International Journal of Mechatronics and Applied Mechanics
2025, Issue 21, Vol. 1

A LIGHTWEIGHT CNN-LSTM HYBRID MODEL FOR ANKLE MOTION RECOGNITION AND CLASSIFICATION BASED sEMG

Chang Li[0009-0001-5887-235X], Yi Zheng[0000-0002-7853-8279], Youqiang Wang[0000-0001-5932-1085], Maohua Xiao*[0000-0001-5213-1035]*
[1]School of Mechanical and Automotive Engineering, Qingdao University of Technology, Qingdao, 266520, China;
[2]College of Engineering, Nanjing Agricultural University, Nanjing 210031, China;
**Corresponding author: Yi Zheng and Maohua Xiao*
Email: xiaomaohua@njau.edu.cn , 901020210098@qut.edu.cn

Abstract - In recent years, research on decoding human movement intentions using surface electromyography (sEMG) has garnered increasing attention in the field of human-computer interaction control. To enhance the accuracy of controllers, the adoption of effective signal processing methods and prediction models has become necessary. Based on this, a lightweight deep learning algorithm—Convolutional Neural Network (CNN)-Long Short-Term Memory (LSTM) hybrid model—is proposed in this paper, which aims to address three key issues in ankle movement recognition and classification:(1) Most of the existing methods can only recognize four types of ankle joint movements; and to achieve more efficient training and application, the classification of six types of ankle joint movements is necessary. (2) Currently, the CNN-LSTM model based on sEMG not only has high structural complexity but also has unsatisfactory classification accuracy. (3) The existing models have a slow convergence speed and a low utilization rate of computing resources. The lightweight CNN-LSTM hybrid model proposed in this paper can predict and classify six types of ankle movements. After 17 iterations, the model converged to the target state, achieving a classification accuracy of 99.46%. The hybrid model comprises a CNN layer, an LSTM layer, and a fully connected layer. Among them, the CNN layer can automatically extract features from sEMG signals, while the LSTM layer can capture long-term temporal dependencies in sEMG data. By comparing the output results of the CNN and LSTM models with those of the hybrid model, the experimental results show that the hybrid model outperforms both the individual CNN model and the LSTM model in terms of performance. In addition, the proposed CNN-LSTM hybrid model was compared in this paper with other commonly used classification algorithms, including RNN-GRU, RF, SVM, and LDA. Experimental results demonstrate that the CNN-LSTM model achieves the best classification accuracy.

Keywords: Ankle movement classification, Lightweight, Motion capture, CNN-LSTM, sEMG signal.

1. Introduction

With the continuous advancement of human-computer interaction technology, ankle rehabilitation robots have become increasingly regarded as a key focus of medical research in the rehabilitation of stroke and amputee patients. Meanwhile, the significance of accurate and efficient motion pattern recognition technology has been widely acknowledged in the realization of auxiliary functions for ankle rehabilitation devices. Classified as non-stationary bioelectrical signals, sEMG signals are typically generated 30 to 150 milliseconds prior to the initiation of movement. This characteristic enables the direct reflection of human movement intentions, ensuring that movement intentions are captured accurately and without delay or omission [1].Furthermore, as different movements activate distinct muscle groups and are accompanied by varying degrees of muscle activation, sEMG signals exhibit significant differences in temporal and amplitude characteristics [2]. This characteristic enables a novel solution to be obtained for the classification and recognition of ankle movements based on sEMG signals.

Traditional ankle movement recognition methods rely on sEMG signals and require the utilization of machine learning models to achieve relevant recognition tasks[3].specifically, the process can be divided into three main stages: first, sEMG signal

dx.doi.org/10.17683/ijomam/issue21.33

data must be collected; second, relevant feature data are extracted via manual means; finally, the extracted features (encompassing time-domain, frequency-domain, and time-frequency domain features) are employed as input parameters for training the machine learning model. In current research on ankle movement recognition, machine learning algorithms such as Linear Discriminant Analysis (LDA), Support Vector Machines (SVM), Random Forests (RF), K-Nearest Neighbor (KNN), Naive Bayes Method (NBM), and Backpropagation Neural Networks (BPNN) are widely employed.[4-7]. A classification study on ankle-foot movements in patients with foot drop was conducted by Bidabadi et al.[8]using machine learning classifiers, and the results indicated that, after the introduction of the wrapper feature selection technique, the random forest classifier exhibited the highest classification performance with an overall accuracy of 93.18%. However, as data collection in this study relies on Inertial Measurement Unit (IMU) sensors, certain limitations may exist in data quality, which can consequently affect the accuracy of classification results. An ankle motion pattern recognition system based on dual-channel sEMG was constructed by Wang et al. [9], which, through the optimization of channel selection and feature extraction processes and the utilization of LDA, achieved efficient recognition of motion patterns; experimental verification demonstrated that its average recognition rate could reach 98.04%. A SVM classification algorithm integrated with Principal Component Analysis (PCA) technology, proposed by Toledo-Pérez et al. [10], was applied to the classification of ankle-foot movements in 8 healthy subjects, with an ultimate classification accuracy exceeding 90%; however, limitations exist in this study at the data acquisition level, as only 3 or 4 channels of sEMG signals were utilized.

