Hindawi
Mobile Information Systems
Volume 2023, Article ID 8273546, 17 pages
https://doi.org/10.1155/2023/8273546

Research Article

Human Activity Recognition Based on a Modified Capsule Network

Shanying Zhu[1,2] **Wei Chen**[1,2] **Fulong Liu**[1,2] **Xiaotao Zhang**[1,2] **and Xiupeng Han**[1,2]

[1]*Tianjin Key Laboratory for Advanced Mechatronic System Design and Intelligent Control, School of Mechanical Engineering, Tianjin University of Technology, Tianjin 300384, China*
[2]*National Demonstration Center for Experimental Mechanical and Electrical Engineering Education (Tianjin University of Technology), Tianjin, China*

Correspondence should be addressed to Xiaotao Zhang; xiaotaozhang@email.tjut.edu.cn

Received 17 October 2022; Revised 12 January 2023; Accepted 13 January 2023; Published 4 February 2023

Academic Editor: Floriano Scioscia

Human activity recognition (HAR) has attracted considerable research attention in the past decade with the development of wearable sensor technology and deep learning algorithms. However, most of the existing HAR methods ignored the spatial relationship of features, which may lead to recognition errors. In this paper, a novel model based on a modified capsule network (MCN) is proposed to accurately recognize various human activities. This novel model is composed of a convolution block and a capsule block, which can achieve end-to-end intelligent recognition. In the meantime, the spatial information among features is preserved through a dynamic routing process. To validate the effectiveness of the model, a human activity dataset is constructed by placing an inertial measurement unit (IMU) on the calf of the volunteers to collect their activity data in daily life, including walking, jogging, upstairs, downstairs, up-ramps, and down-ramps. The recognition accuracy of this novel approach can reach 96.08%, which performs better than the convolutional neural network (CNN) with an accuracy of 91.62%. In addition, it is evaluated on two public datasets named WISDM and UCI-HAR, and the accuracies achieve 98.21% and 95.28%, respectively, which presents higher accuracy than the reported results obtained from benchmark algorithms like CNN. The experimental results show that the proposed model has better activity detection capability and achieves outstanding performance for HAR.

1. Introduction

Human activity recognition (HAR) is the foundation of many fields and has become a research hotspot in the past decade on account of its significance. At present, this technology has been widely applied in the fields of smart homes [1], indoor navigation [2], identity recognition [3], human-machine interaction [4], gait analysis [5], and the Internet of Healthcare Things [6, 7]. The identification accuracy of corresponding activities has significant effects on these applications. In order to improve the accuracy of recognition, various sensor techniques have been employed to collect activity data and different approaches have been constructed according to data features to identify the activities.

The activity data collection methods are mainly divided into two groups: video images [8] and wearable sensors [9]. The former acquires a series of human motion images through cameras and extracts human motion feature information from these images. The commonly used method is the image processing method based on the Kinect sensor, which can extract the depth image features of the moving target [10]. The latter one is to place the sensors on a specific part of the wearer's body to obtain the movement information. Wearable sensors mainly include surface electromyography (sEMG), plantar pressure sensors, and inertial measurement unit (IMU) or sensor fusion to obtain more comprehensive motion information.

However, it has some disadvantages and limitations to the aforementioned data collection methods [9]. The method

based on video images needs to be completed under laboratory conditions since a specific background is required, and the price of cameras is usually expensive. The sEMG sensors need to be in close contact with the wearer's skin, so it is easily affected by sweat and causes discomfort to wearers. Plantar pressure sensors are susceptible to uneven ground. With the development of sensor technology, IMU is becoming more and more popular due to its advantages of lightness, cheapness, high precision, and easy wearing [11]. In addition, IMU sensors are also embedded in mobile devices such as smartphones and smartwatches that are widely used by people [6, 7, 12]. Obviously, IMU is a good choice to collect activity data.

According to the feature of different activity signals obtained by different approaches, researchers have proposed different activity recognition methods. Machine learning algorithms are popular at the beginning, such as support vector machines (SVM) [13], random forest (RF) [14], linear discriminant analysis (LDA) [13], Gaussian mixture model (GMM) [15], and extreme learning machine (ELM) [16]. Tahir Hussain et al. [13] proposed a new feature extraction method to process sEMG, using two classification models of SVM and LDA to identify the motion intentions of four subjects. The method was more robust compared to the existing methods, but needed eleven sEMG sensors located on the lower limb muscles. Moreover, as mentioned earlier, the approaches of multisensor fusion are also effective for activity recognition [12]. Xi et al. [16] proposed a feature-level data fusion method and double parameter Kernel optimization based on an extreme learning machine (DPK-OMELM) to identify activity types by fusing sEMG signals and plantar pressure signals and achieved high recognition accuracy. Chen et al. [5] proposed a novel activity recognition algorithm based on human gait characteristics to classify six activities using a wearable smart insole system integrating a plantar pressure sensor and IMU, which had a low computational cost and higher accuracy, and introduced gait labs into daily activities.

However, the traditional machine learning methods rely on manually extracting features, which requires researchers to have extensive expertise and experience in related fields, whereas, there is currently no standard on how to manually extract features [17]. As a result, these methods are time-consuming and even unachievable.

Deep learning based on neural networks has been proposed since 2006 [18], and it has achieved outstanding performance in many challenging fields such as computer vision, natural language processing, speech recognition, and autonomous vehicles. Neural networks, including convolutional neural networks (CNN) and long short-term memory networks (LSTM) exhibit powerful feature extraction capabilities and can automatically extract features from raw data for classification and other tasks, bringing great convenience to researchers [19].

Therefore, combining activity information obtained by IMU and CNN or LSTM to study HAR has become a new and promising trend [20]. Chen et al. [21] proposed a recognition method based on LSTM-CNN, which combined the advantages of LSTM and CNN models, and used the collected IMU motion information to classify five common activities and achieved 97.78% average accuracy. Aiming at the complexity of the traditional activity recognition methods, Zhu et al. [22] proposed a new deep convolutional neural network model denoted as DDLMI, which classified five types of terrain by collecting the IMU information on the thigh and calf and the recognition accuracy rate can reach 97.64%. Semwal et al. [23] used artificial neural networks (ANN), extreme learning machines (ELM), and deep neural networks (DNN) to identify six kinds of motion information collected by accelerometers. DNN achieved the best recognition accuracy. Hu et al. [24] used six IMU installed on the body to train three deep-learning models, all of which achieved more than 90% accuracy. The experimental results further proved that the use of deep learning models and wearable IMU sensors had great potential in gait analysis. Semwal et al. [25] proposed a deep learning framework based on ensemble learning to classify the collected IMU information for seven gait activities. The experimental results showed that the framework outperformed other methods. Bozkurt [26] used various machine learning methods and deep learning methods to test the IMU dataset, and the results showed that the deep learning method achieved the best performance.

Although CNN and LSTM have shown excellent performance in the field of HAR, they still have some disadvantages as follows: (1) the pooling layer of CNN loses some information, which may have an impact on the classification results; (2) CNN has translation invariance, so the generalization ability is poor; and (3) LSTM ignores spatial features and parallel processing is poor. In response to these shortcomings, Sabour et al. [27] proposed a capsule network (CapsNet) model in 2017. The so-called capsule is a vector composed of a group of neurons. Traditional neural networks use a single neuron as the inputs and outputs, while capsule networks use vectors as the inputs and outputs. The length of the capsule represents the probability that the entity exists and the direction represents its characteristic properties. The CapsNet realizes the encoding between local features and the whole through a dynamic routing mechanism to preserve the spatial relationship of features, so it has translation homogeneity, which has the ability to overcome the referred shortcomings of CNN and LSTM.

Up to now, only a small number of researchers have applied the CapsNet to HAR. The first work of using the framework of CapsNet for HAR was conducted by Pham et al. [28] who proposed a model named SensCapsNet. The experimental results showed that the method outperformed CNN and LSTM. Shi et al. [29] proposed a HAR system based on capsule and "long range". The system could realize long-distance, low-power consumption, and real-time HAR. The results demonstrated that the method achieved a higher accuracy than CNN and recurrent neural network (RNN). Khaled et al. [30] proposed an enhanced model of CapsNet named 1D-HARCapsNe. This provided an efficient intelligent decision-support approach for HAR. Sun et al. [31] proposed a novel method named CapsGaNet based on capsule and gate recurrent unit (GRU) with attention mechanisms. This method could achieve spatiotemporal multifeature extraction from wearable IMU sensors for HAR.

Meanwhile, it is worth highlighting that the features of different human activities have some similarities, and preserving the spatial relationship between features may be more conducive to distinguishing different activities. In order to further investigate the effectiveness of the capsule network in HAR and overcome the abovementioned shortcomings of CNN and LSTM, this paper proposed a new model based on a modified capsule network (MCN) for HAR. The novel model is composed of a convolution block and a capsule block. The convolution block can extract shallow activity features with small convolution kernels. The weight sharing and small kernels in the convolution layer allow fewer parameters to be trained during backpropagation. Then, the capsule block employs vectors as the inputs and outputs of the network and mines the spatial information of the features. As a result, the model gives full play to the feature extraction capabilities of CNN and CapsNet, which well retains spatial information while digging into in-depth activity features.

To validate the effectiveness of this novel model, experimental studies are conducted which consist of the following two parts: first, a self-collected dataset is applied to test the learning ability of the MCN. The dataset includes three-axis accelerometer information and three-axis gyroscope information, which is collected by ten volunteers. Moreover, the effectiveness of the MCN model is verified by comparing experiments with CNN. Second, the public datasets WISDM and UCI-HAR are employed to further verify the generalization ability of the MCN model.

The main contributions of this paper are summarized as follows:

(1) A novel deep learning model based on a modified capsule network is proposed for human activity recognition. This model can not only realize end-to-end intelligent recognition but also retain the spatial relationship of features.

(2) A human activity dataset based on IMU sensor information is constructed. The proposed model achieves 96.08% recognition accuracy on the dataset, which is higher than 91.62% of the convolutional neural network.

(3) The proposed model achieves 98.21% and 95.28% recognition accuracies on public datasets WISDM and UCI-HAR, respectively.

The structure of the paper is organized as follows: the capsule network model and proposed MCN model are presented in Section 2. Introduction to datasets and data preprocessing are presented in Section 3. Comparative experiments and results are disclosed in Section 4. Corresponding discussions are given in Section 5. The conclusions of this paper are summarized in Section 6.

