Motion Pattern Classification based on Multi-sensors Information Fusion through ResNet and LSTM

Lie Yu, Pengliang Chen and Lei Ding

Abstract—**Gait pattern classification is a comprehensive technique applied in the fields of medical diagnosis, rehabilitation training and human perception. In this paper, we designed a gait phase detection system with six inertial measurement units (IMUs) fixed on the the lower limbs and four force-sensitive resistors (FSRs) embedded in the shoe insole. The sensor data are collected through a wireless sensor system in three walking statuses, such as walking on flat ground, upstairs and downstairs. Then, a hybrid neural network is proposed which combine the advantages of residual networks (ResNet) and long short-term memory (LSTM) to segment the gait phases. The data are first processed through the ResNet to extract the deep features and then analyzed by the LSTM network to fully exploit the sequential trait of gait cycles. Data normalization and sliding window methods are adopted to fit in the ResNet to improve the convergence speed of training. Seven groups of experiments are conducted to analyze the influence of different parts of IMUs on overall accuracy. The model performed best when all six IMUs and four FSRs are utilized, with an average accuracy of 93.67%, which is higher than another hybrid neural network and other traditional algorithms.**

Index Terms—**Gait pattern classification, IMU, FSR, ResNet, LSTM.**

I. Introduction

Walking is a basic activity in human's daily life. As a systematic and comprehensive means in kinesiology, gait analysis is an important component in intelligent surveillance, physical condition monitoring, rehabilitation therapy and intelligent prosthetic leg design. Gait analysis also plays a key role in hot fields like exercise guidance and identity verification. [1-7]

Existing methods for gait data collection can be categorized into direct and indirect approaches. Direct methods involve the use of physical sensors, such as force plates, inertial measurement units (IMUs) and force-sensitive resistors (FSRs), which capture biomechanical signals like ground reaction forces or lower limb kinematics [8, 9]. The indirect approach includes unintrusive sensors, such as video camera [10], infrared camera [11], ultrasound imaging device [12, 13], light detection and ranging device [14, 15]. Based on the two approaches, multiple well-developed systems have been developed such as 3D force plates [16], infrared camera systems, systematic data processors and computers. To achieve a balance between structural simplicity, real-time monitoring capability and high sampling rates, we selected IMUs and FSR as the primary input sensors for our proposed model. This choice ensures efficient gait data acquisition while maintaining portability and practical applicability in real-world settings.

With the advancement of artificial neural networks and their powerful predictive capabilities, recent research has predominantly adopted deep learning architectures as the backbone of gait analysis models [17, 18]. Zhou Zirui et al. uses convolutional neural networks (CNNs) as backbone and gait silhouette as input to classify early scoliosis. The sensitivity score of their model achieved 99.0% compared to 90.6% of conventional scoliosis screening techniques [19]. Zhang Ziyuan et al. disentangle appearance, canonical and pose features from RGB imagery by adopting long short-term memory (LSTM) network as backbone and achieved state of the art performance at multiple public datasets [20]. Zheng Jinkai et al. proposed the 3D skinned multi-person linear model of the human body. The model utilized CNN and 3D-spatial-transformation network to fuse 3D skinned multi-person linear model and silhouette into the gait feature [21]. Xin Chen et al. proposed a multi-gait recognition model to address the challenge of gait interference when people walking together. The model combined a latent conditional random field model and support vector machine to extract gait feature of single people from multi-walking RGB imagery [22]. Shaochen Xu et al. proposed an efficient real-time detection method to detect three gait phases based on a bidirectional long short-term memory network. The model accept data of a single IMU which attached to the shank and achieved 97.40% of average recognition accuracy [23]. More recently, a number of hybrid neural network model have achieved comparable good results in various fields by combining CNN, LSTM and other models such as residual networks (ResNet) [24] and graph neural networks. Xinyu Wu et al. proposed a graph convolutional network model for gait phase classification. The intention of their design was to control a lower extremity exoskeleton by collecting angle and force data from goniometers and FSRs. The model delivered a significantly high prediction accuracy

Manuscript received February 13, 2025; revised June 20, 2025.

This work was supported by the National Natural Science Foundation of China "Research on motion pattern recognition of exoskeleton robot based on curve similarity model" (NO.62106178).

Lie Yu is an Associate Professor at the School of Electronic and Electrical Engineering, Wuhan Textile University, Wuhan, China (corresponding author to provide phone: +86 18607155647; e-mail: lyu@wtu.edu.cn).

Pengliang Chen is a postgraduate student at the School of Electronic and Electrical Engineering, Wuhan Textile University, Wuhan, China (e-mail: 729151570@qq.com).

Lei Ding is an Associate Professor at the School of Computer Science and Artificial Intelligence, Wuhan Textile University, Wuhan, China (e-mail: lding@wtu.edu.cn).

at 96.12% and better robustness on gait phase classification [25].

Despite all the achievements it has made, the CNN model has demonstrated optimal performance primarily in appearance-based gait analysis methods, such as Gait Energy Image (GEI) and Gait Entropy Image (GEnI) [26]. As an advanced variant of CNN, ResNet introduces a skipping connection mechanism to mitigate the vanishing gradient problem and enable deeper architectures. However, when the network depth increases significantly, ResNet requires extensive training time—often spanning several weeks—due to the high computational complexity. Additionally, its effective receptive field tends to be smaller than theoretically expected, which may limit its ability to capture long-term dependencies in gait analysis. This issue becomes particularly evident when modeling complex temporal relationships across gait phases. As another commonly used neural network, LSTM is particularly effective for processing time-series data due to its ability to capture long-term dependencies. Compared to traditional recurrent neural networks, LSTM exhibits higher computational complexity due to the introduction of gate mechanisms (input, forget and output gates) and long-term memory cells. Furthermore, LSTM networks typically require large-scale datasets for training to prevent overfitting, as insufficient data can lead to poor generalization. These limitations highlight the need for hybrid architectures that can leverage the strengths of both spatial and temporal modeling.

In order to avoid the flaws mentioned above, we combined the LSTM network, which is designed for processing sequential data, with ResNet, which focuses on addressing gradient vanishing and gradient explosion. This approach leverages ResNet's ability to extract deep features from multiple IMUs and FSRs sensors for gait segmentation while also addressing the issue of ResNet's limited receptive field. By integrating temporal modeling through LSTM, the hybrid architecture improves the network's ability to learn sequential dependencies critical for accurate gait phase classification. We developed a custom-designed wireless sensor platform to collect force and angle data for gait analysis. The collected data were subsequently partitioned into training and validation datasets to ensure robust model evaluation and reduce overfitting. To assess the model's performance, we conducted seven experimental trials using six IMUs, exploring various configurations. Each configuration tested different IMU placements to examine their individual and combined contributions. In addition, the effectiveness of FSRs was evaluated by analyzing force and angle data separately, allowing for a comparative assessment of their contributions to gait segmentation. This systematic experimental design supports a deeper understanding of sensor relevance and enhances model interpretability.

