WMNN: Wearables-based Multi-column Neural Network for Human Activity Recognition

Chenyu Tang, Xuhang Chen, Jing Gong, Luigi G. Occhipinti, *Senior Member, IEEE*, and Shuo Gao, *Member, IEEE*

***Abstract*— In recent years, human activity recognition (HAR) technologies in e-health have triggered broad interest. In literature, mainstream works focus on the body's spatial information (i.e. postures) which lacks the interpretation of key bioinformatics associated with movements, limiting the use in applications requiring comprehensively evaluating motion tasks' correctness. To address the issue, in this article, a Wearables-based Multi-column Neural Network (WMNN) for HAR based on multi-sensor fusion and deep learning is presented. Here, the Tai Chi Eight Methods were utilized as an example as in which both postures and muscle activity strengths are significant. The research work was validated by recruiting 14 subjects in total, and we experimentally show 96.9% and 92.5% accuracy for training and testing, for a total of 144 postures and corresponding muscle activities. The method is then provided with a human-machine interface (HMI), which returns users with motion suggestions (i.e. postures and muscle strength). The report demonstrates that the proposed HAR technique can enhance users' self-training efficiency, potentially promoting the development of the HAR area.**

***Index Terms*— HAR, wearables, multi-sensor fusion, multi-column neural network, deep learning**

I. Introduction

Human activity recognition (HAR), which monitors human activities via smart sensor technology, has recently become an important task in diverse fields, such as health monitoring, rehabilitation supervision, and home-centered fitness assistance [1-8]. HAR can be implemented by fixed equipment and wearables. The former often utilizes cameras, infrared sensors, and radio frequency identification (RFID) [9-15]; and the latter takes advantage of inertial measurement unit (IMU), plantar pressure sensors, and surface electromyogram (sEMG), etc. [16-25]. Between them, the fixed equipment enjoys the benefit of decent accuracy, while lacking scenario flexibility, i.e. HAR can be merely carried out in a certain room. In contrast, wearables offer users omnipresent utilization conditions, however, the location accuracy is sometimes unsatisfying.

Plenty of HAR applications have been reported. For example, in 2019, Gochoo *et al*. [13] developed an IoT-based yoga posture recognition system based on infrared sensors and achieved an average accuracy of 99.45% in classifying 26 yoga postures. In 2021, Gupta *et al*. [17] proposed a YogaHelp system based on IMUs to classify 12 yoga steps with an accuracy of 83.2%-94.5% and give feedback on the improper motion speed. Existing techniques can strongly support tasks that concentrate on the correctness of posture's spatial status. However, in many cases, the same posture can correspond to different muscle strengths, indicating that current works can hardly analyze users' performances in sports such as Yoga and Tai Chi, in which incorrect muscle strength can result in undesired training effects, even though the posture is reasonable.

To address this issue, in this article, we develop a HAR framework (conceptually shown in Fig. 1), the core of which is a Wearables-based Multi-column Neural Network (WMNN) (demonstrated by Fig. 4), and take the practice of Tai Chi as an example. Here, IMU, plantar stress sensors, and sEMG are employed for retrieving users' bioinformatics. In this work, the term Bioinformatics is used to identify a generic set of physiological data related to a human body, either visible or invisible, such as the actual movements and the muscle activity that preceded or controls the same movement. The output of the presented network is the classification of the standard motion and 17 common mistakes of each Tai Chi Method, which can indicate whether the motion is correct or not, and how to adjust the motion. Signals from the sensors are merged and processed by a deep learning algorithm for determining 144 (18×8) postures and core muscle strength. Eventually, the recognition accuracy of the two metrics can achieve 96.9% and 92.5%, respectively, for the 14 subjects. Compared to the state-of-the-art studies, the proposed method considers broader motion-related features, e.g. center of gravity and muscle status, and refines the evaluation at a higher level, showcasing its prospect in supporting more scenarios.

II. Related Work

In this section, we will review the state-of-the-art work of HAR from the perspectives of fixed equipment and wearables.

A. HAR based on fixed equipment

There are three prevailing types of non-wearable devices based HAR: RGB camera or depth camera-based, infrared sensor-based, and RFID) based. The relevant work will be reviewed in this subsection.

Zerrouki *et al*. [9] developed an adaptive boosting algorithm to analyze six classes of activities and compared them with the state-of-the-art machine learning classifiers. In the work, they employed the RGB-based motion data from two public databases (i.e. URFD and UMAFD) and yielded the accuracy of 96.56% and 93.91% respectively. Maddala *et al*. [10] proposed a joint angular displacement map (JADM) and combined it with a single-stream deep convolutional neural network (CNN) model to recognize yoga asanas. In their work, they created the network with two RGB-D (Red, Green, Blue, and Depth) based 3-D yoga posture datasets (i.e. HDM05 and CMU), achieving the accuracy of 89.15% and 88.67% respectively. The model was tested by the data generated using a nine-camera mocap system from ten subjects. Similarly, Liu *et al*. [11] developed a human posture recognition system using RGB-D color images alone. They utilized Kinect V2 and collected 13800 samples of six subjects with 15 motion postures. The color images were put into proposed generative adversarial networks (GAN) to generate estimated depth

This work was supported in part by the National Natural Science Foundation under grant 61827802 and grant 61803017, and in part by the Beihang University under grant KG12090401 and grant ZG216S19C8 and in part by the Beihang University under First-class Undergraduate Course (Digital Image Processing) Construction Project.

Chenyu Tang is with the School of Instrumentation Science and Optoelectronic Engineering, Beihang University, Beijing 100191, China, and Cambridge Graphene Centre, Department of Engineering, University of Cambridge, 9 JJ Thomson Avenue, Cambridge CB3 0FA, UK. Xuhang Chen is with the School of Instrumentation Science and Optoelectronic Engineering, Beihang University, Beijing 100191, China, and the Electrical and Electronic Engineering Department, Imperial College London, South Kensington, London SW7 2AZ. Jing Gong is with the School of Instrumentation Science and Optoelectronic Engineering, Beihang University, Beijing 100191, China. Luigi G. Occhipinti is with Cambridge Graphene Centre, Department of Engineering, University of Cambridge, 9 JJ Thomson Avenue, Cambridge CB3 0FA, UK. Shuo Gao is with the School of Instrumentation Science and Optoelectronic Engineering, Beihang University, Beijing 100191, China (e-mail: shuo_gao@buaa.edu.cn).

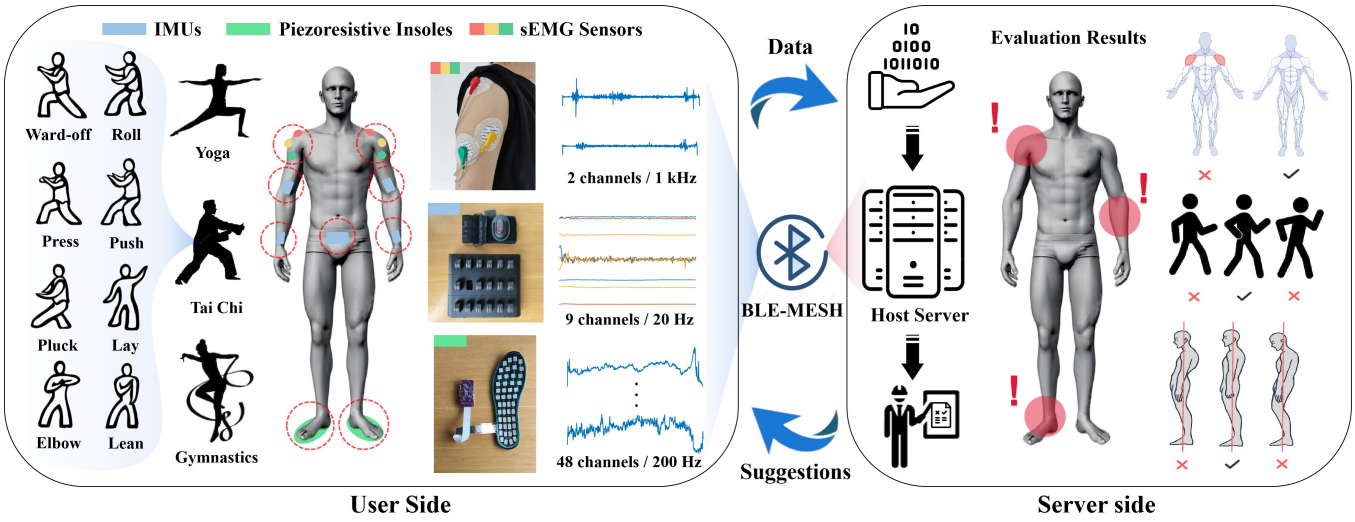


Fig. 1: The proposed HAR framework, which is showcased in home-centered wearable cloud fitness scenario.

images. After feeding the color images and depth images into a developed discriminator network, an accuracy of 96.7% was obtained. Although the above RGB(-D) based posture recognition systems can achieve high accuracy, their operation is contingent on the detection area equipped with cameras, and they underlie the risk of privacy invasion.