2. Methods

2.1. CapsNet Model. The original CapsNet consists of three layers, namely, the convolutional layer, PrimaryCaps layer, and DigitCaps layer [27]. The convolution layer performs feature extraction through 256 convolution kernels with a size of 9×9, a stride of 1, and ReLU activation, which are then input into the PrimaryCaps layer. The PrimaryCaps layer further extracts features through 32×8 convolution kernels with a size of 9×9 and a stride of 2. Then, 1152 capsules with a dimension of 8 are generated, which are input into the DigitCaps layer as low-level capsules. The DigitCaps layer finally generates 10 capsules with a dimension of 16 through a dynamic routing mechanism. The capsule with the largest length is the final classification result. Finally, the correctly predicted capsule is passed through a three-layer fully connected neural network to reconstruct the input.

Dynamic routing is a core part of the CapsNet, in which the inputs and outputs of capsules are vectors and through it, the CapsNet retains the spatial information of features [32]. Figure 1 shows the process of information transfer between capsules.

The input capsule $\mathbf{u}_i$ is multiplied by the weight matrix $\mathbf{W}_{ij}$ to obtain the predicted capsule $\hat{\mathbf{u}}_{j|i}$, which is completed by the following formula:

$$\hat{\mathbf{u}}_{j|i} = \mathbf{W}_{ij}\mathbf{u}_i, \tag{1}$$

where the weight matrix $\mathbf{W}_{ij}$ is updated by backpropagation.

The weighted summation of $\hat{\mathbf{u}}_{j|i}$ and the coupling coefficients c_{ij} can obtain the deep feature capsule $\mathbf{s}_j$. Then, $\mathbf{s}_j$ is squeezed nonlinearly through the activation function, so that the short vector is almost 0 and the long vector is close to 1, and the output capsule $\mathbf{v}_j$ is obtained, as shown in the following equations:

$$\mathbf{s}_j = \sum_i c_{ij}\mathbf{u}_{j|i}, \tag{2}$$

$$\mathbf{v}_j = \frac{\| \mathbf{s}_j \|^2}{1 + \| \mathbf{s}_j \|^2} \frac{\mathbf{s}_j}{\| \mathbf{s}_j \|}, \tag{3}$$

where c_{ij} is the coupling coefficient determined by the iterative dynamic routing process, which can be updated by the intermediate variable b_{ij}, $\mathbf{v}_j$ is the vector output of capsule j, and $\mathbf{s}_j$ is its total input.

Updating b_{ij} and c_{ij} is by calculating the correlation between each output capsule $\mathbf{v}_j$ and prediction capsule $\hat{\mathbf{u}}_{j|i}$, as shown in the following equations:

$$b_{ij} \leftarrow b_{ij} + \mathbf{u}_{j|i} \cdot \mathbf{v}_{j|i}, \tag{4}$$

$$c_{ij} = \mathrm{softmax}\left(b_{ij}\right) = \frac{\exp\left(b_{ij}\right)}{\sum_k \exp\left(b_{ik}\right)}, \tag{5}$$

where the initial value of b_{ij} is 0. After getting b_{ij}, c_{ij} is updated. If the consistency of the two vectors is high, c_{ij} becomes larger, and it becomes smaller when they are inconsistent. Then, $\mathbf{s}_j$ and $\mathbf{v}_j$ will be updated, and the final output capsule $\mathbf{v}_j$ will be obtained after the dynamic routing process.

2.2. Proposed MCN Model. The convolutional neural network (CNN), first proposed in 1998, is a feedforward neural network [33], which has outstanding performance in HAR.

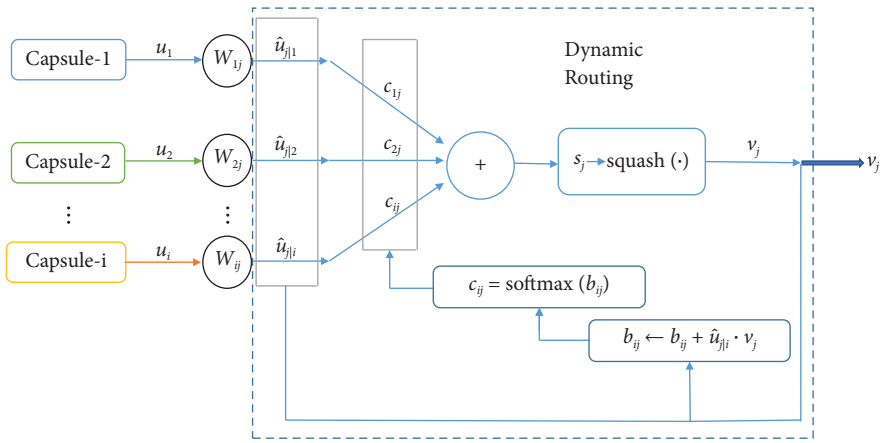

Figure 1: Information transfer between capsules.

But in the process of pooling layer, spatial information of features such as pose and velocity is discarded. However, through the dynamic routing process, the information between local parts and the whole is preserved by the CapsNet. Consequently, the CapsNet can distinguish smaller differences between features of different activities. In addition, the parallel processing of LSTM is poor. Toward the drawbacks of CNN and LSTM, a deep learning model based on a modified capsule network, namely, MCN, is proposed for HAR.

The MCN structure proposed in this paper is shown in Figure 2, which can be divided into two parts: the CNN block and the CapsNet block. Compared to the original capsule network, the network structure is modified as follows: First, the three-layer convolution layer with a kernel size of 3 replaces the one-layer convolution layer with a kernel size of 9. This can realize parameter sharing and effectively reduce the number of parameters in the network. Moreover, a batch normalization (BN) layer is added after each convolution layer. It can speed up the convergence of neural networks [34]. The activation function uses Leaky ReLU instead of ReLU. This increases the nonlinearity of neural networks and gives all negative values a nonzero slope so that all negative values can be preserved. The dropout layer is added to the network to randomly drop some neurons in a certain proportion, which can effectively prevent the model from overfitting. Finally, because the result does not require image reconstruction, the three-layer fully connected layer is discarded, which further reduces the network parameters and is conducive to the lightweight of the network.

The specific process of the proposed MCN model is shown in Figure 2. First, the IMU data are preprocessed and then input into the CNN block. After three 2D convolution layers, they are input into the CapsNet block. After the PrimaryCaps layer and the ActivityCaps layer, six capsules with a dimension of 16 are output. Finally, the length of each capsule is calculated.

For the CNN block, considering the in-depth feature mining capabilities of convolution layers, three Conv2D layers with a kernel size of 3×3 are assigned to extract the shallow features. Using small kernels can effectively reduce the number of parameters in training. After each convolution layer is connected to a BN layer and Leaky ReLU activation layer, the BN layer can speed up network convergence, and the activation layer introduces nonlinear mapping to the network.

For the CapsNet block, include the PrimaryCaps layer and the ActivityCaps layer. The PrimaryCaps layer receives the feature maps from the CNN block, and it is a 2D convolution capsule layer with a kernel size of 2×2 and a stride of 2 to further extract features. After that, they are reshaped to low-level capsules of a dimension of 8 and input them into the ActivityCaps layer. After dynamic routing iterations, six capsules with a dimension of 16 are finally generated. The capsule with the largest length is the human activity recognition result by the neural network. In addition, a dropout layer is added between the PrimaryCaps layer and the ActivityCaps layer to randomly drop some neurons with a probability of 0.5 to prevent overfitting. Owing to the inputs being IMU data, the recognition result does not need to be reconstructed. Detailed network parameters will be given in Section 4.

The loss function is defined as Margin Loss [27]. For each output capsule vector, the loss function is calculated as shown in the following formula:

$$L_k = T_k \max\left(0, m^+ - \|\mathbf{v}_k\|\right)^2 + \lambda\left(1 - T_k\right)\max\left(0, \|\mathbf{v}_k\| - m^-\right)^2, \tag{6}$$

where $T_k = 1$ when the class k activity actually exists, otherwise it is 0, and m^+, m^-, and λ are the hyperparameters during training, which take the values 0.9, 0.1, and 0.5, respectively. The total loss is the sum of the losses of all output capsules.

In addition, a conventional CNN model is designed for comparative experiments. The structure is shown in Figure 3. The model consists of six layers, including three convolution layers, a pooling layer, and two fully connected layers. A BN layer and activation layer are added after each convolution layer. The pooling layer can achieve dimensionality reduction. The fully connected layer is used for classification and outputs the probability value of each

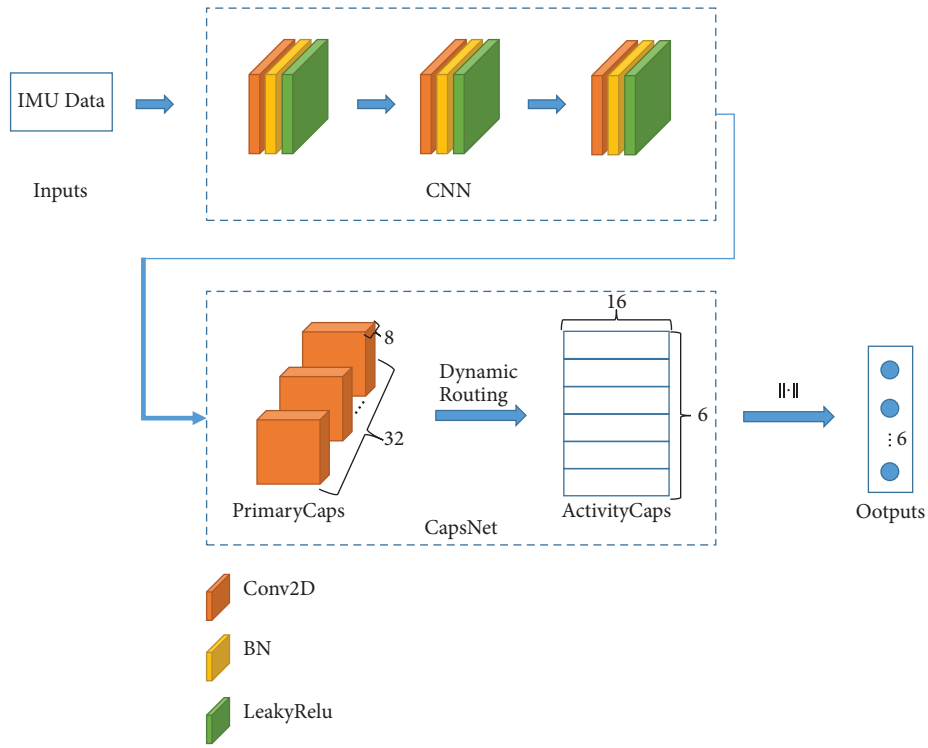

Figure 2: The structure of the proposed MCN model.

category through softmax. The one with the highest probability value is the recognition result of the neural network.