II. Materials and Methods

A. Data Collection

Fifteen healthy men (aged 26 ± 2 years, weight 72 ± 10 kg, height 1.75 ± 0.08 m) completed the protocol. In the experiment, participants were instructed to start walking from a fixed standing position at one end of a 4-meter walkway and climbing a 5-meter-long, 3.2-meter-high stairway. Each trial began from the same initial stance to ensure consistency across measurements. All participant were familiarized with the procedure before the actual experiment start. Participants completed several types of walking movements at a self-selected speed.

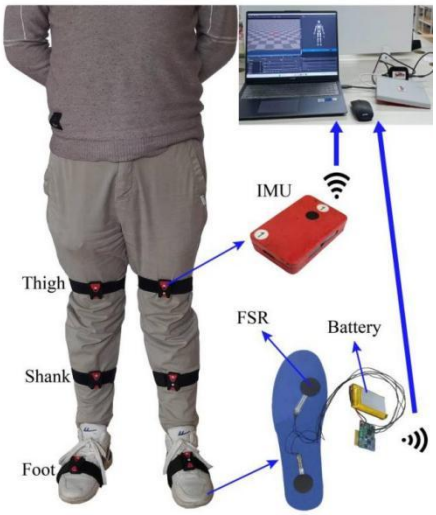

Fig. 1. Overall data collecting process and equipment.

Two sets of data were collected on each end of the path. The data collecting system are shown in Figure 1. The system is composed of four IMUs sensor and two flexible FSRs. Two types of data are then transferred to a computer by a wireless transmitter.

B. Data Preprocessing

In order to reduce computational complexity and improve model performance, the raw data collected by the wireless sensor system underwent a preprocessing stage. Ensuring consistent data dimensions was necessary for compatibility with the residual network, especially since ResNet expects fixed-size input tensors. Additionally, data normalization was applied to prevent features with larger numerical ranges from dominating others, thereby improving the stability and convergence rate of the training process. Noise filtering was also incorporated to remove potential signal artifacts caused by sensor drift or mechanical disturbances. Following these preprocessing steps, the computational cost of training the neural network was significantly reduced, leading to improved gait recognition accuracy. In this study, the raw data were segmented based on different gait phases using the sliding window method, with each sliding window containing 40 samples. The normalization process can be mathematically expressed as follows:

$$\widetilde{x_i} = \frac{x_i - x_{\min}}{x_{\max} - x_{\min}}\left(x_{\max} - x_{\min}\right) + y_{\min} \qquad (1)$$

where x_i is the number of sampled data from input data series. x_{max} and x_{min} are the peak and trough point respectively. y_{max} and y_{min} are the upper and lower edge of the normalized data. In this paper we set y_{max} and y_{min} as 1 and -1 respectively.

(a) (b)

Fig. 2. Data collection. Stair descent (a); Stair ascent (b).

Each experiment is considered to be a continuous process and two stages are included: the gait of level walking and the gait of stair walking. For instance, in one type of experiment, data were collected on two different gaits simultaneously: level walking, the up-stair walking or down-stair walking. The data collecting process is shown in Figure 2.

C. Data Labeling

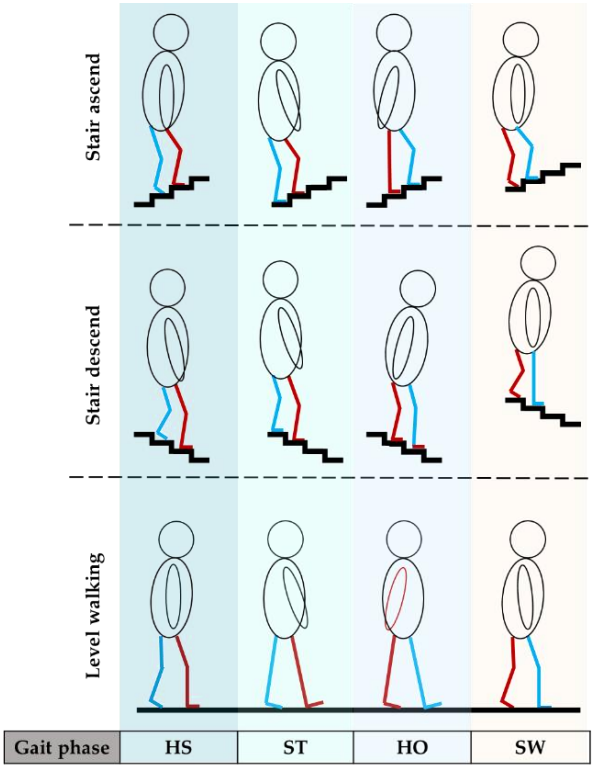

Fig. 3. Four gait segments of three walking status. heel strike (HS), full stance (ST), heel off (HO) and swing (SW).

Although the kinematic characteristics between walking on level ground and stair ambulation differ considerably due to variations in elevation, joint angles, and muscle activation patterns, both scenarios share fundamental gait cycle components. In this study, the gait cycle is segmented into four distinct phases: heel strike (HS), full stance (ST), heel off (HO), and swing (SW), as illustrated in Figure 3. These phases serve as the basis for both temporal segmentation and classification tasks in our model.

To label the dataset accurately, a rule-based segmentation strategy was implemented based on force signals collected from foot pressure sensors. The gait event detection criteria are visualized in Figure 4 and defined as follows :

Rules for HS: As seen in Figure 4, the strike phase lasts from the primary leg touches ground to the secondary leg leaves ground, during which the heel force of primary foot rises to peak and the force of ball for the secondary leg drops to zero.

Rules for ST: As seen in Figure 4, heel force and ball force of the primary leg drops to zero during stance phase. For the secondary leg, the forces of both parts stay zero before it reaches to ground. The end of stance phase can be identified by a flat-zone detection rule.

Rules for HO: As seen in Figure 4, ball force of the primary leg drops near to zero during heel off phase. The end of heel off phase can be identified by a flat-zone detection rule.

Rules for SW: As seen in Figure 4, both forces of two parts stay near zero before another heel strike phase.

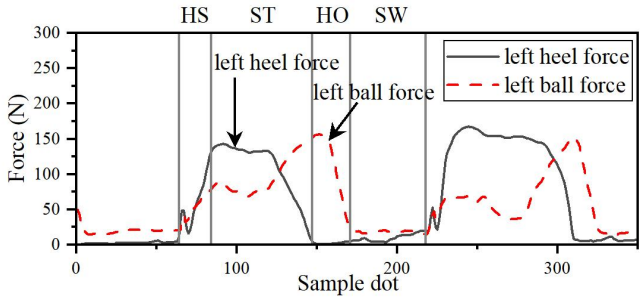

Fig. 4. Segmentation of four gait phases. The sample rate is set as 100Hz.