In recent years, with the rapid development of deep learning technology, new research avenues have been opened up in the field of human movement intention recognition[11]; deep learning, as a machine learning method based on artificial neural networks, processes data and constructs patterns by simulating the operational mechanisms of the human brain, thereby providing support for decision-making [12], with its notable advantages residing in strong model fitting capability and automatic feature extraction capability. As a critical step in model construction, feature extraction can significantly simplify the cumbersome manual feature design processes in traditional methods, an advantage that has been widely recognized [13]; thus, when confronted with challenges such as difficulty in designing novel sEMG features or enhancing the performance of existing classification models, the utilization of deep learning for sEMG signal classification becomes an up-and-coming solution.

A human ankle-foot movement classification method combining a two-dimensional convolutional neural network (2D-CNN) with low-cost EMG sensors, proposed by Chaobankoh et al.[14], achieved a high accuracy of 99.34% in classifying three movement states (Dorsiflexion, Neutral-position, and Plantar flexion); however, limitations exist in the classification scope of this method, as it only includes the above three movements and fails to cover a broader range of ankle-foot movement categories. An ankle-foot movement recognition system based on wireless wearable sEMG and acceleration (ACC) sensors, proposed by Zhou et al.[15],achieved classification of four simple ankle movements (Dorsiflexion, Plantarlexion, Eversion, and Inversion) by fusing the features of sEMG and ACC signals and adopting the Bidirectional Long Short-Term Memory (BiLSTM) algorithm, with a classification accuracy of 99.8%; nevertheless, limitations exist in this study, as it only focuses on these four specific movements and does not involve more complex movement patterns. A hybrid model, CNN-LSTM, proposed by Min et al. [16], for the first time adopted the Boruta algorithm to screen sEMG time-domain features, which were then input into the model (i.e., the "two-step method"); this model achieved the classification of six types of ankle movements under different load conditions, with an accuracy of 93.5%. However, due to the relatively complex structural design of existing models, lightweight optimization of the model is required, which can be achieved by reducing the number of LSTM layers, among other methods, while ensuring classification accuracy.

This paper proposes a lightweight CNN-LSTM hybrid model, which processes five-channel sEMG signals to achieve the classification of six movement patterns of the human ankle-foot joint; the model aims to overcome the limitations of existing methods to improve the accuracy and efficiency of ankle movement recognition, and its research results will support the development of neurally controlled mechanical exoskeletons, thereby enabling stroke patients or disabled individuals to perform rehabilitation training or engage in more daily activities [17-19].

2. Materials and Methods

2.1 Data Acquisition

The experiment is conducted with the eight-channel nine-axis (including accelerometer, gyroscope, magnetometer) sEMG sensor produced by the “Neuracle company”, and the sampling frequency is 1000Hz. Its primary operational process is shown in Figure 1.

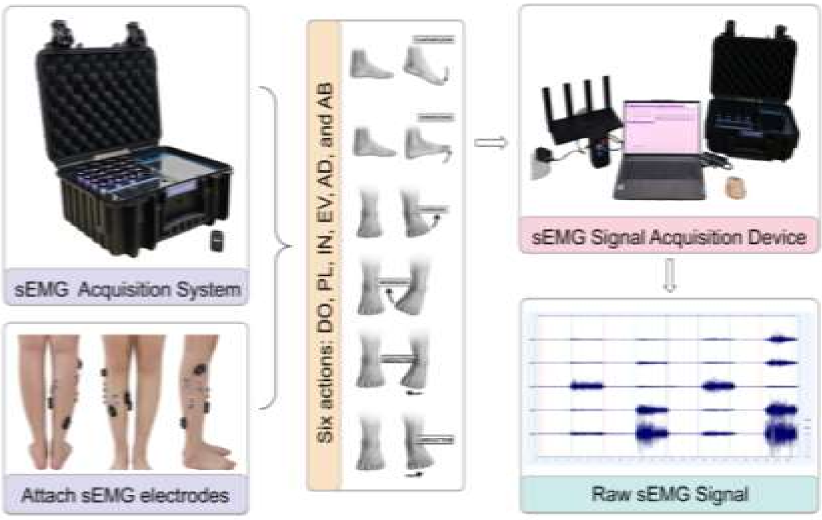

Figure 1: sEMG Acquisition System

In this experiment, six movement patterns of the ankle-foot were classified, including dorsiflexion (DO), plantarflexion (PL), inversion (IN), eversion (EV), adduction (AD), and abduction (AB). For clarity, as illustrated in Figure 2, a three-dimensional coordinate system is established with the center of the ankle joint as the origin, simplifying its movement representation. Specifically, rotation about the X-axis results in PL (toe drop) and DO (in the opposite direction), rotation about the Y-axis produces IN (inside foot lift) and EV (in the opposite direction), while rotation about the Z-axis leads to AD (toe turning inward) and AB (in the opposite direction).