3. Datasets and Preprocessing

3.1. Collected Dataset. To collect experimental data, 10 volunteers (all male, age 25 ± 1 years, weight 70 ± 20 kg, and height 170 ± 10 cm) are invited to school. They are healthy people without any lower limb-related disability. For each volunteer, six different activity experiments (walking, jogging, up-ramps, down-ramps, upstairs, and downstairs) are performed. Before the experiment, the nature of the experiment is informed to each participant and written consent is obtained from each volunteer.

In the process of data collection, the IMU (Model: BWT901BLECL5.0, Shenzhen wit-motion Technology Co. Ltd., Shenzhen, China) is applied to collect activity data, which can collect accelerometer and gyroscope on three orthogonal axes during the activity. The IMU sensor layout and coordinate system are shown in Figure 4(a). The IMU is tied to the outside of the volunteer's right calf with a strap, which does not cause discomfort to the volunteer's activities. Data collection is performed outdoors and indoors, rather than under strictly controlled laboratory conditions. The volunteers walk in their own comfortable way and each activity is shown in Figure 4(b).

During the data collection process, the sampling frequency of the IMU is 100 Hz. The motion information is sent to the laptop through Bluetooth transmission. A text file is generated and stores in the laptop after each activity. Finally, 2,015,766 data samples are collected, and the data distribution of each activity is shown in Figure 5. The number of samples in this dataset is relatively balanced, which is conducive to improving the generalization ability of the neural network.

3.2. Public Datasets

3.2.1. WISDM. The WISDM dataset is collected by 36 volunteers using a built-in three-axis accelerometer in an Android phone in the front leg pocket [35]. The sampling frequency is 20 Hz, and six activities are collected: walking, jogging, upstairs, downstairs, standing, and sitting. A total of 1,098,209 samples are recorded, and the distribution of each activity is shown in Figure 6. It can be seen that the dataset is an imbalanced dataset.

3.2.2. UCI-HAR. The UCI-HAR dataset is built from 30 volunteers [36]. The volunteers, aged 19–48, put the smartphone on their waists and completed a total of six activities in their daily life. The six activities are standing, sitting, laying, walking, upstairs, and downstairs. Accelerometer and gyroscope data are collected for each activity at a sampling frequency of 50 Hz. These experiments are videotaped to facilitate manual labeling of the data. Ultimately, the dataset yields 748,406 samples. The distribution of each activity is shown in Figure 7. It can be seen that the dataset is a balanced dataset.

3.3. Data Preprocessing. In order to facilitate the training of the network model and improve the recognition accuracy, the following preprocessing needs to be performed on the original IMU data.

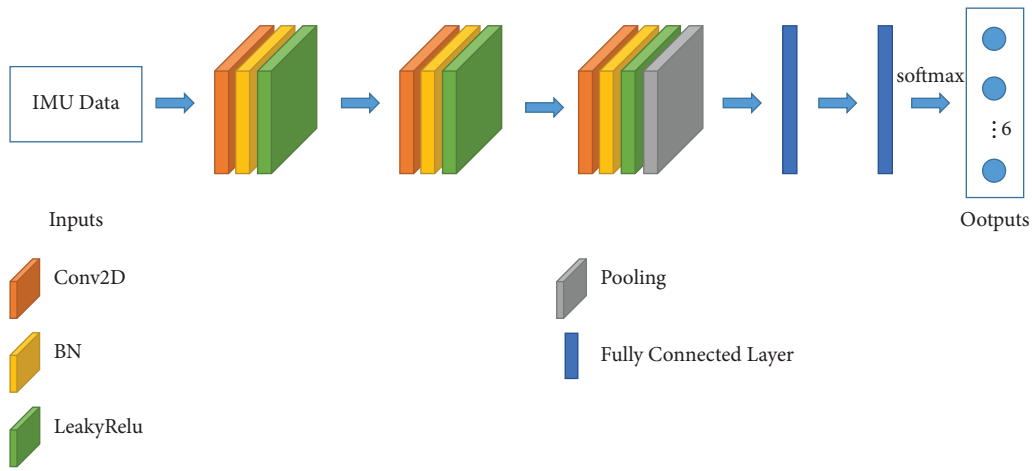

Figure 3: The structure of the CNN model.

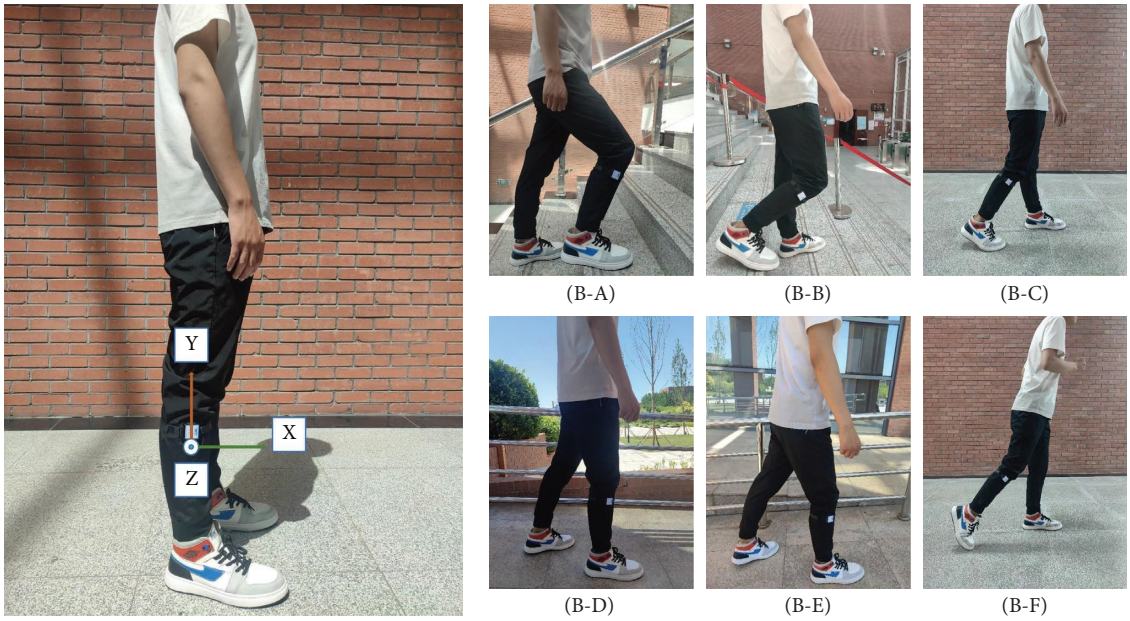

Figure 4: (a) Schematic diagram of sensor location and coordinate system. (b) The volunteers wore an IMU sensor to their right calf and completed six activities. (B-A) Upstairs, (B-B) downstairs, (B-C) walking, (B–D) down-ramps, (B–E) up-ramps, and (B–F) jogging.

3.3.1. Data Normalization. The collection of activity data is wirelessly transmitted through Bluetooth, and data may be lost during the transmission process. First, if there are missing values, the entire row of the data is discarded.

The accelerometer and gyroscope data collected from the IMU sensor have different numerical ranges. Using the data directly to train a neural network model may have poor training results. Therefore, in order to treat each feature equally, it is necessary to normalize the IMU data to a range with a mean of 0 and a variance of 1. The normalization process removes any overall bias and the impact of the different ranges in the IMU data [37, 38]. The normalized formula is shown as follows:

$$Y_i = \frac{X_i - \overline{X}}{\sigma}, \tag{7}$$

where Y_i is the normalized data, X_i is the original data, and $\overline{X}$, σ are the mean and variance of the original data, respectively.

3.3.2. Data Segmentation. For the collected activity data, it is not advisable to directly use each sample for model training, because each sample is 0.01 seconds of data, which represents the instantaneous state of the activity and cannot reflect the characteristics of each activity. Therefore, in this paper, a sliding window with a fixed length is applied to

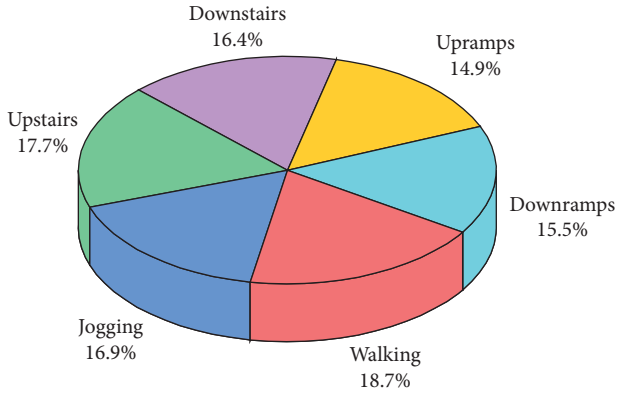

Figure 5: Schematic diagram of the data distribution of the collected dataset.

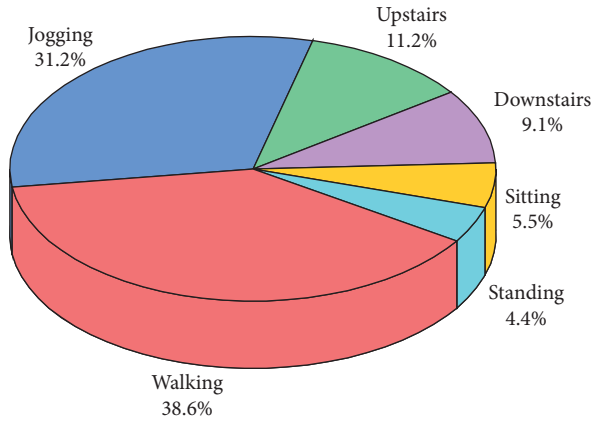

Figure 6: Schematic diagram of the data distribution of the WISDM dataset.