This segmentation method has been proven to be simple and effective, as it provides a clear representation of gait phases. In this study, the labeling criteria for the four gait phases are primarily determined using a force threshold, as illustrated in Figure 4.

For model training, each data point in the time series was assigned a one-hot encoded label representing one of the four gait phases. The output layer of our classification model is configured to produce a probability vector of length four. In this study, we select the label with the highest probability as the predicted output, assigning a value of 1 to indicate selection and 0 to indicate discard. Gait phase of HS, ST, HO and SW are denoted as [1 0 0 0], [0 1 0 0], [0 0 1 0] and [0 0 0 1] respectively.

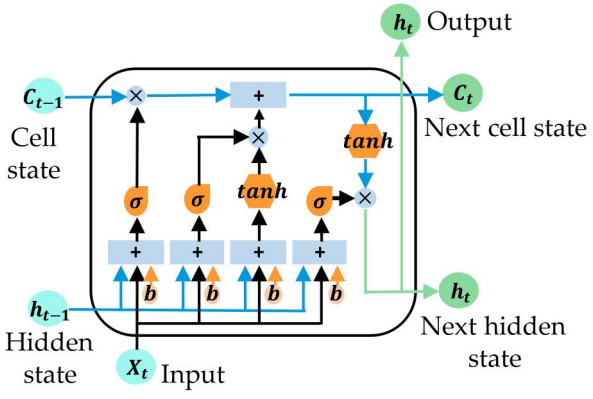

Fig. 5. Internal structure of LSTM unit.

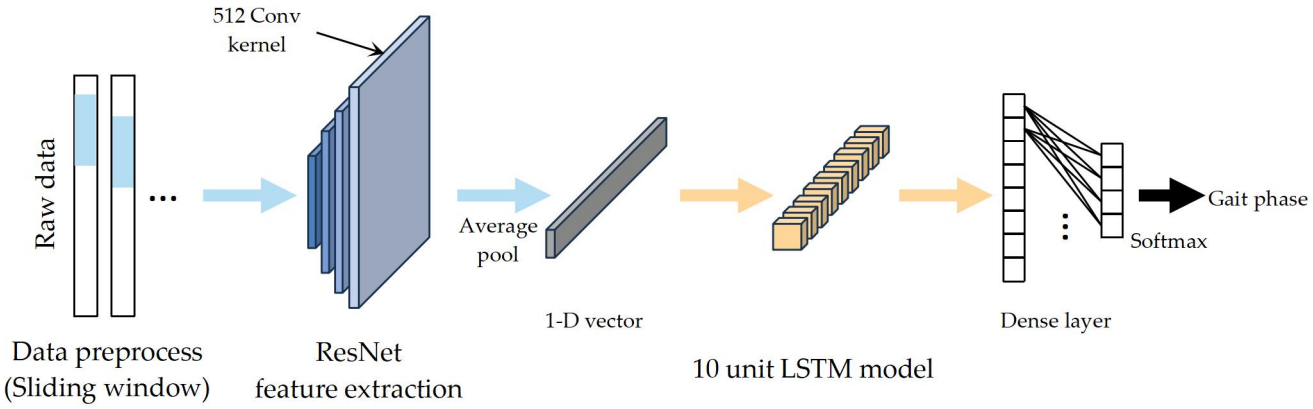

Fig. 6. The structure of ResNet-LSTM model

D. Structure of ResNet-LSTM Neural Network

As a variant of recurrent neural network, LSTM is designed to address the vanishing and exploding gradient problems encountered during the training of deep recurrent models. Specifically, three gates and cell state mechanism are introduced to properly process the long-term dependent feature. The forget gate determines the extent to which previous information should be discarded from the cell state. The input gate regulates the incorporation of new information into the cell state. The output gate determines how much of the cell state contributes to the hidden state and ultimately influences the network's output. Through these gating mechanism, LSTM effectively preserves long-range dependencies while preventing the degradation of gradient magnitudes, making them particularly well-suited for tasks involving sequential data. The internal structure of LSTM unit is shown in Figure 5. Each gate mechanism can be mathematically expressed as follows:

$$\begin{cases} i_t = \sigma(W_i h_{t-1} + W_i X_t + b_i) \\ f_t = \sigma(W_f h_{t-1} + W_f X_t + b_f) \\ o_t = \sigma(W_o h_{t-1} + W_o X_t + b_o) \end{cases} \quad (2)$$

where i_t, f_t and o_t represent the activation of the input gate, the forget gate and the output gate respectively. The update formula and output of the whole unit h_t are as follows:

$$\begin{cases} h_t = O_t \times \tan(C_t) \\ C_t = f_t \times C_{t-1} + i_t \times \tan(W_c h_{t-1} + W_c X_t + b_c) \end{cases} \quad (3)$$

where x_t represents input, h_t represents state of the hidden layer at time t, and w and b represent the trainable weight matrix and bias term of LSTM unit respectively.

ResNet is an advanced deep convolutional neural network to address the degradation problem encountered in very deep networks. Traditional deep learning model such as AlexNet [27], GoogleNet [28] are constructed with 8 layers and 22 layers respectively. Before the introduction of ResNet, the most effective deep learning models typically contained fewer than 30 layers, as deeper architectures often suffered from vanishing or exploding gradients. To mitigate these issues, ResNet incorporates shortcut connections within feed-forward neural networks, allowing the input to bypass one or more layers. This residual learning framework effectively stabilizes gradient propagation without introducing additional trainable parameters, thereby facilitating the training of significantly deeper networks. The internal structure of ResNet unit is shown in Figure 7.

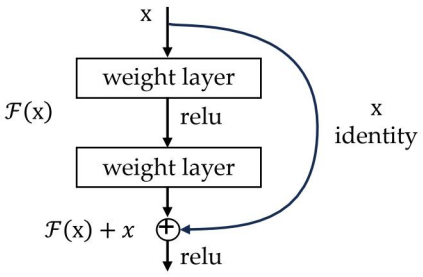

Fig. 7. Basic residual leaning unit.

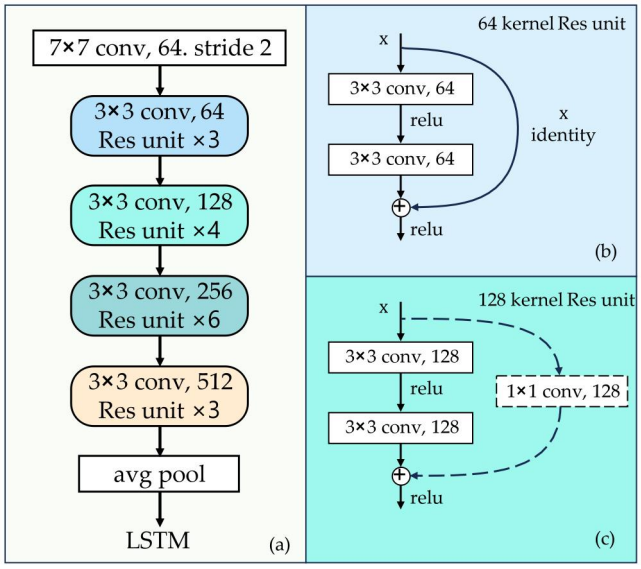

Fig. 8. The structure of 34 layer ResNet (a); Inner structure of residual unit (b); First unit of the second, third and fourth blocks (c).

