



Research Article

Hybrid Deep-Learning Framework Based on Gaussian Fusion of Multiple Spatiotemporal Networks for Walking Gait Phase Recognition

Tao Zhen,^{1,2} Jian-lei Kong^{1,3,4} and Lei Yan^{1,2}

¹School of Artificial Intelligence, Beijing Technology and Business University, Beijing 100048, China

²School of Engineering, Beijing Forestry University, Beijing 100083, China

³National Key Laboratory of Environmental Protection Food Chain Pollution Prevention, Beijing 100048, China

⁴China Light Key Laboratory of Industry Internet and Big Data, Beijing 100048, China

Correspondence should be addressed to Jian-lei Kong; kongjianlei@btbu.edu.cn and Lei Yan; mark_yanlei@bjfu.edu.cn

Received 22 May 2020; Revised 15 August 2020; Accepted 28 August 2020; Published 9 October 2020

Academic Editor: Carlos-Arturo Loredo-Villalobos

Copyright © 2020 Tao Zhen et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Human gait phase detection is a significance technology for robotics exoskeletons control and exercise rehabilitation therapy. Inertial Measurement Units (IMUs) with accelerometer and gyroscope are convenient and inexpensive to collect gait data, which are often used to analyze gait dynamics for personal daily applications. However, current deep-learning methods that extract spatial and the isolated temporal features can easily ignore the correlation that may exist in the high-dimensional space, which limits the recognition effect of a single model. In this study, an effective hybrid deep-learning framework based on Gaussian probability fusion of multiple spatiotemporal networks (GFM-Net) is proposed to detect different gait phases from multisource IMU signals. Furthermore, it first employs the gait information acquisition system to collect IMU data fixed on lower limb. With the data preprocessing, the framework constructs a spatial feature extractor with AutoEncoder and CNN modules and a multistream temporal feature extractor with three collateral modules combining RNN, LSTM, and GRU modules. Finally, the novel Gaussian probability fusion module optimized by the Expectation-Maximum (EM) algorithm is developed to integrate the different feature maps output by the three submodels and continues to realize gait recognition. The framework proposed in this paper implements the inner loop that also contains the EM algorithm in the outer loop and optimizes the reverse gradient in the entire network. Experiments show that this method has better performance in gait classification with accuracy reaching more than 96.7%.

1. Introduction

Robotics exoskeletons has become a burgeoning technology in continuous development in the field of medical, architectural, and military applications. Focusing on the medical domain, lower limb exoskeletons are mainly designed to enhance the patient's mobility in the rehabilitation therapy and strengthen physical performance after undergoing treatment with great expectations of improving his/her living quality as much as possible. This robot should own an intelligent gait phase recognition method to provide cogent means to deal with large amounts of momentary or sequential data and identify different walking styles, one of the

most important features displaying posture and phase of each particular patient [1]. Therefore, accurate classification of changing walking style of human lower limbs' status is urgently required to achieve consistency and coordination of human-machine interaction [2]. An effective analysis of walking style is performed well in athletic performance improvement or disease diagnosis and rehabilitation research, which have been applied in clinical treatment plan with multiple sclerosis, Parkinson, brain trauma, and other diseases [3, 4]. Note that traditional walking analysis is represented by detecting the different gait phase based on motion information (e.g., angles, speed, or acceleration) of the knees, ankles, and hips when walking or running. For

example, Buckley et al. [3] used gait phase detection to diagnose stroke patients, and Achanta et al. [5] used a novel hidden-Markov-based adaptive dynamic time warping to analysis gait for identifying physically challenged persons and providing them with appropriate alerts by monitoring walking. In order to make exoskeleton robots have better human-machine coordination, some researchers have begun to try program human robots, generate gait trajectories of wearers through gait phase recognition technology, and control the movement of wearable auxiliary devices, such as robotic prostheses and orthotics [6, 7]. Berger et al. [8] reported that robot-assisted gait training has proven to be a promising treatment for restoring and improving walking ability. Similarly, Luo et al. [9] also mentioned that the high-quality gait-subphase recognition plays an important role in the synergetic control of lower limb powered exoskeletons.

In recent years, walking gait phase detection has been one of the most important research problems. Many scholars have studied the relevant sensor technologies and methods to distinguish attributes for gait phase recognition. Jin et al. [10] realized the recognition of the frontal gait phase by using deep-learning models to obtain information from multiple cameras. Although the optical gait analysis technology has obtained the accurate enough performance, the full cost of multicamera system is very high for simple walking gait identification. Moreover, the operation process captures spatial trajectory of marking points related to body positions in the indoor infrastructure, which make the identification accuracy easily affected by the instable light intensity and limit the generic application of gait phase detection. To better adapt to environmental factors, the measuring technology of biological signals such as electromyography (EMG) is introduced to evaluate the continuous movement status of the human body. With EMG sensors fixed on the skin surface, the action potential changes are observed after being amplified to reflect muscle activity information when the central nervous system controls lower limb motion. Since the muscle contraction process has a time effect of electromechanical delay, those EMG signals are generated about 40 ms to 100 ms faster than lower limb motion, which is conducive for understanding motion pattern and recognizing gait phases in advance, also making it possible for the real-time control system of exoskeletons. Fei et al. [11] took the multisource EMG signals on the hip joint angle and the skin surface of thigh muscles to realize prediction of the human lower limb motion. To further improve the prediction effect of the model on the gait phase, some studies have also added the plantar pressure signal to the input of the neural network model. Si et al. [12] propose a support vector machine model optimized by fractal analysis algorithm to cope with synchronized information of EMG and foot pressure sensors for gait phase recognition. However, such an acquisition approach of EMG data asks participants wearing a special cumbersome cloth with expensive cost, which is subject to observers' intra- and external variability including weight, muscle content, or load, making it difficult to design progressive therapeutic strategies to improve automatic control performance of lower limb exoskeletons.

At present, many researchers have paid attention to more flexible wearable sensors, typically as IMUs, to handle with the relative issues of gait phase detection. These sensors are relatively lightweight, cheap, easy to use, and unobtrusive compared with aforementioned motion capture systems consisting of RGB-D cameras, EMG, or force plates sensors. With some simple constrained equipment fixed on waist, thigh, calf, or foot, IMUs sensors can provide continuous and high-resolution inertial data, which can well quantify gait phase recognition performance, which is basically unaffected by various human body factors. It is not surprising that there exist the developing trends of human motion analysis using IMUs, for instance, gait phase detection by statistical analysis or machine learning and exoskeleton decision-making control by postural stability metrics. For instance, Kang et al. [13] used inertial information obtained from the IMU on the calf to study whether cognitive impairment increases the fall risk of patients. Gohar et al. [14] used inertia information on the chest to realize ID identification of personnel. However, the local information provided by a single IMU sensor is limited to descript the complexity exercise process of human body. More and more scholars have begun to explore the combined use of inertial and acceleration information from multiple body parts to enhance the reliability and accuracy of gait event detection. For example, Zhou et al. [15] used the inertial information obtained by the IMU on the legs and feet to achieve gait assistance. Yeo and Park [16] used the inertial information of the shin and leg surface through the IMU to provide gait measurements and enable accurate analysis. In another example, Yan et al. [1] used the inertial information on the feet, thighs, and calves to accurately identify the gait phase. The above researches have exhibited that only using a small number of IMU sensors located on the lower limb can provide extremely cost-effective and efficient signals to characterize the periodic cycle of the gait phase and to discriminate different levels of walking ability. Importantly, this sensor application is harmless, comfort, and convenient to the personal activities and daily work, which will help clinicians make more accurate judgments about the early intervention of lower extremity diseases, the treatment plan, and the assessment of the rehabilitation progress of patients.

Although researchers have always been interested in IMU technology, there is still a lack of realistic research that implements multiple wearable sensors validators to monitor walking status and gait stability in actual clinical practice. In the context of multi-IMU-based approaches, the intelligent enough pattern recognition model of diversity gait phases with multisensor data preprocessing analysis and information fusion technology is the most critical issue that needs to be solved for lower limb gait phase detection of various wearable exercise systems or rehabilitation exoskeletons. To achieve this, some statistical learning or machine learning methods are carried to calculate spatial-temporal and biomechanical parameters of the walking gait patterns, achieving a complete assessment of lower limb motion, and providing the potential for rehabilitation training of abnormal body activities. Many researchers have studied the integrated strategy of traditional intelligent algorithms from

different perspectives. The mainstream solution is constructing various shallow-structure models including artificial neural networks (ANN), hidden-Markov model (HMM), AdaBoosting, and support vector machines (SVMs), which carefully select the threshold parameters based on physical and statistical analysis of raw or processed data to divide the gait phases. These methods build feature engineering to adaptively learn model parameters and obtain hidden relationships between historical data and then perform operations based on preset rules and algorithms to identify the subsequent patterns of gait phases. However, the gait phase detecting is still a challenging problem because the high-sampling frequency data collected from sensors always contain complex nonlinear relationships with multiple components, which makes it impossible to apply the traditional models to analyze sensory data and distinguish walking information in real time.

Different from the aforementioned algorithms, the deep-learning neural networks (DNNs) had shown the outstanding ability of handling with the complex temporal relation of gait phase detection. Thanks to the breakthroughs in the design and training of model architecture with complex structures consisting of multiple processing layers or nonlinear transformations, the improved network structure has penetrated into many smart devices including large scale visual classification, natural language processing, and time series predicting. Such fast-pacing progresses in research have also drawn attention of the related researchers and corporations to build software and hardware to recognize walking gait phases snapshotted in real life. Particularly, the convolutional neural network (CNN) and the recurrent neural network (RNN) have been used for extracting the motion features of sequential temporal data obtained by accelerometers and gyroscopes in IMU. For example, Chao et al. [17] Yan use a multitask framework to extract features and perform gait, perspective, and scene recognition. They use the output of gait energy image descriptors as input data for a CNN to detect gait phases in the complete walking cycles. Similarly, Omid et al. [18] use time-frequency expansion in order to capture joint two-dimensional spectral and temporal patterns of gait cycles, which are inserted to train an ensemble CNNs-based classifier, a typical multilayer perceptron consisting of convolutional and full-connection based on multisensor fusion. In addition, the authors in [19] propose a different voting-weighted integrated neural network for training a model for gait recognition, obtaining state-of-the-art results.

Due to the talent in handling two-dimensional signals such as the images, the most CNNs have to translate the time series inertial data as energy image or visual segmentation data. This does help to exploit the characteristics of spatial relationships in gait phase recognition, yet obviously ignoring temporal rule and periodic change when processing sequential time series data captured by IMU sensors, which is often hard to measure continuous motion trajectory and extract quality features of lower limb in unconstrained scenarios. Therefore, RNN and its improved models including long short-term memory (LSTM) network and gated recurrent unit (GRU) network pass time recording order

and parameters of previous hidden layer thought the current output layer to capture the high nonlinearity and sequential relationship of time-serial IMU data, which have attracted extensive attention from researchers [20]. Neverova et al. [21] build a temporal RNN network for active biometric authentication and walking motion analysis with multi-source data provided by accelerometers and gyroscope sensors in smartphone. As an improved version of the RNN, the LSTM is gradually replacing it as a new popular time series data analysis technology for gait phase recognition. For example, Hu et al. [22] trained a deep-learning network with LSTM units to process IMU data segmented by sliding windows and implement gait phase detection. Similar, Zhen et al. [23] proposed an LSTM-based recognition algorithm to perform real-time gait phase detection using absolute heading and angular velocity of IMU sensor mounted on the shank and foot. Although related researches have increased significantly in recent years, it is still difficult to accurately predict the current phase with LSTM alone through long-term sensor data. To this end, there have only been a small number of studies attempting to combine LSTM and CNN to assist patients with severe gait abnormalities and related ethical issues. Jin et al. [24] proposed a deep-learning algorithm based on a LSTM and CNN fusion framework for diagnosis and classification of abnormal gait patterns using Euler angle information of IMU sensor on the patient's legs. In this way, CNN is usually used as a spatial feature extractor and then LSTM is used to further mine the temporal features for ultima gait phases detection. In addition, the GRU unit inherits the advantages of LSTM and can automatically learn features and is an effective model [25], and the AutoEncoder unit also exhibits a significant increase in computational speed and model size compared to the existing deep-learning models [26]. Both of them are introduced as alternative patterns parts of various hybrid models based on deep-learning in many application scenarios, which have been proven effective at improving prediction performance of gait phase recognition for nonlinear time series IMU data.