Figure 2: Ankle joint coordinate system

For the acquisition of sEMG signals, five muscles of the right lower leg were selected: the medial gastrocnemius (MG), lateral gastrocnemius (LG), tibialis anterior (TA), soleus (SOL), and peroneus longus (PL).The placement of bipolar sEMG electrodes was determined by the standards of the Surface EMG for Non-Invasive Assessment of Muscles (SENIAM, seniam.org) [20-21]. The electrodes, numbered 1-5, were connected to the corresponding muscles of each subject, as shown in Figure 3.

Figure 3: EMG electrodes positions

A total of 8 healthy subjects (4 females and 4 males) were enrolled in this study, with an age range of 23-29 years, height of 165-183 cm, and weight of 50-75 kg; all participants signed a written informed consent form in accordance with the revised Declaration of Helsinki [22],all participants signed a written informed consent form. Subjects were instructed to sit upright on a chair, with their lower legs hanging naturally and their bodies relaxed. In this experiment, sEMG signals from the right lower leg were collected uniformly.

To accurately record the six ankle joint movements and eliminate mutual interference among the three degrees of freedom, an ankle joint movement experimental device was designed in this study to ensure the standardized execution of each action. As shown in Figure 4, the device enables independent acquisition of the six movements under the three degrees of freedom by restricting the movement of the other two degrees of freedom when collecting a movement of a specific degree of freedom. During the sEMG signal collection process, the subjects were required to actively operate the experimental device to complete the corresponding movements.

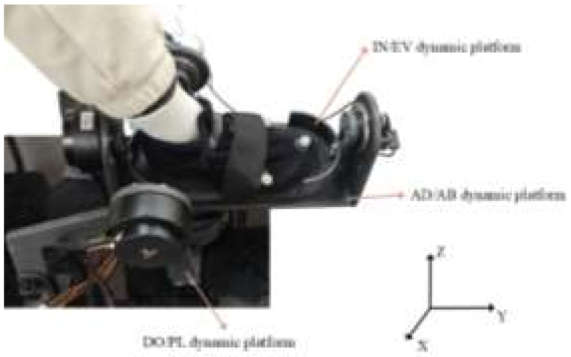

Figure 4: Ankle joint movement experimental device

To ensure the standardization of ankle movement execution and precisely determine whether each action meets the predefined criteria, a NOKOV motion capture system was integrated into the data acquisition process, as shown in Figure 5. This system was utilized to monitor and validate the standardization of the six ankle movements in real time, thereby complementing sEMG signal collection and enhancing dataset reliability.

Furthermore, the motion capture data and sEMG signals were time-aligned via a common trigger signal, ensuring that each segment of sEMG data (3000 sampling points) corresponded to a 3-second window of motion capture data. This synchronization allowed for cross-validation between muscle activation patterns and actual joint movements.

During the experiment, the motion capture software provided real-time data feedback to the subjects, guiding them to adjust their movements to conform to the standard range of motion. This rigorous data collection process not only ensured that the sEMG dataset was associated with uniformly standardized ankle movements but also effectively reduced interference noise caused by movement variability.

Figure 5: NOKOV motion capture system

Prior to the experiment, the hair on the right lower leg was removed, and skin oil was wiped off with alcohol to ensure that the impedance between the detection electrodes and the skin was less than 50 kΩ; after the experiment started, the triggerbox program was run via Python, and a synchronization box transmitted trigger signals to realize serial port labeling. The experiment was set to send nine signals per group (with a 3-s interval, totaling 24 s per group), and each group contained two movements (each repeated twice), resulting in 30 groups in total. The experimental procedure is shown in Figure 6. As each subject completed 360 experiments (6 movements × 60 repetitions), a total of 6 × 60 × 8 = 2880 experiments were completed in this study.

