Article:
Wang, W. orcid.org/0000-0002-2336-0477, Zhao, C. orcid.org/0000-0001-5286-9419, Li, X. orcid.org/0000-0002-0289-6926 et al. (3 more authors) (Cover date: 2023) Research on Multimodal Fusion Recognition Method of Upper Limb Motion Patterns. IEEE Transactions on Instrumentation and Measurement, 72. 4008312. ISSN 0018-9456

https://doi.org/10.1109/tim.2023.3289556

Research on multimodal fusion recognition method of upper limb motion patterns

Wendong Wang, Chengzhi Zhao, Xin Li, Zhi-Qiang Zhang, *Member, IEEE*, Xiaoqing Yuan, Hanhao Li

***Abstract*—In order to solve the problems of single movement pattern recognition information and low recognition accuracy of multi-joint upper limb exoskeleton rehabilitation training, a multimodal information fusion method with human surface electromyography(sEMG) and electrocardiogram(ECG) was proposed, and an Inception-Sim model for upper limb motion pattern recognition was designed. Integrating the advantages of multimodal information, inspired by the convolutional neural network processing image classification problem, the original signal was converted into a Gramian Angular Summation/Difference Fields-Histogram of Oriented Gradient (GASF/GADF-HOG) image based on the principle of Grameen angle superposition/difference field, and the directional gradient histogram feature of the GASF/GADF image was extracted. The Inception-Sim model was constructed based on the Inception V3 model, and the human motion pattern recognition was completed on the basis of the transfer learning network. VGG16, ResNet-50, and other backbone networks were selected as comparison models. The recognition accuracy of each motion pattern for all participants reaches up to 90%, which is better than that of the control model. The average iteration speed of the proposed Inception-Sim model improved by about 21% compared to the control model. The experimental results show that the proposed multimodal information fusion recognition method can improve the accuracy and iteration speed of upper limb motion recognition mode and then improve the effect of upper limb rehabilitation training.**

***Index*: sEMG, ECG, Deep learning, GASF/GADF-HOG, Motion intensity variation interval estimation.**

I. Introduction

The rehabilitation of hemiplegia caused by stroke brings a substantial social burden. As an emerging method of rehabilitation for hemiplegia, the upper limb rehabilitation exoskeleton can provide more scientific training methods for patients while reducing the burden on physical therapists. One of the key technologies of the upper limb rehabilitation exoskeleton is to identify the rehabilitation motion pattern. A reasonable rehabilitation motion pattern can more effectively restore and improve the patient's limb motion ability. The upper limb motion pattern recognition has developed from the traditional single signal recognition to the multimodal [1] information fusion recognition, and the signal processing method has also developed from the extraction of the signal time domain, frequency domain, and other features to the method of converting a one-bit time-series signal into a two-dimensional picture[1]–[3]. Pattern recognition methods have also developed from linear discriminant analysis(LDA), support vector machines(SVM), and neural networks to deep learning models with multi-layer networks(MLP)[4]–[6].

Bioelectrical signals such as surface electromyography(sEMG) play an essential role in characterizing human motion intentions. sEMG signals can better characterize muscle activity levels, and human motion patterns[7] and have been widely used in the field of rehabilitation medicine[8]. Furthermore, sEMG signals are generated by nerve impulses and can be detected before muscle activity[9]. Therefore, the sEMG signal is an excellent choice to ensure the speed and efficiency of the exoskeleton robot controller in the process of human-robot interaction. Since sEMG can be detected non-invasively, it is suitable for the work of the human-machine interface.

However, there are still many problems in the utilization of sEMG signal as the control source. The recognition accuracy and robustness are often reduced in non-ideal conditions. For example, electrode shift[10] occurred in the process of human movement, human movement as well as muscle strength changes[11]. Oluwarotimi[12] identified the negative impact of mobility on the recognition of upper limb movement patterns by collecting accelerometer signals(ACC) and sEMG simultaneously, and presented a solution to integrate ACC and sEMG to increase the accuracy of upper limb movement pattern

*This work is supported by the Science and technology plan project of Xi'an city, Grant no. 21XJZZ0079, and the Natural Science Foundation of Shaanxi Province (Grant No. 2020JM-131 and 2020KW-058)

Wendong Wang is with Northwestern Polytechnical University, Associate professor, (e-mail: wdwang@nwpu.edu.cn)

Chengzhi Zhao currently works on the control system design of exoskeleton robot in Northwestern Polytechnical University, where he is pursuing the master degree in Mechanical Engineering.(e-mail: chengzhizhao@mail.nwpu.edu.cn)

Xin Li currently works on the control system design of exoskeleton robot in Northwestern Polytechnical University, where he is pursuing the master degree in Mechanical Engineering. (e-mail: wqyy@mail.nwpu.edu.cn)

Zhi-Qiang Zhang is currently an Associate Professor in and robotic control with the University of Leeds, Leeds, U.K. (e-mail: z.zhang3@leeds.ac.uk)

Xiaoqing Yuan is with Northwestern Polytechnical University, professor, (e-mail: yuan@nwpu.edu.cn)

Hanhao Li currently works on the control system design of exoskeleton robot in Northwestern Polytechnical University, where he is pursuing the master degree in Mechanical Engineering. (e-mail: lihanhao_1997@mail.nwpu.edu.cn)

recognition. Wu[13] employed high-density electrodes to capture sEMG and transform it into pictures in order to determine the effect of electrode movement on the recognition of upper limb movement patterns. A dilated convolutional neural network(DCNN) is proposed to improve recognition's robustness. Al-Timemy[14] investigated the problem of the robustness of upper limb motion pattern recognition under variable force levels and proposed time-dependent power spectrum descriptors(TD-PSD) features to characterize sEMG activity, thereby minimizing the classification performance degradation brought on by changes in muscle contraction force levels.

As mentioned above, researchers would like to use high-density electrodes or increase electrodes to improve the accuracy of motion recognition, which can provide spatial information of the entire muscle[15]. However, too many channels lead to computational burden and redundant information, requiring additional techniques such as feature projection/selection to solve these problems. Multi-mode information fusion is a better solution.

The change of heart rate during exercise is related to the change of exercise intensity[16], [17], which can be reflected in the electrocardiogram(ECG) signal as the heart rate variability. Through a large number of mathematical model processing and statistical analysis, some researchers found that there is an exponential decay relationship between heart rate(HR) and heart rate variability(HRV)[18]. Studies have proved that the power spectrum of heart rate variability is affected by exercise intensity[19]. Both heart rate and heart rate variability come from the electrical signals of the heart, which indicated that the electrical signals contain movement information. PPG is a cardiac signal but is more susceptible to motion artifacts. Some investigations have calculated heart rate by fusing accelerometers with PPG[20]. Unlike ECG, the diagnostic value of PPG (based on morphological analysis) is still in the research phase, and its primary use is limited to the measurement of vital signs. The signal can be volatile even in noiseless PPG recordings of healthy individuals[21]. Since deep learning has a strong feature extraction ability, we use the combination of sEMG signal and ECG signal to conduct motion pattern recognition through neural network, so as to improve the accuracy and robustness of motion pattern recognition.

Multimodal fusion and deep learning have been proved to improve the accuracy and robustness of upper limb motion pattern recognition. Therefore, our research is conducted from two aspects. The performance of upper limb motion pattern recognition may be improved by the fusion of sEMG and ECG signals containing motion intensity information and the use of CNN as a classifier. The main purpose is to improve the recognition accuracy and robustness.

Based on the above analysis, in order to solve the problem of using sEMG merely, a multimodal information fusion method based on gramian angular summation/difference fields-histogram of oriented gradient (GASF/GADF-HOG) is proposed. Besides, an upper limb motion pattern recognition model Inception-Sim is proposed. This method converts ECG signals and sEMG signals into two-dimensional GASF-HOG features, and fuses ECG signals with sEMG signals. The fused features are used as inputs into the constructed Inception-Sim convolutional neural network, and the classification of upper limb motion patterns is realized finally.

The purpose of this work is to propose an upper limb motion classification method that can ensure robustness, accuracy and real-time performance with fewer sensors. Compared with the existing methods, the main contributions of this work are proved in three aspects.

(1) The proposed multi-mode information fusion method based on GASF/GADF-HOG can improve the accuracy and robustness of upper limb motion pattern recognition.

(2) The presented Inception-Sim model constructed has fewer parameters and faster calculation speed, while ensuring the identification accuracy.

(3) Using too many electrodes is solved by adding the motion intensity dimension to the upper limb motion pattern and fusing ECG and sEMG with motion intensity information to improve the robustness of upper limb motion pattern recognition.

II. Motion Pattern Classification Strategy.

2.1. Experimental Design.

Traditional upper limb motion pattern recognition is often only aimed at recognizing discrete motions. However, under the same or similar actions, actions with immobile motion intensity represent different motion intentions of the human body, and the motion pattern recognition for discrete actions has certain limitations. Therefore, this paper defines a multi-dimensional motion pattern, including motion action and motion intensity, to characterize the complete motion intention of the human body.