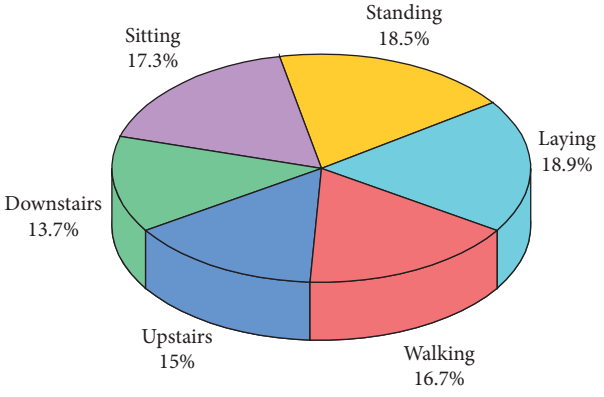

Figure 7: Schematic diagram of the data distribution of the UCI-HAR dataset.

segment the data, and each window contains three-axis accelerometer information and three-axis gyroscope information. A fixed-length window moves from one sampling point to another, moving forward by the same length each time, while retaining a certain proportion of the historical information of the previous window. Then separate each window from the original sequence for feature extraction. The collected data can be represented by the sample matrix S as shown in the following:

$$S = \begin{bmatrix} A_x^0 & A_y^0 & A_z^0 & B_x^0 & B_y^0 & B_z^0 \\ A_x^1 & A_y^1 & A_z^1 & B_x^1 & B_y^1 & B_z^1 \\ A_x^2 & A_y^2 & A_z^2 & B_x^2 & B_y^2 & B_z^2 \\ A_x^3 & A_y^3 & A_z^3 & B_x^3 & B_y^3 & B_z^3 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ A_x^t & A_y^t & A_z^t & B_x^t & B_y^t & B_z^t \end{bmatrix}, \tag{8}$$

where A and B represent the accelerometer and gyroscope; x, y, and z represent the three channels of the accelerometer and gyroscope; t represents the total number of rows of the sample matrix. Assuming that the size of the sliding window is m, and the step size of each window is step (step < m), the sample matrix S can be divided into $S_1, S_2, \cdots, S_n$ with the same size, and the segmentation result is shown in Figure 8.

3.3.3. Data Labels. The "one-hot" encoding method is applied to encode six activities, that is, in each column vector, except one is 1, and the rest are 0, which can solve the discrete value problem of categorical data. The encoding result of six activities is as shown in the following equation:

$$\begin{cases} \text{Downramps} = [1\ 0\ 0\ 0\ 0\ 0] \\ \text{Downstairs} = [0\ 1\ 0\ 0\ 0\ 0] \\ \text{Jogging} = [0\ 0\ 1\ 0\ 0\ 0] \\ \text{Upramps} = [0\ 0\ 0\ 1\ 0\ 0] \\ \text{Upstairs} = [0\ 0\ 0\ 0\ 1\ 0] \\ \text{Walking} = [0\ 0\ 0\ 0\ 0\ 1] \end{cases}. \tag{9}$$

4. Results

The experiments use the PyTorch deep learning framework to implement the MCN model, which supports C++, Python, and other programming languages, and can run on CPU and GPU. The experimental hardware configuration is Intel i5-8300 CPU, NVIDIA GeForce GTX 1050 graphics card, and 8 G RAM. The experiments are performed on Windows 10 system. The software is anaconda3, Python 3.10, PyTorch 1.11, and CUDA 11.3.

In order to evaluate the performance of the MCN model, accuracy, precision, recall, $F1$–score, and confusion matrix (CM) are used as evaluation metrics, and the calculation formulas are shown in the following equations:

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}}, \tag{10}$$

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}}, \tag{11}$$

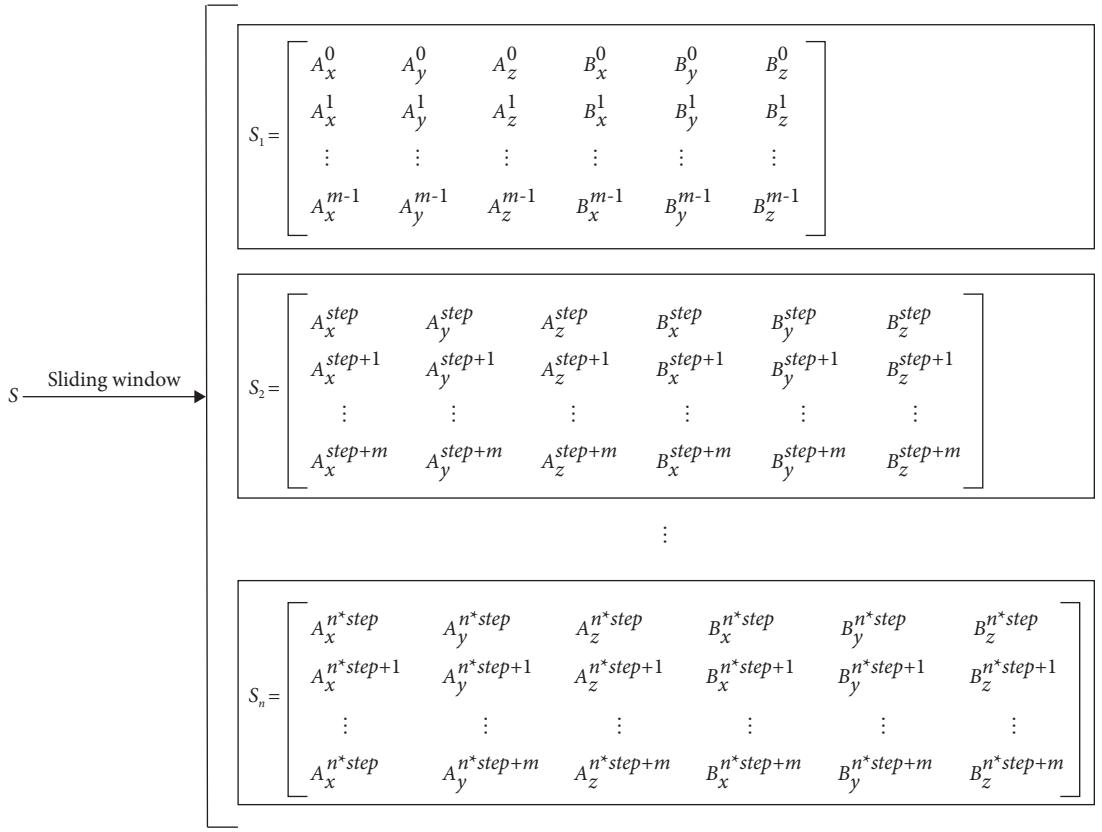

Figure 8: Schematic diagram of the IMU data segmentation.

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}}, \tag{12}$$

$$F1 - \text{score} = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}, \tag{13}$$

where TP is the number of the model correctly predicts the positive class, FP is the number of the model mistakenly predicts the positive class, TN is the number of the model correctly predicts the negative class, and FN is the number of the model mistakenly predicts the negative class. $F1$-score comprehensively considers the precision and recall, so it is a fairer evaluation index. The value range is [0, 1]. The larger the value, the better the model output is.

It can be seen from the CM that the number of actual labels is misidentified as the other labels by the model. The CM is defined as follows:

$$CM = \begin{bmatrix} C_{11} & C_{12} & C_{13} & C_{14} & C_{15} & C_{16} \\ C_{21} & C_{22} & C_{23} & C_{24} & C_{25} & C_{26} \\ C_{31} & C_{32} & C_{33} & C_{34} & C_{35} & C_{36} \\ C_{41} & C_{42} & C_{43} & C_{44} & C_{45} & C_{46} \\ C_{51} & C_{52} & C_{53} & C_{54} & C_{55} & C_{56} \\ C_{61} & C_{62} & C_{63} & C_{64} & C_{65} & C_{66} \end{bmatrix}, \tag{14}$$

where the horizontal axis represents the true label, and the vertical axis represents the predicted label. The diagonal elements represent the number of correct recognition of each type of activity, while the off-diagonal elements represent the number of activities of each type that are incorrectly identified as other activities. Therefore, the larger the number of diagonal elements and the smaller the number of off-diagonal elements, the better the recognition results of the model.

4.1. Results in the Collected Dataset

4.1.1. MCN Model Evaluation and Results. In this paper, the sliding window size is 128 and the step size is 64. That is, the data of 1.28 s are taken, and the overlap rate is 50%. The six activities recognized are all periodic and 1.28 s data can contain one activity cycle [14, 21]. Therefore, this paper selects 1.28 s of data as the training data of the classifier to ensure that the window data contains at least one complete activity cycle, so as to retain all information about each activity. After the collected IMU data are preprocessed, 31,495 single-channel "images" are generated, and the size of the "images" is 128×6. 80% are randomly taken as the training set, and the remaining 20% are used as the test set for evaluating model performance. 20% of the training set is randomly selected as the validation set, which is employed to monitor the effect of the model during the training process. The recognition flow chart is shown in Figure 9.

The input dimension of the MCN model is 128×6. After three convolution layers, it is input into the PrimaryCaps

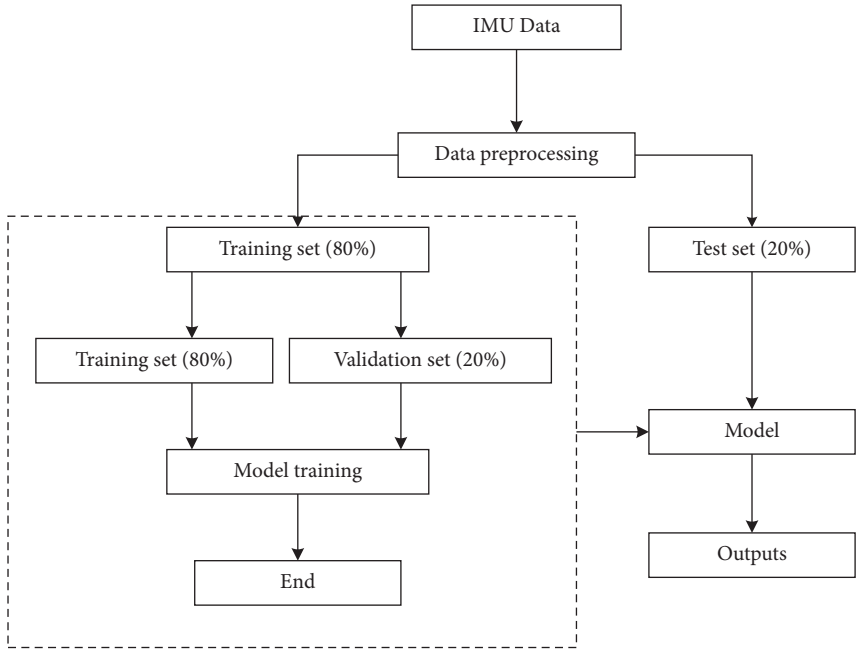

Figure 9: The MCN model activity recognition flow chart.

layer. After the PrimaryCaps layer, $32 \times 62 \times 1$ capsules with a dimension of 8 are generated and input into the ActivityCaps layer. After passing through the ActivityCaps layer, 6 capsules with a dimension of 16 are output.