In the previous studies mentioned above, it can be found that, with the development of wearable IMU biotechnology, the human motion status can be estimated by a comprehensive combination of signal processing approaches and intelligent pattern recognition algorithms to extract quantitative features of walking gait and distinguish the categories of different gait phases. For a realistic implementation of exercise rehabilitation therapy in the clinical setting, various machine learning algorithms have been developed to handle with a mass of time series data offered by multisource IMU sensors on the lower limb for quantifying gait phase and balance. Particularly, because of the excellent ability in high-dimensional feature representation of spatial and temporal characteristics, many sorts of deep-learning neural network are selected for their relevance to the sensor data types and used together in a parallel or serial structure adaptive to data structures, which have been proven effective at improving the performance for walking gait phase detection. However, the combination framework of these networks still needs to be developed to obtain better prediction effect and recognition for the nonlinear time series IMU data in order to

meet the requirement of effective rehabilitation therapy and real-time exoskeleton control.

Researchers agree that one of the reasons for the decline in recognition performance is that the data from different IMU sensors always contain multiple component signals, which is complex nonlinearity to a single deep-learning model or simple combination of different networks [27]. On the one hand, each single network is likely to focus on the constant feature of specific domain, such as spatial feature related to CNN and temporal feature associated with RNN, which lacks the capability of distinguishing small variances between similar gait phases in a global perspective. On the other hand, the simple combination solution of multiple networks lacks the guidance of effective fusion strategy and fails in leverage information complementation and comprehensive decision-making, which will reduce the overall classification accuracy especially when low-quality interference noise occurs. To augment algorithm performance currently used in the IMU-based gait phase recognition, we propose an effective hybrid deep-learning framework based on Gaussian probability fusion of multiple spatiotemporal networks (named as GFM-Net) for recognizing discriminative parts of various walking gait phases. In detail, the framework consisted of three components: a spatial feature extractor with AutoEncoder and CNN modules and temporal feature extractor with three collateral modules combining RNN, LSTM, and GRU modules. In addition, a novel classifier equipped with Gaussian probability fusion module optimized by Expectation-Maximum (EM) algorithm is developed to integrate different feature maps of components model for the ultimate gait phase recognition [28]. Different from the previous studies, the end-to-end network adaptively selects spatiotemporal feature vectors from different IMU sensors and absorbs the vast quantities of hybrid complementary knowledge available in the training corpora, which show the better promotion effect in terms of gait phase recognition accuracy in sequential walking cycle. Such an approach would likely help exoskeletons make informed control decisions about patients' treatment efficacy and recovery progress.

The remainder of this study is organized as follows. Section 2 introduces the data source and preprocess technology and then describes each part of the hybrid framework. Section 3 presents experimental results of the proposed model evaluated with correlation comparison methods. Section 4 discusses the advantages and disadvantages of our work, and finally Section 5 presents our conclusion.

2. Materials and Methods

2.1. Data Collection. In terms of experimental data, 16 volunteers with body weight ranging from 46 kg to 70 kg and height ranging from 158 cm to 177 cm were selected to collect IMU data. The height and weight distribution of the subjects is shown in Figure 1. The subjects are all healthy participants and have no physical or nerve injury to their legs or feet, which may affect walking gait phase detection. In

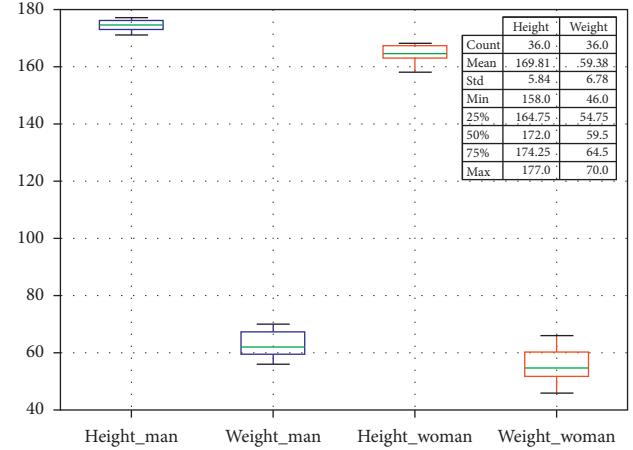


FIGURE 1: Weight and height information about volunteers participating.

addition, it should be noted that the participants are between 20 and 26 years old.

With the advancement of sensor processing technology and algorithms, this study used three IMU modules to collect the corresponding inertial information. Input data in this work only include lower leg calf acceleration signals. To collect lower limb calf acceleration signals, the JY901 nine-axis angle sensor (Uxin Electronics Co., Ltd., Gansu, China) with Kalman filtering algorithm is used in this paper. We can choose two modes of serial communication and I2C communication. In order to cooperate with the microprocessor, we chose the serial communication mode when the system was built. Connect the TX, RX, VCC, and GND pins on the JY901 sensor to the corresponding pins on the microcontroller. The microcontroller selected is STM32C8T6, which is a 32-bit microcontroller based on the ARM Cortex-M core STM32 series, the program memory capacity is 64 KB, the required voltage is 2V~3.6V, the operating temperature is -40°C~85°C, and the operating frequency is 72 MHz.

The inertial sensor module is placed outside the lower leg. The arrangement of acceleration sensors for calf monitoring lower limb movement is shown in Figure 2 that shows the system flow of the entire experimental data collection, processing, and application. The acceleration resolution of the nine-axis inertial sensor module (MPU9250) used in the experiment is 0.0005 g, the stability of the attitude measurement is 0.05°, and the transmission baud rate in the experiment is set to 115200 bps.

During the experiment, all participants were required to walk normally on the same treadmill at a speed of 0.78 m/s, 1.0 m/s, and 1.25 m/s for at least 120 s. All participants were asked to walk normally 3 times at each speed. All participants have the same sports environment in the same state. In order to prevent participants from affecting the later movement gait due to continuous exercise, the experiment requires all participants to rest for 2 minutes after completing the designated walking test each time to alleviate the possible impact of exercise fatigue on walking gait. In addition, when collecting data, it should be noted that we only

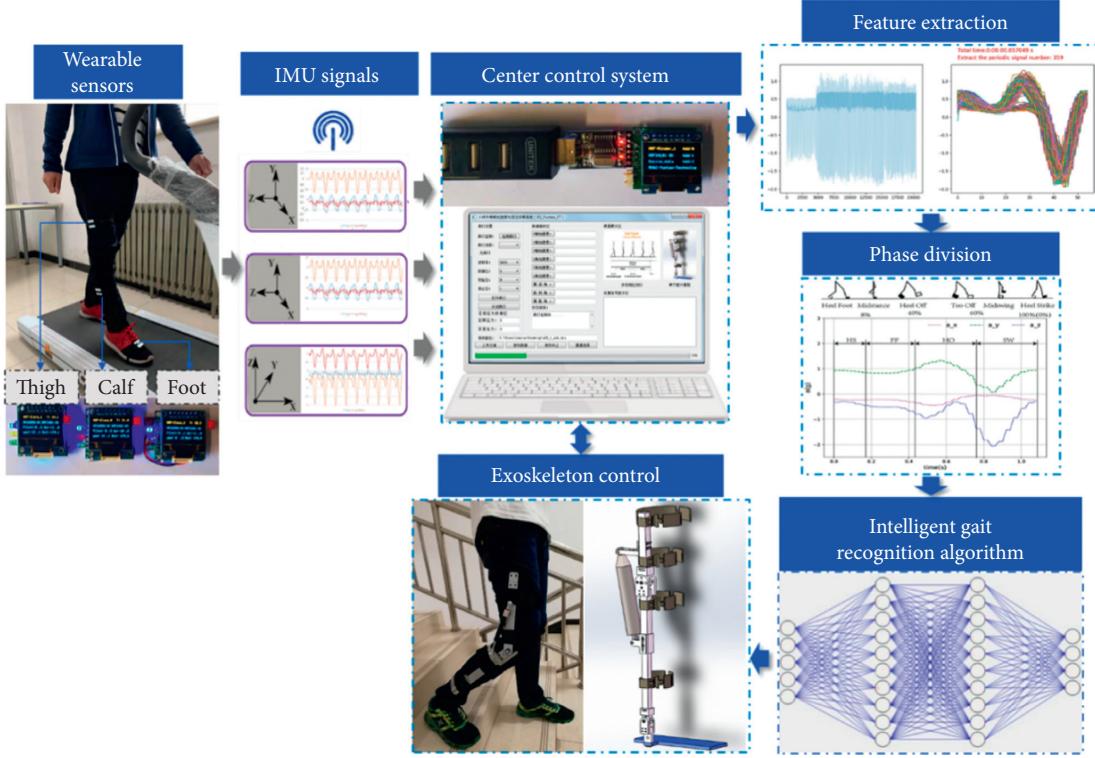


FIGURE 2: Human gait information acquisition system.

start saving data after the running speed of the treadmill reaches the set speed. When the treadmill starts to slow down, we stop collecting data and complete the data collection process. The information acquisition process of human gait is shown in Figure 3.

2.2. Data Preprocessing. Each data sample contains multiple features from three sensors and each data includes acceleration and angular velocity data in X , Y , and Z directions. Let two sequences of gait data be the input of the network, which is expressed as

$$\vec{x} = (a_x, a_y, a_z, g_x, g_y, g_z), \quad (1)$$

where a_x , a_y , and a_z represent the acceleration signals in the X , Y , and Z directions, respectively, and g_x , g_y , and g_z represent the angular velocity in the three-dimensional space, respectively. Based on the above operation, we can get the combined curve of the acceleration and angular velocity in the X , Y , and Z directions obtained by three IMU sensors as shown in Figure 2.