The ResNet-LSTM neural network model structure used for gait recognition in this paper is demonstrated in Figure 6. Our model consists of a 34-layer ResNet and a 2-layer LSTM network. The ResNet architecture is divided into five main components as shown in Figure 8: initial convolution layer is a 7x7 convolutional layer with 64 filters and a stride of 2 is used to unify the dimensions of the input data. residual blocks: the model passes through four residual blocks, each with a different number of convolutional kernels: 64, 128, 256 and

512, progressively extracting deeper feature representations. In the second, third and fourth blocks, the first unit includes a pooling layer that adjusts the dimensions of the data. Global average pooling is after the residual blocks, an average pooling layer converts the extracted features into a one-dimensional vector. Following the feature extraction stage, the resulting feature vector is input into a two-layer LSTM. By effectively capturing long-range dependencies in time-series data, the LSTM part enhances the model's predictive accuracy, making it particularly suitable for gait segmentation tasks.

E. Evaluation Method

In this paper, we employed accuracy, Micro-F1, and Macro-F1 as evaluation metrics to assess the performance of the ResNet-LSTM model. The Micro-F1 and Macro-F1 scores are particularly relevant for multi-class classification tasks, as they provide a comprehensive measure of model performance across different categories. Given the imbalanced duration of the four gait phases, the Macro-F1 score is specifically computed to account for this discrepancy. As shown in Figure 4, ST and SW phases last approximately twice as long as the HS and HO phases, which can lead to model bias toward the longer-duration classes if not properly addressed. To mitigate the potential statistical bias introduced by this imbalance, we first calculate the precision and recall for each class individually based on their true positive, false positive, and false negative rates. Meanwhile, the F1-score for each class is determined, and the Macro-F1 score is obtained by averaging these individual F1-scores equally, regardless of class size. This approach ensures that the Macro-F1 score provides a balanced evaluation of model performance across gait phases with both high and low representation, effectively capturing variations in recall and precision across all categories. Moreover, using both Macro- and Micro-F1 scores allows for complementary insights into how well the model performs on rare versus frequent phases. The formula of accuracy, recall, and precision is as follows,

$$\begin{cases} A_{cc} = \dfrac{TP + TN}{TP + TN + FP + FN} \\ R_{ec} = \dfrac{TP}{TP + FN} \\ P_{re} = \dfrac{TP}{TP + FP} \end{cases} \tag{4}$$

where TP, TN, FP and FN represent true positive, true negative, false positive and false negative respectively. The formula of Macro-F1 is as follows.

$$\begin{cases} F1_i = 2\dfrac{R_{eci} \times P_{rei}}{R_{eci} + P_{rei}} \\ F1_{marco} = \dfrac{1}{k}\sum_{i=1}^{k} F1_i \end{cases} \tag{5}$$

where k represents the number of classes and $F1_i$ represent the F1 score of class i. The formula of Micro-F1 is as follows.

$$\begin{cases} P_{micro} = \dfrac{\sum_{i=1}^{k} l_i}{\sum_{i=1}^{k} m_i} \\ R_{micro} = \dfrac{\sum_{i=1}^{k} l_i}{\sum_{i=1}^{k} n_i} \\ F1_{micro} = \dfrac{2}{\dfrac{1}{P_{micro}} + \dfrac{1}{R_{micro}}} \end{cases} \tag{6}$$

where k represents the number of classes; l_i and m_i represent the true positive and positive number respectively of class i.

III. Experiment and Results of Gait Recognition

The partial walking data of subject 1 in three walking status are shown in Figure 9. In this paper, PyCharm2023 was used as the training software for the neural network model. The hardware configuration includes i5-7300k, GTX1070ti graphics card and 32 GB memory

A. The Training of ResNet-LSTM

The force and angle data collected by IMUs in Figure 9 were preprocessed through two methods: dimension normalization and one-hot encoding. The IMUs were attached to the lower limb on shank and thigh, which generate four groups of raw data demonstrated in Figure 9. We conducted two experiments using different combinations of sensor types and locations. Additionally, a comparative experiment was performed to evaluate gait recognition across three different terrains. In each experiment, the dataset was split at a 7:3 ratio, meaning 70% of the total data was allocated for training and 30% for testing. The training process employed cross-entropy as the loss function and adaptive movement estimation algorithm (Adam) as the optimizer. Training was considered complete when the accuracy reached its peak for 10 consecutive iterations. The corresponding learning curve is presented in Figure 10.

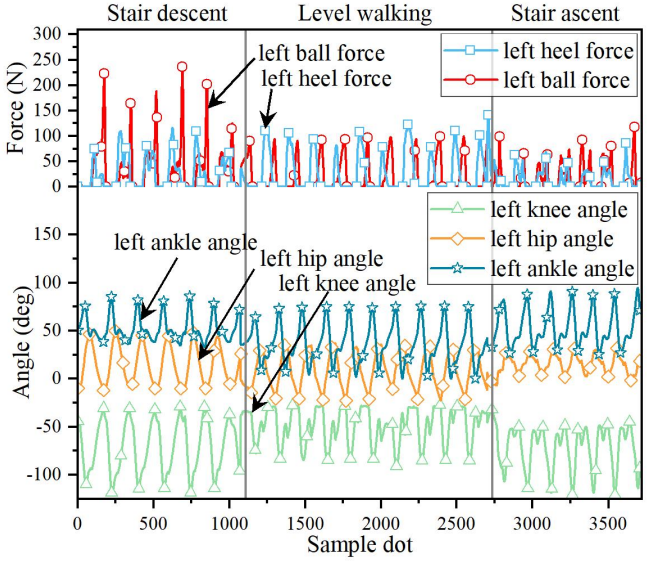

Fig. 9. IMUs data of left leg (primary leg). The sample rate is set as 100Hz.

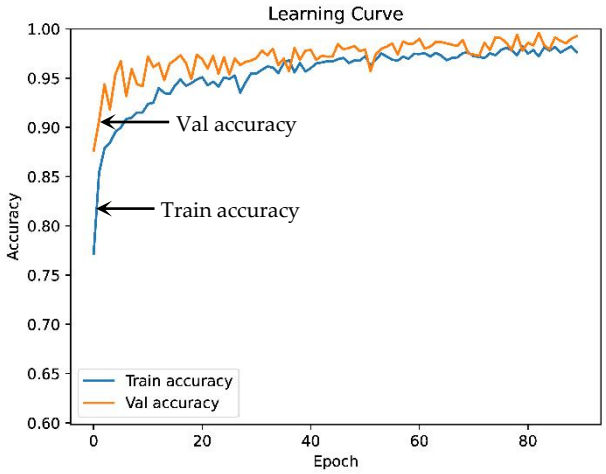

Fig. 10. Training accuracy and validation accuracy during the training process.