Mashiyama *et al*. [12] put forward an activity recognition method using a low-resolution infrared array sensor. In their study, they collected the infrared data of five activities (No event, Stopping, Walking, Sitting, and Falling) and obtained an average accuracy of 94.6% by support vector machine (SVM). Gochoo *et al*. [13] proposed an IoT-based yoga posture recognition system employing a low-resolution infrared sensor-based wireless sensor network (WSN). The authors collected 93200 posture images of 26 yoga postures from 18 volunteers and trained a deep convolutional neural network (DCNN), achieving an average accuracy of 99.45%. Though the aforementioned infrared-based posture recognition systems can effectively solve the privacy problem, they also depend on the relatively fixed detection area. Besides, the systems are easily disturbed by environmental noise such as heat source, light source, and RF radiation.

Yao *et al*. [14] proposed an RFID-based posture recognition system. The writers employed nine RFID tags and collected the motion data of 12 orientation-sensitive postures from three subjects. After training the Dirichlet Process Gaussian Mix- ture Model (DPGMM) based Hidden Markov Model (HMM), they yielded a classification accuracy of 99%. Ding *et al*. [15] designed an automatic, non-invasive and light-weight Free-weight Exercise MOnitoring (FEMO) system to recognize the exercising postures. The authors recruited 15 volunteers and collected data from ten activities. When adopting multiple antennas, the recognition accuracy of the system can maintain 90% with multiple users. Despite the high recognition accuracy achieved by the RFID systems mentioned above, the systems suffer from the complexity of system assembly.

Although non-wearable devices based HAR systems can obtain a decent recognition effect, they have the inherent defect of being limited by the site. For scenarios that highlight the arbitrariness of application place, they are outperformed by the wearable systems.

B. HAR based on wearables

In this subsection, the three most prevailing means of wearables-based HAR: IMUs based, sEMG based, and plantar pressure sensors based, as well as their fusion are reviewed.

Xu *et al*. [16] proposed a Tai Chi action recognition algorithm, using node trajectory features. In the study, the authors collected data from 26 independent actions with 17 IMU sensors and processed the multiple axes data via DTW (Dynamic Time Warping) algorithm. They chose ten actions and information from five sensors to train the SVM model, yielding a ten-classification accuracy of 90.45%. Gupta *et al*. [17] developed a YogaHelp system based on IMU sensors to help amateurs for learning the correct execution of yoga without any supervision of a trainer. The writers collected the training data from eight professional yoga trainers with four sensor units. The designed deep neural network obtained an accuracy of 83.2%-94.5% for 12 yoga steps and efficiently give feedback on whether the motion speed is improper.

Ryu *et al*. [18] presented a lower limb human motion detection system using a sEMG with a TAS (Top And Slope) feature extraction algorithm. In their research, motion data from ten subjects were collected via MP150, a commercial sEMG device. The TAS features were fed into the LDA (Linear Discriminate Analysis) classifier and the accuracy of 91.4% and 94.26% were achieved in the tasks of four gait subphases detection and five locomotion modes detection respectively. Li *et al*. [19] designed an sEMG-based lower arm gesture recognition system, proposing a SEAR (Shifts Estimation and Adaptive coRrection) method to calibrate the electrode shifts' disturbance introduced to the system (not existed in all types of sEMG devices). The authors arbitrarily chose the positions of Myo armband and collected eight gestures from ten subjects. The raw data were processed by the SEAR method and SVM, and an average accuracy of 79.32% was obtained.

Wang *et al*. [20] presented a wearable plantar pressure measurement system for locomotion recognition. In their work, four force sensors were integrated in each smart shoe and six modes of locomotion data were collected from a below-knee amputee subject. The LDA algorithm was employed for classification and an average accuracy of 98.92% was obtained. Jeong *et al*. [21] proposed a method for classifying ambulatory activities using eight plantar pressure sensors. They collected 210 samples of three ambulatory activities with three participants. By using SVM to classify the walking activities, an accuracy of 95.2% was achieved.

Haque *et al*. [22] presented a real-time classification method of ground-level walking and stair climbing. The walking and stair data of two volunteers were collected by two IMUs and two FSRs (Force

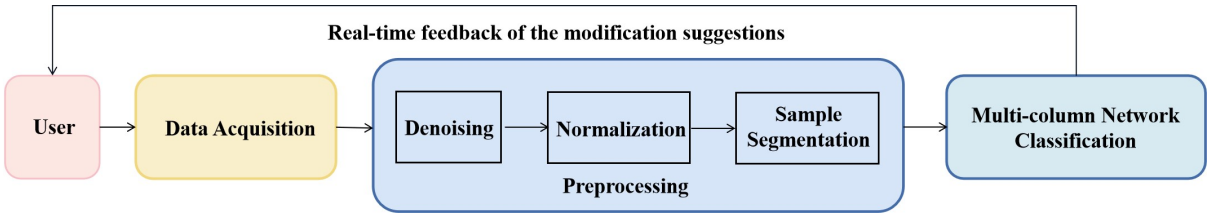


Fig. 2: The block diagram of the proposed system.

Sensing Resistor) and tested in real-time. By using the LDA classifier, the accuracy of 96.50% and 87.21% were obtained in training and testing respectively. Luo *et al*. [23] developed an end-to-end gait sub-phase recognition system based on sEMG and plantar pressure sensors. By utilizing four-channel sEMG sensors and two-channel plantar pressure sensors, three male subjects' locomotion data of four gait sub-phases were collected. The combination of LSTM (Long and Short Term Memory) and MLP (Multilayer Perceptron) was proposed to classify the sub-phases and an average accuracy of 90.79% was achieved. Chang *et al*. [24] proposed a hierarchical hand motions recognition method based on one IMU and two sEMG sensors. The motion data of six motion gestures were collected from ten subjects. By training SVM for sEMG signals and decision-tree classifier for IMU signals before fusing the classification decision, the overall average accuracy of 95.6% was yielded.

C. Issues to be addressed

Though fruitful research has been conducted in the field of activity recognition, researchers are inclined to focus on the recognition of visible postures such as limb, gait, hand gesture, etc. For the training of yoga, gymnastics, Tai Chi, and other sports that lay emphasis on the whole body coordination, apart from visible postures, recognition of the invisible bioinformatics associated with muscle activity is imperative. For whole-body coordinated sports, visible body movements can be easily corrected, while invisible body micro-movements and muscle force errors are often difficult to detect but have a significant impact on the effect of the sports. The system framework proposed in this study can evaluate sports in a comprehensive way, which contains not only the general limb postures but also the deviation of the body's center of gravity and the force of specific muscles. This framework can help users perform sports to a standard even when the coach is absent.

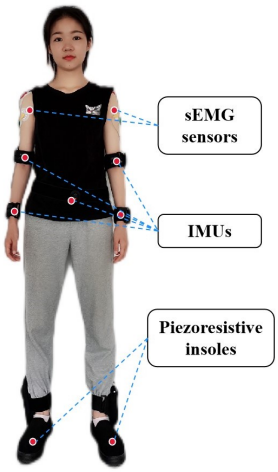
III. Methodology

Fig. 2 shows the system's block diagram, and in this section, we will explain the details of the system.

A. Hardware description

For the purpose of evaluating the fitness activities combining body movements and the muscular forces behind them, the developed system contains five IMUs, two piezoresistive arrays, and two sEMG sensors mainly to analyze limb motions, the center of gravity, and muscle forces, respectively (Fig. 3). Five IMUs are necessary for the analysis of users' hands motion. Different users hold different body features, hence the 3 IMUs deployed at the waist and upper arms serve as the references for the 2 IMUs located in the lower arms, to self-calibrate the system for differences in the human body topology. Though the centre of gravity can also be analyzed by more IMUs fixed on legs, we introduce plantar pressure insoles to reduce the obstacles caused by excessive IMU wearing to the user's movement. As for the use of sEMG, although muscle force can be detected by pressure sensors, their detection of muscle force is indirect, thereby vulnerable to factors such as the tightness of binding with the human body.

The IMUs in this system consist of a 3-axis accelerometer, 3-axis gyroscope, and 3-axis magnetometer (Perception Neuron 3, Noitom Co., Ltd), which can be mounted on limb body segments and the waist. The piezoresistive arrays in this system contain 48 distributed channels each and the whole piezoresistive insole device is a 5-layer structure, whose detailed structure is showcased in our previous paper [26]. The sEMG sensors in this system (Bipolar sEMG, Sichiray Co., Ltd) consist of three Ag/AgCl electrodes each (a bipolar electrodes pair and a reference electrode), which are utilized to measure the EMG signal of deltoid in this work. The sampling rates of the sensors are demonstrated in Fig. 1, and all of the devices are synchronized through the Network Time Protocol (NTP).