Then, the second preprocess step is to extract the periodic feature by splitting the data. The data collected by the inertial sensors is a data stream distributed over time. Features cannot be directly extracted and classified, so the data needs to be segmented. At present, the data segmentation method is a multidimensional sliding window segmentation method. The acceleration signal is cut into periodic signals by using a sliding window segmentation method [29]. However, this signal extraction method

requires a lot of tentative experiments to try different sizes of sliding windows on the extraction of gait signals, which is difficult for us to guarantee the quality of the extracted signals. In order to improve the adaptivity and efficiency of sliding window, this paper uses Pearson's correlation coefficient and significance level P [30] value to extract gait periodic signals. Since the significance level is the P value, the sliding window is associated with significance without being affected by the high correlation coefficients. However, relevance may be caused by accident. So, how much is significant; generally P value is less than 0.05 which is significant; if it is less than 0.01, it is more significant. In addition, according to Pearson's correspondence, the relationship between the correlation coefficient value and the degree of correlation is defined as follows: the correlation coefficient range of 0.0–0.2 means “very weakly correlated or uncorrelated” in degree of relevance, the range of 0.2–0.4 means “weak correlation,” the range of 0.4–0.6 means “moderately relevant,” the range of 0.6–0.8 means “strong correlation,” and the range of 0.8–1.0 means “very strong correlation.” In this work, Pearson's correlation coefficient is chosen as 0.87 and P value is set as 0.01 according to actual condition and repeated tests. Based on the selected parameters, a partial signal extraction diagram is shown in Figure 4.

Next, we need to perform phase division. The human walking process is a rhythmic movement, and a complete gait cycle definition is from one-sided heel landing to the same-sided heel landing again. Two phase model recognition systems are sufficient to control active knee orthosis [31].

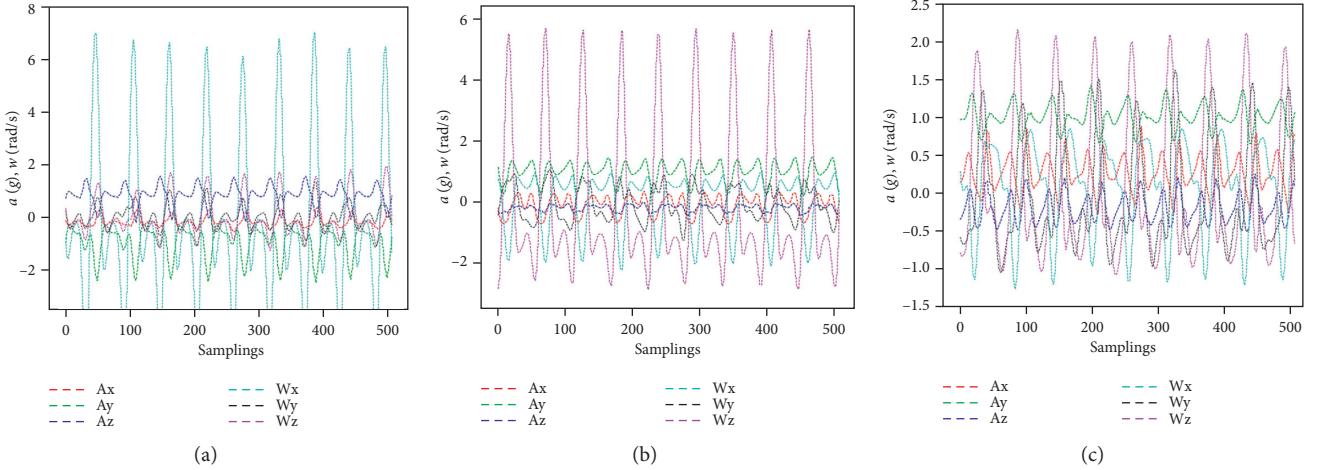


FIGURE 3: Acceleration and angular velocity data collected under three body parts covering foot (left), calf (center), and thigh (right). In the figures, Ax, Ay, Az, Wx, Wy, and Wz represent the curves of acceleration and angular velocity in the X, Y, and Z directions over time during the acquisition process, respectively.

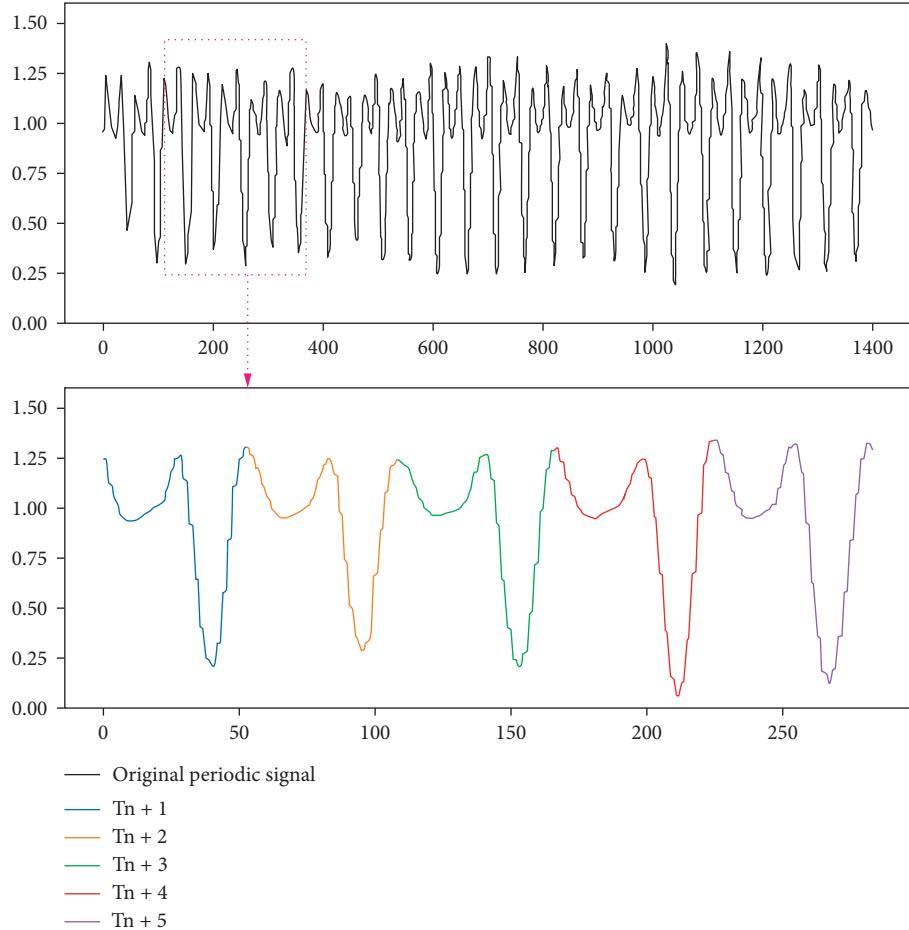


FIGURE 4: Example of periodic signal extraction results. Periodic signals $T_n + 1$, $T_n + 2$, $T_n + 3$, $T_n + 4$, and $T_n + 5$ are extracted from the pink box of original periodic signal using Pearson's correlation coefficient and P value.

However, the most widespread method currently relies on four-phase identification technology [32], which is represented as heel strike (HS), load response phase or flat foot

(FF), heel lift or heel disengagement (HO), and initial swing phase (SW). This four-phase gait partition model has been used to drive multiple robotic ankle-foot orthosis robots

[33]. Referring to previous studies, this study also divides the gait cycle into HS, FF, HO, and SW phases. During normal walking, the acceleration signals and angle signals on the feet, thighs, and lower legs are strongly periodic. Studies have shown that the swing phase accounts for 40% of the entire gait cycle, while the standing phase accounts for 60% of the entire gait cycle [34]. According to the previous analysis, the schematic diagram of the gait cycle division is shown in Figure 5.

2.3. GFM-Net Gait Phase Detection. As the analysis of the ability differences of individual models for different categories and the fusion advantage of multiple models, we find that the reasonable integration way will fuse the different feature extraction capabilities of various models for better gait phase recognition. Therefore, we design a hybrid deep-learning framework to converge the spatiotemporal feature vectors of the multistream networks through the Gaussian probability layer output, so that the submodels complement each other and improve the whole accuracy. The architecture of the hybrid framework is shown in Figure 6.

2.3.1. AutoEncoder-CNN Spatial Feature Extractor. The spatial feature extractor is composed of an automatic encoder and CNN. In order to enhance the adaptive ability of the model, we adjusted the parameters such as the stride of the convolution layer or the size of the convolution kernel. This adjustment usually hardly increases the complexity of the network model, but the prediction accuracy of the model is improved. First, we briefly introduce the AutoEncoder architecture as a coarse feature extractor to study model adjustment, which covers an encoding layer, an intermediate layer, and a decoding layer with the Leaky-ReLU activation function. The input channel of this article is 60 and the output channel is 72. Then, the feature vector f extracted by AutoEncoder is input to the “sub-CNN” structure, which consists of four convolutional layers. In this paper, we select the 3×3 kernels convolution layer with the stride 1 and padding 0 to build the first conv1 pipeline. Then, a 2×2 max pooling layer with a stride of 2 is followed as the conv2 stage, where the input stem begins with a downsampling block. And the conv3 and conv4 layers are, respectively, selected as a 2×2 convolution kernel and 1×1 convolution kernel with the same stride 1. The activation function of each convolution layer is set as the ReLU function in this network. To reduce the risk of the gradient disappearing, we introduce the batch-normalization module right after conv3 and conv4 layers. The period operation process is shown as follows:

$$\begin{aligned} f &= H_{\text{auto}}(\vec{x}, \{W_1, B_1, \delta_{\text{leaky-relu}}\}), \\ F &= H_{\text{cnn}}(f, \vec{y}, \{W_2, B_2, \delta_{\text{relu}}\}), \end{aligned} \quad (2)$$

where the function H_{auto} can represent the encoding and decoding layers of the AutoEncoder architecture with the inputs \vec{x} denoted to the raw IMU data. W_1 denotes the weight matrix of the network convolution kernel. B_1 is the bias of the network convolution kernel. And $\delta_{\text{leaky-relu}}$

denotes the nonlinear activation functions. Then, the feature map f and the related gait phase division result \vec{y} are combined as the input of H_{cnn} function, which indicated the multiple convolutional layers with the weight matrix W_2 and the bias vector B_2 under the inspiration of the ReLU activation function δ_{relu} . We randomly initialize the weights and start training all networks with a momentum of 0.9. With the learning rate 0.05, the spatial feature vector F with 300 channels is obtained to extract the further temporal features.

2.3.2. Multistream Temporal Feature Extractor. $x_t \in F(f_1, f_2, \dots, f_n)$ is the input of the three subnetworks; P_t , S_t , and Q_t ($t = 1, 2, \dots, n$) are the relevant outputs of sub-GRU, sub-LSTM, and sub-RNN models, respectively. The GRU uses an update gate to control the degree to which the previous state affects the current state. The reset gate is equivalent to the forget gate in LSTM for controlling the degree of the previous moment. The forward propagation of each submodel is as follows:

$$\begin{aligned} P &= R_{\text{gru}} \left\{ \sum_{t=1}^n [x_t, h_t, z_t, r_t, (w_1, b_1, \delta_{\text{tanh}})] \right\}, \\ S &= R_{\text{lstm}} \left\{ \sum_{t=1}^n [x_t, h'_t, g_t, (w_2, b_2, \delta_{\text{tanh}})] \right\}, \\ P &= R_{\text{rnn}} \left\{ \sum_{t=1}^n [x_t, a_t, (w_3, b_3, \delta_{\text{tanh}})] \right\}, \end{aligned} \quad (3)$$

where x_t is the input vector to each submodel. h_t , z_t , and r_t stand for the active state, update gate, and reset gate of the current hidden node in GRU cells at time t ; w_1 and b_1 are, respectively, the weight matrices and bias vectors to be learned during model training; and δ_{tanh} are the tanh activation functions. Similarly, h'_t and g_t represent the active state and the forget gate in LSTM cells at time t with the weight matrices w_2 and the related bias vectors b_2 . In addition, a_t , w_3 , and b_3 indicate the network parameters of the RNN models. Finally, the model is trained by gradient descent algorithm and the parameters such as weight and bias are constantly updated.