B. *Experiment and Results*

To evaluate the effectiveness of angle and force data in gait recognition, we calculated the accuracy, macro-F1, and micro-F1 scores based on the formulas described in Section II. Table I presents a comparative analysis of force and angle sensors, demonstrating that both modalities contribute significantly to improving gait recognition accuracy. The integration of these two sensor types offers complementary information, with angle data providing dynamic movement patterns and force data capturing contact characteristics with the ground. Table II compares the performance of angle sensors placed at three different lower limb locations, revealing that placement on the thigh and shank provides the most discriminative features due to the consistent motion patterns and clearer joint rotation data at these sites. Table III summarizes the results of seven contrast experiments across four gait phases—HS, ST, HO, and SW—reporting accuracy, macro-F1, and micro-F1 scores. Among all evaluated models, ResNet-LSTM achieved the highest performance. The precision scores of all four gait phases exceeded 90%, with the SW phase reaching 97% in the seventh experimental group, indicating excellent sensitivity to the swing motion and effective sequence modeling by the LSTM module. However, the model exhibited its lowest performance in the third contrast experiment, where both precision and recall fell below 80%, likely due to insufficient or redundant sensor data, such as missing IMU channels or suboptimal placement.

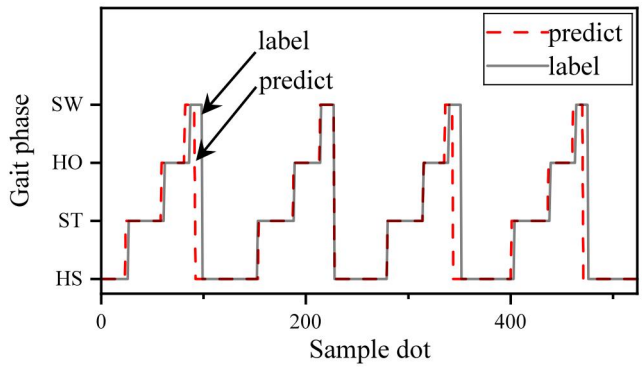

Fig. 11. Gait phase classification output. HS, ST, HO and SW denote heel strike, full stance, heel off and swing respectively.

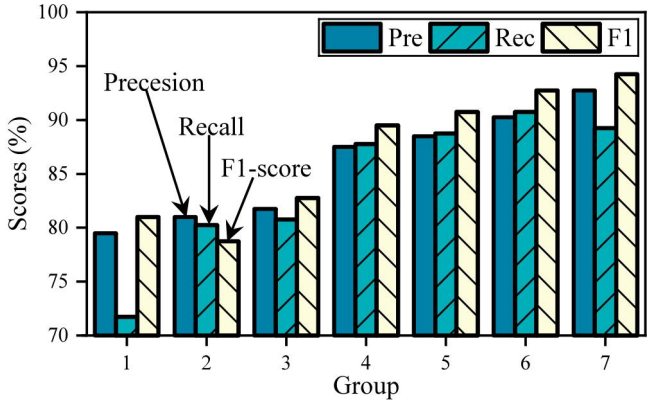

Fig. 12. Results of the presented ResNet-LSTM in 7 experiments.

TABLE I
ACCURACY, MACRO-F1 AND MICRO-F1 SCORE OF GAIT RECOGNITION USING TWO TYPES OF SENSORS.

Data type	Accuracy	Macro-F1	Micro-F1
Force + Angle	92.47%	95.33%	94.73%
Force	86.86%	90.60%	87.26%
Angle	85.59%	82.97%	86.90%

In addition, the F1-score of HO and HS are the only ones that fall below 80% across all evaluation metrics, suggesting that these phases may be more ambiguous and brief, making them harder to detect accurately. Comparing the results from group 1 to group 3, the performance of ResNet-LSTM improved significantly when all three IMUs were used as inputs, supporting the importance of multi-sensor fusion. From group 4 to group 7, the model achieved an average F1-score of 91.81%, reflecting the stability and robustness of the proposed framework under varied sensor configurations. Figure 12 presents the average precision, recall, and macro-F1 scores across seven groups, with group 7 showing the best performance. The highest F1-score was observed in group 7, reaching an average of 94.25%. The segmentation output for group 7 is illustrated in Figure 11, demonstrating high temporal alignment between predicted labels and ground truth. Additionally, we conducted a comparative experiment across three different terrains, with the results summarized in Table IV. The accuracy for stair ascent was 3% lower than that for level ground and stair descent, possibly due to greater biomechanical variability and altered postural control when climbing stairs. The classification performance was further analyzed using a confusion matrix, as shown in Figure 13, which helped identify common misclassification trends and assess model robustness phase by phase.

TABLE II
THE GROUPS OF EXPERIMENTS DEPEND ON THE IMU POSITION COMBINATION.

Grou	Position Combination of IMUs
1	Left and right knee
2	Left and right hip
3	Left and right ankle
4	Left and right knee and hip
5	Left and right knee and ankle
6	Left and right hip and ankle
7	Left and right knee, hip and ankle

TABLE III.
COMPARISON OF PRECISION, RECALL AND F1-SCORE OF DIFFERENT GROUPS.

Groups	Gait Phases	Precision (%)	Recall (%)	F1-score (%)
1	HS	84	77	85
	ST	82	72	83
	HO	71	75	73
	SW	81	63	83
2	HS	87	85	75
	ST	82	80	82
	HO	75	76	75
	SW	80	80	83
3	HS	85	81	86
	ST	84	82	84
	HO	75	76	76
	SW	83	84	85
4	HS	88	89	90
	ST	87	87	89
	HO	85	84	86
	SW	90	91	93
5	HS	93	93	94
	ST	86	87	89
	HO	85	85	88
	SW	90	90	92
6	HS	88	87	91
	ST	90	91	93
	HO	87	88	90
	SW	96	97	97
7	HS	90	91	93
	ST	91	92	95
	HO	93	93	94
	SW	97	81	95

TABLE IV
RESULTS OF GAIT RECOGNITION FOR THREE TERRAINS.