There are 20,157 samples for training, which is absolutely sufficient. The BN layers are in the model structure. Moreover, the dropout layer between the PrimaryCaps layer and the ActivityCaps layer randomly discards some neurons with a probability of 0.5. These strategies can prevent the model from overfitting during training.

The optimizer is Adam, the initial learning rate is 0.001, the batch size is 128, and the number of training epochs is 100. In the process of model training, the method of dynamically adjusting the learning rate provided by PyTorch is applied to ensure that the model is closer to the optimal solution in the late training period.

The dynamic routing process retains the spatial relationship of features. The number of iterations of the dynamic routing is an important parameter of the MCN model. The number of iterations is set to 2, 3, and 4. The accuracy of 93.92%, 96.08%, and 94.09% is achieved on the test set, respectively, so the number of iterations of the dynamic routing is 3. The detailed structural parameters of the MCN model are shown in Table 1.

When the model training is completed, the test set is used for evaluation. The accuracy reaches 96.08%, which has achieved a good recognition result. Other evaluation indicators and CM are shown in Table 2 and Figure 10.

From Table 2, it can be seen that the values of precision, recall, and $F1$-score of each activity all exceed 0.9. For the $F1$-score, the minimum value is 0.929 for down-ramps and the maximum value is 0.990 for jogging. The average value of the $F1$-score is 0.960, which indicated that the MCN model proposed in this paper has achieved a good recognition result in this collected dataset.

Table 1: Detailed structural parameters of the MCN model.

Model	Layers	Parameters
MCN	Conv2d-1	Kernel size = 3×3 step = 1 filters = 32 padding = 1
	Conv2d-2	Kernel size = 3×3 step = 1 filters = 64
	Conv2d-3	Kernel size = 3×3 step = 1 filters = 128
	Primary caps	Kernel size = 2×2 step = 2 filters = 32×8
	Activity caps	Routing iterations = 3

Table 2: Evaluation metrics of the MCN model on the test set.

	Precision	Recall	$F1$-score
Walking	0.912	0.971	0.941
Jogging	0.986	0.995	0.990
Down-ramps	0.952	0.908	0.929
Up-ramps	0.951	0.944	0.947
Upstairs	0.982	0.975	0.978
Downstairs	0.987	0.962	0.974

In Figure 10, it is not difficult to find that 59 down-ramps samples are misidentified as walking and 41 up-ramps samples are misidentified as walking. So, down-ramps and up-ramps are easy to be confused with walking. The possible reason is that at the beginning and the end of these two activities, the slope becomes smaller, which makes it difficult to distinguish it from the level ground. There are also some down-ramps that are incorrectly identified as up-ramps and

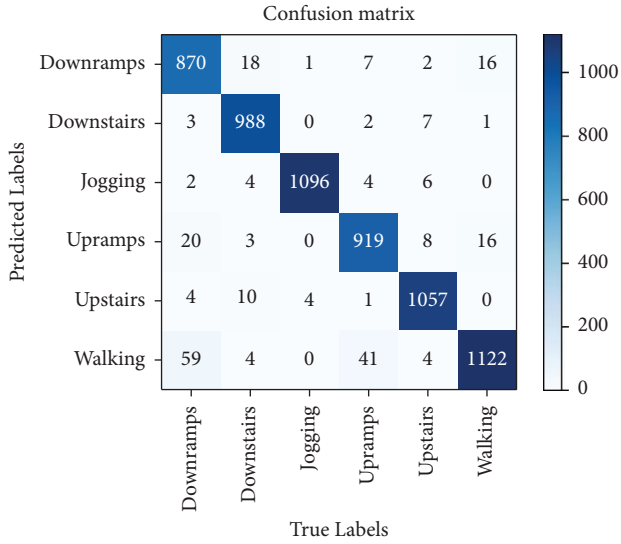

Figure 10: CM of the MCN model on the test set.

downstairs incorrectly identified as down-ramps, probably because these have similar features.

To display the original features and the features captured by each layer of the MCN, the t-SNE dimensionality reduction algorithm is applied for visualization [39]. The visualization results are shown in Figure 11.

In Figure 11, the dots of different colors represent the extracted features from different activities. It can be found that with the deepening of the network, similar activity features are gradually gathered. The original data are diffuse, and after three convolutional layers, activities of the same type show a tendency to cluster together. After passing through the CapsNet layer, the characteristics of the six activities are basically separated.

4.1.2. CNN Model Evaluation and Results. In order to verify the effectiveness of the MCN model, the CNN model is used for comparative experiments. The model structure of CNN is shown in Figure 3. In order to ensure the fairness of the comparative experiment, the structure of the first three 2D convolution layers is the same as the MCN. After three convolutional layers, the CNN model achieves dimensionality reduction through the max pooling layer, and the kernel size is 2×2 with a stride of 2. Then, two fully connected layers are connected to replace the CapsNet of the MCN. The dropout layer is also used in the fully connected layer, and some neurons are randomly discarded according to the probability of 0.5. Finally, the CNN model output the probability value of each activity through softmax. During training, except for using the cross entropy loss function, the rest of the settings are the same as the MCN, and finally, the accuracy on the test set is 91.62%, which is lower than 96.08% of MCN.

4.2. Results in the Public Datasets. To further validate the generalization capability of the MCN model, evaluation experiments are also performed on the public datasets WISDM and UCI-HAR.

4.2.1. WISDM Dataset. The window size is 128 and the step size is 64. After data preprocessing, 17,158 samples are generated. 70% of the samples are used for training and the rest are used for testing. The input dimension of the MCN model is 128×3. The optimizer, initial learning rate, and batch size are the same as the self-collected dataset. The detailed structural parameters are shown in Table 3. After 100 epochs of training, the accuracy rate on the test set is 98.21%. Other evaluation indicators and CM are shown in Table 4 and Figure 12. The t-SNE dimensionality reduction algorithm is also employed to visualize the features extracted by each layer, as shown in Figure 13.

In Table 4, the standing achieves the best recognition effect which all evaluation metrics are 1. The dataset is an imbalanced sample dataset, but the $F1$-score of each activity exceeds 0.9 and the average is 0.978, which further proves the effectiveness of the MCN model.

From Figure 12, it can be seen that the standing and the sitting are all correctly identified. 15 downstairs samples are misidentified as upstairs and 23 upstairs samples are misclassified as downstairs, probably, because the features of the two activities are too similar.

4.2.2. UCI-HAR Dataset. The sliding window size is 128 (2.56 s), and the overlap rate is 50%. Therefore, 10,299 samples are generated. Among them, 7,352 samples from 21 volunteers are used for training, and 2,947 samples from 9 volunteers are used for testing. The training set is input into the MCN model with the dimension of 128×9. The optimizer, initial learning rate, and batch size are the same as the self-collected dataset. The detailed structural parameters are shown in Table 5. After the same training, the accuracy rate of 95.28% is obtained on the test set. Other evaluation indicators and confusion matrix are shown in Table 6 and Figure 14. The feature visualization results of each layer are shown in Figure 15.

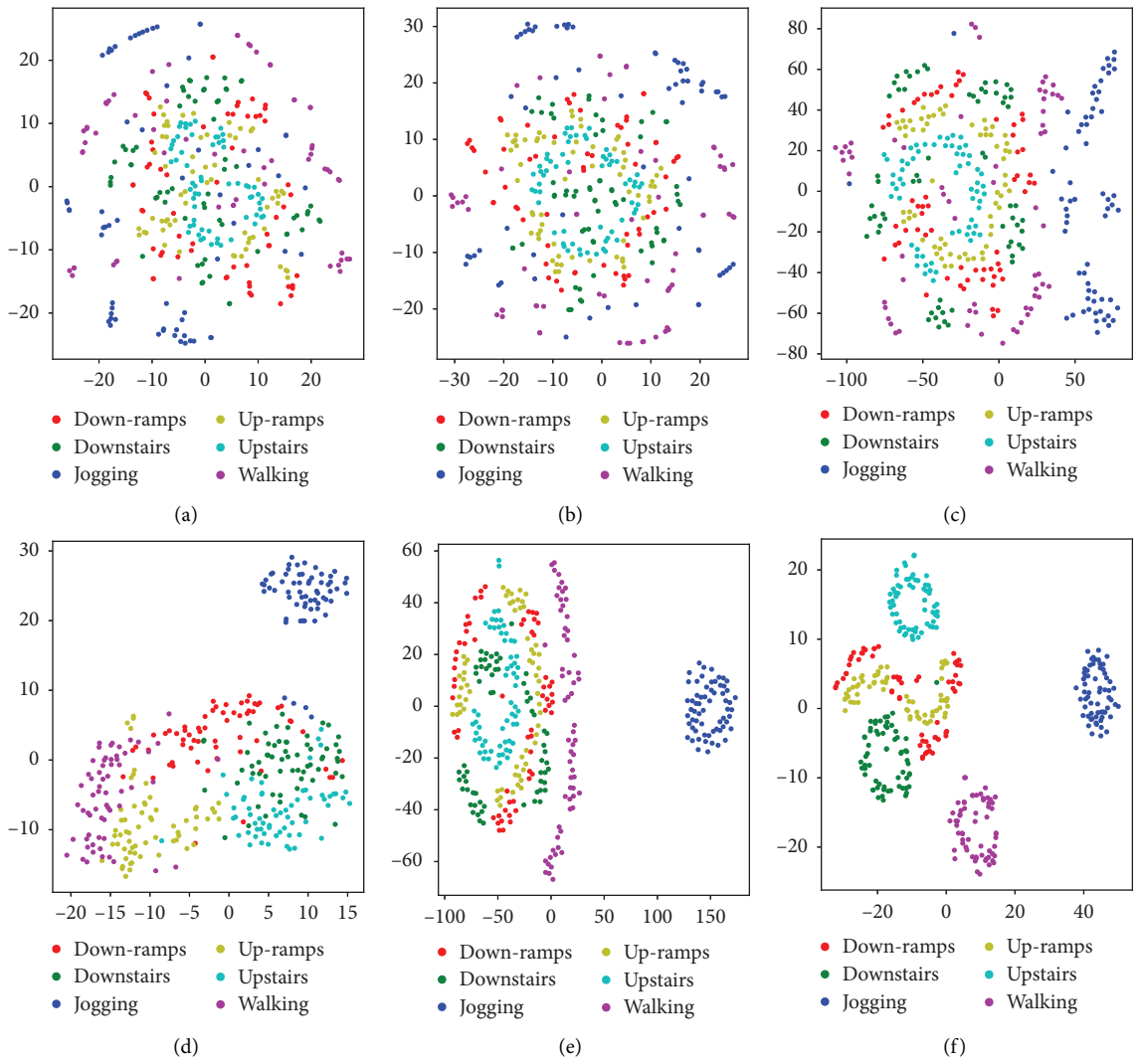

Figure 11: Features visualization of each layer of the MCN model via t-SNE: (a) origin, (b) conv 1, (c) conv 2, (d) conv 3, (e) primary caps, and (f) activity caps.