Sensor	Parameter	Value
IMU	Accelerometer Accuracy	±0.0005 g
	Gyroscope Accuracy	±0.061 °/s
	Magnetometer Accuracy	±0.0667 Gauss
Piezoresistive Array	Response Time	15 ms
	Measurement Accuracy	±2‰
	Resistive Range of Interest	4.7 kΩ - 47 kΩ
sEMG	Response Time	20 ms
	Input Impedance	10 kΩ
	Magnification	500

Fig. 3: The full set of hardware equipment wearing schematic and the specific parameters of the sensors.

B. Network development

On balance, the multi-sensor data in a system tend to maintain coupled and complementary information (e.g. in this system, although the plantar piezoresistive arrays are designed as sensors to analyze the center of gravity of the human body, IMUs also contain certain such information). Besides, in the tasks of multi-sensor data fusion, data from various types of sensors often display diverse characteristics, so the effect of data layer fusion is usually outperformed by feature layer fusion. Therefore, we developed the WMNN based on feature layer fusion and CNN-based feature extractor (Fig. 4). In a specific posture, the data of IMUs and piezoresistive arrays can be regarded as static data, while sEMG sensing data needs to be analyzed in the form of sequence due to the inherent properties of EMG signal. We modified the AlexNet structure [27] and applied it to the WMNN to extract

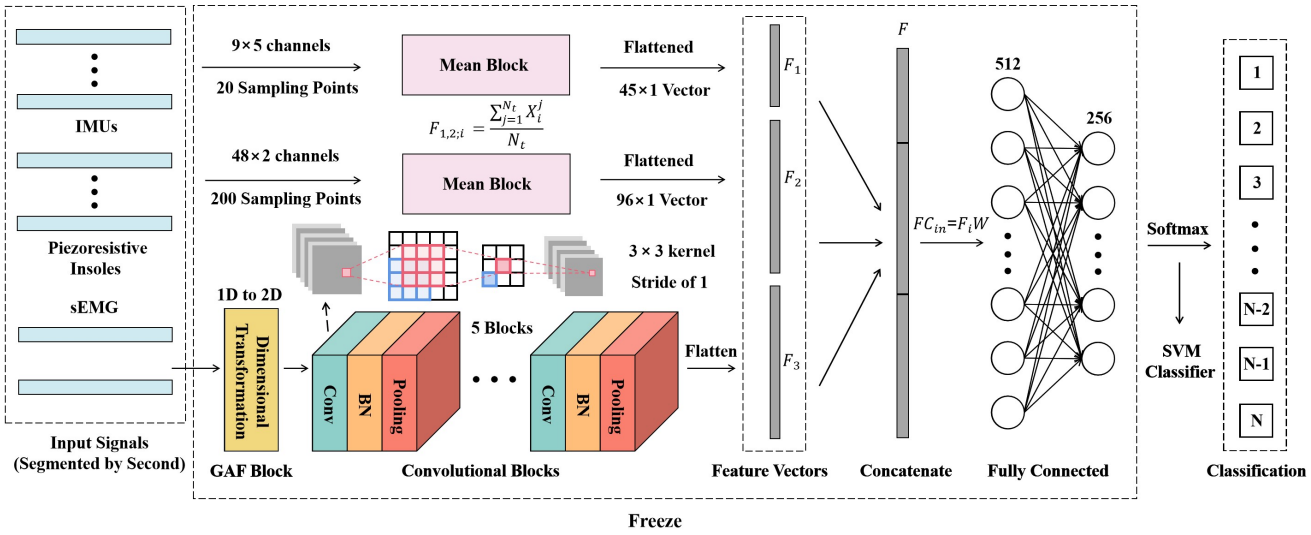


Fig. 4: The architecture of the proposed WMNN.

the features of sEMG signals, and used simple mean processing to extract the features of static IMUs and piezoresistive arrays signals.

1) GAF Block: As CNN is a two-dimensional inputs based neural network (though one-dimensional convolution exists, its ability of information extraction is overwhelmed by two-dimensional convolution) while sEMG signals are one-dimensional, it is necessary to encode the time series of the signals to images. In this research, we utilized Gramian Angular Field (GAF) proposed by Wang *et al*. [28] to generate sEMG images. In the GAF block, the sEMG time series are rescaled in the interval [-1,1] before being transformed into the polar coordinate system. The rescaled series is denoted as: $X = \{x_1, x_2, \cdots, x_n\}$. If the time stamp of each point is denoted as t_i, and the series span is N, the signal series in polar coordinate can be obtained by:

$$\begin{cases} \theta_i = arccos(x_i), -1 \le x_i \le 1, x_i \in X \\ r_i = \frac{t_i}{N}, t_i \in \mathbb{N} \end{cases} \tag{1}$$

θ_i and r_i are points' polar angle and polar radius respectively. Then the GAF images are yielded by considering the trigonometric sum between each point. The Gramian matrix G is defined as:

$$G = \begin{bmatrix} cos(\theta_1 + \theta_1) & \cdots & cos(\theta_1 + \theta_n) \\ cos(\theta_2 + \theta_1) & \cdots & cos(\theta_2 + \theta_n) \\ \vdots & \ddots & \vdots \\ cos(\theta_n + \theta_1) & \cdots & cos(\theta_n + \theta_n) \end{bmatrix} \tag{2}$$

By introducing the GAF method, sEMG signals can be encoded into images without losing the temporal dependency. When passing through GAF transformation, the time series with length N will be converted into $N \times N$ images. To avoid the influence of images' oversizing on the subsequent convolution efficiency, the dimension of the sEMG signals is reduced by segmented averaging at the beginning of the GAF block. The time window for averaging is chosen as 5, so the output dimension of the GAF block is $200 \times 200 \times 2$, where 2 denotes the two channels of the sEMG sensors' inputs (i.e. left and right deltoid muscles).

Although it will increase the complexity of the operation, the advantages of GAF transform will cover this defect: first, the GAF block retains the signals' temporal dependency when transforming the time series into two-dimensional images. The main diagonal of the G matrix contains all the information of the time series. At the same time, the generation of other elements of the matrix is essentially a kernel trick different from convolution, which can increase the dimensionality of features. In addition, converting a one-dimensional sequence into a two-dimensional image can enable many state-of-the-art algorithms in the field of computer vision to be used. Therefore, using the GAF block can obtain a more stable performance than using raw time series.

TABLE I: Tensors' dimension in convolutional blocks

Block	Layer	Output Tensor	Block	Layer	Output Tensor
Block1	Conv1	$32 \times 200 \times 200$	Block3	BN3	$256 \times 50 \times 50$
	BN1	$32 \times 200 \times 200$		Pool3	$256 \times 25 \times 25$
	Pool1	$32 \times 100 \times 100$	Block4	Conv6	$512 \times 25 \times 25$
Block2	Conv2	$64 \times 100 \times 100$		Conv7	$512 \times 25 \times 25$
	BN2	$64 \times 100 \times 100$		BN4	$512 \times 25 \times 25$
	Pool2	$64 \times 50 \times 50$		Pool4	$512 \times 12 \times 12$
Block3	Conv3	$128 \times 50 \times 50$	Block5	Conv8	$512 \times 12 \times 12$
	Conv4	$256 \times 50 \times 50$		BN5	$512 \times 12 \times 12$
	Conv5	$256 \times 50 \times 50$		Pool5	$512 \times 6 \times 6$

2) Mean Block: In a specific posture, signals from IMUs and piezoresistive arrays can be regarded as static information (i.e. within a time window of a certain length, the signal value in a channel can be considered to be ideally constant). Therefore, in the mean block, the mean value of the signal in each sampling time window is calculated according to the channel and taken as the features of this posture. To flatten the signals, the time window for the Mean Block is the same as the sampling points in a sample segment, so the time window N_t for IMU data and Piezoresistive data are 20 and 200 respectively (as the sampling frequencies are 20 Hz and 200 Hz respectively).

3) Convolutional Block: In each convolutional block, the input feature maps go through the convolution layer, batch normalization (BN) layer, and max-pooling layer. For convolution layers, 3×3 filters with a stride of 1 are tested as the overall optimum hyperparameters in this study. BN layers are set to accelerate the training by reducing internal covariate shift [29]. The BN layer is placed before the activation function, BN operation is denoted as:

$$BN_{\gamma,\beta}(x_i) = \gamma \frac{x_i - \mu_B}{\sqrt{\sigma_B^2 + \varepsilon}} + \beta \tag{3}$$

Where μ_B and σ_B are mini-batch mean and mini-batch variance

respectively, γ and β are learnable parameters: the scale and shift, and ε is a minute constant introduced for calculating stability.