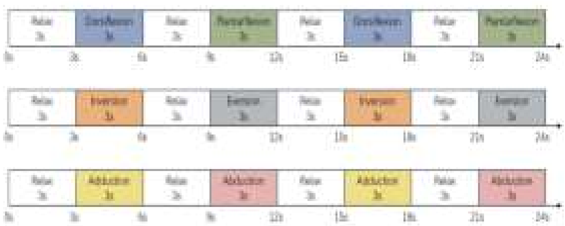

Figure 6: Experimental procedure

2.2 Data Pre-processing

sEMG signals, as complex and weak bioelectrical signals, are prone to distortion due to acquisition equipment, environmental factors, and the complexity of human physiology, with power frequency interference, high-frequency interference, and motion artifacts being the leading causes; to retain valuable information in the signals and eliminate various noises, a 6th-order Butterworth filter was employed to remove frequency components below 20Hz and above 450Hz, a 50Hz notch filter (with a quality factor set to 30) was used to remove power frequency interference. A bidirectional filter was applied for zero-phase filtering of the signals. The fixed length of each EMG data segment was defined as 3000 sampling points. Annotation timestamps were read from the evt.bdf file, and EMG data were segmented into fixed-length segments (via cropping or padding) according to event timestamps, with corresponding labels assigned to each segment.

In data segmentation, the window length was set to 350 milliseconds with a step size of 50 milliseconds. A longer window length can capture signal changes over a more extended period, as it covers a sufficient time range, while a shorter step size can capture finer-grained changes by providing higher temporal resolution. Subsequently, all valid data files were merged into a complete dataset (including feature matrix X and label vector Y) and converted into a feature matrix in the format (N, T, C) and a label vector in the format (N) to serve as input to the model, where N is the total number of samples, T is the time step, C is the number of channels, and the final number of samples was 310608.

3. Structure of the CNN-LSTM Hybrid Model

3.1 Structure of the CNN

As a feedforward neural network, CNN is widely applied for extracting local features of input data due to its advantages, performing particularly well in processing one-dimensional time series signals; the CNN model designed in this paper contained only one convolutional layer, which performed one-dimensional convolution operations on input signals by setting a 5-point convolution kernel, a padding parameter of 2, and 64 output channels. Convolved features were normalized via batch normalization, an operation that not only reduces internal covariate shift but also accelerates model convergence and enhances its generalization ability.

In addition, to introduce nonlinear characteristics, the Rectified Linear Unit (ReLU) was adopted by the model as the activation function, whose form is; its role is to significantly enhance the expressive ability of the model by retaining positive signals and mapping negative values to zero, while the following formula can express the operation of CNN:

$$y_t = \sigma\left(\sum_{i=0}^{k-1} W_i \cdot x_{t+i} + b\right) \tag{1}$$

Where: y_t denotes the output feature， σ denotes the nonlinear activation function, k denotes the size of the convolution kernel, W_i denotes the weight of the convolution kernel， x_{t+i} denotes the time - series input data observed at time step $t+i$ and b denotes the bias term.

3.2 Structure of the LSTM

LSTM is a type of recurrent neural network Recurrent Neural Network (RNN) specifically designed to handle long-term dependencies in time series data. Unlike traditional RNNs, LSTM can effectively alleviate the problems of gradient vanishing or gradient explosion [23-24], thereby capturing dynamic changes in the temporal dimension more accurately. The LSTM model designed in this paper, aimed at extracting deep-level feature representations from time series data and applying them to specific tasks, had its input feature dimension set to 64 (consistent with the output dimension of the preceding CNN layer) to ensure the coherence and compatibility of data flow. The hidden state dimension of the model's LSTM was set to 128, which gives the feature vector at each time step a higher dimension; However, a higher hidden state dimension can more effectively characterize complex patterns in time series, thereby improving the model's expressive ability and prediction accuracy, an excessively high dimension may increase computational overhead and risk overfitting in practical applications, thus requiring trade-offs based on the requirements of specific tasks. The specific update method of LSTM is as follows:

$$f_t = \sigma\left(W_f \cdot [h_{t-1}, x_t] + b_f\right) \tag{2}$$

$$i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i) \tag{3}$$

$$\tilde{C}_t = tanh(W_c \cdot [h_{t-1}, x_t] + b_c) \tag{4}$$

$$C_t = f_t * C_{t-1} + i_t * \tilde{C}_t \tag{5}$$

$$o_t = \sigma(W_o \cdot [h_{t-1}, x_t] + b_o) \tag{6}$$

$$h_t = o_t * tanh(C_t) \tag{7}$$

Where: f_t denotes the forget gate, i_t denotes the input gate, $\tilde{C}_t$denotes the candidate memory, C_t denotes the output gate, o_t denotes the memory cell state at the current moment, h_t denotes the output at the current moment, W_f , W_i , W_c , W_o denotes the corresponding weight and b_f , b_i , b_c denotes the corresponding bias.

3.3 Structure of the CNN-LST Mhybrid Model

CNN excels at extracting local features from short-time segments, while LSTM excels at tracking the dynamic changes of features in the temporal dimension and capturing long-term dependencies. Their combination can leverage unique advantages in spatiotemporal joint modeling, with particularly remarkable performance when processing sequence data with complex spatiotemporal characteristics.

