Online-Journals.org

iJOE International Journal of Online and Biomedical Engineering

iJOE | eISSN: 2626-8493 | Vol. 21 No. 13 (2025) | OPEN ACCESS

https://doi.org/10.3991/ijoe.v21i13.56587

PAPER

TCN-LSTM Fusion for Lower Limb Joint Angle Prediction Under Multimodal Signals

Quan Chen[1], Yongxian Song[2](✉), Qi Zhang[1], Yan Yan[2], Yuanlin Fang[1], Xuenian Zheng[1]

[1]Jiangsu Ocean University, Lianyungang, China

[2]Nanjing Xiaozhuang University, Nanjing, China

q825255301@163.com

ABSTRACT

This study addresses the limitations of single-modal input and insufficient temporal feature extraction in traditional deep learning models by proposing a multimodal framework for lower-limb joint-angle prediction. Using Inertial Measurement Unit (IMU), Surface Electromyography (sEMG), and goniometer data as inputs, the model cascades a temporal convolutional network (TCN) with a long short-term memory (LSTM) network. The TCN first extracts complex spatial features from the multimodal signals, and this is followed by the LSTM capturing their temporal dependencies to map input sequences to multiple joint angles. Experimental results demonstrate that, compared to TCN, LSTM, Bi-LSTM, and GRU benchmarks, the proposed TCN-LSTM model reduces average RMSE by 61.54%, 21.13%, 21.45%, and 16.37%, respectively, and reduces average MAE by 13.48%, 7.25%, 2.97%, and 6.15%, respectively. At the same time, it improves R^2 by 3.73%, 1.08%, 1.00%, and 0.84%, respectively. Overall, the TCN-LSTM model delivers superior prediction accuracy, demonstrating significant practical value in the control of lower-limb rehabilitation robots.

KEYWORDS

temporal convolutional network (TCN), long short-term memory (LSTM), multimodal, joint angle prediction, lower limb rehabilitation robot

1 INTRODUCTION

With the aging global population and shifts in lifestyle in modern society, the incidence of acute cerebrovascular diseases, such as strokes and spinal cord injuries, has risen significantly. Consequently, the number of patients experiencing lower limb motor dysfunction, including hemiplegia and paraplegia, resulting from these conditions has continued to grow [1] [2]. As this issue intensifies, lower limb rehabilitation robots are increasingly becoming essential tools for assisting hemiplegic patients in restoring motor function and enhancing their quality of life. By simulating human lower limb movements, these rehabilitation robots help patients perform daily activities and offer personalized rehabilitation therapy, which has

Chen, Q., Song, Y., Zhang, Q., Yan, Y., Fang, Y., Zheng, X. (2025). TCN-LSTM Fusion for Lower Limb Joint Angle Prediction Under Multimodal Signals. *International Journal of Online and Biomedical Engineering (iJOE)*, 21(13), pp. 63–81. https://doi.org/10.3991/ijoe.v21i13.56587

Article submitted 2025-06-05. Revision uploaded 2025-09-02. Final acceptance 2025-09-02.

yielded remarkable results in clinical settings [3] [4]. However, achieving real-time and accurate predictions of patients' movement intentions remains a significant challenge in the control of exoskeleton robots.

Joint angle prediction, a fundamental aspect of rehabilitation robot control, captures the patient's movement status in real time, predicts dynamic changes in the joints, and optimizes the robot's assistive functions by integrating intelligent control strategies. This approach aims to achieve personalized and efficient rehabilitation training [5] [6]. In the realm of joint angle prediction, some researchers have relied on single-modal input signals, such as Surface Electromyography (sEMG) signals or Inertial Measurement Unit (IMU) signals [7] [8]. Although single-modal signals offer a certain degree of motion information, they are often hindered by noise and the inherent incompleteness of the signals in practical applications. This is particularly true for hemiplegic patients, as a single-modal signal source frequently fails to accurately represent the motion status of the affected limbs. To address the challenges of noise and insufficient information associated with single-modal signals, some researchers have concentrated on the fusion of multimodal signal inputs [9] [10] [11]. Sivakumar et al. [12] integrated plantar pressure data with multiple sources of information, including sEMG signals and hip angles, to develop a deep learning approach for predicting lower limb kinematics. Huang et al. [13] integrated three-leg EMG signals with IMU data from the leg, demonstrating that the fused data provided greater accuracy in predicting joint angles compared to a single modality. Wang et al. [14] combined the vibroarthrogram (VAG) of the affected leg with the fused EMG signals from the sound leg. Su et al. [15] merged sEMG signals with multimodal data from physical sensors (e.g., accelerometers and gyroscopes) to achieve real-time predictions of ankle joint torque.

Joint angle prediction models primarily consist of statistical methods, as well as machine learning and deep learning approaches. Traditional statistical models include regression models [16], Kalman filtering [17], and autoregressive integration [18]. However, these methods are constrained by linear assumptions and static data, which hinder their ability to capture the nonlinear time-series characteristics of human motion, resulting in low prediction accuracy. In recent years, artificial intelligence techniques, particularly those based on machine learning and deep learning, have attracted widespread interest. Li et al. [19] utilized sEMG signals to estimate knee joint angles using a multicore correlation vector regression method, while Lu et al. [20] employed an extreme gradient boosting (XGBoost) regression model to predict lower limb joint angles based on surface sEMG signals. Nevertheless, these models often exhibit limitations in capturing spatial and temporal patterns within historical data, reducing their effectiveness in extracting crucial information. As a result, deep learning models tailored for time series analysis and feature extraction have become a focal point of research. TCNs excel at capturing local features and short-term dependencies; for instance, Wang et al. [14] introduced a TCN-based algorithm for predicting lower limb joint angles during human movement. However, the localized nature of the convolutional operation may overlook complex correlations between distant time steps, particularly limiting performance in ultra-long sequences [21]. In contrast, LSTM reinforces long-term pattern modeling and offers significant advantages in time series prediction. Ma et al. [22] employed EMG signals and a short-connected autoencoder long short-term memory model (SCA-LSTM) to estimate the three joint angles of the upper limbs. Additionally, Wen et al. [23] proposed a variational modal decomposition (VMD) and wavelet packet transform (WPT)-based multidecomposition feature extraction method. To accurately capture future trends in time series data, a bidirectional long short-term memory (BiLSTM) network was implemented to model the relationship between sEMG signals and joint angles, enabling the continuous estimation of upper limb kinematics. In order to enhance the interpretability

of predictive models, some researchers have tried to introduce the Transformer model [24]. Although previous studies have primarily focused on single-model-based joint angle prediction, some research suggests that integrating multiple models can further improve the accuracy of time series prediction [25] [26].

This study focuses on enhancing the prediction model by fusing signals from multiple modalities to achieve a more comprehensive understanding of human motion states. By leveraging the complementary strengths of different models and multimodal signal fusion, we employ TCN as a feature extractor and LSTM networks as a subsequent temporal modeling module. This study introduces a multimodal approach for multi-joint angle prediction by integrating TCN and LSTM. Leveraging inputs from IMU, sEMG, and Gon, the method aims to predict hip, knee, and ankle joint angles in the lower limbs. This method integrates TCN's strength in spatial feature extraction with LSTM's proficiency in time series modeling, leading to a substantial improvement in prediction accuracy. The main goal of this study is to enhance the accuracy of joint angle predictions and deliver more precise motion feedback for controlling lower limb rehabilitation exoskeleton systems, ultimately improving rehabilitation outcomes for hemiplegic patients.