In this experiment, limited by the freedom of motion of the exoskeleton manipulator, six types of common motion patterns are selected as the target classification motion. The motion pattern has two dimensions: sEMG is mainly used for the dimension of muscle space motion direction, and ECG is mainly used for the Exercise intensity dimension.

TABLE. I

THE SIX MOTION MODES CORRESPOND TO DIMENSION LABELS

Joint direction \ Exercise intensity	High intensity	Low intensity
Lateral joint flexion and extension	Grip(PP)	Drink water(DW)
positive joint flexion and extension	Weightlift(WL)	Horizontal action(WU)
reverse joint flexion and extension	Pull bar(TB)	Horizontal lift(EQ)

The corresponding relations generated by the motion mode dimension labels are shown in TABLE. I. For brevity, the six motion mode categories are replaced by abbreviations for their motion modes.

Before data collection, each participant was asked to perform six types of target classification exercises sequentially, and the degree of muscle contraction should meet the requirements to ensure familiarity with the target classification motions. The specific action classification is shown in Fig. 1. sEMG and ECG

were recorded for 40 seconds per motion (each motion lasted about 4 seconds and was repeated ten times).

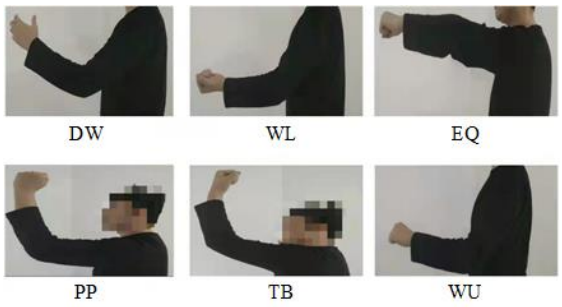

Fig. 1 Classification of six motion modes based on shoulder and elbow motion

2.2. Signal Acquisition.

The sEMG sensor is a six-channel sEMG sensor manufactured by SICHIRAY. The product size is 11 cm x 4 cm. The operating voltage is 5 V, the output signal is analog/digital, and it supports up to six channels for simultaneous acquisition with a maximum sampling frequency of 1000 Hz. The heart sensor is the BMD101 Bluetooth ECG sensor produced by NeuroSky. Product size is 1.86 x 6.45 cm, charging voltage is 5 V, working voltage is 3.3 V, and wireless Bluetooth mode communication method. The sampling frequency is 512 Hz. The physical diagram of the sEMG sensor and the ECG sensor is shown in Fig. 2.

(a)sEMG sensor (b)ECG sensor

Fig. 2 Physical view of sEMG sensor and ECG sensor

At the same time, the Bmd101 Bluetooth ECG sensor and the 6-channel sEMG sensor were activated, and the data of 10 healthy subjects were collected at the sampling frequency of 512 Hz and 1000 Hz. We sample the signal to the same frequency by resampling and then process it through a sliding window. The dataset is divided into two sub-datasets, of which 80% of the data is used as the training data set for modeling and parameter optimization, and the other 20% of the data is used as the test data set for algorithm verification. A total of 10 healthy subjects (including two women and eight men aged 20-41 years) participated in the bioelectric signal experiments. Subjects did not have cardiovascular disease or muscle disease and were informed of study objectives, methods and risks prior to participating in the test. The day before the test, subjects should avoid high-intensity training and avoid ingesting stimulant substances such as nicotine and alcohol. In order to record sEMG and ECG signals related to the target upper limb motions, sEMG electrodes were placed on the biceps, triceps, anterior deltoid, and posterior deltoid muscles of the subject's arm[11]. The sEMG electrode uses a round AgCl wet electrode with a diameter of 5 mm and spacing of 10 mm.[22] The ECG electrodes were placed on the 12-lead international chest leads V3, V4 (marked by black circles), as shown in Fig. 3. Before placing the electrodes, thoroughly clean the skin surface where the electrodes are placed with alcohol to ensure good contact with the underlying arm muscles. If the participant's skin is too dry, apply medical gel to the site to ensure good contact between the electrodes and the skin, helping to improve the quality of the recorded signal[23].

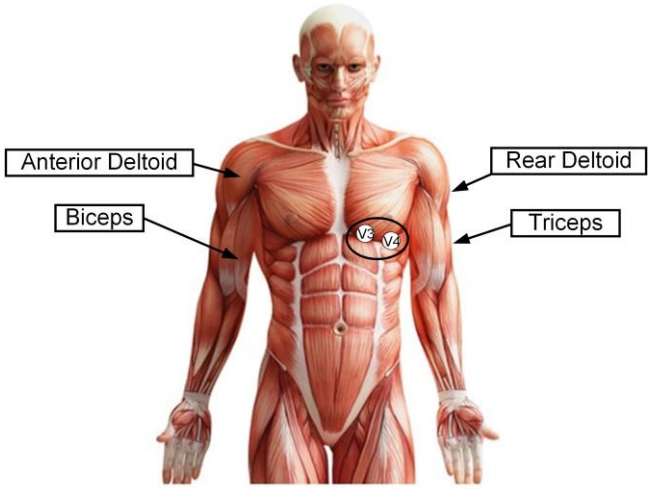

Fig. 3 Position of ECG and sEMG electrodes

III. Signal Pre-processing.

3.1. sEMG Signal Pre-processing.

Use filtering and fitting baselines to eliminate baseline shifts. After the baseline shift removal is done, build high-pass and low-pass filters. The acquired sEMG signal was filtered using a fourth-order Butterworth filter (20 Hz ~ 500 Hz) and corrected by taking its absolute value.

Low-pass filtering produces a "smooth" sEMG signal, as shown in Fig. 4, with a more pronounced shape (envelope) than when a higher cutoff frequency is used. Under different conditions (such as continuous contraction, gait, etc.), it is still worthy of further exploration to select an appropriate low-pass cutoff frequency suitable for the construction of the sEMG envelope. 2 Hz was finally chosen as the cutoff frequency in a limited range of choices.

To extract the envelope of a signal using low-pass filtering, we use the following formula

$$y(t) = \mathrm{abs}(h(t) \bullet x(t)) \tag{1}$$

Where $x(t)$ is the input signal, $h(t)$ is the impulse response of the low-pass filter, ' • ' denotes convolution, abs() denotes the absolute value.

The impulse response h(t) should be a low-pass filter that attenuates high-frequency components and passes low-frequency components. One commonly used filter is a first-order low-pass filter with transfer function

$$H(s) = 1/(1 + s/\omega) \tag{2}$$

where s is the Laplace variable and ω is the cutoff frequency. The impulse response of this filter is

$$h(t) = (1/\omega)\bullet\exp(-t/T) \tag{3}$$

where $\tau = 1/\omega$ is the time constant of the filter.

By convolving the input signal $x(t)$ with the impulse response $h(t)$ and taking the absolute value, we get the envelope of the signal. The resulting signal $y(t)$ will be a low-pass filtered version of the input signal with the high-frequency components removed, leaving only the envelope.

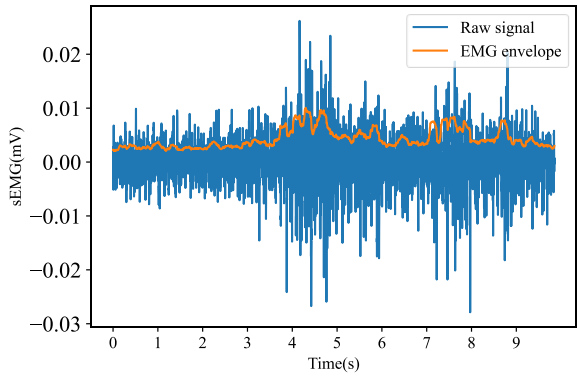

Fig. 4 The raw sEMG and envelope of raw sEMG evaluated using low-pass filtering method

3.2. ECG Signal Pre-processing.

3.2.1 ECG Denoising Method Based on Unbiased FIR.

After the fast Fourier transform of the signal, a Butterworth high-pass filter is used to eliminate baseline drift (0 Hz~0.5 Hz), and a 50 Hz notch filter is used to eliminate power frequency interference. In the time domain space, its baseline drift has been completely eliminated, and the base point shifts from the range of -0.2 mv to 0.2 mv to the 0 baselines, as shown in Fig. 5.