A lightweight CNN-LSTM hybrid model is proposed in this paper, consisting of three main modules: a CNN layer, an LSTM layer, and a fully connected layer. Among them, the CNN layer is used to extract local spatial features from input data and efficiently captures short-term patterns within time segments through convolution operations; the LSTM layer receives the output features of the CNN layer and models them in the temporal dimension to capture long-term dependencies; the fully connected layer, taking the 128-dimensional hidden state output by the LSTM layer as input, further integrates features via linear transformation and nonlinear activation functions, then calculates the probability distribution of each category via the softmax function, and finally achieves 6-classification tasks. Figure 7 illustrates the CNN-LSTM hybrid model framework for the 6-class classification method of the ankle joint.

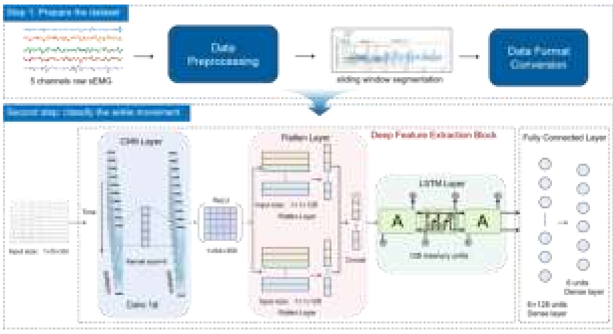

Figure 7: CNN-LSTM hybrid model framework for 6-class classification method

Through a lightweight architecture, the design of the hybrid model not only gives full play to the respective advantages of CNN and LSTM but also reduces computational cost, thereby improving the model's efficiency and generalization ability; the softmax function, in turn, can be expressed by the following formula：

$$Softmax(x_t) = \frac{e^{x_i}}{\sum_{j=1}^{K} e^{x_j}} \qquad (8)$$

Where: x_t denotes the raw output of the t category, K denotes the total number of categories, x_i denotes the raw output score (logit) for the i-th class before normalization and x_j denotes the raw output score (logit) for the j-th class, used in the denominator to normalize the probabilities.

In addition to the neural network architecture, the selection of hyperparameters also has a significant impact on model performance, with the finally determined hyperparameters including a batch size of 64, 30 training epochs, a learning rate of 0.001, and the Adam optimizer and cross-entropy loss function adopted; the dataset was divided into a training set and a validation set at an 8:2 ratio, with the batch sizes of both data loaders set to 64. The training data was set to be randomly shuffled to help the model generalize better and avoid overfitting. In contrast, the validation data was not shuffled, as the primary purpose of validation is to evaluate the model, and maintaining the data order unchanged facilitates tracking model performance.

The training and validation of the model were conducted on a computer with the following configuration: the CPU was 13th Gen Intel(R) Core(TM) i7-13700KF, with a central frequency of 3.40 GHz and a memory of 64 GB; the graphics card was NVIDIA GeForce RTX 4060 Ti, with a video memory of 39.9 GB.

4. Experiment and Results

This study employed a hybrid architecture combining CNN-LSTM to extract features from EMG signals and achieve the classification task of six ankle joint movements. The overall training process of the model was stable and efficient, reaching a converged state after approximately 17 epochs and ultimately exhibiting excellent classification performance on the validation set, as shown in Figure 8.

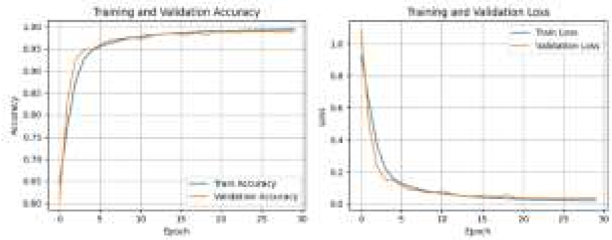

Figure 8 :Training and Validation Accuracy Curves and Loss Value Curves

During the model training process, the loss function value was observed to decrease gradually with training epochs, exhibiting rapid convergence characteristics; specifically, the significant reduction in loss value within the first five epochs indicated that the model could quickly adapt to the training data, while the loss value entered a steady decline stage from the 6th epoch and stabilized by the 17th epoch. This phenomenon verified the model's good generalization ability. Additionally, the adaptive adjustment strategy employed for the learning rate (Adam optimizer) further accelerated the model's convergence speed.

4.1Training performance with cross-validation

Five-fold cross-validation adopted in the training phase was used to evaluate model performance, specifically: the dataset was randomly divided into 5 equally sized subsets, with 1 subset selected as the test set and the remaining 4 merged into the training set each time; this process was repeated 5 times, with a distinct test set used each time. Finally, the average of the 5 test results was employed as the final evaluation metric. The performance metrics of the model in the five-fold cross-validation are presented in Table 1, and the five-fold cross-validation accuracy curve is shown in Figure 9.