2 METHODS

2.1 Data sources

The dataset for this study was obtained from the publicly available Exoskeleton and Prosthetic Intelligent Control (EPIC) Lab dataset. It involved 22 healthy subjects who were examined using wearable sensors and motion capture technology to measure motion information during various gait activities, including walking on a treadmill, on flat ground, and on stairs [27]. The experimental data were carefully selected from wearable sensor and motion capture recordings obtained during flat ground walking. The flat-floor walking protocol included five clockwise and five counterclockwise cycles at three self-chosen speeds: slow, normal, and fast, resulting in a total of 30 cycles per subject. The pelvic line speed was estimated by averaging the velocities of four hip motion capture markers during the steady-state walking phase, providing an approximation of each subject's walking speed. The mean pelvic velocities for all subjects were recorded as 0.88 ± 0.19 m/s for slow, 1.17 ± 0.21 m/s for normal, and 1.45 ± 0.27 m/s for fast.

Inertial Measurement Units, sEMG, and goniometer angles were selected as inputs for this study. An optical motion capture system, using a high-performance infrared camera, was utilized to track reflective markers attached to the limbs. This system facilitated the real-time generation of precise 3D models for extracting joint angles (IK_offset). The output joint angles were assessed in comparison to predicted values, and the locations of signal acquisition are illustrated in Figure 1. The IMU data were gathered from sensors placed on the anterior surfaces of the thighs, calves, and feet. Acceleration signals were extracted specifically from the Z-axis, as it aligns with the sagittal plane and the vertical orientation of the mounted sensors. The signals were sampled at 200 Hz and processed with a sixth-order Butterworth filter set to a 100 Hz cutoff. Likewise, sEMG signals from the medial gastrocnemius, tibialis anterior, and soleus muscles were recorded at 1000 Hz and filtered within the 20–500 Hz range using the same filter order. Joint angular data for the hip, knee, and ankle were collected at 1000 Hz and refined with a fourth-order Butterworth filter at a 20 Hz cutoff. The IK_offset was derived from 32 bilaterally placed motion capture markers and recorded at 200 Hz.

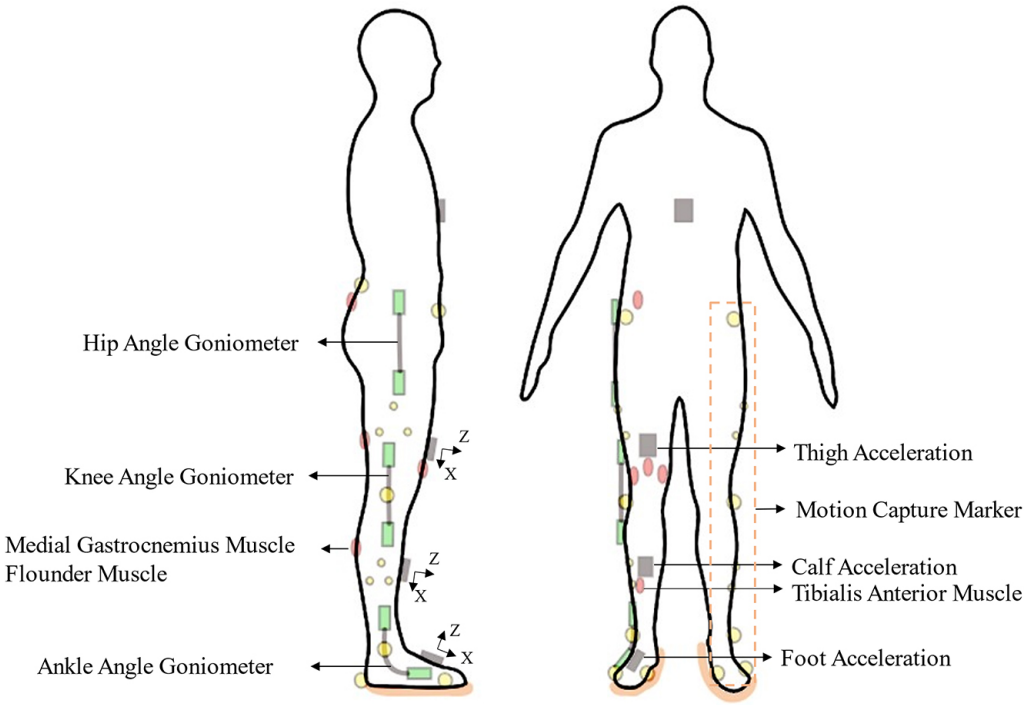

Fig. 1. Signal acquisition location

2.2 Input signal selection basis

Lower limb gait is a highly coordinated motor process involving the ordered activation of multiple muscles. The functional synergy of muscles directly determines the accuracy of joint angle prediction, which is essential for the control of rehabilitation robots. Focusing on 11 key lower limb muscles, this section first clarifies their functional positioning in the gait cycle, then quantifies the correlation between sEMG signals of muscles in combination with the characteristics of multimodal input signals, and finally screens out the optimal muscle combination that balances the completeness of motion information and the efficiency of model input. Meanwhile, the selection logic for IMU and goniometer (Gon) signals is also illustrated.

According to the annotation standards of the EPIC dataset [27], the gait cycle is divided into the stance phase (accounting for approximately 60%, including initial contact, mid-stance, and terminal propulsion) and the swing phase (accounting for approximately 40%, including initial swing, mid-swing, and terminal swing). The 11 muscles exhibit significant functional specialization. Among them, the medial gastrocnemius dominates ankle plantarflexion during the terminal stance phase, providing forward propulsive force for the body and showing a strong correlation with the ankle angle (Pearson correlation coefficient $r \approx 0.72$). The tibialis anterior is responsible for ankle dorsiflexion throughout the swing phase to prevent foot drop and also exhibits a strong correlation with the ankle angle ($r \approx 0.68$). The soleus assists in ankle plantarflexion throughout the stance phase to maintain standing balance and has a moderate-to-strong correlation with the ankle angle ($r \approx 0.65$). In contrast, muscles such as the semitendinosus ($r \approx 0.41$ with the knee angle) and the right external oblique ($r < 0.30$ with lower limb joint angles) primarily play auxiliary stabilizing roles, contributing less to the dynamic prediction of core joint (ankle, knee, hip) angles. To avoid information redundancy caused by overlapping muscle functions,

a correlation analysis was conducted on the sEMG signals of five highly relevant muscles: medial gastrocnemius, tibialis anterior, soleus, vastus medialis, and vastus lateralis. The results showed that the medial gastrocnemius and tibialis anterior ($r = –0.43$), as well as the tibialis anterior and soleus ($r = –0.39$), exhibited low redundancy due to functional antagonism, with complementary signals and no temporal overlap. In contrast, the medial gastrocnemius and soleus ($r = 0.66$) and the vastus medialis and vastus lateralis ($r = 0.71$) showed high redundancy due to their similar functional roles, resulting in significant overlap of signal information. By fixing the sensors to the target joints and calibrating them, reliable reference data on joint angles can be obtained, creating multimodal complementarity with sEMG and IMU signals. Considering functional correlation, signal redundancy, and the model's input dimension requirements, the medial gastrocnemius, tibialis anterior, and soleus muscles were ultimately selected as the sEMG input sources. The combination covers 92% of ankle angle dynamics and adapts to the three walking speeds in the dataset; it also enables the sensors to be worn centrally in the lower leg area, reducing displacement noise interference and improving wearability for hemiplegic patients.