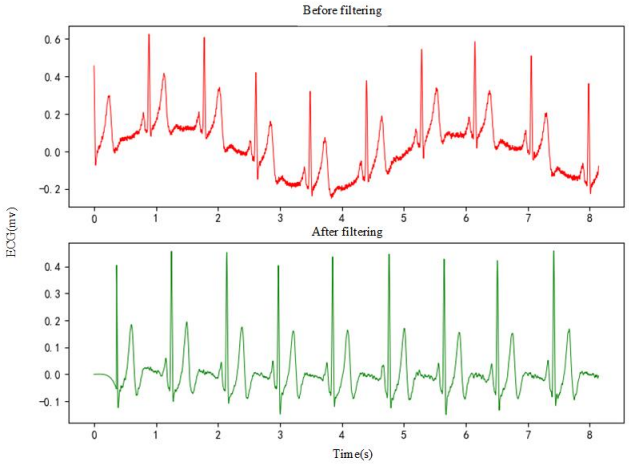

Fig. 5 ECG time domain sequence after 0.5 Hz high-pass filtering and 50 Hz notch

3.2.2 ECG Smoothing Based on p-shift UFIR.

Different artifacts and measurement noise often affect accurate feature extraction when extracting ECG signal features. One of the standard methods of ECG signal processing is linear prediction. For frequency bands that do not need prediction in ECG signal processing, smoothing can improve the prediction efficiency of ECG signals. A p-shift finite-length unit impulse response filter (p-shift UFIR) was used for ECG signal smoothing[24].

Suppose that the ECG signal x_n is added with unknown random noise v_n with mean 0. Then the measured value s_n can be summed under the discrete time series of the two

$$s_n = x_n + v_n \tag{4}$$

Given that random noise v_n may not be of a white Gaussian distribution, and its statistics are generally not known, the best way to avoid significant estimation errors is to use filters that do not require a strong dependence on noise information. Given the advantages of the above filters, noise and initial conditions can be ignored entirely using a p-shift UFIR filter.

In a finite sequence [m-p, n-p], the ECG signal after denoising by p-shift UFIR can be represented by a polynomial. The value of $\hat{x}_{n|n-p}$ estimated by the p-shift UFIR filter can be obtained in the form based on the convolution taken from s_n

$$\begin{aligned}\hat{x}_{n|n-p} &= \sum_{i=p}^{N-1+p} h_{li}(p)s_{n-i} \\ &= \mathbf{w}_l^T(p)\mathbf{s}_N(p)\end{aligned} \tag{5}$$

where $h_{\ln}(p) \triangleq h_{\ln}(N,p)$ is the UFIR variable impulse response of the {N, p} sequence; the extended measurement vector $\mathbf{s}_N$ is

$$\mathbf{s}_N(p) = \left[s_{n-p}, s_{n-p}\cdots s_{m-p}\right]^T \tag{6}$$

The filter gain matrix is given by

$$\mathbf{w}_l^T(p) = \left[h_{lp}(p)h_{l(1+p)}(p)\cdots h_{l(N-1+p)}(p)\right] \tag{7}$$

Satisfy the unbiased condition

$$E\left\{\hat{x}_{n|n-p}\right\} = E\left\{x_n\right\} \tag{8}$$

where $E\{z\}$ is the average value of z; $h_{\ln}(p)$ is

$$h_{li}(p) = \sum_{j=0}^{l} a_{jl}(p)i^j \tag{9}$$

where $i \in [p, N-1+p]$; $a_{jl}(p)$ is

$$a_{jl}(p) = (-1)^j \frac{M_{(j+1)l}(p)}{|\mathbf{D}(p)|} \tag{10}$$

where $|\mathbf{D}(p)|$ is the determinant of the matrix $\mathbf{D}(p) = \mathbf{V}^T(p)\mathbf{V}(p)$; $\mathbf{V}(p)$ is the $N\times(l+1)$ -order Vandermonde matrix, and its Vandermonde matrix is expressed as

$$\mathbf{V}(p)=\begin{bmatrix} 1 & p & p^2 & \cdots & p^l \\ 1 & 1+p & (1+p)^l & \cdots & (1+p)^l \\ 1 & 2+p & (2+p)^l & \cdots & (2+p)^l \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & N-1+p & (N-1+p)^l & \cdots & (N-1+p)^l \end{bmatrix} \quad (11)$$

where $M_{(j+1)l}(p)$ is the sub-determinant of $|\mathbf{D}(p)|$, and the function $h_{li}(N,p)$ has the following basic properties

$$\begin{cases} h_{li}(N,p)=\begin{cases}\text{non-trivial value}, & p\le i\le N-1+p\\ 0, & \text{else}\end{cases} \\ \sum_{i=p}^{N-1+p} h_{li}(N,p)=1 \\ \sum_{i=p}^{N-1+p} h_{li}(N,p)i^u=0, \quad 1\le u\le l \end{cases} \quad (12)$$

For low-order $h_{li}(N,p)$ functions, when l=1,2 there is a simpler formula. For higher orders of l, $h_{li}(N,0)$ can be computed using the recurrence relation in the Gamboa-Rosales[25] study, and then $h_{li}(N,p)$ is obtained by its projection. The important thing is that the UFIR-based $\hat{x}_{n|n-p}$ estimate (calculated from equation (5)) does not require noise statistics and initial values. Under the condition of this filter, the computing environment allows zero-mean noise to have any distribution and covariance, which is a fundamental difference from the best estimation method.

TABLE. II

BASIC UFIR SIGNAL SMOOTHING ALGORITHM

Algorithm: Basic UFIR Signal Smoothing Algorithm
Input: $\mathbf{S}_N, l, N$ Output： $\hat{x}_{n\|n-p}$
1. $\mathbf{G}=Calculate\mathbf{G}(l)$
2. $p=-\frac{N-2}{2}$
3. $\mathbf{V}=Calculate\mathbf{V}(p,l,N)$
4. $\mathbf{B}=(\mathbf{V}\mathbf{V}^{\mathbf{T}})\mathbf{G}$
5. $\mathbf{W_l}(p)=\mathbf{V}^{\mathbf{T}}\mathbf{B}$
6. $\hat{x}_{n\|n-p}=\mathbf{W_l}(p)\mathbf{S_l}(p)$

The pseudo-code calculations of the basic UFIR smoothing algorithm are shown in TABLE II. The required inputs to the algorithm calculate the ECG signal estimate for the filter according to the formula described above.

3.2.3 Evaluation of Denoising Methods.

The Signal-to-Noise Ratio(SNR) improvement value is defined as the difference between the SNR of the signal after denoising and the SNR of the signal before denoising, indicating the denoising ability and denoising effect of the denoising method. Let SNR_{imp}, SNR_d, and SNR_n be the SNR boost value, the SNR after denoising, and the SNR before denoising, respectively.

$$\begin{cases} SNR_n=10\lg\dfrac{\sum_{n-1}^{N}s^2(n)}{\sum_{n-1}^{N}(s_n(n)-s(n))^2} \\ SNR_d=10\lg\dfrac{\sum_{n-1}^{N}s^2(n)}{\sum_{n-1}^{N}(s_d(n)-s(n))^2} \\ SNR_{imp}=SNR_d-SNR_n \end{cases} \quad (13)$$

where $s(n)$ is the source signal without noise; $s_n(n)$ is the measurement signal with noise; $s_d(n)$ is the signal after denoising.

The magnitude of the correlation coefficient can only reflect the similarity in the waveform between the denoised signal and the original signal. The size and position (interval) of the P wave, T wave, and QRS wave of the ECG signal are essential indicators of cardiac pathological judgment. To this end, the ratio of the average energy of the denoised signal to the source signal is used as an evaluation index for the change in the amplitude of the ECG signal after denoising, which energy ratio(ER) represents. When ER is equal to 1, it indicates that the amplitude of the denoised signal is consistent with the source signal; when ER is greater than 1, there is residual noise in the denoised signal; when ER is less than 1, there is a loss of useful information in the denoised signal. The formula for calculating the energy ratio ER is

$$\mathrm{ER}=\frac{\frac{1}{N}\sum_{n-1}^{N}{s_d}^2(n)}{\frac{1}{N}\sum_{n-1}^{N}s^2(n)} \quad (14)$$

The accurate ECG signals were derived from the records of the MIT-BIH ECG database, and ten records belonging to 10 subjects were randomly selected. Since these clinical data have been de-noised, the de-noised and corrected, real ECG signals can be superimposed as simulated signals. Ten recordings from 10 subjects were used as noise-free accurate ECG signals, and baseline drift, sEMG noise, white noise, and motion artifacts were superimposed to establish a noisy signal. Then apply p-UFIR filtering for denoising.