Table 3: Detailed structural parameters of the MCN model on the WISDM dataset.

Model	Layers	Parameters
MCN	Conv2d-1	Kernel size = 3 × 3 step = 1 filters = 32 padding = 1
	Conv2d-2	Kernel size = 3 × 3 step = 1 filters = 64 padding = 1
	Conv2d-3	Kernel size = 3 × 3 step = 1 filters = 128
	Primary caps	Kernel size = 2 × 2 step = 2 filters = 32 × 8
	Activity caps	Routing iterations = 3

Table 4: Evaluation metrics of the MCN model on the WISDM dataset.

	Precision	Recall	$F1$-score
Walking	0.992	0.993	0.992
Jogging	0.985	0.992	0.988
Sitting	0.997	1.000	0.998
Standing	1.000	1.000	1.000
Upstairs	0.963	0.923	0.943
Downstairs	0.937	0.960	0.948

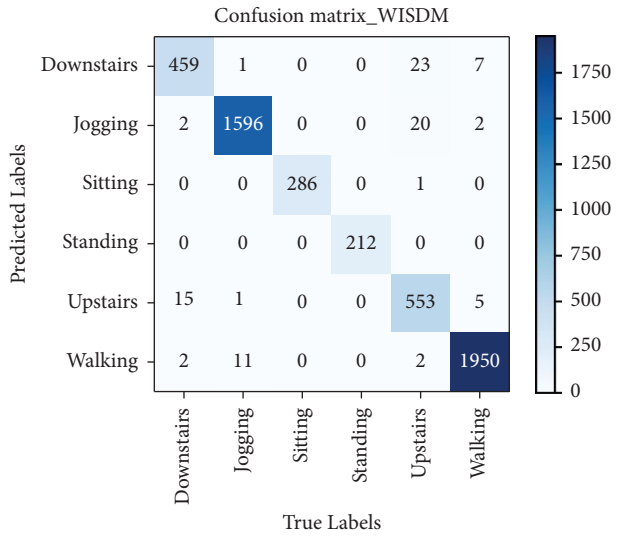

Figure 12: CM of the MCN model on the WISDM dataset.

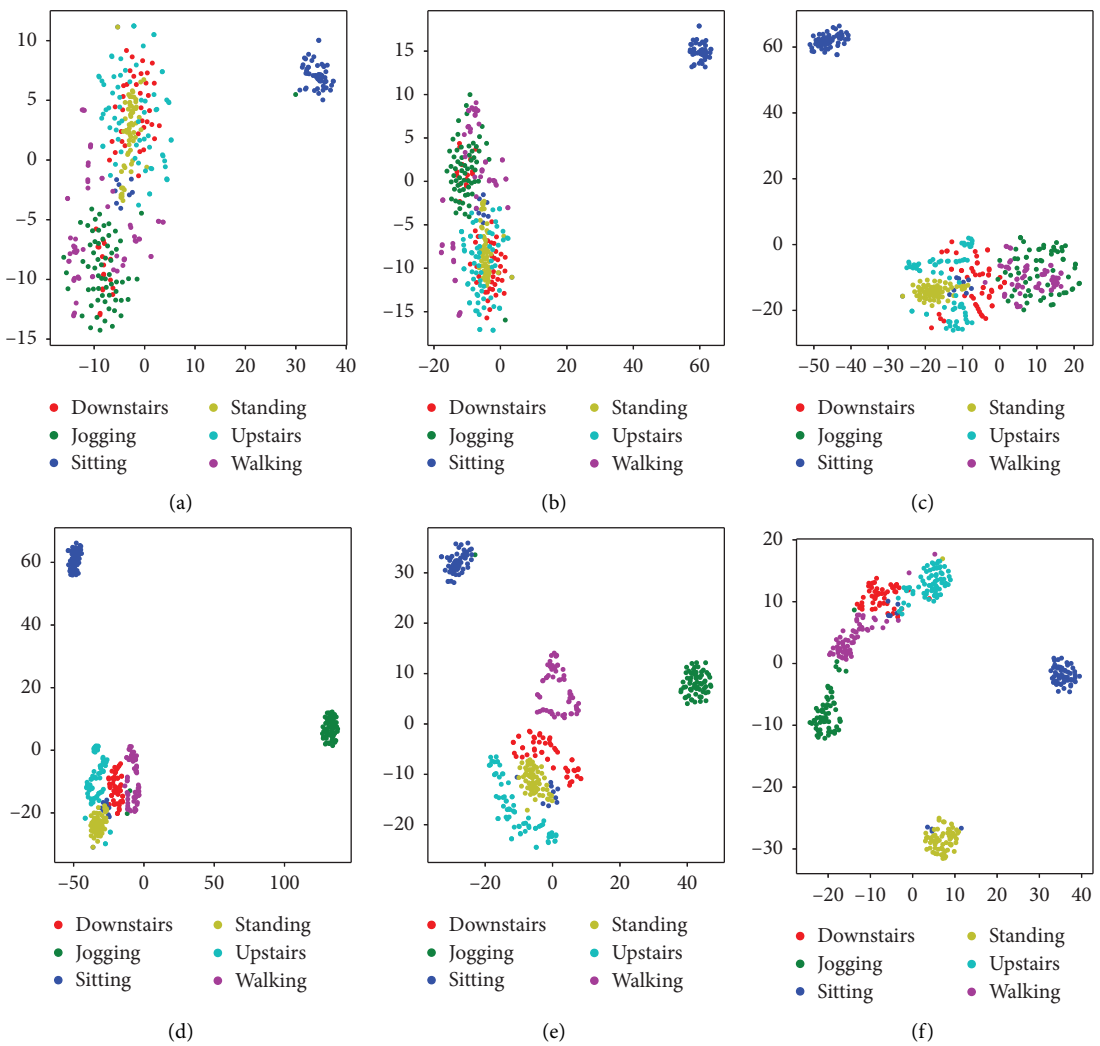

Figure 13: Features visualization of each layer of the MCN model via t-SNE on WISDM.

Table 5: Detailed structural parameters of the MCN model on the UCI-HAR dataset.

Model	Layer	Parameters
MCN	Conv2d-1	Kernel size = 3 × 3 step = 1 filters = 32
	Conv2d-2	Kernel size = 3 × 3 step = 1 filters = 64
	Conv2d-3	Kernel size = 3 × 3 step = 1 filters = 128
	Primary caps	Kernel size = 3 × 3 step = 3 filters = 32 × 8
	Activity caps	Routing iterations = 3

Table 6: Evaluation metrics of the MCN model on the UCI-HAR dataset.

	Precision	Recall	*F*1-score
Walking	0.998	0.986	0.992
Sitting	0.955	0.825	0.885
Standing	0.893	0.959	0.925
Laying	0.993	1.000	0.996
Upstairs	0.951	0.949	0.950
Downstairs	0.931	1.000	0.964

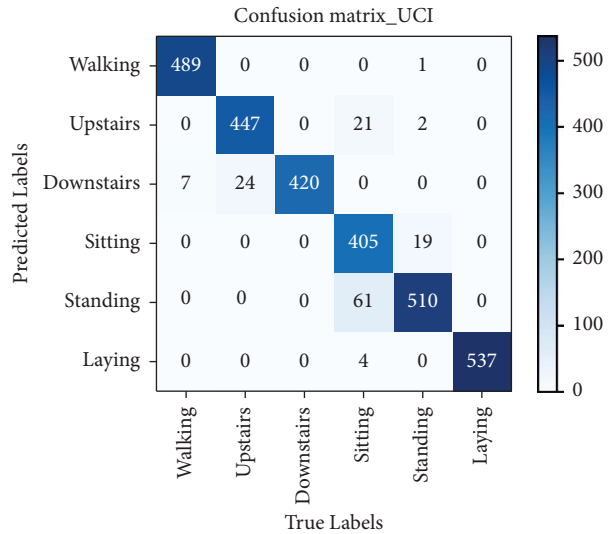

Figure 14: CM of the MCN model on the UCI-HAR dataset.

As can be seen from Table 6, the recall of laying and downstairs are all 1. The *F*1-score of sitting is 0.885, which is slightly lower, and the rest are also above 0.9. Walking has the highest *F*1-score of 0.992 and the average *F*1-score is 0.952.

In Figure 14, it can be seen that downstairs and laying are all correctly identified, but 61 sittings are misclassified as standing, 24 upstairs are misidentified as upstairs, and 21 sittings are misclassified as upstairs.

5. Discussion

In this study, the CapsNet is investigated to apply for HAR and a model based on a modified capsule network called MCN is proposed. To illustrate the effectiveness of the model, data collection is carried out using an IMU sensor and a dataset is built. This work is carried out under natural conditions and the volunteers moved in their own comfortable way under different terrain conditions. It is somewhat arbitrary and not collected in a controlled laboratory. Finally, the model achieved 96.08% accuracy on this dataset. The effectiveness of the MCN method is verified by a comparative experiment with CNN and the recognition effect is better than CNN. Table 7 lists the recognition methods and accuracy rates of some other researchers. As can be seen, this result is similar to other references. However, this accuracy is achieved with only one IMU sensor. And it does not involve the fusion of different types of sensor data.

In order to further verify the effectiveness of the MCN, experiments are also carried out on public datasets WISDM and UCI-HAR. Finally, the accuracy is 98.21% in WISDM and 95.28% in UCI-HAR. Table 8 lists the methods and results of some other studies on WISDM. Among them, CapsNet is also applied for HAR in references [29–31]. The accuracy is higher than references [29, 31] but slightly lower than reference [30].