2.3.3. Gaussian Probability Fusion Module. Next, we further propose a probabilistic fusion module that uses the multi-stream spatiotemporal feature maps generated by the submodules for the final fusion decision of the model. In the fusion module, the output is the fused classification probability, and the input is the various features of sub-RNN, sub-GRU, and sub-LSTM models. In order to further improve the disappearance of the gradient, this paper adds a convolutional layer and a batch-normalization layer to follow the results of three submodel outputs. This operation is to unify the output vector of different models into the same dimensional space layer, which facilitates subsequent fusion operations. The entire operation process is described as follows:

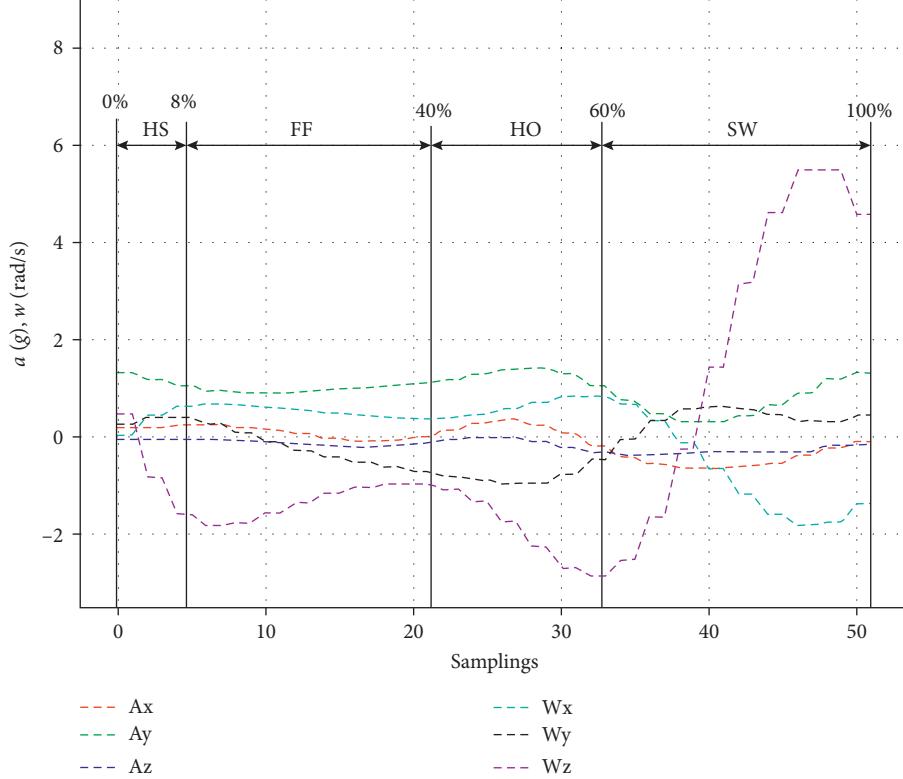


FIGURE 5: Phase division diagram of gait.

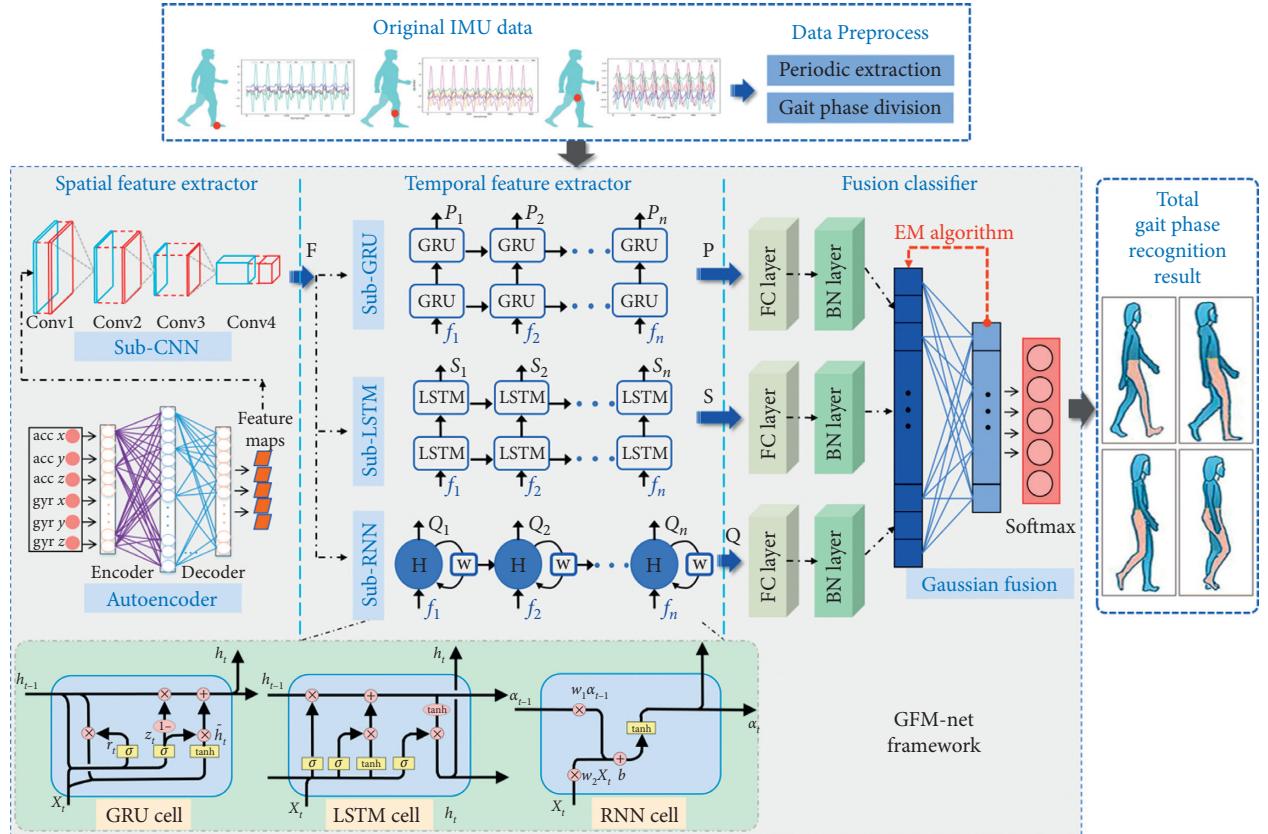


FIGURE 6: The proposed architecture of GFM-Net framework for gait phase recognition.

$$F' = \sum_{i=1}^3 H_{BN}^i * \delta_{\text{relu}}(W_{\text{full}}^i [P, S, Q] + b_{\text{full}}^i). \quad (4)$$

The function H_{BN}^i can represent i -th batch-normalization layer, and the vectors P, S, Q denote the output feature maps of three submodels. W_{full}^i denotes i -th square weight matrix asymptotically approximating complicated combination of the full-connection layer. b_{full}^i can perform i -th biases of the convolution kernel parameters. And δ_{relu} denotes the nonlinear activation functions, which was selected as Leak-ReLU. With initializing a weight decay of 0.0001 and a momentum of 0.9, we obtain the prediction score vector F' representing the probability that the input of i -th extractor belongs to the corresponding category. After normal distribution analysis, we choose the Gaussian distribution functions to fit multimodal distributions E_{ij} of each submodel. The Gaussian distribution of i -th component model is defined as follows:

$$E_{ij}(\tilde{y} | \theta_i) = \frac{1}{\sqrt{2\pi}\sigma_i} \exp\left(-\frac{(F' - \mu_i)^2}{2\sigma_i^2}\right), \quad (5)$$

where $\theta_i = (\mu_i, \sigma_i^2)$ is the estimated parameter, which consists of the mean vector μ_i and covariance matrix σ_i^2 , respectively. $F' = \{F_{\text{gru}}, F_{\text{lstm}}, F_{\text{rnn}}\}$ is the output label vector corresponding to each submodel, which reflects the essential characteristics of the original IMU data. Then, the Gaussian mixture method is used to construct a connection layer based on probabilistic fusion. Its purpose is to cluster similar features adaptively at different submodel clustering centers and further combine the probability sizes of the three submodels to determine the final recognition estimate. S denotes the final output probability, which is calculated as follows:

$$S(F' | \theta) = \sum_{i=1}^3 S_i = \prod_{j=1}^N \prod_{i=1}^3 E_{ij}(F' | \theta_i), \quad (6)$$

where $\theta = (\theta_1, \theta_2, \theta_3)$ represents the Gaussian mixture parameters of the fusion layer. γ_{ji} represent implicit variables. By integrating the joint probabilities between F' and γ_{ji} , the fusion score S in (6) is redefined to solve the log-likelihood expression as follows:

$$\log S(F', \gamma | \theta) = \sum_{i=1}^3 \sum_{j=1}^N \gamma_{ji} \left[\log\left(\frac{1}{\sqrt{2\pi}}\right) - \log \sigma_i - \frac{1}{2\sigma_i^2} (F'_j - \mu_i)^2 \right]. \quad (7)$$

Then, we choose the EM algorithm to estimate the hyperparameters of the fusion layer. Each iteration of the EM algorithm is divided into two steps: the expected step (E-step) and the maximum step (M-step). The E-step calculates the expectation of implicit variables γ_{ji} defined as $\hat{\gamma}_{ji} = E(\gamma_{ji} | \tilde{y}, \theta)$, which denotes the responsivity of i -th component model to the label y_j . Subsequently, step M updates the corresponding parameters by maximizing the expected value of the given log-likelihood function in (7). After several iterations of the EM algorithm, the parameters gradually converge. The detailed process is shown as the following (Algorithm 1).

Note that the internal EM algorithm iteration loop is used for corresponding parameter estimation and the EM algorithm iteration loop is performed inside the external network loop. Specifically, when the network is in training, each external loop will be accompanied by a lot of internal EM algorithm iteration loops. Through the above network structure, our proposed GFM-Net obtains a better prediction effect from the perspective of decision fusion. Finally, we add a softmax layer after the fusion layer to normalize the output result into a probability classification result and output the final classification result through the argmax function.

We use the cross entropy (CE) loss function to evaluate the degree of inconsistency between the predicted probability obtained by softmax and the true label. Then, the gradient descent method is used to update the model parameters so that the two probability distributions are similar to each other. The loss expression is shown as

$$\begin{aligned} \text{softmax}(q)_i &= \frac{e^{q_i}}{\sum_{i=1}^n e^{q_i}}, \\ l &= -\sum_{i=1}^4 y_i \log((q)_i), \\ o &= \max(q), \end{aligned} \quad (8)$$

where y_i denotes the indicative variable (0 or 1); if the category is the same as the sample category, it is 1; otherwise it is 0; q_i denotes the predicted probability that the observation sample belongs to category i . For each input sample x , the predicted output of the network is $q = (q_0, q_1, q_2, q_3)$. Then, the value of q_i is between 0 and 1, and the larger the value, the greater the probability that x belongs to the real label. Based on the output q_i , we can get the class label as O. Then, we can construct the training loss with cross entropy, as formulated by (8); the cross entropy is a positive number. When the probability value of the true label q_i in the vector q is smaller, larger difference between q_i and y_i will result in a larger cross entropy value. This property will help the convergence of the network in the training. In order to avoid overfitting, we chose 70% of the sample set for training and 30% of the samples for testing. After using the same training set to train different models 10,000 times, use the same test set to test the trained model and record the classification accuracy and macro-F value of each classifier after testing the classification model with the test machine. Then, evaluate the performance of all models based on these three indicators.