TABLE. III

CHANGE TABLE OF DENOISING SIGNAL INDEX OF SUPERIMPOSED MIXED NOISE

Original signal interval（dB）	Denoised signal SNR_{imp} mean	Clean signal SNR_{imp} mean	Similarity（%）	ER mean
0~5	10.309	10.817	95.3	0.902
5~10	9.380	9.551	98.2	0.932
10~15	8.223	8.289	99.2	0.958
15~20	5.532	5.538	99.9	0.985

The results of the variation table of denoised signal indicators with superimposed mixed noise are listed in

TABLE. III. In the signal-to-noise ratio range of 0dB~20dB, the denoising results of the accurate signal and the simulated signal are consistent. The baseline drift denoising effect is the best among the accurate ECG signal denoising results. The signal-to-noise ratio is improved by more than 10dB, and the correlation coefficient is more significant than 0.99, indicating that the similarity between the denoised signal and the source signal reaches 99%. The energy ratio is greater than 1, indicating that the denoising. The signal's energy is greater than that of the source signal, and the denoised signal does not lose valuable information. An energy ratio less than 1.017 indicates that the residual noise component in the denoised signal is no more than 1.7%. The denoising effect of sEMG noise is relatively poor, the improvement of the signal-to-noise ratio is relatively small, the correlation coefficient is more significant than 0.89, and the energy ratio is greater than 0.93. The white noise denoising results are better than the sEMG noise in improving the signal-to-noise ratio and the correlation coefficient. Compared with the clean signal, its average similarity rate is 98.15%, and the average energy ratio is 0.944.

The filtered signal in the previous section is subjected to *p*-UFIR smoothing and then zoomed in for observation. The waveforms of the signal before and after noise filtering and the filtered noise curve are shown in Fig. 6. The denoised signal can reflect the actual ECG characteristics, and the random noise in the time-domain sequence is eliminated and replaced by the smoothed ECG signal.

Meanwhile, we compared the ECG signal smoothing methods, such as median filtering, Kalman filtering, and Gaussian filtering, with the *p-shift* UFIR method and used the signal-to-noise ratio improvement value as the evaluation index. It was found that the signal-to-noise ratio improvement values of all three methods were below 10 dB, which was less than that of the *p-shift* UFIR method. Moreover, the median filtering will delay the signal and not remove the high-frequency noise. The low-pass filter will affect the morphological characteristics of the signal and may produce amplitude distortion. The Gaussian filter needs to choose the appropriate filter parameters. Otherwise, it will affect the filtering effect. Moreover, from Fig. 6, it can be seen that the p-shift method maintains the morphological characteristics of the signal well while smoothing the ECG.

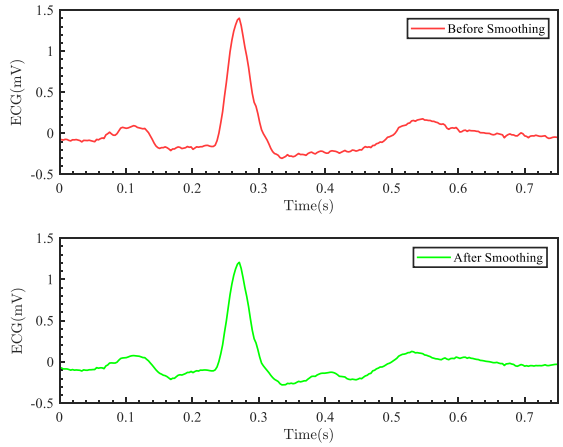

Fig. 6 *p*-UFIR denoising effect display

IV. INCEPTION-SIM MOTION PATTERN RECOGNITION MODEL.

4.1 GASF/GADF-HOG Feature Extraction.

Inspired by machine vision image processing technology, this paper selects sEMG and ECG as the source signals of motion intent, and the overall method is shown in Fig. 7. Convert raw sEMG and ECG to Grammy angle stack/difference field images. Then, the direction gradient(HOG) feature histogram of the corresponding GADF and GASF images was extracted, and the extracted features were named GASF-HOG and GADF-HOG, respectively.

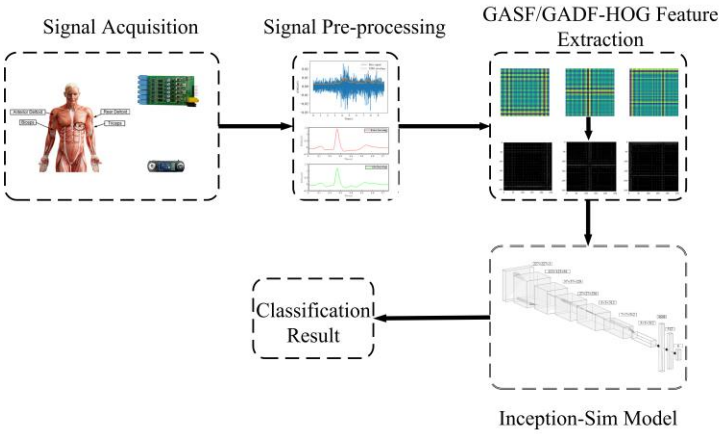

Fig. 7 The overall approach

4.1.1 GASF/GADF Principle.

GASF/GADF are methods for converting time series into images[26]. The GASF/GADF method has the advantages of handling multivariate time series data, retaining original information, handling nonlinear time series, high computational efficiency, and strong interpretation, which makes it widely used in time series data analysis. Rewrite the preprocessed sEMG to a polar representation

$$\begin{cases} \phi = \arccos(\tilde{x}_i) \\ r = t / n \end{cases} \tag{15}$$

where $\tilde{x}_i$ is the preprocessed sEMG signal; ϕ and r are the radians and radii in polar coordinates, respectively; t is the timestamp; n is a constant factor that adjusts the span of the polar coordinate system.

The following operators are defined

$$GASF = [\cos(\varphi_i + \varphi_j)] = x_i \bullet x_j - \sqrt{1 - x_i^{\,2}} \bullet \sqrt{1 - x_j^{\,2}} \tag{16}$$

$$GADF = [\sin(\varphi_i - \varphi_j)] = x_j \sqrt{1 - x_i^{\,2}} - x_i \sqrt{1 - x_j^{\,2}} \tag{17}$$

Then, the following matrix can be obtained

$$\mathrm{GASF\_M} = \begin{bmatrix} \cos(\varphi_1 + \varphi_1) & \cdots & \cos(\varphi_1 + \varphi_n) \\ \vdots & \ddots & \vdots \\ \cos(\varphi_n + \varphi_1) & \cdots & \cos(\varphi_n + \varphi_n) \end{bmatrix} \tag{18}$$

$$\mathrm{GADF\_M} = \begin{bmatrix} \sin(\varphi_1 + \varphi_1) & \cdots & \sin(\varphi_1 + \varphi_n) \\ \vdots & \ddots & \vdots \\ \sin(\varphi_n + \varphi_1) & \cdots & \sin(\varphi_n + \varphi_n) \end{bmatrix} \tag{19}$$

In addition, the GASF_M and GADF_M matrices are converted into images, and image features are extracted from them.

As derived from the above equations, the advantage of GASF/GADF methods is that they provide a way to preserve temporal dependence as time increases as the position moves from the top left to the bottom right. GASF/GADF includes temporal correlation because $G_{(i,j|\,|\,i-j|=k)}$ represents the correlation by superposition/difference relative to the time interval k. The main diagonal $G_{i,i}$ is a case when k = 0, containing the original value/angle information. From the main diagonal, we can reconstruct the time series from the high-level features learned by the deep neural network.

4.1.2 HOG Principle.

HOG feature is a feature descriptor used in computer vision and image processing for object detection, proposed by Deniz[27]. It remains well invariant to the geometric and optical properties of the image and is a commonly used feature to improve the robustness of image classification. Let the gradient of any pixel in the image be a vector

$$\nabla f(x, y) = \left[G_x G_y \right]^T = \left[\frac{\partial f}{\partial x} \frac{\partial f}{\partial y} \right]^T \tag{20}$$

where G_x is the gradient along the x-axis direction; G_y is the gradient along the y-axis direction; the magnitude and direction angle of the gradient can be expressed as

$$\begin{cases} \left|\nabla f(x, y)\right| = mag(\nabla f(x, y)) = (G_x^{\,2} + G_y^{\,2})^{1/2} \\ \phi(x, y) = \arctan(G_y / G_x) \end{cases} \tag{21}$$

The gradient of the pixels in the digital image is calculated by difference

$$\left|\nabla f(x, y)\right| = \sqrt{[f(x, y) - f(x+1, y)]^2 + [f(x, y) - f(x, y+1)]^2} \tag{22}$$

Calculate the gradient information of the image simply, quickly and efficiently

$$\begin{cases} G_x(x, y) = H(x+1, y) - H(x-1, y) \\ G_y(x, y) = H(x, y+1) - H(x, y-1) \end{cases} \tag{23}$$

where $G_x(x, y)$, $G_y(x, y)$, and $H(x, y)$ are the gradients of the pixel (x, y) in the horizontal and vertical directions and the gray value of the pixel, respectively.

The calculation of HOG features is based on the histogram of local gradient direction and size, and the gradient is calculated with good noise immunity. Even if there are some noise points in the image, the HOG features can effectively extract the shape and edge information of the object. Also, effective object detection and recognition can be performed on images of different sizes and rotations.