Table 1: Classification Report of Five-Fold Cross-Validation

Category	Precision (%)	Recall (%)	F1-Score (%)
Adduction	99.72	99.17	99.45
Inversion	98.39	99.76	99.07
Abduction	99.91	99.51	99.71
Eversion	99.41	99.27	99.34
Dorsiflexion	99.65	99.57	99.61
Plantarflexion	99.74	99.49	99.61

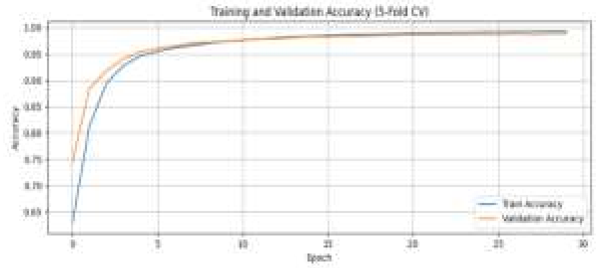

Figure 9: Accuracy curve of five-fold cross-validation

Results of the five-fold cross-validation demonstrated that the model exhibited high stability in performance across all data splits; in the classification report, Precision is defined as the proportion of samples predicted by the model as a specific category that belong to it (serving to measure prediction accuracy), Recall is defined as the proportion of samples belonging to a specific category that are correctly identified (serving to measure coverage capability for specific categories), and F1-score, as the harmonic mean of the two, comprehensively reflects the model's balanced performance across all categories. Experimental results demonstrated that the model performed excellently in the foot movement classification task, achieving not only a significantly high classification accuracy on the test set (with Accuracy at 99.46%, thus demonstrating outstanding performance) but also that results of the Macro Avg calculation showed the average of Precision, Recall, and F1-score across all categories reached 99.47%—a result that further confirmed the model's high stability in performance across categories. Meanwhile, the Weighted Avg was close to the Macro Avg (both at 99.46%), indicating that the imbalance in category distribution did not significantly affect the model evaluation results.

Further observation of the accuracy curves from the five-fold cross-validation revealed that the accuracy of both the training set and validation set exceeded 99%, indicating that the model not only has a strong generalization ability but also exhibits no significant overfitting. In addition, the high consistency of the five-fold cross-validation results confirms the model's insensitivity to data partitioning and its good stability. Notably, the model converged to the target state in only 17 epochs, and this high training efficiency may be attributed to the rationality of the optimizer, learning rate adjustment strategy, or model structure design.

It can be concluded from the above analysis that the model exhibits extremely high overall performance in the foot movement classification task: it not only significantly outperforms the requirements of general industrial applications but also possesses good class balance, uniform distribution of dataset samples, reliable evaluation results, and excellent robustness, with consistent results from the five-fold cross-validation further verifying its generalization ability. However, it should be noted that although cross-validation has been proven to reduce the risk of overfitting effectively, its performance estimation is still limited by the distribution characteristics of the existing dataset; thus, future research should be extended to validation on datasets collected from multi-center and multi-device sources to ensure the model's universality and practicality.

4.2 Test Performance with Confusion Matrix

The classification results of all subjects are visually presented via a confusion matrix, as shown in Figure 10, where row vectors represent predicted classes, column vectors correspond to true classes, and the color depth of its elements intuitively reflects the numerical values of the classification results; the evaluation results of the classification performance of six types of ankle joint movements in this study are reflected through this matrix.