However, this selection also has three limitations: First, the range of joint coverage is restricted, as all three muscles are directly associated with ankle joint movement. Changes in the ankle-knee-hip kinetic chain can only indirectly reflect dynamic changes in the knee and hip joints (such as hip abduction controlled by the gluteus medius) and affect the knee extension function (e.g., controlled by the quadriceps femoris) through the "ankle-knee-hip" kinetic chain. This necessitates heavy reliance on supplementary IMU and Gon signals for predicting knee and hip angles. If multimodal signal synchronization deviates (e.g., IMU delay >50ms), the RMSE of knee and hip angle predictions increases. Second, limitations in adapting to pathological states: for individuals with muscle atrophy (e.g., patients with >30% reduction in gastrocnemius volume) or nerve damage (e.g., tibialis anterior paralysis due to common peroneal nerve palsy), the activity patterns of the three muscles undergo significant alterations (e.g., compensatory overactivation of the soleus muscle). At this point, the correlation between sEMG signals and joint angles drops to $r = 0.4–0.5$, significantly reducing signal validity. Thirdly, limitations in movement scenarios: while stable during level walking, this combination fails to maintain linear correlation between calf muscle activity and ankle angle during non-uniform gaits like running, jumping, or stair climbing (e.g., non-linear relationship between gastrocnemius explosive force output and ankle angle during jumping). The model demonstrates insufficient adaptability to specialized gait patterns.

Meanwhile, the selection of IMU and Gon signals adheres to the principle of simplicity and efficiency. Studies have demonstrated that accurate joint angle prediction does not necessarily require an increased number of IMU sensors; a single or a few strategically placed IMUs can provide precise predictions [28]. Therefore, in this study, IMU sensors were placed only on the anterior surfaces of the thighs, calves, and feet, and the Z-axis acceleration signal was extracted. This axis aligns with the human sagittal plane and the vertical installation direction of the sensors, allowing it to directly reflect gait characteristics and assist in determining the gait cycle, stride length, and gait coordination [29]. Gon signals refer to the hip, knee, and ankle angles directly measured by goniometers.

2.3 Data preprocessing

Due to the varying sampling frequencies of different modes in the input signal, it is essential to align the time labels across the various modal signals. Consequently, the

sEMG signal (sampled at 1000 Hz) and the goniometer (Gon) joint angle signal are both resampled to 200 Hz to ensure time label alignment. To expedite the computation and eliminate the scaling relationship between the features, the aligned data are normalized and standardized. Next, a sliding window technique is applied to align the input data with the corresponding target output sequences. Finally, the dataset is partitioned into training and test sets according to the predefined ratio, as illustrated in Figure 2.

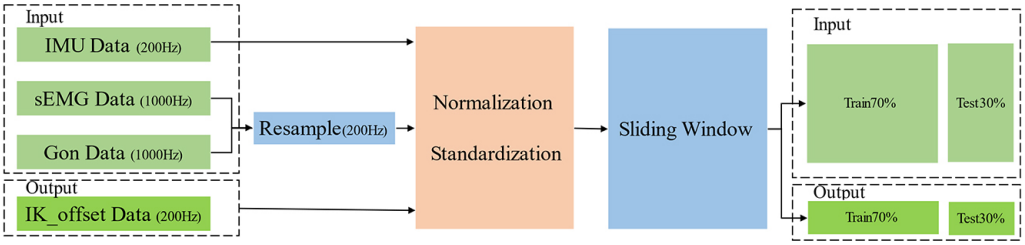

Fig. 2. Signal pre-processing

After resampling, the signal is denoted as $X \in \mathbb{R}^{T \times D_{in}}$, T represents time, and D_{in} represents the dimensions of the input features. Following resampling, the data is normalized and standardized [30]. Normalization scales the data to the range [0, 1], as shown in Equation (1).

$$X_{\text{normalized}} = \frac{X - X_{\min}}{X_{\max} - X_{\min}} \tag{1}$$

where $X_{\min}$ is the minimum value of X and $X_{\max}$ is the maximum value of X.

Standardization adjusts the data to have a mean of zero and a standard deviation of one, as demonstrated in Equation (2).

$$X_{\text{standardized}} = \frac{X_{\text{normalized}} - \mu}{\sigma} \tag{2}$$

where μ, σ represent $X_{\text{normalized}}$ mean and standard deviation, respectively.

To develop the angle prediction model, a sliding window was applied to maintain alignment between input and target sequences, as shown in Figure 3. A 200-point multimodal signal served as the input window for forecasting three lower limb joint angles 10 steps ahead (50 ms). Each subject's walking data were partitioned into training and test sets at a 7:3 ratio, and signals from all subjects were merged to construct the final dataset.

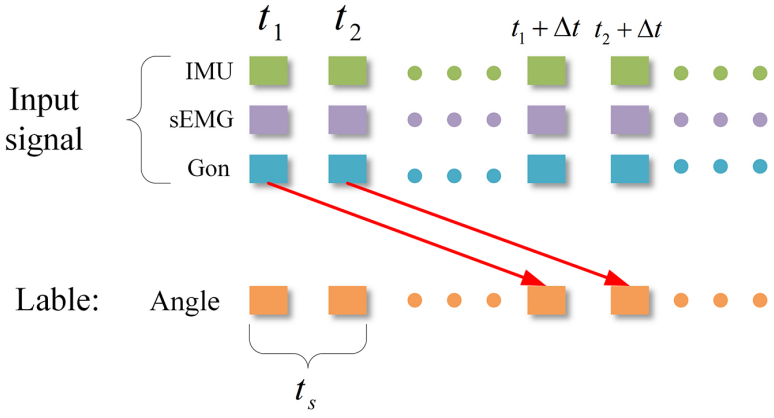

Fig. 3. Sliding window schematic

2.4 TCN multivariate feature extraction module

Temporal convolutional network module is a one-dimensional convolution-based time series feature extraction network that focuses on extracting temporal dependencies and feature relationships from multivariate sequences [31] [32]. TCN captures temporal dependencies in a time series by leveraging causal convolution, which prevents data leakage by ensuring that future information is not incorporated. Additionally, dilated convolution expands the receptive field by adjusting the dilation rate, enabling the model to capture long-range dependencies while maintaining parameter efficiency. As illustrated in Figure 4, the convolution kernel 3, with an expansion rate of s = [1, 2, 4], padding = (K – 1) × s.

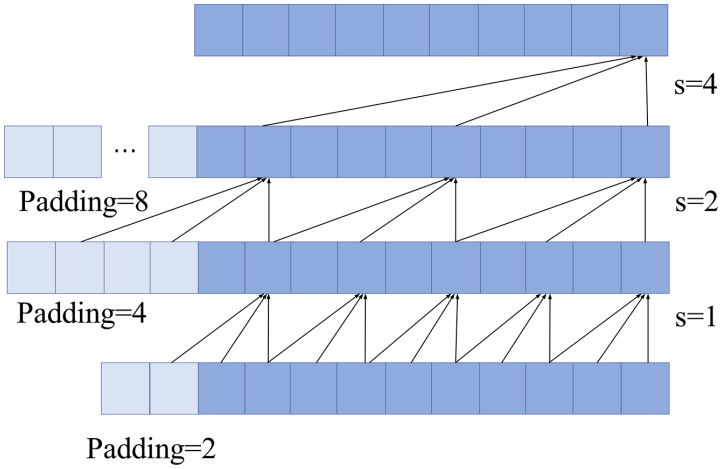

Fig. 4. Expanded causal convolution with 3 convolution kernels

In TCN, dilated causal convolution is employed to expand the receptive field while maintaining computational efficiency. This technique can be understood as incorporating a dilation rate s into the standard one-dimensional convolution, allowing the convolution kernel to be sampled at larger intervals on the input. The formulation of dilated causal convolution is presented in Equation (3), where $Y_{n,k,t}$ represents the convolution output for the n th sample, output channel k, and time step t. The convolution output corresponds to the result for the n th sample.

$$Y_{n,k,t} = \sum_{i=0}^{K-1} W_{k,c,i} \cdot X_{n,c,t-s\cdot i} \tag{3}$$

where $W_{k,c,i}$ is the convolution kernel weight, c is the input channel, i is the convolution kernel position, K is the convolution kernel size, the expansion rate $s = 2^{l-1}$, $X_{n,c,t-s\cdot i}$ is the input, and $t–s\cdot i$ is the time step.