4.1.3 GASF/GADF-HOG Feature Fusion.

The corresponding GASF image of the 4-channel EMG drinking action(DW) is shown in Fig. 8. In order to show a clear channel combination image (the actual input does not contain a white dotted line), the white dotted line divides each channel, and the pixels of each image are p4 (256) × p4. The GASF or GADF image corresponding to the 4-channel EMG signal is stitched into a 4p4×p4 image, replacing the original single-channel image.

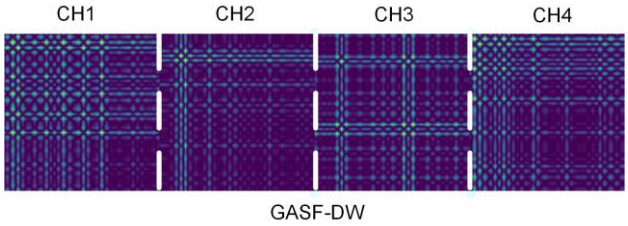

Fig. 8 4-channel EMG drinking water action (DW) corresponding GASF image (enlarged mode)

The GASF/GADF graph is the eccentric covariance matrix between features. The color of each grid in the graph represents the correlation between two features, which is used to measure the characteristics of each dimension and the relationship between each dimension. The diagonal elements provide the respective information of different feature maps, and the remaining elements provide related information between different feature maps, which can simultaneously reflect the feature information and the closeness between features.

After GASF and GADF conversion of bioelectrical signals, HOG feature extraction is performed on the corresponding images. Taking the GADF image as an example, the default cell is set to 8 × 8 pixels. Suppose a histogram of 8 bins is used to count the gradient information of these 6×6 pixels. The cell units are grouped into large blocks, and the normalized gradient histogram within the block makes the variation range of the gradient intensity tremendous due to the change in local illumination and the change of contrast. This requires normalization of the gradient intensity, which can further compress the illumination, shadows, and edges. The comparison diagram before and after feature extraction is shown in Fig. 9.

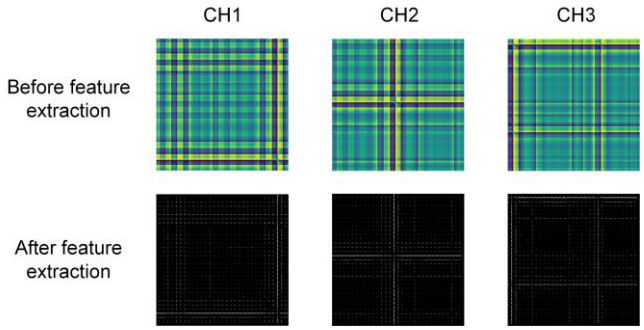

Fig. 9 Comparison diagram before HOG feature extraction (taking GADF as an example)

The comparison shows that the normalization of color and gamma reduces the brightness of the image and improves the texture intensity of the image and the proportion of the local surface exposure contribution. Due to the sliding property of the window and the sliding property of the block, the window and the block will have different degrees of overlap (determined by the step size). At this time, the cells divided in the block will appear multiple times, which means that the output of each cell unit acts on the final descriptor multiple times, which is of great help to the feature extraction of the classification model. To fit the model input, we reshape the final features.

4.2 Inception-Sim Model Design

The programming uses the Pytorch framework, training the model in Python 3.8, using Nvidia's CUDA parallel computing architecture for hardware acceleration, and using the C++ language API interface in the control software to connect to the actuator's control program.

The Inception-Sim proposed in this paper is based on the InceptionV3 model (which also conforms to the design criteria and structure of InceptionV3) and integrates the typical "tandem decomposition structure" and "asymmetric decomposition structure" to make the 3-layer Inception structure merged into a layer of Inception-Sim structure.

The model uses the Zero-phase Component Analysis (ZCA) whitening method to reduce redundant information, takes the eigenvalue overlay map as input, and then sends it to the convolutional layers of 5×5, 3×3 convolution kernels to extract the fluctuation trend information of low-frequency eigenvalues. DropBlock[28] layers can be used to model noise and improve generalization. The traditional dropout regularization technique is not used in the proposed network because for convolutional layers, the adjacent location elements of the feature map share semantics in the spatial patch region. And the structured DropBlock performs better in the convolutional network, and its parameter calculation formula is

$$\gamma = \frac{1-kp}{bs^2} \frac{fs^2}{(fs-bs+1)^2} \tag{24}$$

where kp is the probability of one unit to keep the traditional dropout state (0 to 0.95, the final value is 0.9); fs is the size of the feature map; bs is the block size. In the model parameters, 7

is substituted for bs , 0.9 is substituted for kp , 4 is substituted for fs , and the calculated value of γ is 3.373×10^{-3}.

The topological relationship of Inception-Simflow is viewed from left to right. The first sub-column and the second sub-column keep the original InceptionV3 structure, the third sub-column represents the "tandem decomposition structure" module, and the fourth sub-column represents the "asymmetric decomposition structure" module, as shown in Fig. 10. After the first two convolutional layers, the advantage of the Inception structure is used to reduce the number of parameters. Based on the InceptionV3 convolutional network structure, the Inception-Sim structure is proposed for a small amount of classification (only six classifications are required in the full study).

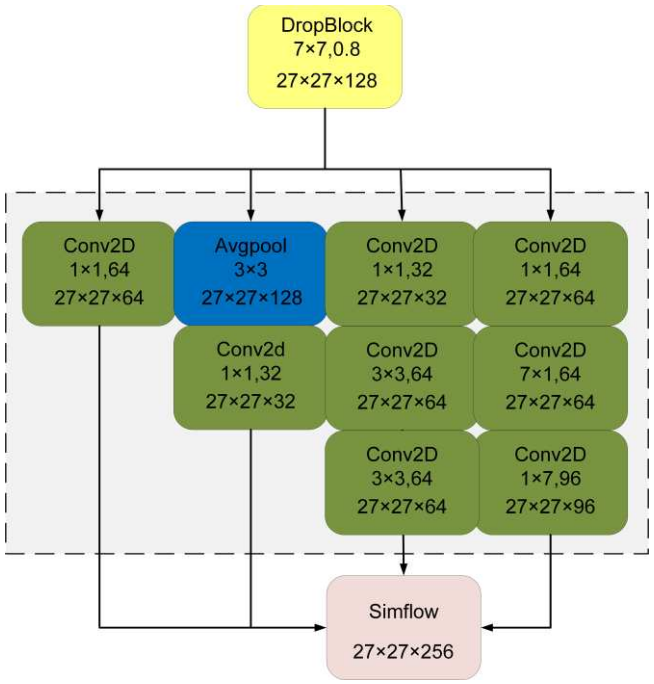

Fig. 10 Inception-Simflow layer topology

Compared with the traditional Inception-V3[29] module, i.e. “one 5×5 convolution replaced by two 3×3 convolutions” module (concatenated decomposition structure). The Inception-Simdeep layer also adds “N×1 and 1×N” modules (asymmetric decomposition structure) to reduce network overfitting and speed up the network. The specific acceleration fusion module is shown in Fig. 11.

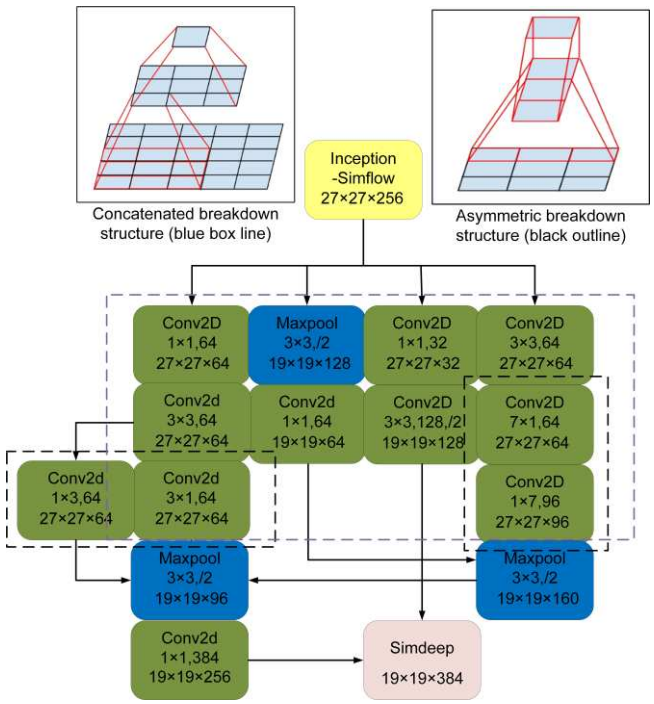

Fig. 11 Inception-Simdeep layer topology

TABLE. IV

INCEPTION-SIM NETWORK PARAMETERS OF MOTION PATTERN RECOGNITION MODEL

Layers	Output layer dimension	Convolution kernel dimension	Parameter
Feature-in	227×227×3	N/A	N/A
Conv1	113×113×64	5×5	Leaky_relu
Maxpool1	57×57×64	2×2	Stride=2
Conv2	57×57×128	3×3	Leaky_relu
Maxpool2	27×27×128	2×2	Stride=2
DropBlock	27×27×128	7	γ =3.373e-3
Inception-Simflow	27×27×256		Relu
Inception-Simdeep	19×19×384	1×1/3×3 /7×1/1×7	Relu
Inception-Simdeep	7×7×512		Relu
AvgPool	3×3×512	2×2	Stride=2
DropBlock	3×3×512	1	γ =3.373e-3
FC1	1×1×4608	N/A	Dropout=0.9
FC2	1×1×512	N/A	Dropout=0.7
Softmax	1×1×6	N/A	N/A