Data type	Accuracy	Macro-F1	Micro-F1
Level ground	93.67%	93.62%	92.93%
Stair ascent	90.20%	87.50%	87.28%
Stair descent	93.54%	70.43%	75.47%

TABLE V
PERFORMANCE OF DIFFERENT ALGORITHMS FOR LEVEL WALKING

Algorithm	Accuracy	Precision	F1-score
LSTM-DNN (2 phase)	91.80%	93.70%	94.03%
LSTM	89.10%	91.10%	90.53%
KNN	72.66%	74.67%	74.00%
SVM	74.33%	77.67%	75.57%
ResNet-LSTM (ours)	92.67%	93.85%	95.63%

Table V presents the performance comparison of five gait analysis algorithms. The scores for four of these algorithms correspond to gait detection during level-ground walking. LSTM-DNN [29], apart from the other three algorithms, detects only two gait phases (stance and swing phase), offering a simplified classification approach that may not generalize well to more detailed tasks. According to Table V, the ResNet-LSTM algorithm achieved the highest gait phase accuracy and F1-score, reaching 92.67% and 95.63% respectively, demonstrating strong capability in handling multi-phase gait segmentation. Additionally, both ResNet-LSTM and LSTM demonstrate high gait phase recognition performance, with both accuracy rates and F1-scores exceeding 89%, highlighting their effectiveness in capturing temporal dynamics from sensor data. In contrast, the k-nearest neighbor classification (KNN) and support vector machine (SVM) algorithms exhibit relatively low recognition rates, with accuracy and F1-scores below 86%, which indicates their limitations when applied to complex, high-dimensional gait signals.

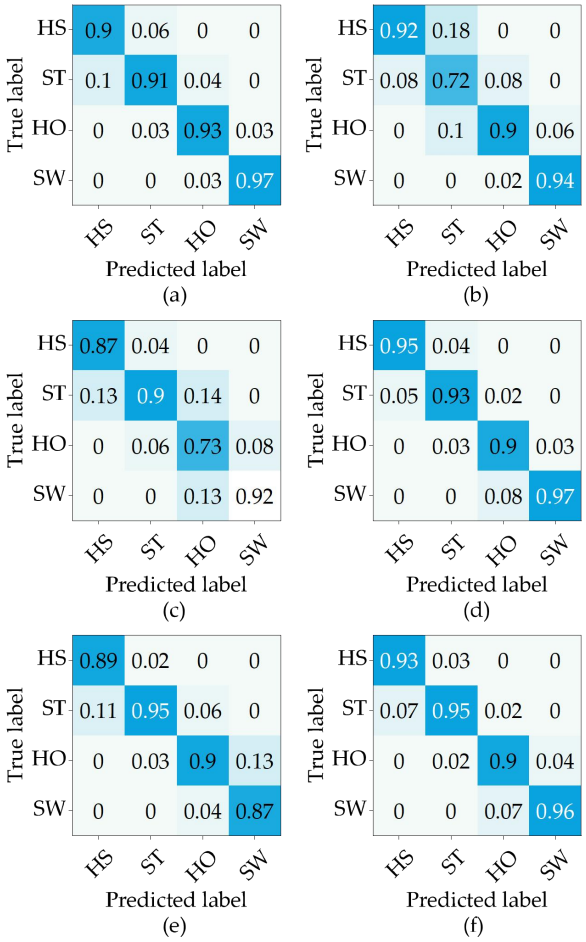

Fig. 13. Confusion matrix of sensor and terrain comparing experiment. Force and angle sensors (a); Force sensors (b); Angle sensors(c); Level ground (d); Stair ascent (e); Stair descent (f).

C. *Experiment and Results Based on ResNet-LSTM-TPA*

We have demonstrated the regulations of four gait phase in Chapter II, Section C. Each gait phase is labeled according to certain gait event; For example, heel strike starts at the moment when primary leg touches ground. For the purpose of gait phase segmentation, the model doesn't need to pay attention to whole time series but only to the key moment. LSTM tends to treat all time steps equally when processing variable-length sequences (or summarizes information solely through the final hidden state), which may lead to the neglect of critical time points within the sequence. Therefore, we added a temporal attention (TPA) [30] layer after LSTM to dynamically assign different levels of importance to the hidden states at each time step, enabling the model to focus more effectively on moments that are highly relevant to the classification task.

As shown in Figure 14, the LSTM produces hidden states at each time step, denoted as $h_1, h_2, ..., h_T$. These hidden states are passed to the attention module along with the final hidden state h_T, which serves as the query vector q. Each LSTM output h_T and the query q are independently projected using learned weight matrices, then combined through a non-linear activation function, typically tanh. The resulting vectors are further projected using a learned attention vector v to compute a scalar score $score_t$, representing the relevance of each time step t to the final output. These scores are then normalized using the softmax function to generate attention weights. The attention weights

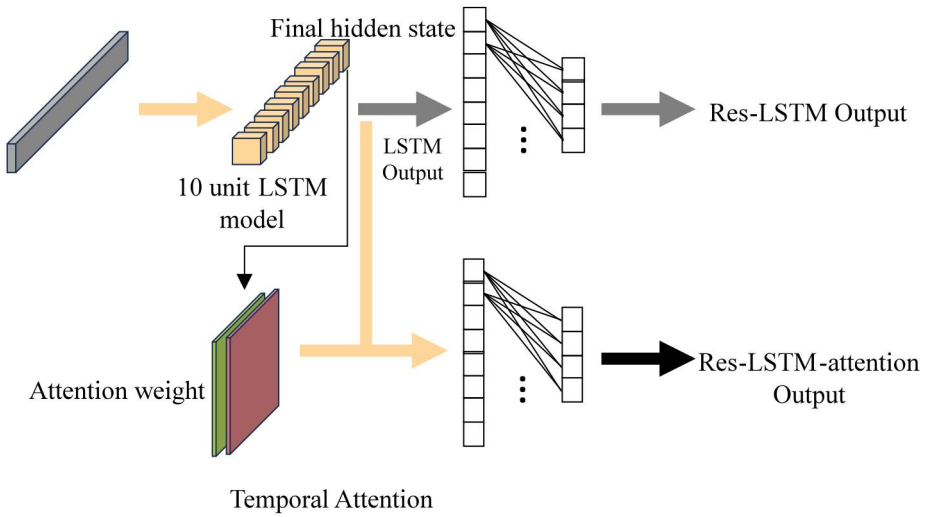

Fig. 14. ResNet-LSTM-TPA model

are used to compute a weighted sum of all LSTM outputs, resulting in a context vector that aggregates information from the entire sequence while emphasizing the most informative time steps. This context vector is then passed to subsequent layers (typically a fully connected layer followed by a softmax layer) for final classification or prediction.

Figure 15 shows the internal structure of TPA layer. The TPA attention score at time step t can be denoted as:

$$score_t = v^T \tanh(W_1 h_t + W_2 q) \quad (7)$$

where h_t represents he hidden state output from the LSTM at time step t; q represents the query vector, in this paper we choose the final hidden state as the query vector; W_1 and W_2 are learnable weight matrices. Both map to a common attention space. v^T is a learnable vector which is used to project the combined representation to a scalar score.