3. Results and Discussion

3.1. Evaluation Methods. In this paper, we propose the GFM-Net network. In order to prove its performance in classification, we need to draw the corresponding conclusion through corresponding indicators. As we all know, accuracy is a good comprehensive indicator, which is widely used in evaluation indicators. However, in the classification, it is difficult to characterize the performance of a certain model simply by relying on accuracy, and we have to choose other

```

Initialize  $\gamma_{ji} = 1$  and set estimation parameter  $\theta$ 
For T iterations do
  Procedure M-step:
    Calculate the expectations of (7):
    
$$M(\theta, \theta^{(t)}) = E[\log S(F', \gamma | \theta) | y_j, \theta^{(t)}] = \sum_{i=1}^3 \sum_{j=1}^N \hat{\gamma}_{ji} [\log(1/\sqrt{2\pi}) - \log \sigma_i - 1/2\sigma_i^2 (y_j - \mu_i)^2],$$

    Calculate partial derivatives of  $\theta(\mu_i, \sigma_i^2)$ :
    
$$(\partial M(\theta, \theta^{(t)}) / \partial \mu_i) \longrightarrow \hat{\mu}_i = \sum_{j=1}^N \hat{\gamma}_{ji} y_j / \sum_{j=1}^N \hat{\gamma}_{ji},$$

    
$$(\partial M(\theta, \theta^{(t)}) / \partial \sigma_i^2) \longrightarrow \hat{\sigma}_i^2 = \sum_{j=1}^N \hat{\gamma}_{ji} (y_j - \mu_i)^2 / \sum_{j=1}^N \hat{\gamma}_{ji},$$

  Return  $\hat{\mu}_i, \hat{\sigma}_i^2$ 
  Procedure E-step:
    Calculate  $i$ -th submodel responsibility:
    
$$\hat{\gamma}_{ji} = S(\gamma_{ji} = 1, \bar{y}_j | \theta) / \sum_{i=1}^3 S(\gamma_{ji} = 1, \bar{y}_j | \theta) = E_{ij}(y_j | \theta_i) / \sum_{i=1}^3 E_{ij}(y_j | \theta_i),$$

  Return  $\hat{\gamma}_{ji}$ 
  Update the estimation parameter:
    
$$\theta^{(t)}(\hat{\mu}_i, \hat{\sigma}_i^2),$$

    
$$\theta^{(t+1)} = \arg \max_{\theta} M(\theta, \theta^{(t)}),$$

End For till the parameter threshold:

$$\theta^{(t+1)} > \theta_{\text{th}} \text{ or } T > T_{\text{th}}$$

  End Procedure M-step and E-step
  Return  $\theta^{(t+1)}$  and the fusion score S

```

ALGORITHM 1: Training process of the Gaussian fusion layer based on EM-Algorithm.

indicators to comprehensively characterize the classification performance of a certain model. In classification problems, commonly used classification performance indicators also include precision (P), recall (R), and F1. Among them, P and R are widely used to evaluate the quality of the model results. P is used to measure the accuracy of the retrieval system. R is used to measure the recall of the retrieval system. Of course, we hope that P and R results are as high as possible. Generally speaking, if both P and R are high, we can conclude that this model performs well in this classification task. Hence, the F1 indicator is chosen since it can represent the performance of the model by combining the results of both P and R. When F1 performs well, it means that both P and R will perform well. However, this paper studies a multi-classification task and cannot directly use F1. The most direct method is to calculate macro-F1 [35]. Accuracy reflects the ratio of correctly classified samples to total samples. The definition equation of the above evaluation factors can be seen in [1]; we can easily calculate the accuracy, macro-P, macro-R, and macro-F1 evaluation factors.

In multiclassification tasks, we also often use the area AUC under the ROC curve to measure the classification effect. The ROC curve was first publicly proposed to verify machine learning in model evaluation [36]. In recent years, it has been widely used in the fields of machine learning and deep-learning. The larger the AUC, the more reliable the model's recognition of the target.

3.2. Results. The confusion matrix (Figures 7–9) provides the performance of visual gait-subphase recognition. The

vertical axis of the matrix represents the actual classification category of the test, and the horizontal axis represents the corresponding predicted classification category. In addition, in the confusion matrix diagram, “0.0” represents the “HS” stage, “1.0” represents the “FF” stage, “2.0” represents the “HO” stage, and “3.0” represents the “SW” stage. These nine matrices are the average recognition results of all subjects under different walking steps. The value in the main diagonal is the proportion of correctly classified samples. As shown in Figure 7, except for the HS stage, all confusion matrices perform well. The HS stage is mostly erroneously classified as FF and SW stages. In order to verify the effectiveness of the proposed recognition model, we compared two other integrated algorithms to identify gait phases, namely, AdaBoosting and Bagging. The corresponding confusion matrix is shown in Figures 8 and 9. As can be seen from Figures 8 and 9, the Bagging model cannot identify the HS phase, and most HS phases are erroneously directly classified as adjacent FF phases. The AdaBoosting algorithm can identify most of the HS phase. The AdaBoosting and Bagging models have achieved good recognition results in other stages.

From the confusion matrix, we can get Table 1. As shown in Table 1, the F1 of the four groups (HS, FF, HO, and SW) of the GFM-Net model differ greatly. When the walking speed is 0.78 m/s, F1 is 65.9%, 96.4%, 97.6%, and 98.8%, respectively; when the walking speed is 1.0 m/s, F1 is 70.8%, 97.0%, 98.1%, and 98.9%, respectively; when the walking speed is 1.25 m/s, F1 is 53.6%, 96.8%, 97.4%, and 98.1%. As can be seen from the above data, FF, HO, and SW have the best effect and obtain better recognition effect (over 96%). The

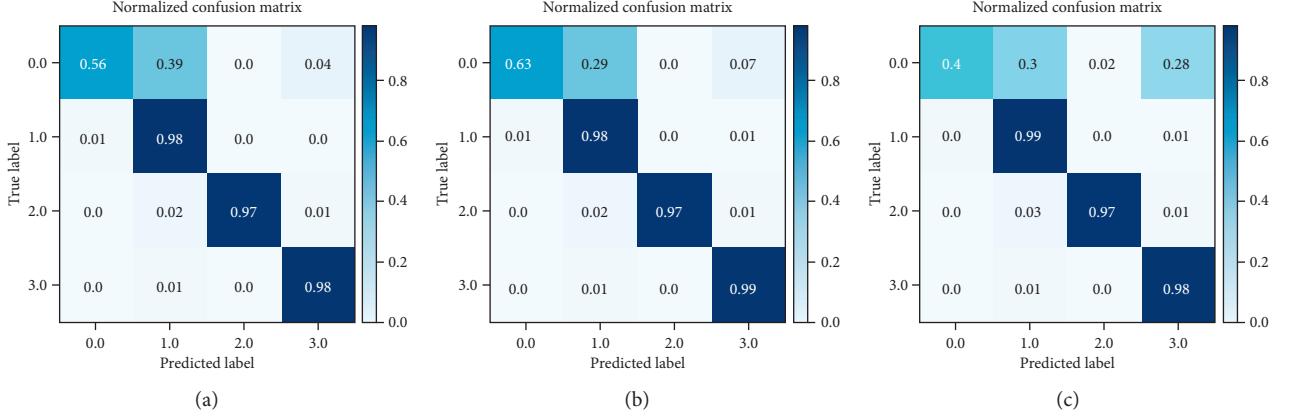


FIGURE 7: Confusion matrix of three gait patterns derived from Bagging classification under three paces. Settings: 0.78 m/s (a), 1.0 m/s (b), and 1.25 m/s (c).

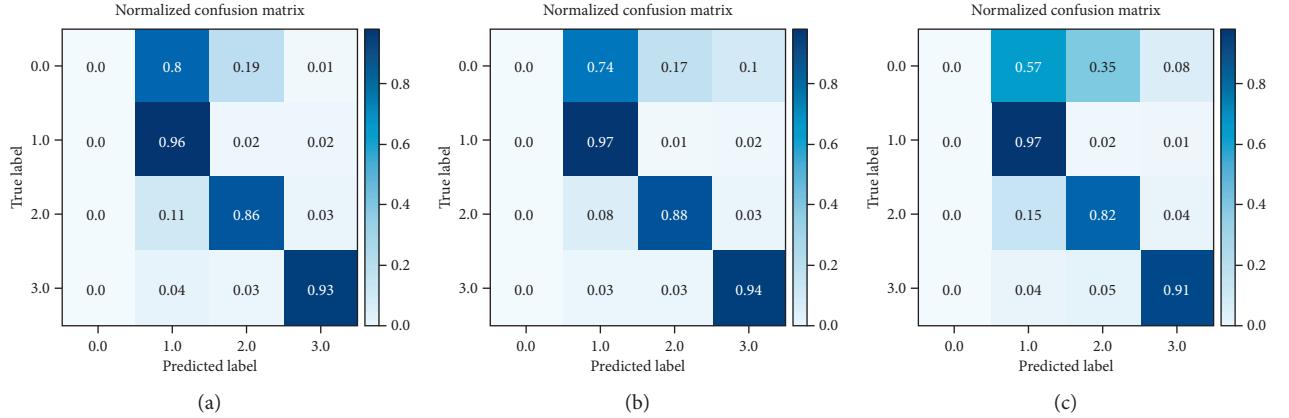


FIGURE 8: Confusion matrix of three gait patterns derived from AdaBoosting classification under three paces. Settings: 0.78 m/s (a), 1.0 m/s (b), and 1.25 m/s (c).

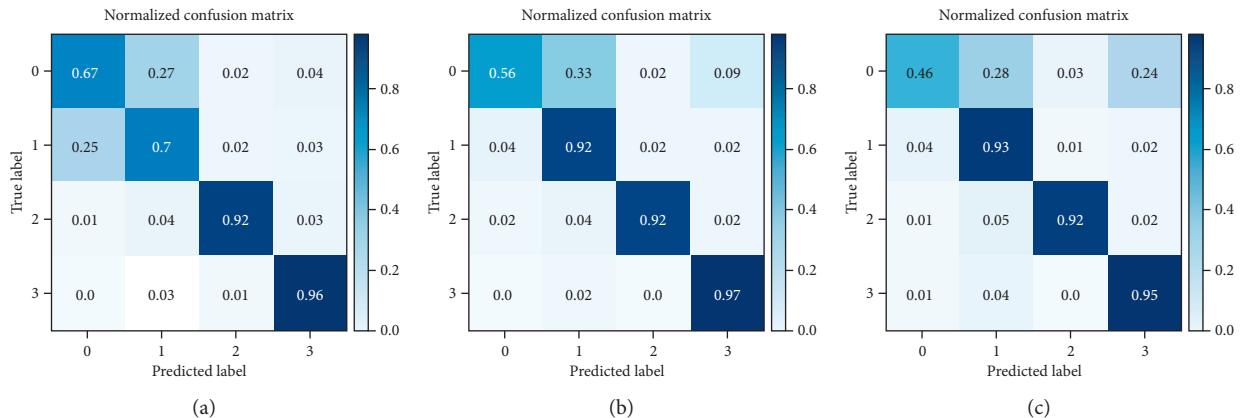


FIGURE 9: Confusion matrix of three gait patterns derived from GFM-Net classification under three paces. Settings: 0.78 m/s (a), 1.0 m/s (b), and 1.25 m/s (c).

performance of HS phase recognition is the worst. As for the recognition accuracy of each substage, the swing stage (SW) performed best, with a maximum value of 1.0 m/s (98.9%).