The overall framework of the model shown is still based on the InceptionV3 framework, and the acceleration is achieved through

the cooperation of the Inception-Sim layer and the DropBlock layer. The specific parameters of the network are shown in

TABLE. IV. The pooling layer is divided into Maxpool and Avgpool. First, use Maxpool in the shallow layer of the network model to improve the model operation speed, and then use Avgpool after passing through the Inception-Sim layer. Using the pattern of "Inception-Sim followed by Avgpool" is beneficial to maintain feature stability[30], but its computational complexity is increased by 3. Finally, six motion intensity patterns are output through the fully connected layer Softmax, and the three-dimensional network structure visualization image is shown in Fig. 12.

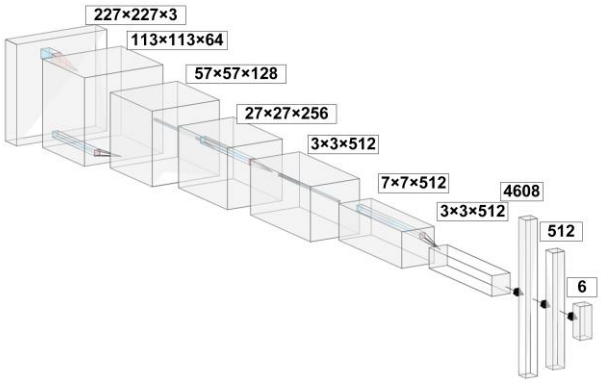

Fig. 12 Inception-Sim network structure of motion pattern recognition model

V. Model Performance Evaluation

5.1 Baseline and Evaluation Metrics

In order to verify the performance of the proposed model in individual motion pattern recognition, multiple models were selected for comparison.

MLP-CNN[31]: This model includes two algorithms that are integrated in a concise and efficient manner using a rule-based decision fusion approach. ECG-CNN[32]: Deep 2D CNN for ECG arrhythmia classification using Xavier initialization and exponential linear units. VGGNET[33]: This model uses very small (3×3) convolutional filters to fully evaluate networks with a depth of 16~19 weight layers. 34-layer CNN[34]: This network consists of 33 layers of convolution followed by a fully connected layer and a Softmax. In addition to the above networks, comparisons are also made with widely used image classification frameworks such as VGG16, Resnet50[35], EfficientNet[36], DenseNet[37], MobileNet V2[38], and Inception-V3[39].

Precision, Recall, average iteration speed per round (ACR), and F1-score are used as network performance evaluation indicators, which are expressed as

$$\begin{aligned} \text{Precision} &= \frac{TP}{TP+FP} \\ Recall &= \frac{TP}{TP+FN} \\ F1 &= 2\times\frac{\text{Precision}\times Recall}{\text{Precision}+Recall} \\ ACR &= \frac{1}{epoch\ time} \end{aligned} \tag{25}$$

where *epoch time* is the average time for the loss function value to iterate an epoch stability; *TP*, *FP*, *TN*, and *FN* are true positives, false positives, true negatives, and false negatives, respectively.

5.2 Time window size

The time window length determines the size of the GASF or GADF image. The size of GADF and GASF images affects the final classification accuracy and computation time. Therefore, the classification accuracy results corresponding to different sizes are given here, as shown in Fig. 13. As mentioned above, the GADF or GASF images of 4 channels are combined into a composite image with an image size of 4px*px. In Fig. 13, the px values corresponding to P1~P8 are P1 (32), P2 (64), P3 (128), P4 (256), P5 (384), P6 (512), P7 (640), and P8 (768). According to the results shown in Fig. 13, P8 corresponds to the highest classification accuracy. However, the corresponding computation time is also the highest. It can be seen that the recognition accuracy increases slowly with the increase of the image size. Compared to P4 (highest accuracy of 95.23%), the accuracy of P8 is not significantly improved (highest accuracy of 95.51%). Therefore, considering accuracy and computation time, P4 was chosen as the best image size.

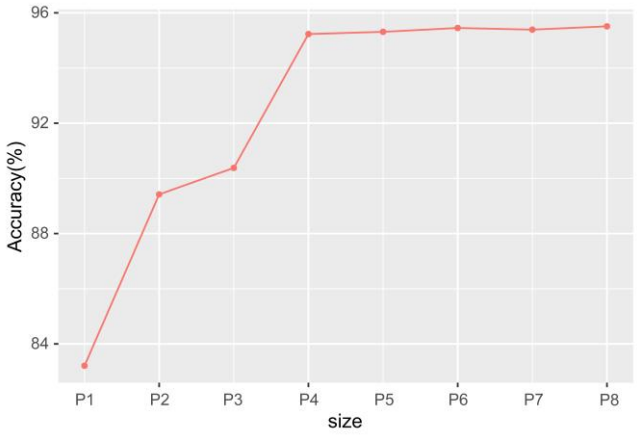

Fig. 13 Corresponding accuracy for various image sizes

5.3 Model Evaluation

TABLE. V

COMPARISON OF THE PERFORMANCE OF THE CONTROL MODEL

Control model	DW Precision /Recall (%)	EQ Precision /Recall (%)	Overall Precision (%)	Overall Recall (%)	F1 (%)	ACR（1/s）
MLP-CNN	93.22/92.68	75.65/72.45	82.45	78.52	80.43	4.724e-2
ECG-CNN	**96.78/95.21**	87.85/67.32	84.25	75.41	79.58	5.584e-2
VGGNET	92.42/92.89	**97.52**/83.54	**96.24**	56.28	69.61	6.426e-2
34-layer CNN	89.52/72.45	80.12/67.23	81.24	68.42	74.28	7.546e-2
VGG16	91.83/74.23	85.23/62.47	83.45	64.23	72.58	5.542e-2
Resnet50	87.45/85.23	91.23/78.25	89.15	83.54	86.25	7.678e-2
Inception-V3	92.24/91.24	92.54/78.52	90.41	81.23	85.57	6.852e-2
EfficientNet	88.34/87.68	85.98/78.59	83.24	76.59	79.78	8.369e-2
DenseNet	92.65/90.43	95.42/82.74	87.19	80.28	83.59	6.879e-2
MobileNet V2	81.54/74.84	79.38/68.21	80.61	69.45	74.62	8.847e-2
Inception-Sim	96.65/91.15	94.57/**86.52**	92.76	**88.63**	**92.79**	**9.132e-2**

As shown in TABLE V, we compare the performance of our proposed model with the control model by calculating the average Precision, Recall, F1, and ACR for all actions on the validation set following five-fold cross-validation. Also, we do not provide the precision rate and recall for all actions to conserve space and use the DW and EQ actions as examples. The table's bolded section displays the best outcomes based on the available metrics.

Our proposed method achieves the best results in most metrics, although it only achieves the best in some. Our proposed method's Overall Recall was 88.63%, which is a 6.09% improvement compared to the second place Resnet50 of 83.54%. The Recall metric represents the degree of accuracy of the predicted negative sample. In the upper limb rehabilitation exercise process, it is essential to ensure that the patient's intention is not misjudged in the upper limb movement pattern recognition to guarantee the safety of the rehabilitation process, and our high Recall can guarantee this. The F1 score of our proposed method is 92.79%, which is an 8.44% improvement compared to the second place Inception-V3's 85.57%. In particular, the ACR of our proposed method is 9.132e-2, which is an improvement of about 21% in the iteration speed compared to models with similar performance.

Moreover, compared to the second place MobileNet V2 with 8.847e-2, the improvement is 3.22%, but our proposed method is higher than MobileNet V2 by more than 10% in all other performance metrics compared to MobileNet V2. This can verify the purpose of our proposed method, which is to improve the network's speed while ensuring accuracy. However, our proposed model does not achieve the best overall accuracy rate, probably because the ECG signal includes not only motion intensity information but also some other information, and the fusion of sEMG signal and ECG signal for actions with different motion intensities but similar trajectories can reduce the misclassification rate, which is one of the reasons for our high Recall. However, for some actions, the information redundancy caused by the fusion will, on the contrary, illuminate the decrease of accuracy for positive samples, thus leading to the accuracy rate not reaching the best result. In contrast, stacking multiple 3×3 convolutions in VGGNET can extract features better and facilitate the information transfer, thus achieving the best Precision. This problem we plan to solve it in two aspects in our future work; one is to improve our Inception-Sim model with reference to the VGGNET network structure, rather, the ECG signal extraction features about motion intensity, which can be extracted by self-encoder or by extracting traditional time-frequency domain features, etc.