The TCN model is constructed using multiple residual blocks, each comprising two key components: a nonlinear mapping module and a dual-layer network architecture. The network incorporates a dilated causal convolution layer, weight normalization, a modified ReLU activation, and a regularization mechanism. To preserve input-output dimensional consistency, the nonlinear mapping module applies a 1D convolution. Within each residual block, data flows through two parallel paths: one utilizing a 1D full convolution layer and the other a dilated causal convolution layer. Their outputs are then merged via a residual connection, as illustrated in Figure 5.

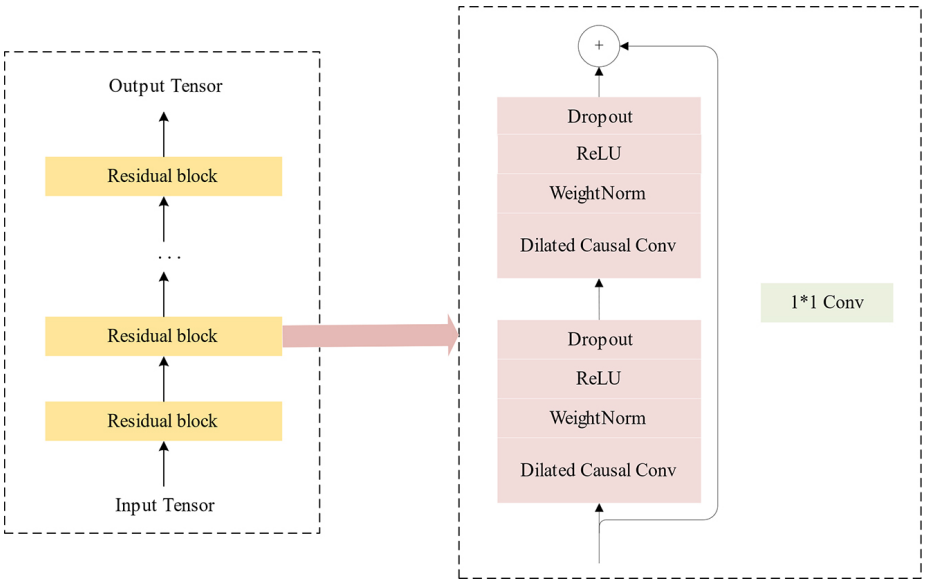

Fig. 5. TCN model architecture

The output $H_{n,k,t}$ of the residual block is shown in equation (4).

$$H_{n,k,t} = \mathrm{ReLU}(Y_{n,k,t} + X_{\mathrm{res},n,k,t}) \tag{4}$$

$Y_{n,k,t}$ is the output of the convolutional layer, the input of the residual link $X_{\mathrm{res},n,k,t}$ is adjusted in dimension by 1*1 convolution, and ReLU() is the activation function. Incorporating a weight normalization layer after the convolution improves training efficiency and enhances model robustness. Meanwhile, the ReLU function is applied to capture nonlinear relationships within the data. To address potential overfitting caused by model complexity, a regularization layer is introduced.

In a multilayer TCN, the convolutional computation process and the residual connection for each layer are illustrated in Equations (5) and (6).

$$Y^{(l)}_{n,k,t} = \sum_{i=0}^{K-1} W^{(l)}_{k,c,i} \cdot H^{(l-1)}_{n,c,t-s\cdot i} \tag{5}$$

$$H^{(l)}_{n,k,t} = \mathrm{ReLU}\left(Y^{(l)}_{n,k,t} + X^{(l)}_{\mathrm{res}}\right) \tag{6}$$

l is the number of TCN layers.

2.5 LSTM as a temporal feature extraction module

Long short-term memory efficiently captures contextual information in time series through feedback loops and addresses the vanishing and exploding gradient issues commonly found in conventional recurrent neural networks by incorporating a memory unit [33]. The memory unit of LSTM consists of a unit state and three gates—input, forget, and output—which employ sigmoid activation functions to control memory cell states and selectively transmit information.

The output sequence $X_t = H^{(3)}_{n,k,t} = [x_1, x_2, \ldots, x_T]$ from the TCN layer is fed into the LSTM model, as illustrated in Figure 6. At each time step t, the memory cell receives signals from the input gate, forget gate, and output gate of the previous time step ($t - 1$), and it combines the current input x_t with the cell state from the previous

moment c_{t-1}. This structural feature enables historical information to be continuously and recursively passed through the loop of the long short-term memory.

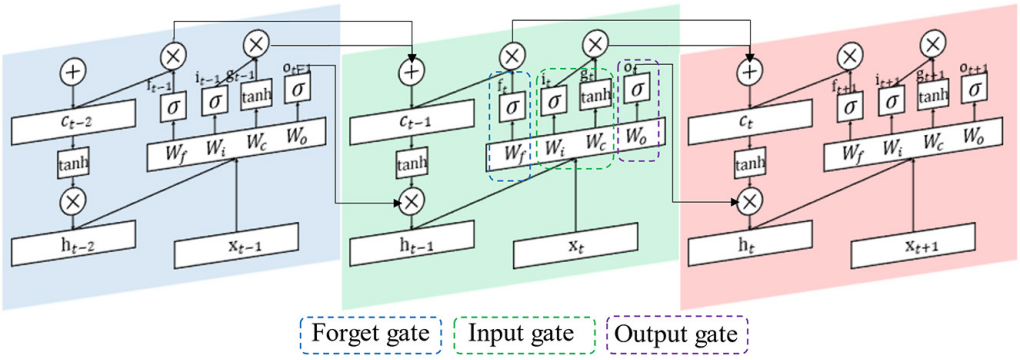

Fig. 6. LSTM internal chain structure

The first step of the LSTM is a forgetting gate, which determines what information to discard from the current state. The forgetting gate evaluates the hidden states h_{t-1} and x_t from the previous time step, and after passing through a sigmoid layer, it outputs a value f_t between 0 and 1. A value of 1 indicates 0 signifies as illustrated in equation (7).

$$f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f) \tag{7}$$

Where W_f and b_f are forgetful gate trainable parameters, $[h_{t-1}, x_t]$ denotes splicing together the hidden state at the current moment and the input at the current moment, and σ denotes the sigmoid function.

The input gate regulates which information is incorporated into the cell state. The variable i_t represents the input gate's output, ranging from 0 to 1, which determines the proportion of information to retain, as defined in equations (8) and (9).

$$i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i) \tag{8}$$

$$g_t = \tanh(W_c \cdot [h_{t-1}, x_t] + b_c) \tag{9}$$

Where W_i and W_c are the input gate weight matrices, b_i and b_c are the corresponding biases, tanh is the hyperbolic tangent function, and g_t denotes the new candidate cell state, each of whose elements is between –1 and 1.

Memory cell C_t is updated as shown in equation (10):

$$c_t = f_t \odot c_{t-1} + i_t \odot g_t \tag{10}$$

Where $\odot$ means Hadamard Product, which is the multiplication of the corresponding elements in the operation matrix.