The recognition effect of FF phase and HO phase is also very good. Obviously, the performance of the HS phase recognition effect is the worst, the F1 value has not reached 71%,

TABLE 1: Summary of classification performance of different models at unsynchronized speed.

| Model | Speed Phase | 0.78 m/s | | | | 1.0 m/s | | | | 1.25 m/s | | | |
|-------------|---------------|----------|------|------|------|---------|------|------|------|----------|------|------|------|
| | | HS | FF | HO | SW | HS | FF | HO | SW | HS | FF | HO | SW |
| Bagging | Precision (%) | 0 | 84.7 | 85.9 | 98.1 | 0 | 86.5 | 87.5 | 97.2 | 0 | 84.3 | 80.4 | 97.3 |
| | Recall (%) | 0 | 96.2 | 85.1 | 94.6 | 0 | 97.5 | 88.4 | 93.4 | 0 | 97.6 | 81.0 | 91.3 |
| | F1 (%) | 0 | 90.1 | 85.5 | 96.3 | 0 | 91.7 | 88.0 | 95.3 | 0 | 90.5 | 80.7 | 94.2 |
| AdaBoosting | Precision (%) | 48.8 | 91.8 | 92.2 | 96.9 | 42.1 | 90.8 | 94.2 | 97.2 | 40.7 | 90.4 | 96.0 | 96.3 |
| | Recall (%) | 52.6 | 90.4 | 93.4 | 97.0 | 53.7 | 91.7 | 91.9 | 94.7 | 54.5 | 92.1 | 93.8 | 94.0 |
| | F1 (%) | 50.7 | 91.1 | 92.8 | 97.0 | 50.7 | 91.3 | 93.0 | 95.9 | 46.6 | 91.3 | 94.9 | 95.1 |
| GFM-Net | Precision (%) | 79.1 | 94.5 | 98.2 | 99.2 | 80.6 | 95.8 | 99.0 | 98.9 | 80.9 | 95.0 | 97.9 | 97.9 |
| | Recall (%) | 56.4 | 98.3 | 97.0 | 98.4 | 63.1 | 98.4 | 97.1 | 98.9 | 40.1 | 98.6 | 96.8 | 98.4 |
| | F1 (%) | 65.9 | 96.4 | 97.6 | 98.8 | 70.8 | 97.0 | 98.1 | 98.9 | 53.6 | 96.8 | 97.3 | 98.1 |

and the F1 with the lowest HS phase recognition is only 53.6%. For the Bagging and AdaBoosting models, at three paces, the average value of F1 in the HS stage is 0 and 49.3%, respectively, indicating that the recognition effect of the HS stage is relatively poor; for SW phase recognition, the minimum value of F1 is 94.2% and 95.1%, showing strong recognition performance for SW phase.

In order to verify the effectiveness of the proposed GFM-Net, this paper also compares the existing deep-learning algorithm model and obtains Table 2. According to Table 2, the recognition accuracy of the GFM-Net algorithm is as high as 96% or higher, while the recognition accuracy of the other two integrated algorithm models is lower. In particular, the accuracy of Bagging's phase recognition in three steps is lower than 92%. In addition, we can also see that the GFM-Net algorithm is also better than the existing CNN + LSTM model. The CNN + GRU model and CNN + RNN model are superior in recognition accuracy and macro-F1, although the gap between them is not counted. In addition, we can easily conclude that the GFM-Net model has the highest AUC at asynchronous speed, which is higher than the other two models. It can be clearly seen from Figures 10 and 11 that, at any step, the accuracy and macro-F1 of the GFM-Net algorithm are higher than those of the other five algorithms. Finally, in order to show that the proposed method not only has better performance, but also has significant differences compared with other methods, we tested the significance of the results of each model and the GFM-Net results and got the results shown in Table 3. As shown, P value of the GFM-Net result and any other model result is less than 0.01, which is “very significant.”

4. Discussion

This study proves that the proposed system can effectively detect the corresponding gait phase based on a single IMU sensor. In order to support this hypothesis, this paper proposes using the GFM-Net algorithm to detect the gait phase and comparing it with other timing algorithms to verify the effectiveness of the algorithm. The core technology of gait phase recognition system is the design of recognition algorithm model. The predecessors generally used machine learning or deep-learning methods to detect HS, FF, HO, and

SW from IMU signals. This paper proposes the GFM-Net model and uses it to identify HS, FF, HO, and SW phases by the collected acceleration and angle data. The data obtained by the JY901 sensor shows that when pedestrians walk normally on flat ground, the acceleration signals and angle signals on the thighs, calves, and feet have low variability and stability. Moreover, this study can obtain a better recognition effect, and this study also proposes a more complex neural network model to achieve this effect, so the possible result is that it takes longer training time. Some portable gait event detection devices require accurate gait biofeedback information and dynamic gait monitoring devices, but there are currently no wearable sensors to meet these requirements.

The GFM-Net algorithm is an integrated algorithm, and its result depends on the fusion of three subneural networks. The three subneural networks in this paper are RNN, GRU, and LSTM. However, whether the three subnetworks are the best choice needs further research. The recognition performance of the three subnetworks should not be too different to avoid the fact that the final classification result of the model only depends on the classification result of one of the subnetworks. In terms of fusion, this paper uses fully connected layers to connect the obtained results with the model input values and then performs Gaussian fusion. In order to avoid the problem that the gradient of the model becomes smaller because the network is too deep, this paper uses the BN network. In terms of feature extraction, AutoEncoder is first used to implement the GFM-Net algorithm with denoising and functional improvements. Then, we use CNN network to extract spatial scale features. It can be seen from Table 1 that neither Bagging nor AdaBoosting algorithms can identify the HS phase, but the GFM-Net algorithm can identify most HS phases. The GFM-Net algorithm has an HS phase recognition rate of 63.4% at three paces. Although the GFM-Net algorithm has greatly improved compared with the other two algorithms, the algorithm still needs further optimization and improvement. But the three models showed better recognition performance for the other three phases. As can be seen from Figures 12 to 14 the macro-AUC of the GFM-Net algorithm is also the best performing of the three algorithms. The GFM-Net algorithm based on the voting fusion mechanism can effectively detect the HS, FF, HO, and SW phases, has high recognition accuracy, and is also macro-F1 and

TABLE 2: Summary of classification performance for different training functions.

| Speed (m/s) | Training function | Classification result | | |
|-------------|-------------------|-----------------------|--------------|-----------|
| | | Accuracy (%) | Macro-F1 (%) | Macro-AUC |
| 0.78 | Bagging | 90.5 | 67.7 | 0.97 |
| | AdaBoosting | 93.1 | 82.9 | 0.92 |
| | CNN + LSTM | 95.2 | 87.8 | 0.98 |
| | CNN + GRU | 94.7 | 87.3 | 0.97 |
| | CNN + RNN | 93.6 | 86.6 | 0.95 |
| | GFM-Net | 97.0 | 89.7 | 0.99 |
| 1.0 | Bagging | 91.4 | 68.7 | 0.95 |
| | AdaBoosting | 92.4 | 82.7 | 0.91 |
| | CNN + LSTM | 97.2 | 90.6 | 0.99 |
| | CNN + GRU | 96.6 | 89.3 | 0.98 |
| | CNN + RNN | 95.6 | 88.5 | 0.97 |
| | GFM-Net | 97.5 | 91.2 | 0.99 |
| 1.25 | Bagging | 89.2 | 66.3 | 0.95 |
| | AdaBoosting | 92.3 | 82.0 | 0.92 |
| | CNN + LSTM | 95.7 | 85.2 | 0.99 |
| | CNN + GRU | 95.3 | 84.6 | 0.99 |
| | CNN + RNN | 94.2 | 83.2 | 0.98 |
| | GFM-Net | 96.7 | 86.5 | 1.0 |

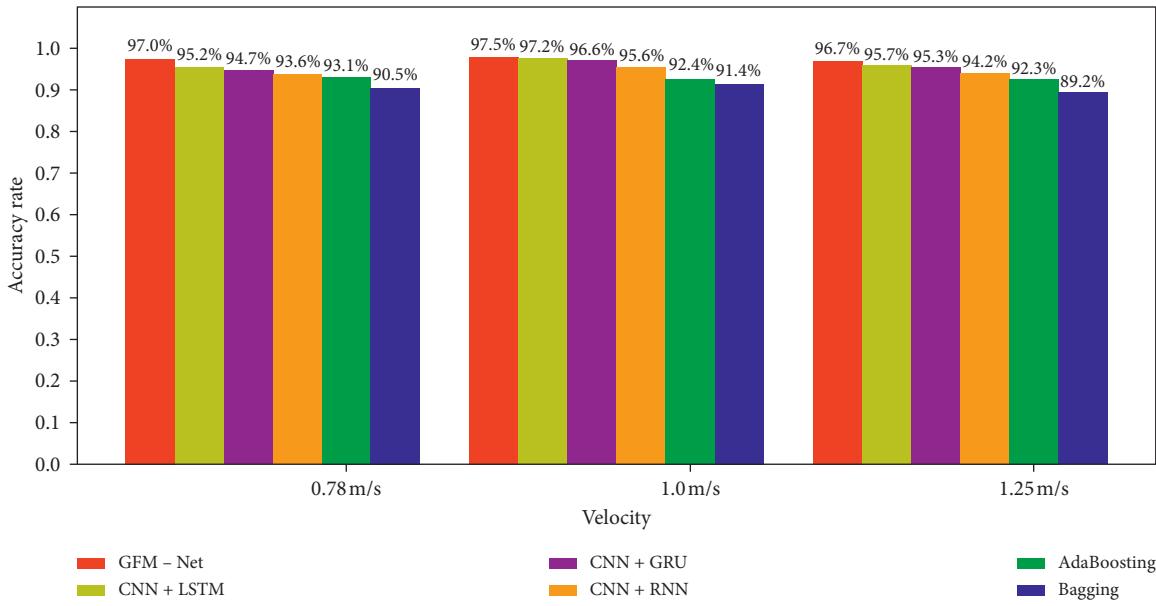


FIGURE 10: More tag set classifier accuracy distribution at three paces.

macro-AUC compared to the existing Bagging and AdaBoosting, being the best performer. We can get from Table 2 that the GFM-Net algorithm has the best recognition accuracy, followed by AdaBoosting, and Bagging is the worst. And we can clearly find that the GFM-Net algorithm improves the recognition accuracy by almost 5 percentage points compared with AdaBoosting. In order to propose an acceleration data acquisition system suitable for the general public, we tested the recognition effect of the FGFM-Net algorithm proposed in this paper on unlearned acceleration signals and angular velocity data. This study found that the proposed system can successfully predict the gait

events of unlearned data, and the phase recognition accuracy of HS, FF, HO, and SW is as high as 97.1%, so the result is relatively reliable but still needs to be optimized.