In order to observe the results more intuitively, the model performance evaluation is transformed into a radar chart through data visualization, as shown in Fig. 14. The lines of our proposed method in Fig.14 have been bolded.

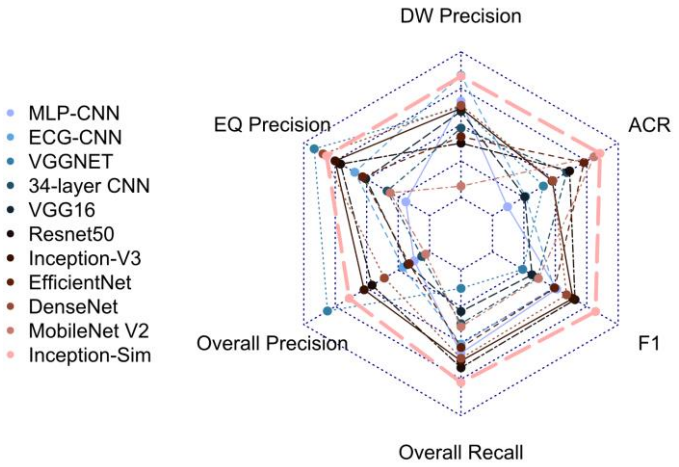

Fig. 14 Performance evaluation radar diagram

Overall, the poor performance of time series-based classification methods (ECG-CNN and 34-layer CNN) can be observed in both datasets. Notably, general image classification frameworks (VGG16, inception-v3, Resnet50) perform better than these time series-based methods. This situation is in line with the original intention of converting time-series signals into two-dimensional images; that is, the application of time series data in image classification frameworks will be limited. When the proposed method is compared with these traditional image classification frameworks, precision, recall, and F1 score metrics remain high. However, the most prominent performance indicator is the average iteration speed, about 21% faster than the average speed of mainstream methods.

As shown in Fig. 12, we can see that the average iteration time has a particular relationship with the performance parameters of the model because the higher the complexity of the model, the longer the training time spent, but the more robust the model fitting ability and our proposed method uses "tandem decomposition structure" and "asymmetric The proposed method uses "tandem decomposition structure" and "asymmetric decomposition structure" to provide good data response and transmission speed for motion pattern recognition, which saves a lot of recognition time while guaranteeing accuracy and ensures real-time upper limb motion pattern recognition in the rehabilitation process.

VI. Conclusion

This paper presents a multimode fusion method of sEMG and ECG signals based on GASF-HOG. Moreover, the effectiveness of this method in the recognition of human upper limb movement patterns in rehabilitation exercise is verified, which can improve the accuracy and robustness of recognition. In addition, an upper limb motion pattern classification model named Inception-Sim is proposed based on multi-mode information fusion method. With fewer electrodes, the "tandem decomposition structure" and "asymmetric decomposition structure" are fully utilized to achieve consistent stability and high performance in classification tasks. In order to adapt to the deployment of the single chip computer, it maintains few parameters and faster reasoning speed, which is helpful to realize the control of the upper limb exoskeleton rehabilitation robot.

The limitation of this work is that this method is not applicable to amputees with prosthetics, and it will be applied to rehabilitation of stroke patients in the future. At the same time, the redundant information contained in the ECG signal may cause the model overfit and reduce the recognition accuracy of some actions. In future work, we will look into signal processing techniques such as wavelet decomposition and other techniques to lessen information redundancy. In future research, we intend to look more closely at how the quantity of sensors affects the precision of recognition. In addition, we plan to explore the effect of additional measurement factors, such as sampling frequency and time window length, on the results in future work. To address the issue that high-intensity muscle activity influences other physiological parameters, the extended model considers using more advanced filtering algorithms and a multi-scene training strategy to reduce the effects of high-intensity muscle activity[11].

In conclusion, this paper proposes a new multi-mode information fusion method for upper limb motion pattern recognition. The proposed model can be used to improve the quality of life in patients with hemiplegia, upper limb impairment, and others. In the future, this work will be applied to the rehabilitation exercise of actual patients and wearable exoskeleton control.

Acknowledgment

This work is supported by the Science and technology plan project of Xi'an city, Grant no. 21XJZZ0079, and the Natural Science Foundation of Shaanxi Province (Grant No. 2020JM-131 and 2020KW-058).

Ethics approval and consent to participate

The data acquisition protocol was approved by Medical and Experimental Animal Ethics Committee of Northwestern Polytechnical University, Xi'an, China (approbation number: 6101030222595-202001001).

Reference

[1] L. Hao, Z. Zhao, X. Li, M. Liu, H. Yang, and Y. Sun, "A safe human–robot interactive control structure with human arm movement detection for an upper-limb wearable robot used during lifting tasks," *Int. J. Adv. Robot. Syst.*, vol. 17, no. 5, p. 1729881420937570, Sep. 2020, doi: 10.1177/1729881420937570.

[2] Z. Li, Z. Huang, W. He, and C.-Y. Su, "Adaptive Impedance Control for an Upper Limb Robotic Exoskeleton Using Biological Signals," *IEEE Trans. Ind. Electron.*, vol. 64, no. 2, pp. 1664–1674, Feb. 2017, doi: 10.1109/TIE.2016.2538741.

[3] W. Wang, J. Zhang, X. Wang, X. Yuan, and P. Zhang, "Motion intensity modeling and trajectory control of upper limb rehabilitation exoskeleton robot based on multi-modal information," *Complex Intell. Syst.*, vol. 8, no. 3, pp. 2091–2103, Jun. 2022, doi: 10.1007/s40747-021-00632-2.

[4] L. Clemmensen, T. Hastie, D. Witten, and B. Ersbøll, "Sparse Discriminant Analysis," *Technometrics*, vol. 53, no. 4, pp. 406–413, Nov. 2011, doi: 10.1198/TECH.2011.08118.

[5] U. Côté-Allard *et al.*, "Deep Learning for Electromyographic Hand Gesture Signal Classification Using Transfer Learning," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 27, no. 4, pp. 760–771, Apr. 2019, doi: 10.1109/TNSRE.2019.2896269.

[6] V. Mondéjar-Guerra, J. Novo, J. Rouco, M. G. Penedo, and M. Ortega, "Heartbeat classification fusing temporal and morphological information of ECGs via ensemble of classifiers," *Biomed. Signal Process. Control*, vol. 47, pp. 41–48, Jan. 2019, doi: 10.1016/j.bspc.2018.08.007.

[7] A. Phinyomark, P. Phukpattaranont, and C. Limsakul, "Feature reduction and selection for EMG signal classification," *Expert Syst. Appl.*, vol. 39, no. 8, pp. 7420–7431, Jun. 2012, doi: 10.1016/j.eswa.2012.01.102.

[8] C. Shen, Z. Pei, W. Chen, J. Wang, J. Zhang, and Z. Chen, "Toward Generalization of sEMG-Based Pattern

Recognition: A Novel Feature Extraction for Gesture Recognition," *IEEE Trans. Instrum. Meas.*, vol. 71, pp. 1–12, 2022, doi: 10.1109/TIM.2022.3141163.

[9] P. Sri-Iesaranusorn *et al.*, "Classification of 41 Hand and Wrist Movements via Surface Electromyogram Using Deep Neural Network," *Front. Bioeng. Biotechnol.*, vol. 9, p. 548357, 2021, doi: 10.3389/fbioe.2021.548357.

[10] A. Ameri, M. A. Akhaee, E. Scheme, and K. Englehart, "A Deep Transfer Learning Approach to Reducing the Effect of Electrode Shift in EMG Pattern Recognition-Based Control," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 28, no. 2, pp. 370–379, Feb. 2020, doi: 10.1109/TNSRE.2019.2962189.

[11] O. W. Samuel *et al.*, "Intelligent EMG Pattern Recognition Control Method for Upper-Limb Multifunctional Prostheses: Advances, Current Challenges, and Future Prospects," *IEEE Access*, vol. 7, pp. 10150–10165, 2019, doi: 10.1109/ACCESS.2019.2891350.

[12] O. W. Samuel *et al.*, "Resolving the adverse impact of mobility on myoelectric pattern recognition in upper-limb multifunctional prostheses," *Comput. Biol. Med.*, vol. 90, pp. 76–87, Nov. 2017, doi: 10.1016/j.compbiomed.2017.09.013.

[13] L. Wu, X. Zhang, K. Wang, X. Chen, and X. Chen, "Improved High-Density Myoelectric Pattern Recognition Control Against Electrode Shift Using Data Augmentation and Dilated Convolutional Neural Network," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 28, no. 12, pp. 2637–2646, Dec. 2020, doi: 10.1109/TNSRE.2020.3030931.

