. F. A., Sotelo, J. L. R., Gonzales, M. H., & Hernandez, G. A. M. (2014). Emg dataset in lower limb data set. *UCI Machine Learning Repository*, 2.
- Sheng, B., Wang, X., Xiong, S., Hou, M., & Zhang, Y. (2019). Kinematic Metrics for upper-limb functional assessment of stroke patients. *International Conference on Intelligent Informatics and Biomedical Sciences (ICIBMS)*, 2019, 45–51.
- Si, X., Dai, Y., & Wang, J. (2022). Recognition of Lower Limb Movements Baesd on Electromyography (EMG) Texture Maps. *2022 IEEE 5th International Conference on Electronics Technology (ICET)*, 1091–1095.
- Singh, P., Shahnavazuddin, S., & Pradhan, G. (2018). An efficient ECG denoising technique based on non-local means estimation and modified empirical mode decomposition. *Circuits, Systems, and Signal Processing*, 37, 4527–4547.
- Snoek, J., Larochelle, H., & Adams, R. P. (2012). Practical Bayesian optimization of machine learning algorithms. *Advances in Neural Information Processing Systems*, 25.
- Szegedy, C., Ioffe, S., Vanhoucke, V., & Alemi, A. (2017). Inception-v4, inception-resnet and the impact of residual connections on learning. *Proceedings of the AAAI Conference on Artificial Intelligence*, 31(1).
- Thukral, R., Kumar, A., & Arora, A. S. (2019). Effect of different thresholding techniques for denoising of emg signals by using different wavelets. *2019 2nd International Conference on Intelligent Communication and Computational Techniques (ICCT)*, 161–165.
- Van der Maaten, L., & Hinton, G. (2008). Visualizing data using t-SNE. *Journal of Machine Learning Research*, 9(11).
- Vijayvargiya, A., Gupta, V., Kumar, R., Dey, N., & Tavares, J. M. R. S. (2021). A hybrid WD-EEMD sEMG feature extraction technique for lower limb activity recognition. *IEEE Sensors Journal*, 21(18), 20431–20439.
- Vijayvargiya, A., Kumar, R., Dey, N., & Tavares, J. M. R. S. (2020). Comparative analysis of machine learning techniques for the classification of knee abnormality. *2020 IEEE 5th International Conference on Computing Communication and Automation (ICCCA)*, 1–6.
- Vijayvargiya, A., Kumar, R., & Dey, N. (2021). Voting-based 1D CNN model for human lower limb activity recognition using sEMG signal. *Physical and Engineering Sciences in Medicine*, 44, 1297–1309.
- Vijayvargiya, A., Prakash, C., Kumar, R., Bansal, S., & Tavares, J. M. R. S. (2021). Human knee abnormality detection from imbalanced sEMG data. *Biomedical Signal Processing and Control*, 66, Article 102406.
- Vijayvargiya, A., Singh, P., Kumar, R., & Dey, N. (2022). Hardware implementation for lower limb surface EMG measurement and analysis using explainable AI for activity recognition. *IEEE Transactions on Instrumentation and Measurement*, 71, 1–9.
- Vijayvargiya, A., Singh, P., Kumar, R., & Tavares, J. M. R. S. (2022). Human lower limb activity recognition techniques, databases, challenges and its applications using sEMG signal: An overview. *Biomedical Engineering Letters*, 12(4), 343–358.
- Wang, J., Dai, Y., & Si, X. (2021). Analysis and recognition of human lower limb motions based on electromyography (EMG) signals. *Electronics*, 10(20), Article 2473.
- Wang, Q., Wu, B., Zhu, P., Li, P., Zuo, W., & Hu, Q. (2020). ECA-Net: Efficient channel attention for deep convolutional neural networks. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 11534–11542.
- Wang, X., Wang, X., Lv, T., Jin, L., & He, M. (2021b). Harnas: Human activity recognition based on automatic neural architecture search using evolutionary algorithms. *Sensors*, 21(20), Article 6927.

- Wang, Y., Cang, S., & Yu, H. (2019). A survey on wearable sensor modality centred human activity recognition in health care. *Expert Systems with Applications*, 137, 167–190.
- Wang, Y., Wu, Q., Dey, N., Fong, S., & Ashour, A. S. (2020). Deep back propagation–long short-term memory network based upper-limb sEMG signal classification for automated rehabilitation. *Biocybernetics and Biomedical Engineering*, 40(3), 987–1001.
- Wei, C., Wang, H., Hu, F., Zhou, B., Feng, N., Lu, Y., ... Jia, X. (2022). Single-channel surface electromyography signal classification with variational mode decomposition and entropy feature for lower limb movements recognition. *Biomedical Signal Processing and Control*, 74, Article 103487.
- Xiong, D., Zhang, D., Zhao, X., & Zhao, Y. (2021). Deep learning for EMG-based human-machine interaction: A review. *IEEE/CAA Journal of Automatica Sinica*, 8(3), 512–533.
- Zhang, P., Zhang, J., & Elsabbagh, A. (2022). Lower limb motion intention recognition based on sEMG fusion features. *IEEE Sensors Journal*, 22(7), 7005–7014.