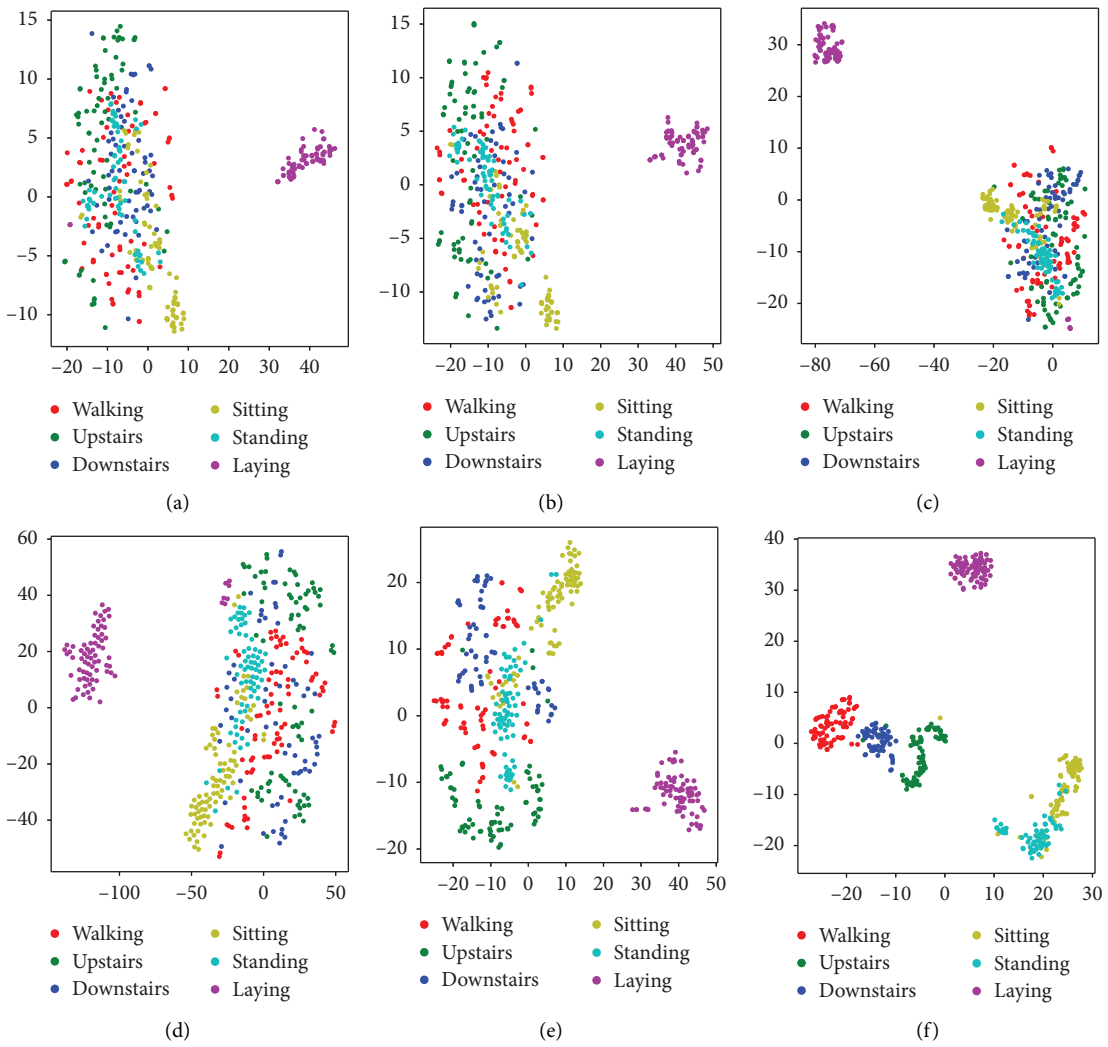

Figure 15: Features visualization of each layer of the MCN model via t-SNE on UCI-HAR.

Table 7: Comparison of the methods and experimental *results*.

References	Sensor	Model	Accuracy (%)
[21]	7 IMU	LSTM-CNN	97.78
[22]	4 IMU	DDLMI	97.64
[40]	IMU EMG	BP	93.76
[41]	EEG sEMG	EDMEFNet	88.44
[42]	8 sEMG	GA-DANN	94.89
[37]	5 IMU	LSTM	>95
This paper	1 IMU	MCN	96.08

Because the WISDM dataset is a very imbalanced dataset, the random SMOTE algorithm was employed in reference [30] to handle the imbalanced issue of the dataset, which is more conducive to training the network model. Table 8 also shows that the MCN model proposed in this study outperforms CNN, LSTM, and their combination on WISDM.

Table 9 lists the methods and results of some other studies on UCI-HAR. It can be seen that the MCN model also achieved higher accuracy than some other researchers. In summary, through a series of comparative experiments, it is verified that the proposed MCN model has high recognition accuracy and good robustness.

Table 8: A comparison of the proposed approach with other studies on WISDM.

Datasets	Model	Accuracy (%)
WISDM	DNN [23]	90
	Capsule and LoRa [29]	95.2
	1D-HARCapsNet [30]	98.67
	CapsGaNet [31]	96.8
	2D CNN [43]	89.67
	Multichannel CNN-GRU [44]	96.41
	LSTM-CNN [45]	95.85
	Multi-input CNN-GRU [46]	97.21
	DCNN [47]	94.18
	Multihead convolutional attention [48]	96.4
	MCN (this paper)	98.21

Table 9: A comparison of the proposed approach with other studies on UCI-HAR.

Dataset	Model	Accuracy (%)
UCI-HAR	CNN [49]	92.50
	Attention-based CNN [50]	93.16
	iSPLInception [51]	95.09
	Ensem-HAR [52]	95.05
	CNN-GRU [53]	94.50
	Fusing 2-D FFT and WVT [54]	93.45
	Deep CNN-LSTM with self-attention [55]	93.11
	CNN-BAOA [56]	95.23
	MCN (this paper)	95.28

6. Conclusion

In this paper, a deep learning model based on a modified capsule network named MCN is proposed. In comparison to traditional machine learning methods, this model can save the complicated process of manual feature extraction from raw IMU data. Compared with CNN and LSTM, this model preserves the spatial information of features, which may be more conducive to activity recognition. Contrast experiments have been conducted on three datasets to evaluate the effectiveness of the model. The first dataset is collected by ourselves, which is a balanced dataset collected under natural conditions using a single IMU sensor. The recognition accuracy of the proposed model is 96.08%, which is 4.46% higher than CNN. Moreover, the $F1$-score is 0.960. The second dataset is the public dataset named WISDM, which is an imbalanced dataset. The proposed model achieves an accuracy of 98.21% and an $F1$-score of 0.978. This accuracy is higher than most similar types of models. The third dataset is the public dataset named UCI-HAR, which is a balanced dataset. The proposed model achieves an accuracy of 95.28% and an $F1$-score of 0.952. Satisfactory results are obtained on the three datasets. Through the t-SNE dimensionality reduction algorithm, the extracted features of each layer by the MCN model are visualized. By comparing with the results of some other researchers, it further shows that the proposed MCN model can achieve higher recognition accuracy and have better activity detection ability.

The proposed model has achieved satisfactory performance in the experimental process, but has not been tested in real life. Therefore, in the future work, optimization and light weight of the model parameters will be considered to deploy the model in embedded devices to detect the actual recognition effect.

Data Availability

The websites of the WISDM and UCI-HAR datasets are https://www.cis.fordham.edu/wisdm/dataset.php;https://archive.ics.uci.edu/ml/datasets/Human+Activity+Recognition+Using+Smartphones.

Consent

Informed consent was obtained from all subjects involved in the study.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This research was funded by Tianjin Technical Expert Project with grant no. 22YDTPJC00480. The authors would like to thank the editors and reviewers for their valuable comments on this article and the volunteers who participated in the data collection.

References

[1] A. Mihoub and A. Nayyar, "A deep learning-based framework for human activity recognition in smart homes," *Mobile Information Systems*, vol. 2021, Article ID 6961343, pp. 1–11, 2021.

[2] S. Guo, H. Xiong, X. Zheng, and Y. Zhou, "Activity recognition and semantic description for indoor mobile localization," *Sensors*, vol. 17, no. 3, p. 649, 2017.

[3] M. A. M. Hasan, F. A. Abir, M. A. Siam, and J. Shin, "Gait recognition with wearable sensors using modified residual block-based lightweight CNN," *IEEE Access*, vol. 10, pp. 42577–42588, 2022.

[4] Y. Zheng, Q. Song, J. Liu, Q. Song, and Q. Yue, "Research on motion pattern recognition of exoskeleton robot based on multimodal machine learning model," *Neural Computing & Applications*, vol. 32, no. 7, pp. 1869–1877, 2019.

[5] D. Chen, Y. Cai, X. Qian et al., "Bring gait lab to everyday life: gait analysis in terms of activities of daily living," *IEEE Internet of Things Journal*, vol. 7, no. 2, pp. 1298–1312, 2020.

[6] M. E. Issa, A. M. Helmi, M. A. A. Al-Qaness, A. Dahou, M. Abd Elaziz, and R. Damaševičius, "Human activity recognition based on embedded sensor data fusion for the Internet of Healthcare Things," *Healthcare*, vol. 10, no. 6, p. 1084, 2022.

[7] A. M. Helmi, M. A. A. Al-Qaness, A. Dahou, R. Damasevicius, T. Krilavicius, and M. A. Elaziz, "A novel hybrid gradient-based optimizer and grey wolf optimizer feature selection method for human activity recognition using smartphone sensors," *Entropy*, vol. 23, no. 8, p. 1065, 2021.

[8] B. Degardin and H. Proença, "Human behavior analysis: a survey on action recognition," *Applied Sciences*, vol. 11, no. 18, p. 8324, 2021.

[9] Z. Meng, M. Zhang, C. Guo et al., "Recent progress in sensing and computing techniques for human activity recognition and motion analysis," *Electronics*, vol. 9, p. 1357, 2020.

[10] J. Sun, Y. Wang, J. Li, W. Wan, D. Cheng, and H. Zhang, "View-invariant gait recognition based on kinect skeleton feature," *Multimedia Tools and Applications*, vol. 77, no. 19, Article ID 24909, 2018.

[11] F. Demrozi, G. Pravadelli, A. Bihorac, and P. Rashidi, "Human activity recognition using inertial, physiological and environmental sensors: a comprehensive survey," *IEEE Access*, vol. 8, Article ID 210816, 2020.

[12] G. Şengül, E. Ozcelik, S. Misra, R. Damaševičius, and R. Maskeliūnas, "Fusion of smartphone sensor data for classification of daily user activities," *Multimedia Tools and Applications*, vol. 80, no. 24, Article ID 33527, 33546 pages, 2021.