At the same time, in order to verify the effectiveness of the model proposed in this paper, this paper still compares our model with CNN + LSTM, CNN + GRU, and CNN + RNN. According to Table 2, we can still see that the GFM-Net model has the best performance in terms of accuracy, macro-F1, and macro-AUC. Through Figure 14, we can see the recognition accuracy of the FMS-Net algorithm. Macro-F1 and AUC basically remain stable as the pace increases. But we can also find that the macro-F1 has been

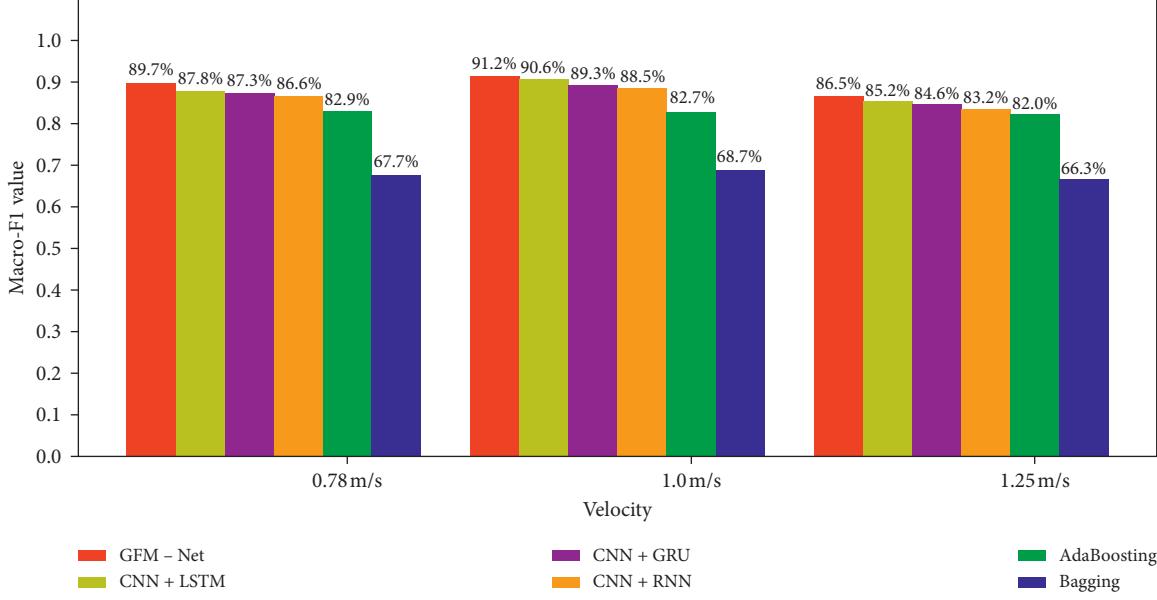


FIGURE 11: More tag data set classifier macro-F1 distribution under sync speed.

TABLE 3: The difference between GFM-Net results and other model results is significant P value.

| Model | P value |
|------------------------|------------|
| (GFM-Net, Bagging) | $1.37e-6$ |
| (GFM-Net, AdaBoosting) | $1.61e-11$ |
| (GFM-Net, CNN + LSTM) | $7.40e-15$ |
| (GFM-Net, CNN + GRU) | $4.67e-15$ |
| (GFM-Net, CNN + RNN) | $7.59e-15$ |

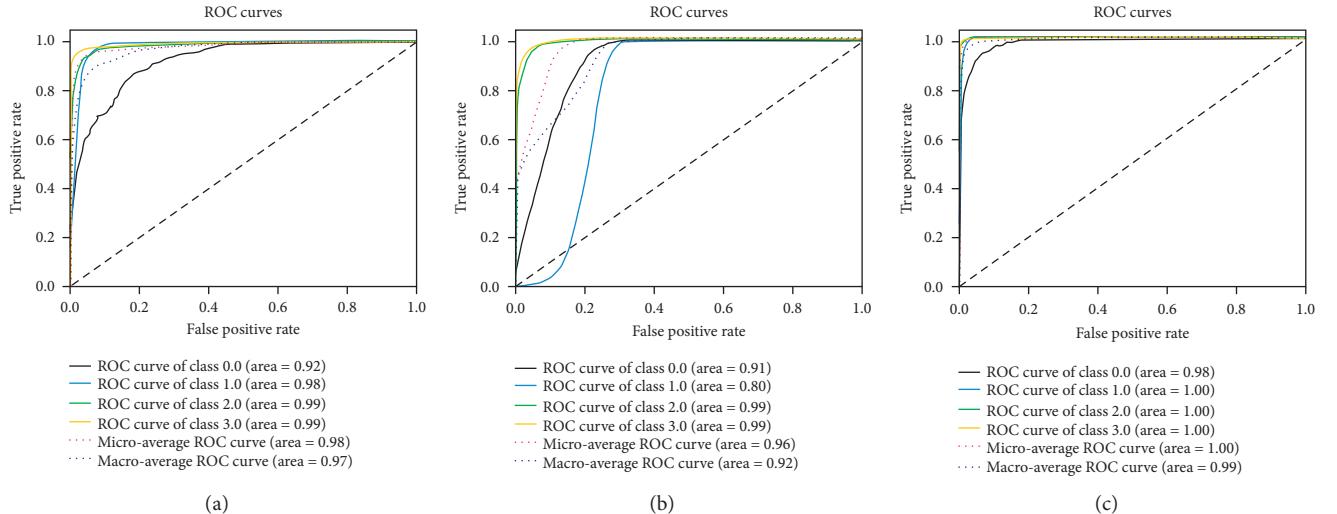


FIGURE 12: ROC curve performance in three models: Bagging (a), AdaBoosting (b), and GFM-Net. (c) The treadmill speed is set to 0.78 m/s.

declining as the pace increases, which needs our attention. In addition, we need to add more pace control experimental groups for further exploration and draw more reliable conclusions. Even though GFM-Net has shown its

usefulness in classifying acceleration and angular velocity signals detected by gait events, other machine learning methods are needed for further evaluation. Future work should improve classification accuracy by improving feature

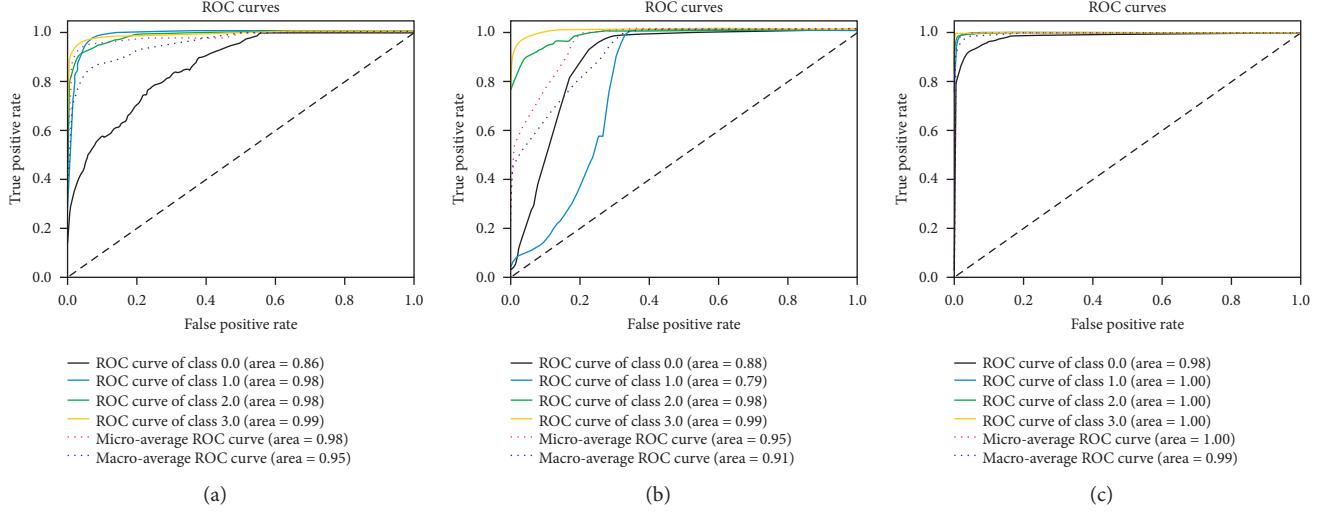


FIGURE 13: ROC curve performance in three models: Bagging (a), AdaBoosting (b), and GFM-Net (c). The treadmill speed is set to 1.0 m/s.

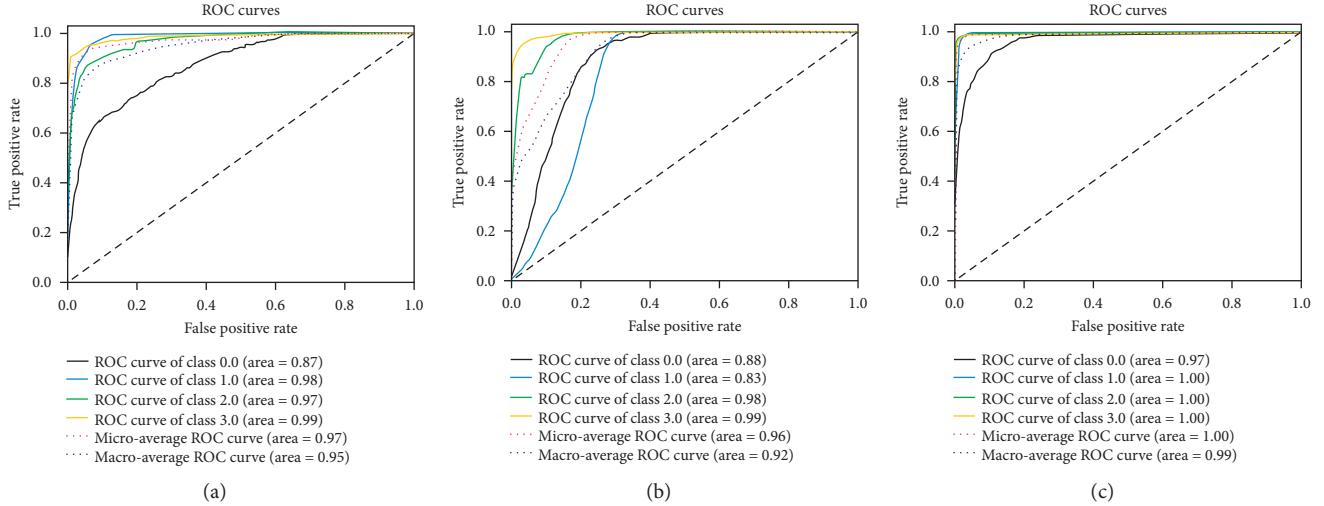


FIGURE 14: ROC curve performance in three models: Bagging (a), AdaBoosting (b), and GFM-Net (c). The treadmill speed is set to 1.25 m/s.

extraction and gait phase recognition algorithms. In this study, the three inertial sensors are transmitted to the host computer wirelessly, so we think that the three wearable inertial sensor modules are acceptable wearable sensors and will not have obvious impact on the subject's walking gait influences. However, in practice, the wearing of sensors may have a potential impact on the gait of people who have not yet been investigated. In the future, we will explore the use of fewer inertial units to identify gait phases and minimize the impact on the human body.