Max pooling layers reduce the dimensions of features and facilitate feature extraction. The non-linear ReLU function is the activation function of this network. After five consecutive convolutional blocks, two fully connected layers are employed before the final classifier. Since the introduction of BN layers, the dropout operation in AlexNet to prevent overfitting is not performed in fully connected layers. The transformation of the tensor's dimension in the convolutional blocks is demonstrated in Table I.

4) Freeze and Classifier: Since the dataset of this study is relatively small, considering that using softmax as the classifier often generates overfitting in the training of small datasets, we learned from the strategy of R-CNN algorithm [30]: after training the convolutional layers as the feature extractor, we substituted the softmax layer with SVM to obtain the activity recognition results. Specifically, after completing the first step of training (using softmax as the classifier), we froze all the hyperparameters and parameters before the softmax layer and fed the extracted features to SVM for classifier training [31-33]. In this step, to obtain the optimal SVM models under different Tai Chi Methods, we changed the kernel function, which is used to map the eigenvector to Hilbert space, in the model. For datasets in feature space $T = \{(x_1, y_1), (x_2, y_2), ..., (x_m, y_m)\}$, the optimization function of SVM can be expressed as:

$$\begin{aligned} &\min_{\alpha} \frac{1}{2} \sum_{i=1}^{m} \sum_{j=1}^{m} \alpha_i \alpha_j y_i y_j K(x_i)^T K(x_j) - \sum i = 1^m \alpha_i \\ &\text{s.t.} \sum_{i=1}^{m} \alpha_i y_i = 0, 0 \le \alpha_i \le C, i = 1, 2, 3, ..., m \end{aligned} \tag{4}$$

Where α is the support vector coefficient and C is the penalty term of the relaxation coefficient. For the FC layer between features and the SVM classifier, although decent accuracy is yielded by directly sending features into SVM, the time complexity is high. FC layer is used here to reduce the dimension of features, thereby improving the time complexity.

C. Experimental details

1) Data Acquisition: In this study, a total of 14 subjects were selected for sports data collection, including ten subjects as the training set and four subjects as the test set. The subjects were 18-38 years old, with different heights (ranging from 157cm-185cm) and body shape, and had no history of musculoskeletal diseases. The relevant information on the subjects is shown in Table II. The study was exempt from IRB approval by Ethics Review Committee of Beijing University of Traditional Chinese Medicine.

TABLE II: Relevant information on subjects

Subject	Gender	Age	BMI	Subject	Gender	Age	BMI
A1	F	18	21.2	A8	M	25	26.3
A2	F	23	18.4	A9	M	33	29.4
A3	F	24	24.8	A10	M	37	31.2
A4	M	21	28.3	B1	F	21	27.4
A5	M	22	23.6	B2	F	24	26.2
A6	M	22	23.7	B3	M	23	22.4
A7	M	23	19.4	B4	M	38	29.1

[1] A and B represent training group and test group respectively.
[2] BMI (Body Mass Index) is the weight divided by the square of height (kg/m^2).

As mentioned in the introduction section, we chose the Eight Methods of Tai Chi and their common mistakes as the postures to be recognized in this study. Through the communication with Tai Chi experts, we selected 17 wrong postures for each method (including the forward and backward of the human body's center of gravity, the high and low of the main hand, and the excessive force of the deltoid and the combination of these errors). In addition, we set the sensor devices on the subjects as shown in Fig. 1 according to the advice of experts. With the help and supervision of the same experts, we collected the data of 14 subjects' standard movements and 17 wrong postures of the Eight Methods. For each sub-posture, we collected data for about three minutes for each subject. After the signal was segmented based on second with 50% overlap (i.e. one segment per second, and for every two seconds, three segments are yielded), 320-400 segments were obtained for each posture. In total, the training set contained 65000 samples, and the test set contained 25000 samples.

2) Labeling protocol: Each Method has 18 classification categories: three postures for the main hand multiplied by three postures for the center of gravity multiplied by two statuses for the deltoid force. To accurately label the postures, we asked three Tai Chi experts to aid the labeling work. Because the offset of the center of gravity and the force exerted by the muscles are difficult to evaluate by observation, three experts offered the results of posture evaluation by touching the subjects. By synthesizing the evaluation results of the three experts, we annotated the corresponding labels.

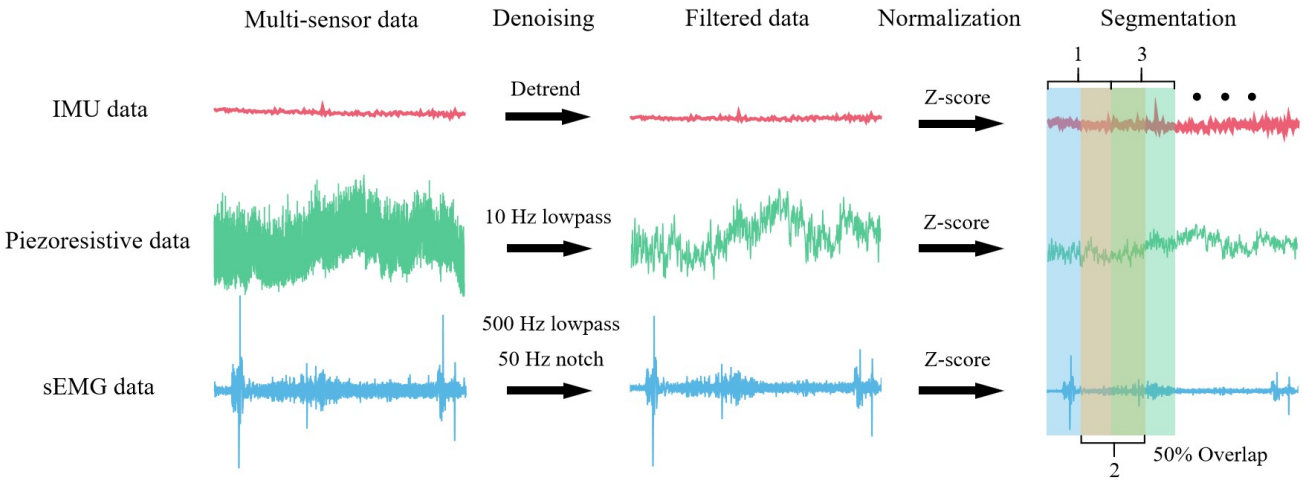


Fig. 5: The flow of the signals in preprocessing step.

3) Preprocessing: Where needed, the aim of this step is to remove the spurious components (e.g. noise, trends) from, and to normalize the signals to improve the classification accuracy of the network. After preliminary analysis of the signals, we found out that signals generated by commercial IMUs with integrated digital filtering do not need additional preprocessing apart from detrending, while signals from piezoresistive arrays and sEMG need to be denoised. For piezoresistive output signals, we first removed the DC component by subtracting the signals with the average value of a blank group (without pressure). Then, a Butterworth low-pass filter with a cut-off frequency of 10 Hz and an order of 4 was used for denoising. For sEMG signals, we firstly low-pass filtered them to extract the passband with Butterworth filter using a cut-off frequency of 500 Hz and an order of 4. Then, a digital IIR (Infinite Impulse Response) notch filter was utilized to remove the 50Hz power-line interference. The system function of the second-order notch filter is:

$$H(z) = \frac{1 - (2cos\omega_0)z^{-1} + z^{-2}}{1 - (2rcos\omega_0)z^{-1} + r^2z^{-2}} \tag{5}$$

Where $w_0 = 2\pi f_0/f_s$ is the notch digital frequency, f_0 is the stopband center frequency, f_s is the sampling frequency, and r is the notch constant, which determines the stopband range and attenuation characteristics.

As data from diverse subjects may present an overall variance originating from various body characteristics, the z-score normalization was conducted on each sample between the denoising and the WMNN. Fig. 5 illustrates the flow of signals during preprocessing (one channel was selected as a sample in each modal). After the preprocessing steps, the fine data are fed into the WMNN as the inputs.

4) Model Training and Testing: Due to the relatively small amount of data in the dataset of this study, we applied the method of cross-validation to train and validate the model [34]. After building the network, we fed the data from ten subjects in group A to the network for ten-fold cross-validation training: after obtaining the optimal structure of the sEMG convolutional feature extractor with softmax as the classifier, we locked the network's hyperparameters and parameters. Then we changed the classifier to the three most prevailing classifiers in the field of motion recognition: LDA [35-37], KNN [38-40], and SVM. The classifiers were also trained with ten-fold cross-validation to acquire the optimal hyperparameters and parameters in recognition within each Method of Tai Chi. To evaluate the generalization of the model, we utilized the network trained by data from group A to recognize the postures of subjects from group B and compared the results with the standard labels. For each Method, we trained a set of parameters for the network to classify 18 classification categories (i.e. one standard posture and 17 common wrong postures).