[14] A. H. Al-Timemy, R. N. Khushaba, G. Bugmann, and J. Escudero, "Improving the Performance Against Force Variation of EMG Controlled Multifunctional Upper-Limb Prostheses for Transradial Amputees," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 24, no. 6, pp. 650–661, Jun. 2016, doi: 10.1109/TNSRE.2015.2445634.

[15] A. Del Vecchio, A. Holobar, D. Falla, F. Felici, R. M. Enoka, and D. Farina, "Tutorial: Analysis of motor unit discharge characteristics from high-density surface EMG signals," *J. Electromyogr. Kinesiol.*, vol. 53, p. 102426, Aug. 2020, doi: 10.1016/j.jelekin.2020.102426.

[16] K. J. Hunt and J. Saengsuwan, "Changes in heart rate variability with respect to exercise intensity and time during treadmill running," *Biomed. Eng. OnLine*, vol. 17, no. 1, p. 128, Sep. 2018, doi: 10.1186/s12938-018-0561-x.

[17] N. A. Jamnick, R. W. Pettitt, C. Granata, D. B. Pyne, and D. J. Bishop, "An Examination and Critique of Current Methods to Determine Exercise Intensity," *Sports Med.*, vol. 50, no. 10, pp. 1729–1756, Oct. 2020, doi: 10.1007/s40279-020-01322-8.

[18] J. Sacha and W. Pluta, "Different methods of heart rate variability analysis reveal different correlations of heart rate variability spectrum with average heart rate," *J. Electrocardiol.*, vol. 38, no. 1, pp. 47–53, Jan. 2005, doi: 10.1016/j.jelectrocard.2004.09.015.

[19] B. Wang and H. Zhu, "The Recognition Method of Athlete Exercise Intensity Based on ECG and PCG," *Comput. Math. Methods Med.*, vol. 2022, p. e5741787, May 2022, doi: 10.1155/2022/5741787.

[20] V. Monasterio, F. Burgess, and G. D. Clifford, "Robust classification of neonatal apnoea-related desaturations," *Physiol. Meas.*, vol. 33, no. 9, p. 1503, Aug. 2012, doi: 10.1088/0967-3334/33/9/1503.

[21] C. Orphanidou, "Quality Assessment for the Photoplethysmogram (PPG)," in *Signal Quality Assessment in Physiological Monitoring: State of the Art and Practical Considerations*, C. Orphanidou, Ed., in SpringerBriefs in Bioengineering. Cham: Springer International Publishing, 2018, pp. 41–63. doi: 10.1007/978-3-319-68415-4_3.

[22] B. Afsharipour, S. Soedirdjo, and R. Merletti, "Two-dimensional surface EMG: The effects of electrode size, interelectrode distance and image truncation," *Biomed. Signal Process. Control*, vol. 49, pp. 298–307, Mar. 2019, doi: 10.1016/j.bspc.2018.12.001.

[23] W. Wang, H. Li, D. Kong, M. Xiao, and P. Zhang, "A novel fatigue detection method for rehabilitation training of upper limb exoskeleton robot using multi-information fusion," *Int. J. Adv. Robot. Syst.*, vol. 17, no. 6, p. 1729881420974295, Nov. 2020, doi: 10.1177/1729881420974295.

[24] S. Zhao, Y. S. Shmaliy, and F. Liu, "Batch Optimal FIR Smoothing: Increasing State Informativity in Nonwhite Measurement Noise Environments," *IEEE Trans. Ind. Inform.*, pp. 1–9, 2022, doi: 10.1109/TII.2022.3193879.

[25] L. J. Morales-Mendoza, H. Gamboa-Rosales, and Y. S. Shmaliy, "A new class of discrete orthogonal polynomials for blind fitting of finite data," *Signal Process.*, vol. 93, no. 7, pp. 1785–1793, Jul. 2013, doi: 10.1016/j.sigpro.2013.01.023.

[26] Z. Wang and T. Oates, "Imaging Time-Series to Improve Classification and Imputation," in *Proceedings of the Twenty-Fourth International Joint Conference on Artificial Intelligence (ijcai)*, Q. Yang and M. Wooldridge, Eds., Freiburg: Ijcai-Int Joint Conf Artif Intell, 2015, pp. 3939–3945.

[27] O. Déniz, G. Bueno, J. Salido, and F. De la Torre, "Face recognition using Histograms of Oriented Gradients," *Pattern Recognit. Lett.*, vol. 32, no. 12, pp. 1598–1603, Sep. 2011, doi: 10.1016/j.patrec.2011.01.004.

[28] J. Wang, F. Gao, J. Dong, and Q. Du, "Adaptive DropBlock-Enhanced Generative Adversarial Networks for Hyperspectral Image Classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 6, pp. 5040–5053, Jun. 2021, doi: 10.1109/TGRS.2020.3015843.

[29] X. Xia, C. Xu, and B. Nan, "Inception-v3 for flower classification," in *2017 2nd International Conference on Image, Vision and Computing (ICIVC)*, Jun. 2017, pp. 783–787. doi: 10.1109/ICIVC.2017.7984661.

[30] Z. Liu, C. Yang, J. Huang, S. Liu, Y. Zhuo, and X. Lu, "Deep learning framework based on integration of S-Mask R-CNN and Inception-v3 for ultrasound image-aided diagnosis of prostate cancer," *Future Gener. Comput. Syst.*, vol. 114, pp. 358–367, Jan. 2021, doi: 10.1016/j.future.2020.08.015.

[31] H. Yang *et al.*, "Denoising of 3D MR Images Using a Voxel-Wise Hybrid Residual MLP-CNN Model

to Improve Small Lesion Diagnostic Confidence," in *Medical Image Computing and Computer Assisted Intervention – MICCAI 2022*, L. Wang, Q. Dou, P. T. Fletcher, S. Speidel, and S. Li, Eds., in Lecture Notes in Computer Science. Cham: Springer Nature Switzerland, 2022, pp. 292–302. doi: 10.1007/978-3-031-16437-8_28.

[32] M. Hammad, S. Zhang, and K. Wang, "A novel two-dimensional ECG feature extraction and classification algorithm based on convolution neural network for human authentication," *Future Gener. Comput. Syst.*, vol. 101, pp. 180–196, Dec. 2019, doi: 10.1016/j.future.2019.06.008.

[33] H. Salem, K. R. Negm, M. Y. Shams, and O. M. Elzeki, "Recognition of Ocular Disease Based Optimized VGG-Net Models," in *Medical Informatics and Bioimaging Using Artificial Intelligence : Challenges, Issues, Innovations and Recent Developments*, A. E. Hassanien, R. Bhatnagar, V. Snášel, and M. Yasin Shams, Eds., in Studies in Computational Intelligence. Cham: Springer International Publishing, 2022, pp. 93–111. doi: 10.1007/978-3-030-91103-4_6.

[34] R. Avram *et al.*, "Predicting diabetes from photoplethysmography using deep learning," *J. Am. Coll. Cardiol.*, vol. 73, no. 9_Supplement_2, pp. 16–16, Mar. 2019, doi: 10.1016/S0735-1097(19)33778-7.

[35] D. Theckedath and R. R. Sedamkar, "Detecting Affect States Using VGG16, ResNet50 and SE-ResNet50 Networks," *SN Comput. Sci.*, vol. 1, no. 2, p. 79, Mar. 2020, doi: 10.1007/s42979-020-0114-9.

[36] B. Koonce, "EfficientNet," in *Convolutional Neural Networks with Swift for Tensorflow: Image Recognition and Dataset Categorization*, B. Koonce, Ed., Berkeley, CA: Apress, 2021, pp. 109–123. doi: 10.1007/978-1-4842-6168-2_10.

[37] Y. Zhu and S. Newsam, "DenseNet for dense flow," in *2017 IEEE International Conference on Image Processing (ICIP)*, Sep. 2017, pp. 790–794. doi: 10.1109/ICIP.2017.8296389.

[38] M. Sandler, A. Howard, M. Zhu, A. Zhmoginov, and L.-C. Chen, "MobileNetV2: Inverted Residuals and Linear Bottlenecks," presented at the Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018, pp. 4510–4520. Accessed: Apr. 07, 2023. [Online]. Available: https://openaccess.thecvf.com/content_cvpr_2018/html/Sandler_MobileNetV2_Inverted_Residuals_CVPR_2018_paper.html

[39] M. A. S. Al Husaini, M. H. Habaebi, T. S. Gunawan, M. R. Islam, E. A. A. Elsheikh, and F. M. Suliman, "Thermal-based early breast cancer detection using inception V3, inception V4 and modified inception MV4," *Neural Comput. Appl.*, vol. 34, no. 1, pp. 333–348, Jan. 2022, doi: 10.1007/s00521-021-06372-1.