5. Conclusion

This paper studies the method of gait recognition using three inertial sensors on the treadmill. A hybrid deep fusion learning method is proposed, which seamlessly combines GRU, LSTM, and RNN to achieve a robust representation of

spatiotemporal features of inertial gait. In order to accurately identify walking gait, this paper proposes an effective hybrid deep-learning framework GFM-Net based on Gaussian probability fusion of multiple spatiotemporal networks to analyze multidimensional acceleration signals and detect out-of-sync events including HS, FF, HO, and SW. It consists of three main parts: data preprocessing, multistream integrated neural network, and fusion model. Data preprocessing uses automatic encoders to select key features, while CNN extracts more spatial information. In addition, we use three parallel modules RNN, LSTM, and GRU as multistream time feature extractor. The network uses mixed particle information to form high-dimensional time-scale features. Finally, a Gaussian fusion module was developed to fuse different submodels. It uses the EM algorithm to optimize Gaussian probability fusion of different submodels and proves that it is a practical method to increase model

capacity on a large scale. Experiments and discussions prove that GFM-Net has higher accuracy up to 96.7% and the effectiveness of macro-F1 is up to 86.5%, which is superior to that of other integrated algorithm models.

There are a lot of variables and hyperparameters in the network structure proposed in this paper, which will undoubtedly cause a lot of time for the model to be trained. Therefore, this paper strongly recommends training the model on the GPU, which will improve the model training efficiency. Our future work is to try to design a lightweight network to recognize human gait phases and achieve online gait phase recognition with high efficiency, which is of great significance for medical rehabilitation training robots and gait disease diagnosis.

Data Availability

In terms of experimental data, 16 volunteers between 20 and 26 years old with body weight ranging from 46 kg to 70 kg and height ranging from 158 cm to 177 cm were selected to collect IMU data. All data can be obtained by sending e-mail to the first author (Tao Zhen).

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

Authors' Contributions

Tao Zhen and Jian-lei Kong proposed the conception and design of this research, acquired the data, and drafted this article. Lei Yan helped with the data analysis and model optimization. All authors have read and agreed to the published version of the manuscript.

Acknowledgments

This work was supported financially by National Natural Science Foundation of China (no. 61903009), Beijing Municipal Education Commission (no. KM201910011010), Beijing Excellent Talent Training Support Project for Young Top-Notch Teams (2018000026833TD01), and Fundamental Research Funds for the Central Universities (no. 2015ZCQ-GX-03).

References

- [1] L. Yan, T. Zhen, J.-L. Kong, L.-M. Wang, and X.-L. Zhou, "Walking gait phase detection based on acceleration signals using voting-weighted integrated neural network," *Complexity*, vol. 2020, no. 1, 14 pages, Article ID 4760297, 2020.
- [2] L. Rong, D. Zhiguo, Z. Jianzhong, and L. Ming, "Identification of individual walking patterns using gait acceleration," *International Conference on Bioinformatics and Biomedical Engineering*, vol. 142, pp. 543–546, 2007.
- [3] C. Buckley, M. Encarna Micó-Amigo, M. Dunne-Willow et al., "Gait Asymmetry Post-Stroke: determining valid and reliable methods using a single accelerometer located on the trunk," *Sensors*, vol. 20, no. 1, p. 37, 2020.
- [4] A. Picelli, C. Melotti, F. Origano et al., "Robot-assisted gait training in patients with Parkinson disease," *Neurorehabilitation and Neural Repair*, vol. 26, no. 4, pp. 353–361, 2012.
- [5] S. D. M. Achanta, T. Karthikeyan, and R. Vinothkanna, "A novel hidden Markov model-based adaptive dynamic time warping (HMDTW) gait analysis for identifying physically challenged persons," *Soft Computing*, vol. 23, no. 18, pp. 8359–8366, 2019.
- [6] H. Vu, F. Gomez, P. Cherelle, D. Lefeber, A. Nowé, and B. Vanderborght, "ED-FNN: a new deep learning algorithm to detect percentage of the gait cycle for powered prostheses," *Sensors*, vol. 18, no. 7, p. 2389, 2018.
- [7] S. Murray and M. Goldfarb, "Towards the use of a lower limb exoskeleton for locomotion assistance in individuals with neuromuscular locomotor deficits," in *Proceedings of the 2012 Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, pp. 1912–1915, San Diego, CA, USA, August–September 2012.
- [8] A. Berger, F. Horst, F. Steinberg et al., "Increased gait variability during robot-assisted walking is accompanied by increased sensorimotor brain activity in healthy people," *Journal of NeuroEngineering and Rehabilitation*, vol. 16, no. 1, pp. 1–13, 2019.
- [9] R. Luo, S. Sun, X. Zhang, Z. Tang, and W. Wang, "A low-cost end-to-end sEMG-based gait sub-phase recognition system," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 28, no. 1, pp. 267–276, 2019.
- [10] X. Jin, H. Sun, and F. Yang, "Advances in multi-sensor information fusion: theory and applications," *Sensors*, vol. 18, no. 4, p. 1162, 2018.
- [11] W. Fei, T. Yin, C. Lei, Y. Zhang, Y. Wang, and J. Liu, "Prediction of lower limb joint angle using sEMG based on GA-GRNN," in *Proceedings of the 2015 IEEE International Conference on Cyber Technology in Automation, Control, and Intelligent Systems (CYBER)*, Shenyang, China, June 2015.
- [12] W. Si, G. Yang, X. Chen, and J. Jia, "Gait identification using fractal analysis and support vector machine," *Soft Computing*, vol. 23, no. 19, pp. 9287–9297, 2019.
- [13] G. E. Kang, J. Yang, and B. Najafi, "Does the presence of cognitive impairment exacerbate the risk of falls in people with peripheral neuropathy? An application of body-worn inertial sensors to measure gait variability," *Sensors*, vol. 20, no. 5, p. 1328, 2020.
- [14] I. Gohar, Q. Riaz, M. Shahzad, M. Zeeshan Ul Hasnain Hashmi, H. Tahir, and M. Ehsan Ul Haq, "Person re-identification using deep modeling of temporally correlated inertial motion patterns," *Sensors*, vol. 20, no. 3, p. 949, 2020.
- [15] Z. Zhou, S. Yang, Z. Ni, W. Qian, C. Gu, and Z. Cao, "Pedestrian navigation method based on machine learning and gait feature assistance," *Sensors*, vol. 20, no. 5, p. 1530, 2020.
- [16] S. S. Yeo and G. Y. Park, "Accuracy verification of spatio-temporal and kinematic parameters for gait using inertial measurement unit system," *Sensors*, vol. 20, no. 5, p. 1343, 2020.
- [17] Y. Chao, F. Coenen, and B. Zhang, "Multi-attributes gait identification by convolutional neural networks," in *Proceedings of the 2015 8th International Congress on Image and Signal Processing (CISP)*, Shenyang, China, October 2015.
- [18] D. Omid, T. Mojtaba, and C. V. Raghvendar, "IMU-based gait recognition using convolutional neural networks and multi-sensor fusion," *Sensors*, vol. 17, no. 12, p. 2735, 2017.
- [19] T. Zhen, L. Yan, and J.-L. Kong, "An acceleration based fusion of multiple spatiotemporal networks for gait phase detection,"

- International Journal of Environmental Research and Public Health*, vol. 17, no. 16, p. 5633, 2020.
- [20] X.-B. Jin, T.-L. Su, J.-L. Kong, Y.-T. Bai, B.-B. Miao, and C. Dou, "State-of-the-art mobile intelligence: enabling robots to move like humans by estimating mobility with artificial intelligence," *Applied Sciences*, vol. 8, no. 3, p. 379, 2018.
 - [21] N. Neverova, C. Wolf, G. Lacey et al., "Learning human identity from motion patterns," *IEEE Access*, vol. 4, pp. 1810–1820, 2016.
 - [22] B. Hu, P. C. Dixon, J. V. Jacobs, J. T. Dennerlein, and J. M. Schiffman, "Machine learning algorithms based on signals from a single wearable inertial sensor can detect surface- and age-related differences in walking," *Journal of Biomechanics*, vol. 71, pp. 37–42, 2018.
 - [23] D. Zhen, C. Yang, K. Xing et al., "The real time gait phase detection based on long short-term memory," in *Proceedings of the 2018 IEEE Third International Conference on Data Science in Cyberspace (DSC)*, Guangzhou, China, June 2018.
 - [24] X.-B. Jin, X.-H. Yu, X.-Y. Wang, Y.-T. Bai, T.-L. Su, and J.-L. Kong, "Deep learning predictor for sustainable precision agriculture based on internet of things system," *Sustainability*, vol. 12, no. 4, p. 1433, 2020.
 - [25] Y.-T. Bai, X.-Y. Wang, X.-b. Jin, Z.-y. Zhao, and B.-h. Zhang, "A neuron-based kalman filter with nonlinear autoregressive model," *Sensors*, vol. 20, no. 1, p. 299, 2020.
 - [26] Y. Bai, X. Wang, X. Jin, T. Su, J. Kong, and B. Zhang, "Adaptive filtering for MEMS gyroscope with dynamic noise model," *ISA Transactions*, vol. 101, pp. 430–441, 2020.
 - [27] Y.-T. Bai, X.-B. Jin, X.-Y. Wang, X.-k. Wang, and J.-p. Xu, "Dynamic correlation analysis method of air pollutants in spatio-temporal analysis," *International Journal of Environmental Research and Public Health*, vol. 17, no. 1, p. 360, 2020.
 - [28] Y.-Y. Zheng, J.-L. Kong, X.-B. Jin, X.-Y. Wang, T.-L. Su, and J.-L. Wang, "Probability fusion decision framework of multiple deep neural networks for fine-grained visual classification," *IEEE Access*, vol. 7, pp. 122740–122757, 2019.
 - [29] S. Qiu, H. Wang, J. Li et al., "Towards wearable-inertial-sensor-based gait posture evaluation for subjects with unbalanced gaits," *Sensors*, vol. 20, no. 4, p. 1193, 2020.
 - [30] Coefficient PSC, "Pearson's correlation coefficient," *The New Zealand Medical Journal*, vol. 109, no. 1015, p. 38, 1996.
 - [31] S. Rosati, V. Agostini, M. Knaflitz, and G. Balestra, "Muscle activation patterns during gait: a hierarchical clustering analysis," *Biomedical Signal Processing and Control*, vol. 31, pp. 463–469, 2017.
 - [32] C. Mummolo, L. Mangialardi, and J. H. Kim, "Quantifying dynamic characteristics of human walking for comprehensive gait cycle," *Journal of Biomechanical Engineering*, vol. 135, no. 9, 2013.
 - [33] X. Su, H. Tong, and P. Ji, "Activity recognition with smartphone sensors," *Tsinghua Science and Technology*, vol. 19, no. 3, pp. 235–249, 2014.
 - [34] Q. Zou, Y. Wang, Q. Wang, Y. Zhao, and Q. Li, "Deep learning based gait recognition using smartphones in the wild," 2018, <https://arxiv.org/abs/1811.00338>.
 - [35] Y.-Y. Zheng, J.-L. Kong, X.-B. Jin, X.-Y. Wang, and M. Zuo, "CropDeep: the crop vision dataset for deep-learning-based classification and detection in precision agriculture," *Sensors*, vol. 19, no. 5, p. 1058, 2019.
 - [36] X.-B. Jin, N.-X. Yang, X.-Y. Wang, Y.-T. Bai, T.-L. Su, and J.-L. Kong, "Deep hybrid model based on emd with classification by frequency characteristics for long-term air quality prediction," *Mathematics*, vol. 8, no. 2, p. 214, 2020.