For the first step of training (feature extractor training), the Adam optimizer [41] was used for 50 epochs, the loss function is cross-entropy loss, the batch size is 100, and the learning rate is 0.001. For the second step of training (classifier training), the polynomial kernel, gaussian kernel, and sigmoid kernel were tested, and the gaussian kernel showed the best performance in this task. The learning process was boosted by Apple Metal Performance Shaders (MPS).

IV. Experimental Results

In this section, the results of the five experiments conducted in this study are demonstrated. Besides, Table III compares the results of this study with relevant work in detail.

A. Experiment I: The frozen part training

In order to obtain the optimal microstructure of the feature extractor, we compared different dimensions of the convolution and changed different hyperparameters to train the original network (without changing the classifier but only using softmax) and achieved the optimal hyperparameters under eight Methods respectively. According to the results, the overall optimal combination of hyperparameters is 3×3 convolutional kernel size, the stride of 1, and the max-pooling mode. The data from group A were utilized in this experiment for cross-validation training. The specific results of the training are shown in table IV and V.

TABLE III: Comparison with state-of-the-art work on HAR

Ref.	Task	Hardware	Algorithms	Datasets	Accuracy
[10]	Recognizing 42 yoga poses in 10 orientations	One nine-camera mocap system	A joint angular displacement map(JADM) and convolutional neural network (CNN)	Training set: HDM05 and CMU (public) Test set: obtained from 10 subjects	Training: 89.15% and 88.67% (two public datasets) Test: 94.43%-98.93% (different subjects)
[13]	Recognizing 26 yoga postures	WSN contains three low-resolution infrared sensors	Deep convolutional neural network (DCNN)	Collection of 93200 posture images from 18 volunteers	An average of 99.45%
[15]	Recognizing 10 fitness motions	Four RFID tags and four antennas	KL divergence based segmentation and Doppler profile matching	Data from 15 volunteers	Approximately 90%
[16]	Classifying 10 actions of Tai Chi	Five IMU sensors	Dynamic time warping (DTW) and support vector machine (SVM)	Twenty sets of 26 independent actions with 17 IMUs	An average of 90.45%
[17]	Recognizing 12 yoga steps and give feedback on the improper motion speed	Four IMU sensors	Deep neural network	Data from 8 professional yoga trainers	83.2%-94.4% (different subjects)
[18]	Detection of four gait subphases and five locomotion modes	A commercial sEMG device (MP150)	Top and slope (TAS) and linear discriminate analysis (LDA)	Data from 10 subjects	91.4% and 94.26%
[23]	Classifying four gait subphases	Four-channel sEMG sensors and two-channel plantar pressure sensors	Long and short term memory (LSTM) and multilayer perceptron (MLP)	Data from 3 male subjects	An average of 90.79%
[24]	Recognizing six hand gestures	One IMU and two sEMG sensors	SVM and decision-tree	Data from 10 subjects	An average of 95.6%
[This work]	**Evaluating the body movements and muscle force for Tai Chi Eight Methods training**	**Five IMU sensors, two bipolar sEMG sensors, and two piezoresistive insoles**	**A CNN feature extractor and SVM (WMNN)**	**Training set: data from 10 subjects Test set: data from 4 subjects**	**Training: overall 96.9% Test: overall 92.5%**

[1] Three colors respectively indicate categories of non-wearable based, uni-sensor wearable based, and multi-sensor wearable based.

TABLE IV: The overall accuracy of the original network under different dimensions of convolution

	Ward-off	Roll	Press	Push	Pluck	Lay	Elbow	Lean
1D	93.3%	91.4%	92.8%	92.4%	91.9%	90.1%	91.1%	92.3%
2D (GAF)	95.1%	96.7%	95.8%	95.2%	94.1%	92.9%	92.6%	93.4%

TABLE V: The overall accuracy of the original network under different hyperparameters

Hyperparameters \ Methods		Ward-off	Roll	Press	Push	Pluck	Lay	Elbow	Lean
Kernel Size	**2×2**	**95.1%**	94.3%	93.7%	95.1%	**94.1%**	89.7%	92.3%	90.8%
	3×3	93.9%	**96.7%**	**95.8%**	**95.2%**	93.3%	**92.9%**	**92.6%**	**93.4%**
	5×5	85.6%	89.2%	88.3%	91.2%	90.3%	91.1%	83.9%	92.4%
Stride	**1**	**95.1%**	**96.7%**	**95.8%**	**95.2%**	**94.1%**	92.2%	**92.6%**	**93.4%**
	2	92.1%	95.5%	94.9%	93.6%	92.2%	**92.9%**	90.5%	93.2%
	3	94.2%	95.9%	92.4%	91.4%	91.7%	91.8%	89.1%	92.3%
Pooling	**Mean Pooling**	92.3%	94.5%	94.9%	92.5%	94.0%	89.6%	90.2%	**93.5%**
	Max Pooling	**95.1%**	**96.7%**	**95.8%**	**95.2%**	**94.1%**	**92.9%**	**92.6%**	93.4%
	No Pooling	91.1%	93.8%	95.2%	94.4%	91.2%	90.8%	91.7%	92.4%

[1] Bolded represents the optimal hyperparameters under the corresponding Method.

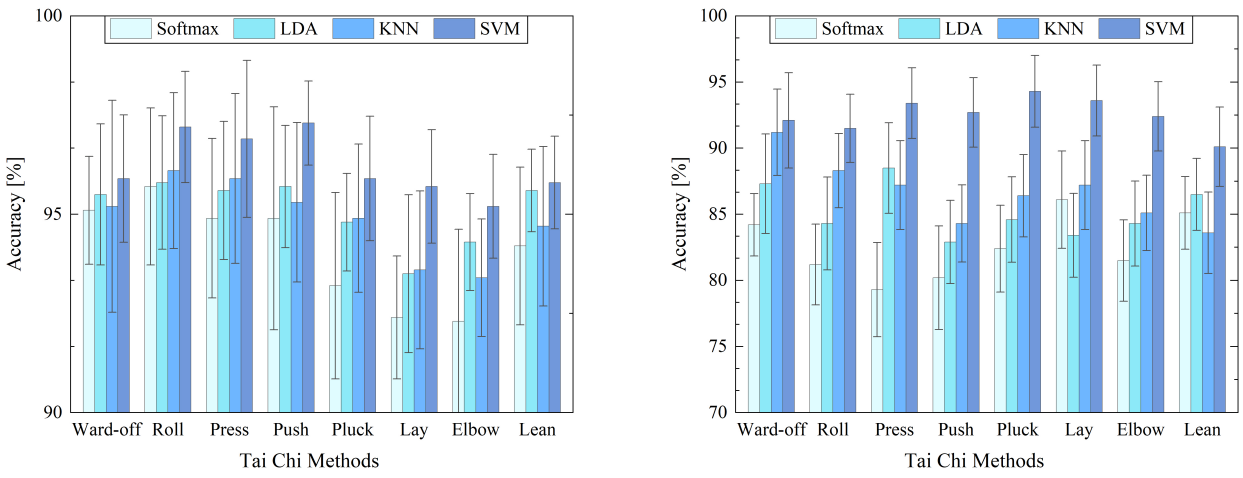


Fig. 6: The recognition accuracy under Eight Methods under different classifiers (error bars indicate standard deviations among subjects): (a) training set (group A) results and (b) test set (group B) results.

B. Experiment II: The classifiers training and generalization optimization

Table VI compares diverse state-of-the-art baseline methods on the dataset and validates the optimal of the proposed method in this task. When no feature extractor is introduced, the results on the training set are decent while a sharp decrease is witnessed in the test set. This is because the classifier only tries to separate the vectors in the hyperspace, and lacks the mining of essential features. Thus, the feature extractor based on deep learning is necessary to extract more deep features to enhance the generalization of the method to new individuals. Although the original network using softmax as the classifier performs well in cross-validation, the recognition accuracy of the model reduces greatly when the test set data (i.e. data from group B) that the model has never learned is introduced. In order to optimize the generalization of the model, we compared the performance of several classifiers frequently-used in activity recognition. As shown in Fig. 6 (a) and Fig. 6 (b), SVM outperforms other classifiers in both the training set and test set. In the training set, the performance of various classifiers is relatively consistent, and they all achieve decent training accuracy. Specifically, in the test set, SVM maintains high classification accuracy while other classifiers, especially softmax, suffer from poor generalization. It can be seen that the system has achieved good accuracy in both the learned training set and the unlearned test set, and is competent for the task of comprehensively analyzing body movements and the muscular forces behind them. The classification confusion matrix of the Ward-