The output gate assesses the stored information in the memory cell alongside the current input to generate the final output. The output gate is computed similarly to the input gate, using a sigmoid function to determine the degree of openness of the gate, followed by a tanh function that produces the output vector. The relevant formulas are presented in equations (11) and (12).

$$o_t = \sigma(W_o \cdot [h_{t-1}, x_t] + b_o) \tag{11}$$

$$h_t = o_t \odot \tanh(c_t) \tag{12}$$

Where W_o, b_o are the output gate weight matrix and bias.

2.6 TCN-LSTM model

The expansive causal convolutional structure of TCN is highly effective in feature extraction, enabling it to convert raw features into high-dimensional abstract representations, thereby facilitating a more thorough exploration of feature information. However, it is less effective than LSTM in preserving long-range dependencies within complex time series, especially for tasks requiring sustained information retention. While LSTM networks demonstrate strong capabilities in time series prediction, they come with higher computational complexity, especially when dealing with large-scale datasets, resulting in slower training and inference processes, as well as increased consumption of storage and computational resources. When combined with TCN, LSTM's parallel processing can enhance the efficiency of its memory unit. Additionally, by appropriately configuring the relevant parameters of the LSTM model, it can effectively capture long-term prediction tasks in time series data. Thus, this study presents a multimodal multi-joint angle prediction approach that combines TCN and LSTM, with the model's overall architecture shown in Figure 7.

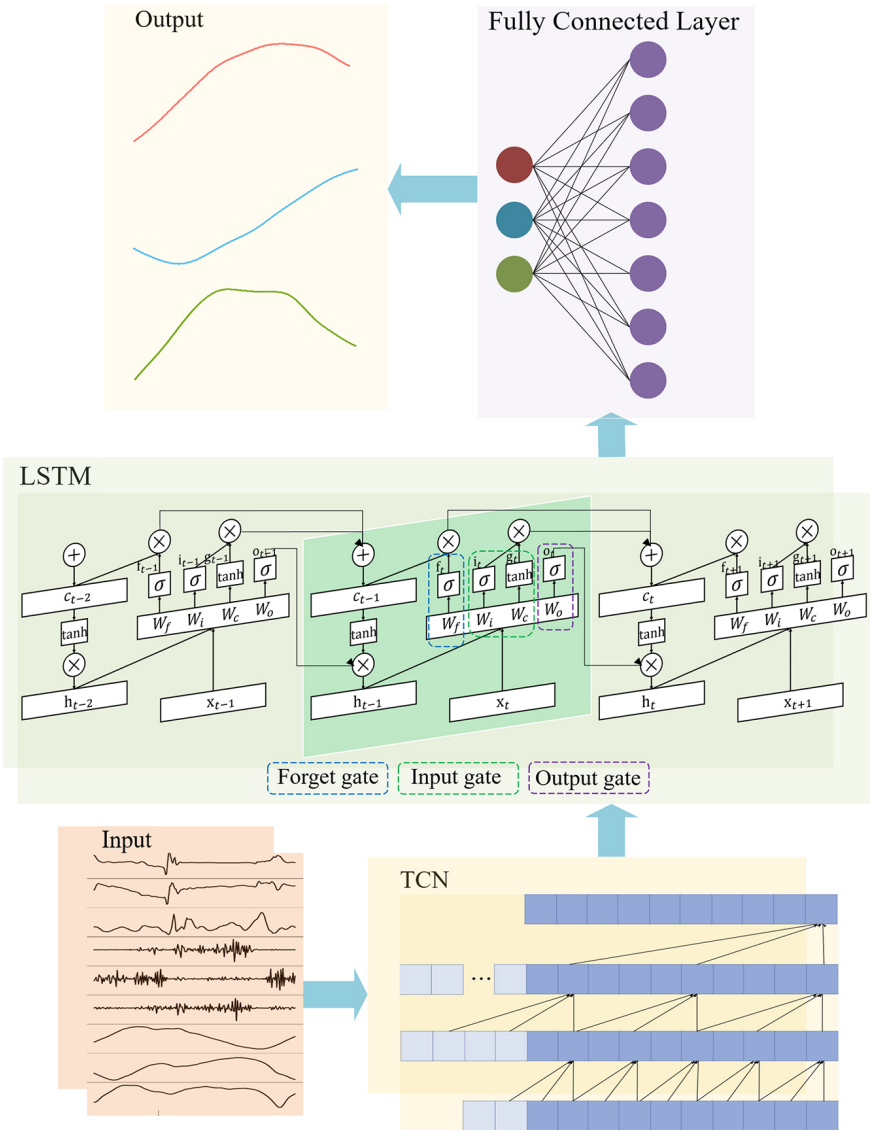

Fig. 7. Overall architecture of the TCN-LSTM model

3 RESULTS

This study utilizes the NVIDIA GeForce RTX 3060 Laptop GPU, featuring a main frequency of 2.30 GHz, along with 16 GB of DDR4 RAM operating at 3200 MHz. The analysis is conducted on a Windows 11 (64-bit) operating system, using Python version 3.8 for data analysis. The associated deep learning model is built on PyTorch version 1.7.1.

3.1 Evaluation indicators

To assess the prediction model's performance, three evaluation metrics were utilized: Root mean square error (RMSE), mean absolute error (MAE), and R-square (R^2) [34] [35].

Root mean square error is a statistical metric that measures the deviation between predictions and actual values. Since it shares the same units as the original data, it offers a clear assessment of model accuracy. The RMSE formula is given in Equation (13).

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2} \tag{13}$$

Where y_i represents the true result, $\hat{y}_i$ represents the predicted result and n represents the total number of samples.

Mean absolute error quantifies the mean absolute difference between predictions and actual values, reflecting the average error magnitude. The corresponding formula is given in Equation (14).

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i| \tag{14}$$

R^2 is a statistical indicator used to assess the goodness of fit of a regression model. The value ranges from 0 to 1, with values closer to 1 indicating that the model is more effective at explaining the target variable and demonstrating a better fit. This relationship is illustrated in Equation (15).

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2} \tag{15}$$

$\bar{y}$ denotes the sample mean.

3.2 Input modal analysis

To evaluate the TCN-LSTM model's effectiveness in joint angle prediction with multimodal inputs, this study compared different input combinations: single-modal inputs (IMU, sEMG, and Gon) and multimodal inputs (IMU and sEMG, IMU and Gon, sEMG and Gon, as well as IMU, sEMG, and Gon) for predicting hip, knee, and ankle joint angles. The evaluation metrics used were RMSE, MAE, and R^2. In this

experiment, the hyperparameters were set as follows: The input feature dimension was 9; the TCN network consisted of three layers with channel sizes of 64, 128, and 256; the convolutional kernel size was 3; the dilation rates were [1, 2, 4]; the LSTM layer had 256 hidden units with a single-layer structure; the dropout rate was 0.2 for both convolutional and LSTM layers; and the output dimension was 3. The model was trained using the Adam optimizer with a learning rate of 0.001 for 800 epochs, utilizing a batch size of 32 per iteration. The results are summarized in Table 1.