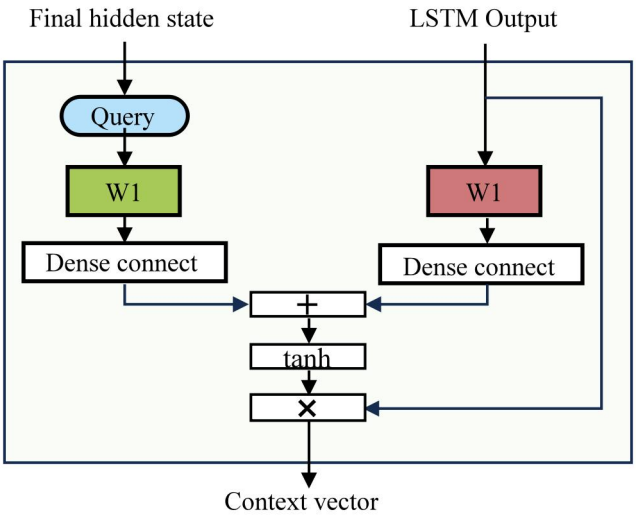

Fig. 15. Internal structure of TPA layer

The data and training methodology remain unchanged and are described in Chapter III, Section A. The learning curve converged at epoch 35, and the validation set achieved an accuracy of 93.33%. The primary attention weights from the TPA mechanism were extracted and visualized as a plot. Every line represents a batch of gait cycle. As shown in Figure 16, a scatter plot of gait phase change points is placed above the attention weight plot, aligned along the time axis. The attention distribution exhibits three distinct peaks, each of which corresponds to a gait phase transition. Notably, the transition from Swing (SW) to Heel Strike (HS) consistently occurs at the start of each gait cycle, and thus does not need to be explicitly marked in the scatter plot. Nevertheless, the attention weight plot still shows a high activation around this transition, indicating that the model has implicitly learned to attend to this critical event. This further demonstrates the model's capability to focus on key temporal dynamics within the gait cycle.

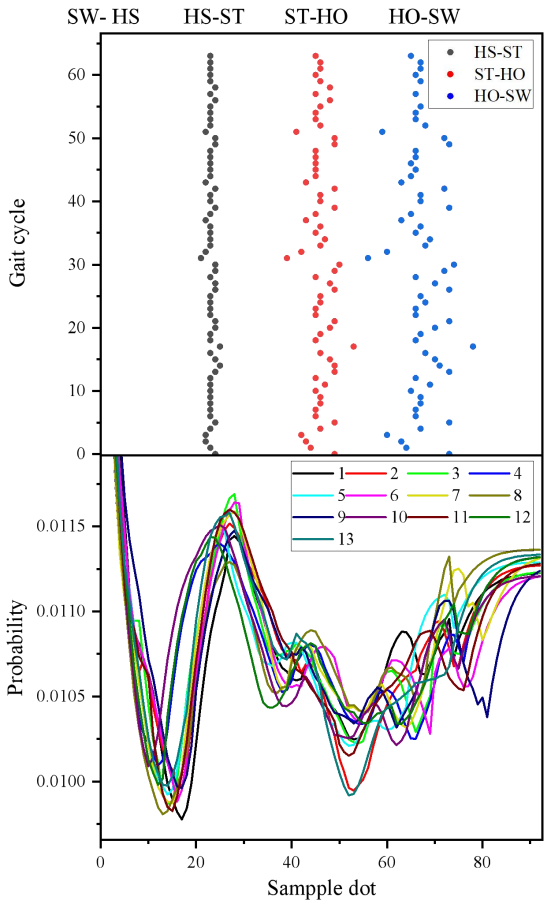

Fig. 16. Gait phase change scatter plot and attention weight

TABLE VI
PERFORMANCE OF DIFFERENT ALGORITHMS FOR LEVEL WALKING

Algorithm	Accuracy	Precision	F1-score
LSTM-DNN (2 phase)	91.80%	93.70%	94.03%
LSTM	89.10%	91.10%	90.53%
KNN	72.66%	74.67%	74.00%
SVM	74.33%	77.67%	75.57%
ResNet-LSTM (ours)	92.67%	93.85%	95.63%
ResNet-LSTM-TPA (ours)	93.33%	93.99%	95.80%

IV. DISCUSSION

In this paper, we propose a novel model that integrates ResNet and LSTM to segment four gait phases across three different terrains, including stair ascent, stair descent and level walking. The proposed approach utilizes FSR and IMU sensors as input to the ResNet-LSTM model. We conducted seven experiments to evaluate the necessity of using three IMUs positioned at different locations. The results presented in Table II indicate that incorporating additional IMUs improves precision, recall and macro-F1 scores. When more than four IMUs were used in the final three experiments, the majority of these metrics exceeded 85%. Notably, in the seventh experiment, all three scores surpassed 90%, leading to the conclusion that all six IMUs placed at different positions contribute to improving gait phase segmentation accuracy. Furthermore, the results of the group 7 experiment demonstrated that the ResNet-LSTM model achieved strong performance in gait recognition in gait recognition, with accuracy, macro-F1 and micro-F1 of 92.67%, 95.33% and 94.73% respectively. The enhanced ResNet-LSTM-TPA model achieved slightly better performance, reaching an accuracy of 93.33%, indicating that the integration of the TPA mechanism further improves the model's ability to capture critical features for gait phase segmentation.

V. CONCLUSION

To sum up, the results of this study show that the proposed ResNet-LSTM model achieved 92.47% accuracy in detecting four gait phases across up-stair, down-stair and level ground walking status. This result suggests that there are underlying similarities in gait locomotion across these three walking terrains. The input data for the model were collected using six IMUs and four FSRs through a self-developed wearable wireless sensor system. This paper conducted a contrast experiment by combining different locations of IMUs into seven groups. The results indicate that the proposed model performed best when all six IMUs were used, achieving an accuracy and macro-F1 of 92.67% and 95.63% respectively. The innovation of this paper is that we integrated three most common walking conditions for gait phase segmentation. The accuracy of level ground, stair ascent and stair descent walking achieved 93.67%, 90.20% and 93.54% respectively with a slight drop for stair ascent. ResNet-LSTM-TPA achieved slightly better performance, reaching an accuracy of 93.33%.

Regarding future trends, we expect an increase of overall accuracy when adopting state of the art technique such as transformer networks which targets data streaming and avoid recursion decay such as the sliding window method that we used in data preprocessing. Furthermore, given the strong correlation between gait event detection and gait phase segmentation, it may be worthwhile to explore adapting our model for gait event detection, potentially enhancing its applicability in real-world scenarios.

REFERENCES

[1] A. Nambiar, A. Bernardino, and J. C. Nascimento, "Gait-based Person Re-identification: A Survey," *ACM Computing Surveys*, vol. 52, no. 2, pp1-34, 2019.

[2] G. Cicirelli and D. Impedovo, "Human Gait Analysis in Neurodegenerative Diseases: A Review," *IEEE Journal of Biomedical and Health Informatics*, vol. 26, no. 1, pp229-242, 2021.

[3] S. Mazilu, U. Blanke, and M. Dorfman, "A Wearable Assistant for Gait Training for Parkinson's Disease with Freezing of Gait in Out-of-the-Lab Environments," *ACM Transactions on Interactive Intelligent Systems*, vol. 5, no. 1, pp1-31, 2015.