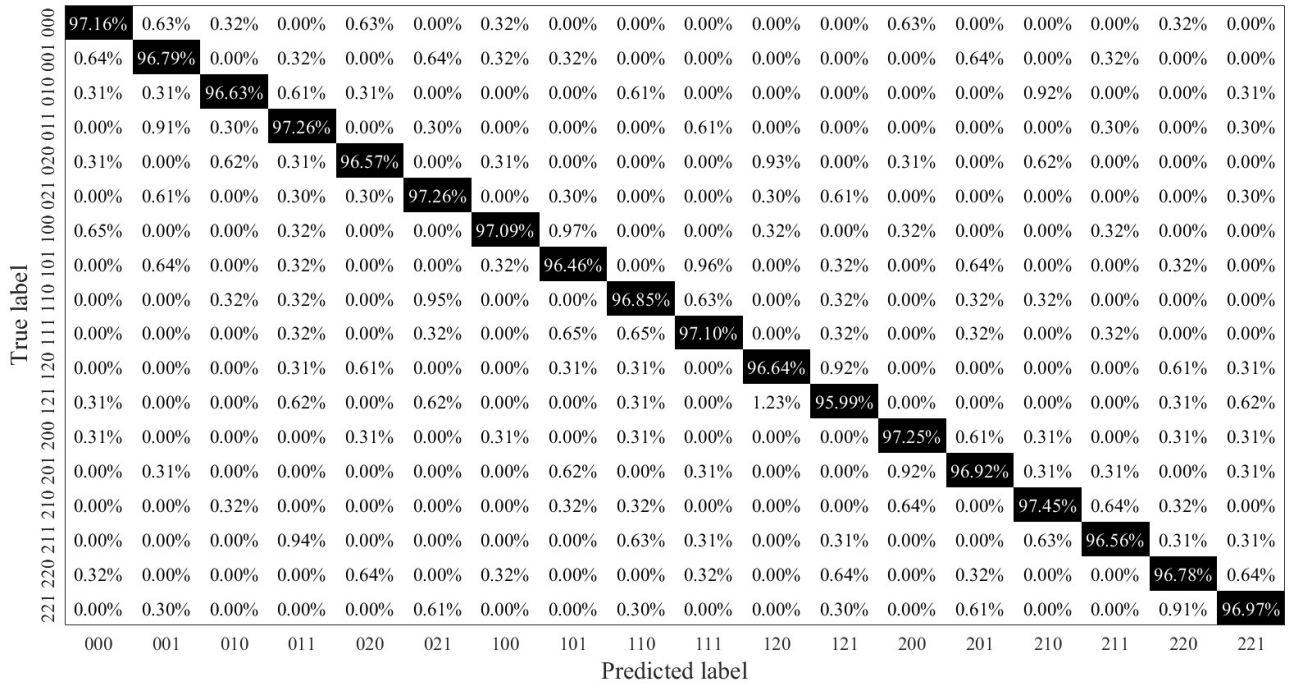


Fig. 7: The confusion matrix of Ward-off Method. For the categories, the first number represents the main hand (0 is standard, 1 is high, 2 is low), the second number represents the centre of gravity (0 is standard, 1 is forward, 2 is backward) and the third number represents the deltoid force (0 is standard, 1 is excessive), which are arranged in Gray code order.

TABLE VI: The comparison among diverse state-of-the-art machine/deep learning backbones on the feature extractor

Backbones	Training accuracy	Test accuracy
No extractor (Only SVM)	94.7%	81.6%
No extractor (Only DNN)	94.5%	87.1%
AlexNet	95.2%	90.5%
GoogleNet	95.9%	90.1%
VGG-11	96.2%	91.7%
ResNet-18	96.1%	92.1%
This method	**96.9%**	**92.5%**

off sub-postures under the SVM classifier is showcased in Fig. 7 as an example, while table III shows a comparison of our work with other state-of-the-art results reported in literature.

C. Experiment III: Multi-sensor data complementarity analysis

Generally, IMU data are mainly used to analyze limb movements (in this study, it is used to analyze whether the main hand is high or low); Piezoresistive arrays data are mainly used to analyze the center of gravity and gait (in this study, it is used to judge whether the center of gravity is forward or backward); sEMG data are mainly used to analyze muscle force (in this study, it is used to estimate whether the deltoid is excessive rigid). For the purpose of analyzing the coupling and complementarity among multi-sensor data, we only utilized uni-modal data to complete its main task, and compare it with the results of introducing all data. The data from group A were utilized in this experiment. Fig. 8 demonstates the samples distribution on the hyperplane mapped by T-SNE algorithm: By fusing the multi-sensor data, boundaries of main hand evalutation and center of gravity evaluation become clearer, which means the reduction of the likelihood of overfitting.

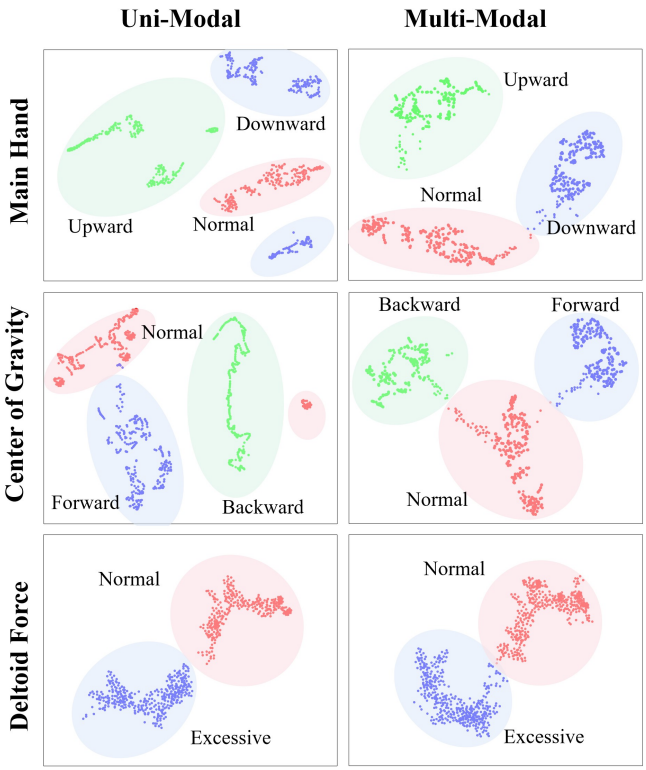


Fig. 8: The comparison of features' hyperplane projections between uni-modal and multi-modal under different sub-tasks.

As showcased in Fig. 9 that IMUs' and piezoresistive arrays' data are highly complementary to data from other sensors. By fusing multi-sensor data, the classification accuracy of single-mode tasks

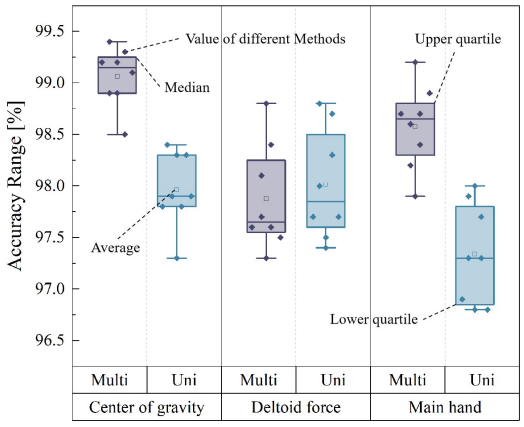


Fig. 9: The comparison of classification accuracy between uni-modal and multi-modal under different sub-tasks.

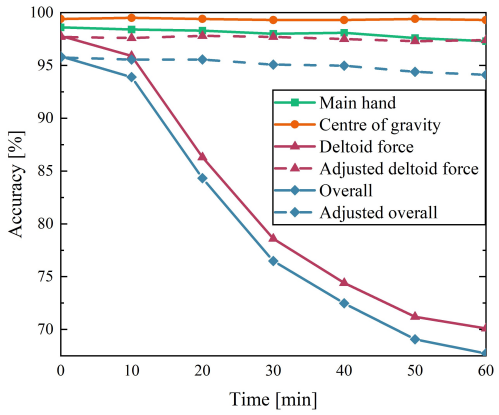


Fig. 10: Recognition accuracy of different sub-tasks with time.

is significantly improved compared with that of using only a single sensor. On the contrary, the modal identification of deltoid force almost completely depends on the sEMG sensors. In this task, the recognition accuracy using multi-sensor data is similar to that of using sEMG sensors only.

D. Experiment IV: System time stability analysis

To test the time stability of the system, we collected the motion data of the subjects from group A after wearing the devices for a certain time and recognized the postures via the trained model. Starting from baseline, the movement data of the Eight Methods were collected every 10 minutes. Fig. 10 displays the change in the average accuracy of three types of sub-tasks (classification of the main hand, center of gravity, and deltoid force) with the continuous wearing time. It can be observed that the recognition accuracy of the main hand and center of gravity has good time stability (reflecting that the time drifts of IMUs and piezoresistive insoles are acceptable); The recognition accuracy of deltoid force has a significant decline with the passage of continuous use time (reflecting the severe time drift of the data collected by sEMG sensors). To study the mechanism of the time drift of the sEMG sensors, the sEMG electrodes of half of the subjects were adjusted (rub alcohol on skin surface and reapply electrodes) before each measurement as a control.