Table 1. Comparison of experimental results with different modality inputs

IMU	sEMG	Gon	RMSE	MAE	R^2
✓			5.62	3.68	93.40%
	✓		10.63	5.39	74.27%
		✓	7.78	3.91	87.27%
✓	✓		4.20	3.02	96.22%
✓		✓	5.57	3.45	93.57%
	✓	✓	4.43	2.79	95.93%
✓	✓	✓	**3.22**	**2.26**	**97.82%**

Based on the comparison of the experimental results, it is evident that the IMU provides the optimal predictive performance with a single-modal input, with an RMSE of 5.62, an MAE of 3.68, and an R^2 of 93.40%. sEMG and Gon methods perform poorly, particularly the single-modal sEMG, which records an RMSE of 10.63, an MAE of 5.39, and an R^2 of 74.27%. These results indicate its limited predictive capability for this task. In the bimodal input, the combination of IMU and sEMG achieved the best results, with an RMSE of 4.20, an MAE of 3.02, and a high R^2 of 96.22%, indicating a significant improvement in performance. The combination of IMU-Gon and sEMG-Gon performs slightly less effectively than expected, yet it still surpasses the performance of the single-modal input. Among the three modal inputs, IMU-sEMG-Gon provided the best predictive performance, with an RMSE of 3.22, an MAE of 2.26, and an R^2 of 97.82%, indicating that the model significantly enhanced prediction accuracy by integrating data from these three distinct modalities, further validating the effectiveness of multimodal inputs. Overall, the use of multimodal inputs markedly improves predictive performance, and the final results demonstrate that this three-modal input approach is the most effective.

3.3 Comparative analysis of models

All models in this section were evaluated using multimodal inputs (IMU, sEMG, and goniometer signals). This process ensures a fair comparison with the proposed TCN-LSTM model. To verify the effectiveness of the TCN-LSTM model, the LSTM and TCN were utilized as baseline models. Additionally, Bi-LSTM, GRU, and CNN-LSTM models were compared for the simultaneous prediction of hip, knee, and ankle angles. The evaluation metrics employed were RMSE, R^2, and MAE. The hyperparameter settings for this experiment are detailed in Table 2.

Table 2. Model hyperparameters

	TCN	LSTM	Bi-LSTM	GRU	TCN-LSTM
Input features	9	9	9	9	9
TCN features	[8, 16, 32, 64]	–	–	–	[64, 128, 256]
convolution kernel	3	–	–	–	3
TCN dilatation	[1, 2, 4, 8]	–	–	–	[1, 2, 4]
Number of hidden units	–	128	128	256	256
hidden layer	–	2	2	2	1
Dropout	0.2	0.2	0.2	0.2	0.2
Output features	3	3	3	3	3
Optimizer	Adam	Adam	Adam	Adam	Adam
learning rate	0.001	0.001	0.001	0.001	0.001
Epoch	800	800	800	800	800
Batch size	32	32	32	32	32

The results are shown in Figure 8; the TCN-LSTM model outperforms the LSTM, TCN, Bi-LSTM, and GRU models in knee angle and ankle angle prediction, with the lowest RMSE and MAE values and the highest R^2 values. In hip angle prediction, the RMSE and R^2 of the TCN-LSTM model are slightly inferior to those of the GRU model, but its MAE is still better than that of the GRU model. Overall, comparing the TCN, LSTM, Bi-LSTM, and GRU models, the TCN-LSTM model showed an average decrease in RMSE [61.54%, 21.13%, 21.45%, and 16.37%], an average decrease in MAE [13.48%, 7.25%, 2.97%, and 6.15%], and an average enhancement in R^2 in the prediction of joint angles [3.73%, 1.08%, 1.00%, and 0.84%].

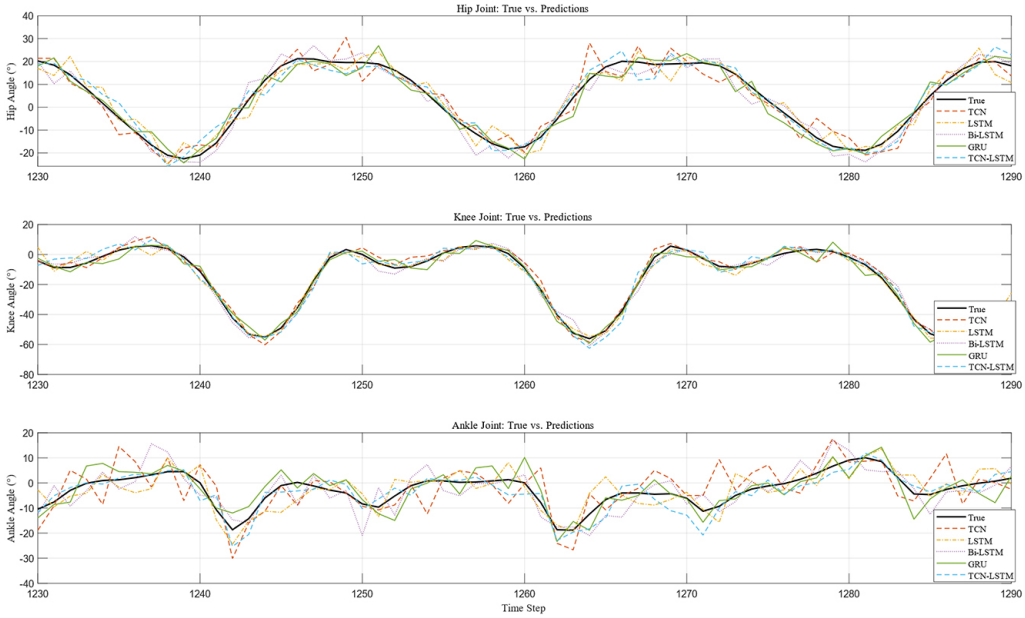

Fig. 8. Hip, knee and ankle joint angle prediction

A comprehensive comparison of the experimental results is presented in Table 3. The findings indicate that the TCN-LSTM model delivers more precise and consistent predictions, particularly for knee and ankle joint estimation, demonstrating strong stability and reliability.

Table 3. Comparison of experimental results of hip, knee, and ankle joint angle prediction under different models

Joint Angle	Models	RMSE	MAE	R^2
Hip angle	TCN	5.52	2.61	93.79%
	LSTM	3.83	2.53	97.01%
	Bi-LSTM	3.37	2.31	97.66%
	GRU	**3.34**	2.30	**97.72%**
	TCN-LSTM	3.36	**2.29**	97.68%
Knee angle	TCN	3.50	2.56	96.87%
	LSTM	3.33	2.48	97.17%
	Bi-LSTM	3.11	2.37	97.61%
	GRU	3.29	2.42	97.25%
	TCN-LSTM	**3.10**	**2.34**	**97.83%**
Ankle angle	TCN	6.56	2.50	92.57%
	LSTM	4.53	2.24	96.47%
	Bi-LSTM	5.24	2.29	95.62%
	GRU	4.60	2.46	96.35%
	TCN-LSTM	**3.18**	**2.14**	**98.26%**
Average	TCN	5.20	2.56	94.41%
	LSTM	3.89	2.42	96.88%
	Bi-LSTM	3.91	2.32	96.96%
	GRU	3.74	2.39	97.11%
	TCN-LSTM	**3.22**	**2.26**	**97.93%**

To assess the robustness of performance differences among the five models (TCN, LSTM, Bi-LSTM, GRU, and TCN-LSTM), Shapiro–Wilk tests (all $p > 0.05$) and Levene's tests (all $p > 0.05$) were first conducted, confirming that the data met the assumptions of normality and homogeneity of variance required for one-way ANOVA. The ANOVA results revealed significant effects of model type on RMSE, MAE, and R^2 across all joints (hip, knee, and ankle) ($p < 0.001$; e.g., hip RMSE: $F = 8.72$, ankle MAE: $F = 6.97$). Subsequent post hoc Tukey HSD analyses demonstrated that TCN-LSTM significantly outperformed the baseline models in most scenarios: across all metrics compared with TCN, on knee and ankle metrics compared with LSTM and GRU, and on ankle metrics compared with Bi-LSTM, with statistical improvements ranging from $p < 0.05$ to $p < 0.001$ (e.g., ankle RMSE was reduced by 51.5% relative to TCN and by 30.9% relative to GRU). For the hip joint, however, its performance was comparable to Bi-LSTM and GRU ($p > 0.05$). Collectively, these findings provide strong statistical evidence supporting the architectural superiority of TCN-LSTM.