[4] S. Ranaldi, M. C. De, and M. Serrao, "Characterization of Prosthetic Knees Through a Low-Dimensional Description of Gait Kinematics," *Journal of NeuroEngineering and Rehabilitation*, vol. 20, no. 1, pp46, 2023.

[5] K. K. Nagwanshi, "Cyber-Forensic Review of Human Footprint and Gait for Personal Identification," *IAENG International Journal of Computer Science*, vol. 46, no. 4, pp645-661, 2019.

[6] I. Baokar, and L. He, "Memristive Neural Networks Application In Predicting of Health Disorders," *Lecture Notes in Engineering and Computer Science: Proceedings of The International MultiConference of Engineers and Computer Scientists 2023*, *IMECS 2023*, 5-7 July, 2023, Hong Kong, pp94-99.

[7] Y. Huang, J. Chen, and S. Fan, "Applying Time-Frequency Image of Convolutional Neural Network to Extract Feature on Long-Term EEG Signals to Predict Depth of Anesthesia," *Lecture Notes in Engineering and Computer Science: Proceedings of The World Congress on Engineering 2019*, *WCE 2019*, 3-5 July, 2019, London, U.K.. pp463-467.

[8] G. P. Panebianco, M. C. Bisi, and R. Stagni, "Analysis of the Performance of 17 Algorithms from a Systematic Review: Influence of Sensor Position," *Gait & Posture*, vol. 66, pp76-82, 2018.

[9] D. Jarchi, J. Pope, and T. K. M. Lee, "A Review on Accelerometry-Based Gait Analysis and Emerging Clinical Applications," *IEEE Reviews in Biomedical Engineering*, vol. 11, pp177-194, 2018.

[10] C. Fan, S. Hou, and J. Wang, "Learning Gait Representation From Massive Unlabelled Walking Videos: A Benchmark," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 45, no. 12, pp14920-14937, 2023.

[11] Z. Xue, D. Ming, and W. Song, "Infrared Gait Recognition Based on Wavelet Transform and Support Vector Machine," *Pattern Recognition*, vol. 43, no. 8, pp2904-2910, 2010.

[12] R. M. Kiss, L. Kocsis, and Z. Knoll, "Joint Kinematics and Spatial-Temporal Parameters of Gait Measured by an Ultrasound-Based System," *Medical Engineering & Physics*, vol. 26, no. 7, pp611-620, 2004.

[13] J. Roggendorf, S. Chen, and S. Baudrexel, "Arm Swing Asymmetry in Parkinson's Disease Measured with Ultrasound Based Motion Analysis During Treadmill Gait," *Gait & Posture*, vol. 35, no. 1, pp116-120, 2012.

[14] M. M. Rahman, D. Martelli, and S. Z. Gurbuz, "Gait Variability Analysis with Multi-Channel FMCW Radar for Fall Risk Assessment," *Proceedings of IEEE 12th Sensor Array and Multichannel Signal Processing Workshop*, *SAM 2022*, 20-23 June, 2022, Trondheim, Norway, pp345-349.

[15] H. Abedi, A. Ansariyan, and P. P. Morita, "AI-Powered Noncontact In-Home Gait Monitoring and Activity Recognition System Based on mm-Wave FMCW Radar and Cloud Computing," *IEEE Internet of Things Journal*, vol. 10, no. 11, pp9465-9481, 2023.

[16] G. S. Faber, C. Chang, and I. Kingma, "A Force Plate Based Method for the Calibration of Force/Torque Sensors," *Journal of Biomechanics*, vol. 45, no. 7, pp1332-1338, 2012.

[17] A. Sepas-Moghaddam and A. Etemad, "Deep Gait Recognition: A Survey," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 45, no. 1, pp264-284, 2022.

[18] G. S. C. Filipi, D. S. Oliveira, and A. P. Leandro, "Gait Recognition Based on Deep Learning: A Survey," *ACM Computing Surveys*, vol. 55, no. 2, pp1-34, 2022.

[19] Z. Zhou, J. Liang, Z. Peng et al., "Gait Patterns as Biomarkers: A Video-Based Approach for Classifying Scoliosis," *Proceedings of*

Medical Image Computing and Computer Assisted Intervention, *MICCAI 2024*, 6-10 Oct, 2024, Marrakesh, Morocco, pp284-294.
[20] Z. Zhang, L. Tran, F. Liu et al., "On Learning Disentangled Representations for Gait Recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 44, no. 1, pp345-360, 2022.
[21] J. Zheng, X. Liu, W. Liu et al., "Gait Recognition in the Wild with Dense 3D Representations and A Benchmark," *Proceedings of IEEE/CVF Conference on Computer Vision and Pattern Recognition*, *CVPR 2022*, 18-20 June, 2022, New Orleans, LA, USA., pp20196-20205.
[22] X. Chen, J. Weng, W. Lu et al., "Multi-Gait Recognition Based on Attribute Discovery," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 40, no. 7, pp1697-1710, 2017.
[23] S. Xu, H. Dong, R. Xu et al., "A Real-Time Gait Phase Detection Method Based on BiLSTM-Attention Model," *Proceedings of 2023 45th Annual International Conference of the IEEE Engineering in Medicine & Biology Society*, *EMBC 2023*, 24-27 July, 2023, Sydney, New South Wales, Australia, pp1-4.
[24] K. He, X. Zhang, S. Ren et al., "Deep Residual Learning for Image Recognition," *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, *CVPR 2016*, 27-30 June, Las Vegas, NV, USA., pp770-778.
[25] X. Wu, Y. Yuan, X. Zhang et al., "Gait Phase Classification for a Lower Limb Exoskeleton System Based on a Graph Convolutional Network Model," *IEEE Transactions on Industrial Electronics*, vol. 69, no. 5, pp4999-5008, 2021.
[26] J. Han and B. Bhanu, "Individual Recognition Using Gait Energy Image," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 28, no. 2, pp316-322, 2005.
[27] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet Classification with Deep Convolutional Neural Networks," *Communications of the ACM*, vol. 60, no. 6, pp84-90, 2012.
[28] C. Szegedy, W. Liu, Y. Jia et al., "Going Deeper with Convolutions," *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, *CVPR 2015*, 7-12 June, 2015, Boston, MA, USA., pp1-9.
[29] T. Zhen, L. Yan, and P. Yuan, "Walking Gait Phase Detection Based on Acceleration Signals Using LSTM-DNN Algorithm," *Algorithms*, vol. 12, no. 12, p253, 2019.
[30] D. Bahdanau, K. Cho, and Y. Bengio, “Neural Machine Translation by Jointly Learning to Align and Translate,” *Proceedings of 3rd International Conference on Learning Representations*, *ICLR 2015*, 7-9 May, 2015, San Diego, CA, USA.