V. Discussion and Limitation

By applying the proposed WMNN, we fused the multi-sensor data in the feature layer. The system recognizes the fitness activities by comprehensively analyzing the visible body movements and the invisible muscular forces behind them for the first time. The system can achieve a decent accuracy in the classification of sub-motions of Eight Methods of Tai Chi and be commendably generalized to the data it has never learned before. In this section, some results are discussed in detail, which draws forth the merits and the limitation of this work.

In Experiment I, though the set of convolution kernel of 3×3 and the stride of 1 brings a relatively small receptive field on one hand, we increase the depth and make it easier to search for details on the other hand. Using less width to increase depth is in line with the perception of the network in the AI field in recent years [42, 43]. For sEMG signals, the pooling layer may reduce the sensitivity of feature perception at different time steps. While in this method, the use of pooling layers improved the performance of the system. This may be due to the difference in the tasks: in tasks such as gesture recognition that recognize motion gestures through sEMG signals, pooling layers tend to blur the features of the signal, thereby decreasing the classification effect. In this work, sEMG is mainly utilized to analyze the strength of deltoid muscle force, and it focuses more on the density and amplitude of the signal. Although the sEMG signal has been denoised before WMNN, it still contains some noise. Using max-pooling can further suppress the noise while improving the saliency of the feature map in the region.

Softmax trains deep features that divide the entire hyperspace or hypersphere according to the number of categories, ensuring that the classes are separable. However, Softmax does not require intra-class compactness and inter-class separation, which often results in poor performance on test sets with relatively small datasets. In Experiment II, when the original network attempted to work on the test set, which it had never learned, the performance degenerated greatly. To address the decline in accuracy, we introduced the SVM to serve as the classifier instead of the Softmax. SVM is an excellent small sample learning method and has good robustness, which is derived from the underlying logic of SVM: In Hilbert space, SVM is committed to making the decision boundary as far away from the samples as possible [31-33]. Therefore, based on the small sample size of this study, SVM effectively ameliorates the generalization of the system. Compared to more complex multimodal architectures such as the Multi-modal Transformer [44], the WMNN proposed in this study focuses more on robustness and generalization on small datasets. When the dataset is larger, normalized data from a large number of subjects can effectively improve the generalization problem presented in the previous section, when a deeper network architecture is expected to achieve higher accuracy in application scenarios where computing power allows, and this is where our future research will be conducted.

In experiment III, the judgment accuracy of the system for the main hand and center of gravity has been significantly improved after using multi-sensor data, which is in line with the advantage of multi-sensor fusion method: when the multi-sensor data is collaborative and comprehensive, the precision and reliability of the system will be improved [46]. In this study, by introducing the multi-sensor fusion method, in addition to expanding the function boundary of a single sensor, it also mines the complementarity between IMUs and piezoresistive insoles' data to enhance the accuracy of the system.

In Experiment IV, the accuracy of muscle force classification analyzed by sEMG sensors decreased significantly with the passage of continuous use time. This is because, with the increase

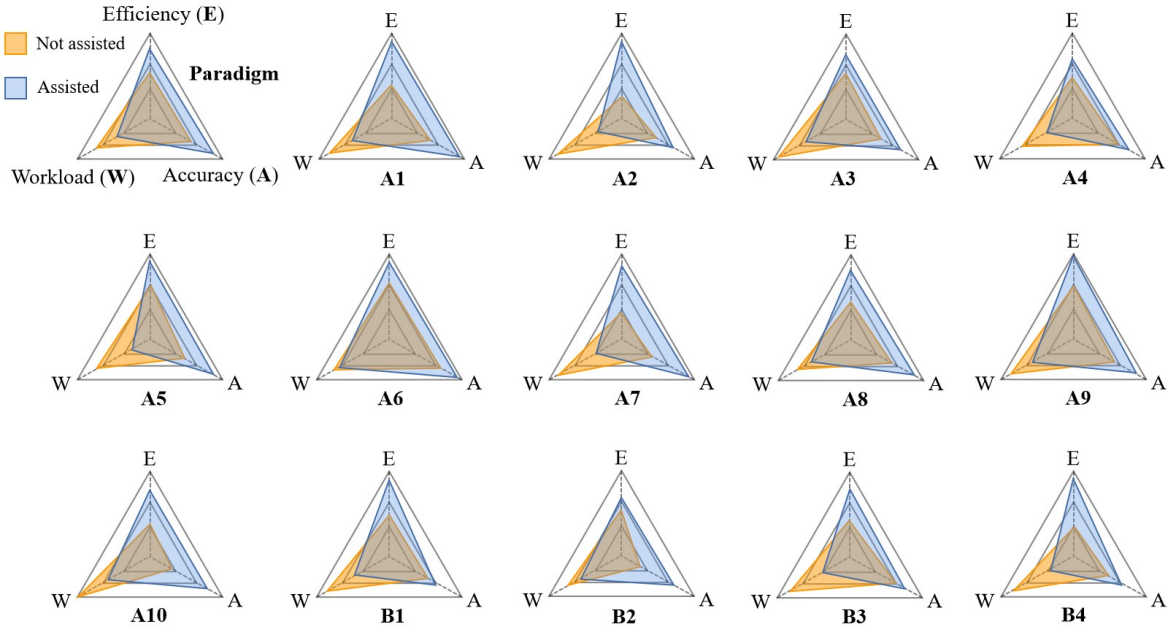


Fig. 11: The radar charts of different subjects' user experience.

of continuous wearing time, the sweat film formed by sweating between the electrodes and the skin and the stretching and contraction of human muscles will impair the performance of sEMG sensors [45]. Therefore, unless optimizing the hardware (e.g. the adhesive biocomposite electrodes proposed by Yang *et al*. [47] and in-sensor adaptive machine learning system developed by Ali *et al*. [48]), it is necessary to frequently readjust the conductive hydrogel on the electrode surface and wipe the skin with alcohol during use. In addition, the development of new algorithms for online, real-time calibration of human impedance to mitigate the bias caused by factors such as sweat is also expected to be feasible. We leave this for future.

To evaluate the experience of using the system, the 14 subjects were invited to perform the Tai Chi Eight Methods with and without system assistance. After the training, we evaluated the efficiency, workload, and accuracy of the training, where efficiency was characterized by the time consumed in training to achieve the same accuracy, accuracy was characterized by the accuracy of the training movements, and the workload was measured by the NASA Task Load Index (TLX) form [49]. The radar charts of user experience for all subjects are shown in Fig. 11. As can be seen from the radar charts, the efficiency and accuracy of completing the training with system assistance increased significantly and the workload decreased greatly. Although subjects in Group B suffered a slightly worse experience overall than subjects in Group A due to their data not being learned online, their E, A, and W all overwhelmed the unaided situation.

The latency of the system mainly derives from the following aspects: communication latency and computation latency. All the sensors are connected with the host server by BLE Mesh and its latency refers to the time interval from when the device puts data into the Bluetooth protocol stack to when it is received by the host. Through the test, it can be seen that the BLE transmission delay fluctuates around 10ms. Computation latency is mainly generated in processing inputs and offering feedback via GUI in our system. In the processing stage, the proposed system based on lightweight WMNN spends less than 40 ms to obtain the classification results. Besides, the time of showcasing the results by GUI can be omitted. Therefore, the overall system latency is about 50ms, which is fully acceptable for users.

VI. Conclusion

In this study, we propose a generalized multi-sensor fusion framework for HAR based on WMNN to comprehensively evaluate motion tasks' quality. The visible movements (limb motion and center of gravity) and associated invisible bioinformatics (muscle force) are firstly analyzed jointly in the framework. Compared to camera-based systems, which often require exercising in an environment with multiple cameras, our system is more adaptable to different scenarios and can be carried more flexibly outdoors. Besides, our system contains rich information (including body movement, the body center of gravity, and muscle force) needed to analyze sports actions, which can facilitate the analysis of sports' correctness. Therefore, the proposed technique can function more conveniently, flexibly, and accurately, which empowers multiple applications like rehabilitation and health monitoring, apart from home-centered fitness, potentially advancing the development of next-generation e-health with the novel framework.

References

[1] Chen, Yi, et al. "Robust activity recognition for aging society." IEEE journal of biomedical and health informatics 22.6 (2018): 1754-1764.
[2] Zhang, Zhe, et al. "A fuzzy kernel motion classifier for autonomous stroke rehabilitation." IEEE journal of biomedical and health informatics 20.3 (2015): 893-901.
[3] Tang, Wenlong, and Edward S. Sazonov. "Highly accurate recognition of human postures and activities through classification with rejection." IEEE journal of biomedical and health informatics 18.1 (2014): 309-315.
[4] Sazonova, Nadezhda, et al. "Posture and Activity Recognition and Energy Expenditure Estimation in a Wearable Platform." IEEE journal of biomedical and health informatics 19.4 (2015): 1339-1346.
[5] Huang, Wenbo, et al. "The convolutional neural networks training with channel-selectivity for human activity recognition based on sensors." IEEE Journal of Biomedical and Health Informatics 25.10 (2021): 3834-3843.
[6] Bi, Haixia, et al. "Human activity recognition based on dynamic active learning." IEEE Journal of Biomedical and Health Informatics 25.4 (2020): 922-934.
[7] Lee, Hong Ji, et al. "Estimation of body postures on bed using unconstrained ECG measurements." IEEE journal of biomedical and health informatics 17.6 (2013): 985-993.