4 DISCUSSION

The prediction method in this study outperforms the traditional unimodal or single-model structure in terms of accuracy. By comparing the model prediction accuracies under different modality combinations, this study proposes a lower limb joint angle prediction method based on multimodal inputs (TCN-LSTM). By leveraging IMU signals, sEMG signals, and joint angle data, the model accurately predicts joint angles during gait motion. Among them, TCN extracts local temporal features, while LSTM captures long-term dependencies, enhancing model performance and ensuring high prediction accuracy across different gait phases.

The exceptional performance observed in this study stems from two key factors. Firstly, integrating TCN with LSTM enables the model to capture multi-scale temporal features effectively, thereby improving prediction accuracy and stability. On the other hand, the fusion of multimodal data provides a richer source of information, enabling the model to understand the gait pattern more comprehensively. Further comparing the model parameters (RMSE, MAE, and R^2) under the same data, it can be determined that the method proposed in this study (TCN-LSTM) is more advantageous than the existing deep learning methods in terms of comprehensive performance. This provides reliable technical support for the development of gait assistive devices.

Future work can further optimize the feature fusion method, incorporating the Transformer structure to enhance the capability of modelling longer time-series dependencies while exploring a wider range of application scenarios, such as prediction schemes for different terrains or special gait patterns. This will help to realize smarter gait assistance systems and advance the field of rehabilitation medicine.

5 CONCLUSIONS

This study proposes a multimodal multi-joint angle prediction method integrating TCN and LSTM, designed to enhance motion feedback accuracy for controlling lower limb rehabilitation exoskeleton systems. By jointly modeling surface EMG signals, IMU acceleration signals, and joint angle signals, this study significantly improves the accuracy and real-time performance of joint angle prediction. The experimental results indicate that the TCN-LSTM model performs effectively across various modal input combinations, with particularly notable improvements when multimodal inputs are used. This approach significantly enhances the accuracy of joint angle predictions. In comparison to traditional unimodal methods, the TCN-LSTM model exhibits superior prediction stability and accuracy, particularly in predicting knee and ankle joint angles. The RMSE, MAE, and R^2 metrics show substantial improvements, highlighting the model's enhanced performance. Comparison results with other deep learning models, including LSTM, TCN, Bi-LSTM, and GRU, further highlight the advantages of the TCN-LSTM model, particularly in predicting knee and ankle angles, where it achieves significantly superior performance.

In conclusion, the TCN-LSTM-based multimodal signal fusion method provides an efficient and accurate joint angle prediction scheme for lower limb motion rehabilitation exoskeleton systems, which helps to capture the patient's motion intention in real-time and accurately and provides more reliable support for personalized rehabilitation therapy. Future research will focus on integrating the lower limb rehabilitation exoskeleton robot system with lower limb intention recognition systems to enhance the flexibility and efficiency of human-robot interaction control.

6 COMPETING INTERESTS

All authors declare no known competing financial interests or personal relationships that could influence the work reported.

7 DATA AVAILABILITY

The dataset used is publicly available from the EPIC Lab dataset (Camargo et al., 2021 [27]). Access can be requested via the EPIC Lab repository: https://www.epic.gatech.edu/opensource-biomechanics-camargo-et-al/.

8 REFERENCES

[1] C. D. Wolfe, "The impact of stroke," *British Medical Bulletin*, vol. 56, no. 2, pp. 275–286, 2000. https://doi.org/10.1258/0007142001903120

[2] J. C. Wu *et al.*, "Increased risk of stroke after spinal cord injury: A nationwide 4-year follow-up cohort study," *Neurology*, vol. 78, no. 14, pp. 1051–1057, 2012. https://doi.org/10.1212/WNL.0b013e31824e8eaa

[3] D. Shi, W. Zhang, W. Zhang, and X. Ding, "A review on lower limb rehabilitation exoskeleton robots," *Chinese Journal of Mechanical Engineering*, vol. 32, no. 1, pp. 1–11, 2019. https://doi.org/10.1186/s10033-019-0389-8

[4] A. Gil-Agudo *et al.*, "Exoskeleton-based training improves walking independence in incomplete spinal cord injury patients: Results from a randomized controlled trial," *Journal of Neuroengineering and Rehabilitation*, vol. 20, no. 1, p. 36, 2023. https://doi.org/10.1186/s12984-023-01158-z

[5] P. Valencia-Londoño, H. Cardona-Rodas, and J. Jiménez-Builes, "An educational inclusion model for adults with diverse neuromuscular conditions through the use of an artificial intelligence algorithm," *International Journal of Emerging Technologies in Learning*, vol. 19, no. 5, pp. 35–56, 2024. https://doi.org/10.3991/ijet.v19i05.46619

[6] S. Dey, T. Yoshida, R. H. Foerster, M. Ernst, T. Schmalz, and A. F. Schilling, "Continuous prediction of joint angular positions and moments: A potential control strategy for active knee-ankle prostheses," *IEEE Transactions on Medical Robotics and Bionics*, vol. 2, no. 3, pp. 347–355, 2020. https://doi.org/10.1109/TMRB.2020.3011841

[7] S. Dey and A. F. Schilling, "A function approximator model for robust online foot angle trajectory prediction using a single IMU sensor: Implication for controlling active prosthetic feet," *IEEE Transactions on Industrial Informatics*, vol. 19, no. 2, pp. 1467–1475, 2022. https://doi.org/10.1109/TII.2022.3158935

[8] H. Lv, Y. Wang, and B. Hao, "Lower limb joint angle estimation based on surface electromyography signals," *Biomedical Signal Processing and Control*, vol. 104, p. 107563, 2025. https://doi.org/10.1016/j.bspc.2025.107563

[9] D. Lahat, T. Adali, and C. Jutten, "Multimodal data fusion: An overview of methods, challenges, and prospects," *Proceedings of the IEEE*, vol. 103, no. 9, pp. 1449–1477, 2015. https://doi.org/10.1109/JPROC.2015.2460697

[10] G. Muhammad *et al.*, "A comprehensive survey on multimodal medical signals fusion for smart healthcare systems," *Information Fusion*, vol. 76, pp. 355–375, 2021. https://doi.org/10.1016/j.inffus.2021.06.007

[11] H. Xie, G. Li, X. Zhao, and F. Li, "Prediction of limb joint angles based on multi-source signals by GS-GRNN for exoskeleton wearer," *Sensors*, vol. 20, no. 4, p. 1104, 2020. https://doi.org/10.3390/s20041104

[12] S. Sivakumar, A. A. Gopalai, K. H. Lim, and D. Gouwanda, "Artificial neural network based ankle joint angle estimation using instrumented foot insoles," *Biomedical Signal Processing and Control*, vol. 54, p. 101614, 2019. https://doi.org/10.1016/j.bspc.2019.101614

[13] Y. Huang *et al.*, "Real-time intended knee joint motion prediction by deep-recurrent neural networks," *IEEE Sensors Journal*, vol. 19, no. 23, pp. 11503–11509, 2019. https://doi.org/10.1109/JSEN.2019.2933603