[13] T. Hussain, N. Iqbal, H. F. Maqbool, M. Khan, M. I. Awad, and A. A. Dehghani-Sanij, "Intent based recognition of walking and ramp activities for amputee using sEMG based lower limb prostheses," *Biocybernetics and Biomedical Engineering*, vol. 40, no. 3, pp. 1110–1123, 2020.

[14] L.-F. Shi, C.-X. Qiu, D.-J. Xin, and G.-X. Liu, "Gait recognition via random forests based on wearable inertial measurement unit," *Journal of Ambient Intelligence and Humanized Computing*, vol. 11, pp. 5329–5340, 2020.

[15] D. Shin, S. Lee, and S. Hwang, "Locomotion mode recognition algorithm based on Gaussian mixture model using IMU sensors," *Sensors*, vol. 21, no. 8, p. 2785, 2021.

[16] X. Xi, W. Jiang, Z. Lü, S. M. Miran, and Z.-Z. Luo, "Daily activity monitoring and fall detection based on surface electromyography and plantar pressure," *Complexity*, vol. 2020, Article ID 9532067, pp. 112, 2020.

[17] M. Zeng, L. T. Nguyen, B. Yu et al., "Convolutional neural networks for human activity recognition using mobile sensor," in *Proceedings of the 6th International Conference on Mobile Computing, Applications and Services (MobiCASE)*, Austin, TX, USA, April 2014.

[18] G. E. Hinton, R. R. Salakhutdinov, and R. R. Salakhutdinov, "Reducing the dimensionality of data with neural networks," *Science*, vol. 313, no. 5786, pp. 504–507, 2006.

[19] F. J. Ordonez and D. Roggen, "Deep convolutional and LSTM recurrent neural networks for multimodal wearable activity recognition," *Sensors*, vol. 16, no. 1, p. 115, 2016.

[20] H. F. Nweke, Y. W. Teh, M. A. Al-garadi, and U. R. Alo, "Deep learning algorithms for human activity recognition using mobile and wearable sensor networks: state of the art and research challenges," *Expert Systems with Applications*, vol. 105, pp. 233–261, 2018.

[21] C.-f. Chen, Z.-j. Du, L. He, Y.-j. Shi, J.-q. Wang, and W. Dong, "A novel gait pattern recognition method based on LSTM-CNN for lower limb exoskeleton," *Journal of Bionics Engineering*, vol. 18, no. 5, pp. 1059–1072, 2021.

[22] L. Zhu, Z. Wang, Z. Ning et al., "A novel motion intention recognition approach for soft exoskeleton via IMU," *Electronics*, vol. 9, no. 12, p. 2176, 2020.

[23] V. B. Semwal, N. Gaud, P. Lalwani, V. Bijalwan, and A. K. Alok, "Pattern identification of different human joints for different human walking styles using inertial measurement unit (IMU) sensor," *Artificial Intelligence Review*, vol. 55, no. 2, pp. 1149–1169, 2021.

[24] B. Hu, S. Li, Y. Chen, R. Kavi, and S. Coppola, "Applying deep neural networks and inertial measurement unit in recognizing irregular walking differences in the real world," *Applied Ergonomics*, vol. 96, Article ID 103414, 2021.

[25] V. B. Semwal, A. Gupta, and P. Lalwani, "An optimized hybrid deep learning model using ensemble learning approach for human walking activities recognition," *The Journal of Supercomputing*, vol. 77, no. 11, pp. 12256–12279, 2021.

[26] F. Bozkurt, "A comparative study on classifying human activities using classical machine and deep learning methods," *Arabian Journal for Science and Engineering*, vol. 47, no. 2, pp. 1507–1521, 2021.

[27] S. Sabour, N. Frosst, and G. E. Hinton, "Dynamic routing between capsules," in *Proceedings of the Advances in Neural Information Processing Systems 30: Annual Conference on Neural Information Processing Systems 2017*, pp. 3859–3869, Long Beach, CA, USA, December 2017.

[28] C. Pham, S. Nguyen-Thai, H. Tran-Quang et al., "SensCapsNet: deep neural network for non-obtrusive sensing based human activity recognition," *IEEE Access*, vol. 8, pp. 86934–86946, 2020.

[29] L. Shi, H. Xu, W. Ji, B. Zhang, X. Sun, and J. Li, "Real-time human activity recognition system based on capsule and LoRa," *IEEE Sensors Journal*, vol. 13, p. 1, 2020.

[30] H. Khaled, O. Abu-Elnasr, S. Elmougy, and A. S. Tolba, "Intelligent system for human activity recognition in IoT environment," *Complex & Intelligent Systems*, vol. 19, pp. 1–12, 2021.

[31] X. Sun, H. Xu, Z. Dong et al., "CapsGaNet: deep neural network based on capsule and GRU for human activity recognition," *IEEE Systems Journal*, vol. 16, no. 4, pp. 5845–5855, 2022.

[32] D. Li, M. Zhang, T. Kang et al., "Fault diagnosis of rotating machinery based on dual convolutional-capsule network (DC-CN)," *Measurement*, vol. 187, Article ID 110258, 2022.

[33] Y. Lecun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," *Proceedings of the IEEE*, vol. 86, pp. 2278–2324, 1998.

[34] S. Ioffe and C. Szegedy, "Batch normalization: accelerating deep network training by reducing internal covariate shift," in *Proceedings of the 32nd International Conference on Machine Learning*, pp. 448–456, Beijing China, June 2015.

[35] J. R. Kwapisz, G. M. Weiss, and S. A. Moore, "Activity recognition using cell phone accelerometers," *ACM SIGKDD Explorations Newsletter*, vol. 12, no. 2, pp. 74–82, 2011.

[36] D. Anguita, A. Ghio, L. Oneto, X. Parra, and J. L. Reyes-Ortiz, "A public domain dataset for human activity recognition using smartphones," *Esann*, vol. 3, p. 3, 2013.

[37] F. Sherratt, A. Plummer, and P. Iravani, "Understanding LSTM network behaviour of IMU-based locomotion mode recognition for applications in prostheses and wearables," *Sensors*, vol. 21, no. 4, p. 1264, 2021.

[38] L. Wang, Y. Sun, Q. Li, T. Liu, and J. Yi, "Two shank-mounted IMUs-based gait analysis and classification for neurological disease patients," *IEEE Robotics and Automation Letters*, vol. 5, no. 2, pp. 1970–1976, 2020.

[39] V. D. M. Laurens and G. Hinton, "Visualizing data using t-SNE," *Journal of Machine Learning Research*, vol. 9, pp. 2579–2605, 2008.

[40] J. Wang, Y. Dai, and X. Si, "Analysis and recognition of human lower limb motions based on electromyography (EMG) signals," *Electronics*, vol. 10, no. 20, p. 2473, 2021.

[41] K. Shi, F. Mu, R. Huang et al., "Multimodal human-exoskeleton interface for lower limb movement prediction

through a dense Co-attention symmetric mechanism," *Frontiers in Neuroscience*, vol. 16, Article ID 796290, 2022.

[42] P. Zhang, J. Zhang, and A. Elsabbagh, "Lower limb motion intention recognition based on sEMG fusion features," *IEEE Sensors Journal*, vol. 22, no. 7, pp. 7005–7014, 2022.

[43] A. Prasad, A. K. Tyagi, M. M. Althobaiti, A. Almulihi, R. F. Mansour, and A. M. Mahmoud, "Human activity recognition using cell phone-based accelerometer and convolutional neural network," *Applied Sciences*, vol. 11, no. 24, Article ID 12099, 2021.

[44] L. Lu, C. Zhang, K. Cao, T. Deng, and Q. Yang, "A multichannel CNN-GRU model for human activity recognition," *IEEE Access*, vol. 10, pp. 66797–66810, 2022.

[45] K. Xia, J. Huang, and H. Wang, "LSTM-CNN architecture for human activity recognition," *IEEE Access*, vol. 8, pp. 56855–56866, 2020.

[46] N. Dua, S. N. Singh, and V. B. Semwal, "Multi-input CNN-GRU based human activity recognition using wearable sensors," *Computing*, vol. 103, no. 7, pp. 1461–1478, 2021.

[47] K. Peppas, A. C. Tsolakis, S. Krinidis, and D. Tzovaras, "Real-time physical activity recognition on smart mobile devices using convolutional neural networks," *Applied Sciences*, vol. 10, no. 23, p. 8482, 2020.

[48] H. Zhang, Z. Xiao, J. Wang, F. Li, and E. Szczerbicki, "A novel IoT-perceptive human activity recognition (HAR) approach using multihead convolutional attention," *IEEE Internet of Things Journal*, vol. 7, no. 2, pp. 1072–1080, 2020.

[49] V. Bianchi, M. Bassoli, G. Lombardo, P. Fornacciari, M. Mordonini, and I. De Munari, "IoT wearable sensor and deep learning: an integrated approach for personalized human activity recognition in a smart home environment," *IEEE Internet of Things Journal*, vol. 6, no. 5, pp. 8553–8562, 2019.

[50] K. Wang, J. He, and L. Zhang, "Attention-based convolutional neural network for weakly labeled human activities' recognition with wearable sensors," *IEEE Sensors Journal*, vol. 19, no. 17, pp. 7598–7604, 2019.

[51] M. Ronald, A. Poulose, and D. S. Han, "iSPLInception: an inception-ResNet deep learning architecture for human activity recognition," *IEEE Access*, vol. 9, pp. 68985–69001, 2021.

[52] D. Bhattacharya, D. Sharma, W. Kim, M. F. Ijaz, and P. K. Singh, "Ensem-HAR: an ensemble deep learning model for smartphone sensor-based human activity recognition for measurement of elderly health monitoring," *Biosensors*, vol. 12, no. 6, p. 393, 2022.

[53] O. Nafea, W. Abdul, and G. Muhammad, "Multi-sensor human activity recognition using CNN and GRU," *International Journal of Multimedia Information Retrieval*, vol. 11, no. 2, pp. 135–147, 2022.

[54] S. Zebhi, "Human activity recognition using wearable sensors based on image classification," *IEEE Sensors Journal*, vol. 22, no. 12, pp. 12117–12126, 2022.

[55] M. A. Khatun, M. A. Yousuf, S. Ahmed et al., "Deep CNN-LSTM with self-attention model for human activity recognition using wearable sensor," *IEEE J Transl Eng Health Med*, vol. 10, pp. 1–16, 2022.

[56] A. Dahou, M. A. A. Al-qaness, M. Abd Elaziz, and A. Helmi, "Human activity recognition in IoHT applications using Arithmetic Optimization Algorithm and deep learning," *Measurement*, vol. 199, Article ID 111445, 2022.