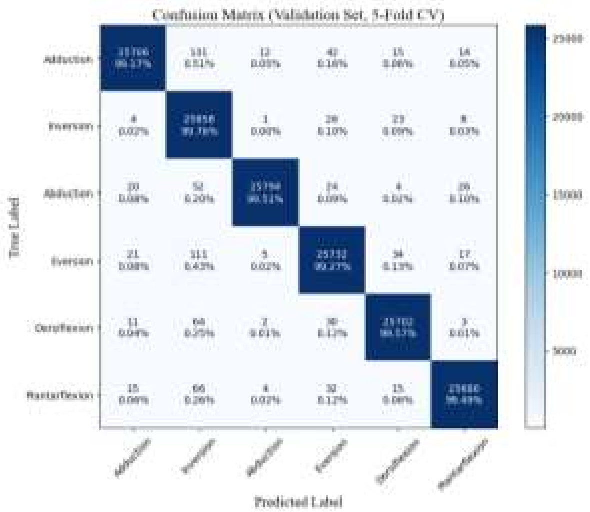

Figure 10: Confusion matrix of five-fold cross-validation

The main diagonal elements, representing the number of correctly classified samples for each category, were found to be significantly higher than the off-diagonal elements, indicating that the CNN-LSTM hybrid model proposed in this paper exhibits excellent performance in the six-class ankle joint movement classification task. Additionally, quantitative analysis revealed that the model achieved an average classification accuracy of 99.46% ± 0.19% across all categories, with a coefficient of variation (CV) of only 0.19% in accuracy among categories, confirming its highly stable classification performance. Specifically, as shown in Figure 11, AB achieved the best classification performance (99.71% ± 0.16%), followed by DO and PL (both 99.61%), AD and EV (99.45% ± 0.22% and 99.34% ± 0.05%, respectively), while IN showed relatively lower performance (99.07% ± 0.56%). This result not only highlights the effectiveness of the CNN-LSTM model but also provides an important reference for research on ankle joint movement recognition based on sEMG signals.

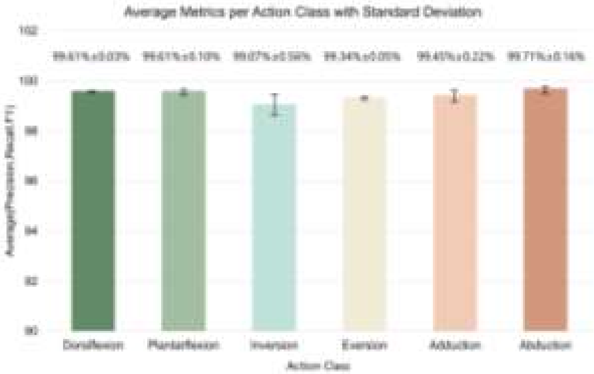

Figure 11: Classification Accuracy of 6 Types of Movements

The relatively low classification performance of inversion and eversion movements (99.07% ± 0.56% and 99.34% ± 0.05%, respectively) was examined. This phenomenon may be attributed to three key factors: first, due to the small joint range of motion (ROM approximately 20-30°), the amplitude variation of their sEMG signals was observed to be weak (<100μV); second, as anatomically antagonistic muscle group movements, their IMU signals were confirmed to have significant correlation in the time domain (Pearson $r=0.72$, $p<0.05$); third, t-SNE dimensionality reduction analysis indicates that there is an approximate 15% overlap region between the two types of movements in the feature space. Thus, future research can enhance classification performance through multimodal signal fusion (e.g., combining plantar pressure distribution) or optimization of attention mechanisms.

4.3 Comparison Among Different Models

To verify the performance of the proposed CNN-LSTM model, several typical model architectures were included in the comparison, including CNN, LSTM, and RNN-GRU in the field of deep learning, as well as RF, SVM, and LDA in traditional machine learning methods; the specific comparison results of all models are shown in Figure 12.

The CNN model architecture proposed in this subsection was more complex than the CNN module in the CNN-LSTM hybrid model presented in this paper, with its core structure consisting of three combinations of convolutional layers and max-pooling layers. Each convolutional layer was followed by Batch Normalization and Dropout operations to enhance generalization performance. Local features of the input data were extracted stepwise by the module through three convolutional layers, with feature dimensions reduced via max-pooling layers. Subsequently, Batch Normalization was introduced during feature extraction to accelerate training and stabilize performance, and the Dropout mechanism was integrated to suppress overfitting further. Finally, the extracted features were mapped to six categories through three fully connected layers to achieve a prediction output for the target task, resulting in an accuracy of 97.79%.

The LSTM structure proposed in this subsection differed from the CNN-LSTM hybrid model presented in this paper in that the CNN-LSTM hybrid model included a CNN layer for feature extraction from raw signals. In contrast, the single LSTM model lacked this capability, thus achieving a final accuracy of 97.82%.

In summary, the CNN-LSTM hybrid model proposed in this paper, by combining the advantages of CNN and LSTM, can not only efficiently extract spatial features of input data but also effectively model temporal dependencies; meanwhile, its lightweight structure design significantly reduces computational costs, thus confirming that it is particularly suitable for tasks with rich spatial features and temporal dependencies, exhibiting superior performance and stronger robustness in complex scenarios.

In addition, typical machine learning models were included in the comparative analysis in this subsection, and experimental results showed that they exhibited relatively poor performance in the foot movement classification task, with a significant gap between their performance and that of deep learning models. This finding further indicates that deep learning, with its strong feature learning ability and model expression capabilities, has become an important direction for solving complex tasks, exhibiting significant advantages, especially in scenarios requiring the processing of high-dimensional data and nonlinear relationships. Consequently, it has broad prospects for development in future research and applications.