[14] C. Wang *et al.*, "Prediction of contralateral lower-limb joint angles using vibroarthrography and surface electromyography signals in time-series network," *IEEE Transactions on Automation Science and Engineering*, vol. 20, no. 2, pp. 901–908, 2022. https://doi.org/10.1109/TASE.2022.3185706

[15] C. Su, S. Chen, H. Jiang, and Y. Chen, "Ankle joint torque prediction based on surface electromyographic and angular velocity signals," *IEEE Access*, vol. 8, pp. 217681–217687, 2020. https://doi.org/10.1109/ACCESS.2020.3040820

[16] B. H. Hayadi and T. Hariguna, "Predictive analytics in mobile education: Evaluating logistic regression, random forest, and gradient boosting for course completion forecasting," *International Journal of Interactive Mobile Technologies*, vol. 19, no. 5, pp. 210–232, 2025. https://doi.org/10.3991/ijim.v19i05.52381

[17] T. Triwiyanto, O. Wahyunggoro, H. A. Nugroho, and H. Herianto, "Evaluating the performance of Kalman filter on elbow joint angle prediction based on electromyography," *International Journal of Precision Engineering and Manufacturing*, vol. 18, pp. 1739–1748, 2017. https://doi.org/10.1007/s12541-017-0202-5

[18] W. Zhang, M. Tomizuka, and J. Bae, "Time series prediction of knee joint movement and its application to a network-based rehabilitation system," in *2014 American Control Conference*, 2014, pp. 4810–4815. https://doi.org/10.1109/ACC.2014.6859402

[19] H. B. Li, X. R. Guan, Z. Li, K. F. Zou, and L. He, "Estimation of knee joint angle from surface EMG using multiple kernels relevance vector regression," *Sensors*, vol. 23, no. 10, p. 4934, 2023. https://doi.org/10.3390/s23104934

[20] Z. Lu, S. Chen, J. Yang, C. Liu, and H. Zhao, "Prediction of lower limb joint angles from surface electromyography using XGBoost," *Expert Systems with Applications*, vol. 264, p. 125930, 2025. https://doi.org/10.1016/j.eswa.2024.125930

[21] J. Yan *et al.*, "Modeling the time evolution of compact binary systems with machine learning," *The Astrophysical Journal*, vol. 973, no. 2, p. 163, 2024. https://doi.org/10.3847/1538-4357/ad6989

[22] C. Ma, C. Lin, O. W. Samuel, L. Xu, and G. Li, "Continuous estimation of upper limb joint angle from sEMG signals based on SCA-LSTM deep learning approach," *Biomedical Signal Processing and Control*, vol. 61, p. 102024, 2020. https://doi.org/10.1016/j.bspc.2020.102024

[23] L. Wen, J. Xu, D. Li, X. Pei, and J. Wang, "Continuous estimation of upper limb joint angle from sEMG based on multiple decomposition feature and BiLSTM network," *Biomedical Signal Processing and Control*, vol. 80, p. 104303, 2023. https://doi.org/10.1016/j.bspc.2022.104303

[24] S. Madan *et al.*, "Transformer models in biomedicine," *BMC Medical Informatics and Decision Making*, vol. 24, no. 1, p. 214, 2024. https://doi.org/10.1186/s12911-024-02600-5

[25] Y. Lu, H. Wang, B. Zhou, C. Wei, and S. Xu, "Continuous and simultaneous estimation of lower limb multi-joint angles from sEMG signals based on stacked convolutional and LSTM models," *Expert Systems with Applications*, vol. 203, p. 117340, 2022. https://doi.org/10.1016/j.eswa.2022.117340

[26] B. Kumar and N. Yadav, "A novel hybrid model combining βSARMA and LSTM for time series forecasting," *Applied Soft Computing*, vol. 134, p. 110019, 2023. https://doi.org/10.1016/j.asoc.2023.110019

[27] J. Camargo, A. Ramanathan, W. Flanagan, and A. Young, "A comprehensive, open-source dataset of lower limb biomechanics in multiple conditions of stairs, ramps, and level-ground ambulation and transitions," *Journal of Biomechanics*, vol. 119, p. 110320, 2021. https://doi.org/10.1016/j.jbiomech.2021.110320

[28] D. Hollinger *et al.*, "The effect of sensor feature inputs on joint angle prediction across simple movements," *Sensors*, vol. 24, no. 11, p. 3657, 2024. https://doi.org/10.3390/s24113657

[29] G. Biagetti, P. Crippa, L. Falaschetti, S. Orcioni, and C. Turchetti, "Human activity monitoring system based on wearable sEMG and accelerometer wireless sensor nodes," *Biomedical Engineering Online*, vol. 17, pp. 1–18, 2018. https://doi.org/10.1186/s12938-018-0567-4

[30] M. Khairy, T. M. Mahmoud, and T. Abd-El-Hafeez, "The effect of rebalancing techniques on the classification performance in cyberbullying datasets," *Neural Computing and Applications*, vol. 36, no. 3, pp. 1049–1065, 2024. https://doi.org/10.1007/s00521-023-09084-w

[31] S. T. Naitik and J. V. Gorabal, "Design and development of multimodal biometric system using finger Veins and Iris by CNN integrated with hybrid SIO and whale optimization techniques," *International Journal of Interactive Mobile Technologies*, vol. 18, no. 22, pp. 97–114, 2024. https://doi.org/10.3991/ijim.v18i22.50865

[32] C. Lea, R. Vidal, A. Reiter, and G. D. Hager, "Temporal convolutional networks: A unified approach to action segmentation," in *Computer Vision–ECCV 2016 Workshops*: *Amsterdam, the Netherlands*, Springer International Publishing, 2016, pp. 47–54. https://doi.org/10.1007/978-3-319-49409-8_7

[33] K. Greff, R. K. Srivastava, J. Koutník, B. R. Steunebrink, and J. Schmidhuber, "LSTM: A search space odyssey," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 28, no. 10, pp. 2222–2232, 2017. https://doi.org/10.1109/TNNLS.2016.2582924

[34] D. S. K. Karunasingha, "Root mean square error or mean absolute error? Use their ratio as well," *Information Sciences*, vol. 585, pp. 609–629, 2022. https://doi.org/10.1016/j.ins.2021.11.036

[35] J. P. Barrett, "The coefficient of determination—some limitations," *The American Statistician*, vol. 28, no. 1, pp. 19–20, 1974. https://doi.org/10.1080/00031305.1974.10479056

9 AUTHORS

Quan Chen is currently a postgraduate student majoring in Control Engineering at Jiangsu Ocean University, with research interests including timing prediction and lower limb rehabilitation robot control.

Yongxian Song serves as a Professor at Nanjing Xiaozhuang University and Jiangsu Ocean University and also holds the position of a master's student supervisor at Jiangsu Ocean University, with research interests covering the Internet of Things and intelligent systems, industrial process detection and optimal control, rehabilitation robot control, and deep learning (E-mail: q825255301@163.com).

Qi Zhang is currently a postgraduate student at Jiangsu Ocean University, whose research interests include image processing and object detection.

Yan Yan is a Lecturer at Nanjing Xiaozhuang University, has obtained a PhD in Information and Communication Engineering from Nanjing University of Information Science and Technology, mainly focuses on signal processing and predictive modeling using deep learning networks for her research, and has published several research papers in these fields.

Yuanlin Fang is currently a postgraduate student majoring in Control Engineering at Jiangsu Ocean University, with research interests in deep learning and lower limb rehabilitation robot control.

Xuenian Zheng is currently a postgraduate student majoring in Control Theory and Control Engineering at Jiangsu Ocean University, whose research interests include deep learning and rehabilitation robotics.