[8] Jellish, Jeremy, et al. "A system for real-time feedback to improve gait and posture in Parkinson's disease." IEEE journal of biomedical and health informatics 19.6 (2015): 1809-1819.

[9] Zerrouki, Nabil, et al. "Vision-based human action classification using adaptive boosting algorithm." IEEE Sensors Journal 18.12 (2018): 5115-5121.

[10] Maddala, Teja Kiran Kumar, et al. "Yoganet: 3-d yoga asana recognition using joint angular displacement maps with convnets." IEEE Transactions on Multimedia 21.10 (2019): 2492-2503.

[11] Liu, Jiaqing, et al. "Single Image Depth Map Estimation for Improving Posture Recognition." IEEE Sensors Journal 21.23 (2021): 26997-27004.

[12] Mashiyama, Shota, Jihoon Hong, and Tomoaki Ohtsuki. "Activity recognition using low resolution infrared array sensor." 2015 IEEE International Conference on Communications (ICC). IEEE, 2015.

[13] Gochoo, Munkhjargal, et al. "Novel IoT-based privacy-preserving yoga posture recognition system using low-resolution infrared sensors and deep learning." IEEE Internet of Things Journal 6.4 (2019): 7192-7200.

[14] Yao, Lina, et al. "Rf-care: Device-free posture recognition for elderly people using a passive rfid tag array." (2015).

[15] Ding, Han, et al. "A platform for free-weight exercise monitoring with passive tags." IEEE Transactions on Mobile Computing 16.12 (2017): 3279-3293.

[16] Xu, Leiyang, et al. "Using trajectory features for tai chi action recognition." 2020 IEEE International Instrumentation and Measurement Technology Conference (I2MTC). IEEE, 2020.

[17] Gupta, Ashish, and Hari Prabhat Gupta. "YogaHelp: Leveraging Motion Sensors for Learning Correct Execution of Yoga With Feedback." IEEE Transactions on Artificial Intelligence 2.4 (2021): 362-371.

[18] Ryu, Jaehwan, Byeong-Hyeon Lee, and Deok-Hwan Kim. "sEMG Signal-Based Lower Limb Human Motion Detection Using a Top and Slope Feature Extraction Algorithm." IEEE Signal Processing Letters 24.7 (2017): 929-932.

[19] Li, Ziyou, et al. "Electrode shifts estimation and adaptive correction for improving robustness of sEMG-based recognition." IEEE Journal of Biomedical and Health Informatics 25.4 (2020): 1101-1110. APA

[20] Wang, Xuegang, et al. "A wearable plantar pressure measurement system: Design specifications and first experiments with an amputee." Intelligent Autonomous Systems 12. Springer, Berlin, Heidelberg, 2013. 273-281.

[21] Jeong, Gu-Min, Phuc Huu Truong, and Sang-Il Choi. "Classification of three types of walking activities regarding stairs using plantar pressure sensors." IEEE Sensors Journal 17.9 (2017): 2638-2639.

[22] Haque, Md Rejwanul, et al. "Real Time Level Ground Walking vs Stair-Climbing Locomotion Mode Detection." 2020 IEEE SENSORS. IEEE, 2020.

[23] Luo, Ruiming, et al. "A low-cost end-to-end sEMG-based gait sub-phase recognition system." IEEE Transactions on Neural Systems and Rehabilitation Engineering 28.1 (2019): 267-276.

[24] Chang, Wennan, et al. "A hierarchical hand motions recognition method based on IMU and sEMG sensors." 2015 IEEE International Conference on Robotics and Biomimetics (ROBIO). IEEE, 2015.

[25] Zhou, Han, et al. "Posture tracking meets fitness coaching: A two-phase optimization approach with wearable devices." 2020 IEEE 17th International Conference on Mobile Ad Hoc and Sensor Systems (MASS). IEEE, 2020.

[26] Y. Zhao et al., "Flexible and Wearable EMG and PSD Sensors Enabled Locomotion Mode Recognition for IoHT-Based In-Home Rehabilitation," in IEEE Sensors Journal, vol. 21, no. 23, pp. 26311-26319, 1 Dec.1, 2021, doi: 10.1109/JSEN.2021.3058429.

[27] Krizhevsky, Alex, Ilya Sutskever, and Geoffrey E. Hinton. "Imagenet classification with deep convolutional neural networks." Communications of the ACM 60.6 (2017): 84-90.

[28] Wang, Zhiguang, and Tim Oates. "Encoding time series as images for visual inspection and classification using tiled convolutional neural networks." Workshops at the twenty-ninth AAAI conference on artificial intelligence. 2015.

[29] Ioffe, Sergey, and Christian Szegedy. "Batch normalization: Accelerating deep network training by reducing internal covariate shift." International conference on machine learning. PMLR, 2015.

[30] Schapire, Robert E., and Yoram Singer. "Improved boosting algorithms using confidence-rated predictions." Proceedings of the eleventh annual conference on Computational learning theory. 1998.

[31] Cortes, Corinna, and Vladimir Vapnik. "Support-vector networks." Machine learning 20.3 (1995): 273-297.

[32] Friedrichs, Frauke, and Christian Igel. "Evolutionary tuning of multiple SVM parameters." Neurocomputing 64 (2005): 107-117.

[33] Mathur, Ajay, and Giles M. Foody. "Multiclass and binary SVM classification: Implications for training and classification users." IEEE Geoscience and remote sensing letters 5.2 (2008): 241-245.

[34] Picard, Richard R., and R. Dennis Cook. "Cross-validation of regression models." Journal of the American Statistical Association 79.387 (1984): 575-583.

[35] Pavlinek, Miha, and Vili Podgorelec. "Text classification method based on self-training and LDA topic models." Expert Systems with Applications 80 (2017): 83-93.

[36] Subasi, Abdulhamit, and M. Ismail Gursoy. "EEG signal classification using PCA, ICA, LDA and support vector machines." Expert systems with applications 37.12 (2010): 8659-8666.

[37] AlOmari, Firas, and Gunhai Liu. "Analysis of extracted forearm sEMG signal using LDA, QDA, K-NN classification algorithms." The Open Automation and Control Systems Journal 6.1 (2014).

[38] Peterson, Leif E. "K-nearest neighbor." Scholarpedia 4.2 (2009): 1883.

[39] Hassanat, Ahmad Basheer, et al. "Solving the problem of the K parameter in the KNN classifier using an ensemble learning approach." arXiv preprint arXiv:1409.0919 (2014).

[40] Dudani, Sahibsingh A. "The distance-weighted k-nearest-neighbor rule." IEEE Transactions on Systems, Man, and Cybernetics 4 (1976): 325-327.

[41] Kingma, Diederik P., and Jimmy Ba. "Adam: A method for stochastic optimization." arXiv preprint arXiv:1412.6980 (2014).

[42] Simonyan, Karen, and Andrew Zisserman. "Very deep convolutional networks for large-scale image recognition." arXiv preprint arXiv:1409.1556 (2014).

[43] Szegedy, Christian, et al. "Inception-v4, inception-resnet and the impact of residual connections on learning." Thirty-first AAAI conference on artificial intelligence. 2017.

[44] Tsai, Yao-Hung Hubert, et al. "Multimodal transformer for unaligned multimodal language sequences." Proceedings of the conference. Association for Computational Linguistics. Meeting. Vol. 2019. NIH Public Access, 2019.

[45] Gravina, Raffaele, et al. "Multi-sensor fusion in body sensor networks: State-of-the-art and research challenges." Information Fusion 35 (2017): 68-80.

[46] Abdoli-Eramaki, Mohammad, et al. "The effect of perspiration on the sEMG amplitude and power spectrum." Journal of Electromyography and Kinesiology 22.6 (2012): 908-913.

[47] Yang, Hui, et al. "Adhesive biocomposite electrodes on sweaty skin for long-term continuous electrophysiological monitoring." ACS Materials Letters 2.5 (2020): 478-484.

[48] Moin, Ali, et al. "A wearable biosensing system with in-sensor adaptive machine learning for hand gesture recognition." Nature Electronics 4.1 (2021): 54-63.

[49] Cao, Alex, et al. "NASA TLX: Software for assessing subjective mental workload." Behavior research methods 41.1 (2009): 113-117.