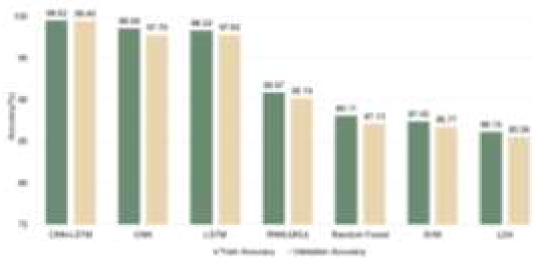

Figure 12: Accuracy of different models

5. Conclusion and Prospects

This study proposes a lightweight CNN-LSTM hybrid model for classifying six types of ankle joint movements based on sEMG signals. Through optimized design and improved training strategies, it is confirmed to achieve high-precision classification in the task of six types of ankle joint movements based on sEMG signals. Experimental results show that the model achieved a classification accuracy of 99.46% on the test set, exhibited high stability in five-fold cross-validation, and demonstrated balanced classification performance across all categories, with both the average accuracy and coefficient of variation at an excellent level.

Although significant results have been achieved by the model in ankle joint movement classification, there remains room for improvement and future research directions:

1. Multimodal signal fusion: The current model mainly achieves classification based on sEMG signals, and further improvement in classification performance can be achieved by fusing other biological signals (e.g., plantar pressure distribution, IMU); future research can explore multimodal signal fusion strategies to enhance the model's ability to recognize complex movement patterns.

2. Cross-center data validation: Although this study has validated the model's effectiveness on a single-center dataset, its universality and practicality still need to be verified using datasets collected from multi-center and multi-device sources; thus, future research can ensure the model's stability across different populations and device conditions by collecting more diverse datasets.

3. Real-time optimization: Although the model exhibits excellent classification performance, real-time applications require high computational efficiency; therefore, future research can optimize the model structure and inference speed to make it more suitable for embedded devices or real-time control systems.

4. Attention mechanism: For movements such as inversion and eversion, the model's relatively low classification performance may stem from the complexity and diversity of signal features; future research can introduce an attention mechanism to further improve classification performance by focusing on key periods or feature regions.

5. Transfer learning and personalized models: Given individual differences, future research can construct personalized models using transfer learning methods to meet the specific needs of different users.

In summary, although the CNN-LSTM hybrid model proposed in this study provides an effective solution for recognizing ankle joint movement intentions, further research is required in key areas, such as multimodal signal fusion, cross-center data validation, and real-time optimization, to facilitate its popularization and application in the field of medical rehabilitation.

Acknowledgements

This work was supported by the Natural Science Foundation of Qing dao (25-1-1-157-zyyd-jch), and Science & Technology Support Project for Young People in Colleges of Shandong Province (Grant No.2023KJ120).

References

[1] Zhuang, Y., Leng, Y., Zhou, J., Song, R., Li, L., & Su, S. W. (2020). Voluntary control of an ankle joint exoskeleton by able-bodied individuals and stroke survivors using EMG-based admittance control scheme. IEEE Transactions on Biomedical Engineering, 68(2), 695-705.

[2] Meng, F.-G., Zhang, B., Zhao, X.-G., & Sui, J. (2020). Estimation of ankle joint continuous motion based on electromyographic signals. 2020 Chinese Automation Congress (CAC). IEEE, 2020: 6399-6406.

[3] Zhao, X., Sun, H., & Ye, D. (2017). Ankle rehabilitation robot control based on biological signals. 2017 29th Chinese Control and Decision Conference (CCDC). IEEE, 2017: 6090-6095.

[4] Zhang, X., Qu, Y., Zhang, G., Wang, Z., Chen, C., & Xu, X. (2025). Review of sEMG for Exoskeleton Robots: Motion Intention Recognition Techniques and Applications. Sensors, 25(8), 2448.

[5] Wei, C., Wang, H., Hu, F., Zhou, B., Feng, N., Lu, Y., Tang, H., & Jia, X. (2022). Single-channel surface electromyography signal classification with variational mode decomposition and entropy feature for lower limb movements recognition. Biomedical Signal Processing and Control, 74, 103487.

[6] Shen, C., Pei, Z., Chen, W., Wang, J., Wu, X., & Chen, J. (2023). Lower limb activity recognition based on sEMG using stacked weighted random forest. IEEE Transactions on Neural Systems and Rehabilitation Engineering, 32, 166-177.

[7] Cai, S., Chen, D., Fan, B., Du, M., Bao, G., & Li, G. (2023). Gait phases recognition based on lower limb sEMG signals using LDA-PSO-LSTM algorithm. Biomedical Signal Processing and Control, 80, 104272.

[8] Bidabadi, S. S., Murray, I., Lee, G. Y. F., Morris, S., & Tan, T. (2019). Classification of foot drop gait characteristic due to lumbar radiculopathy using machine learning algorithms. Gait & Posture, 71, 234-240.

[9] Antuvan, C. W., & Masia, L. (2019). An LDA-based approach for real-time simultaneous classification of movements using surface electromyography. IEEE Transactions on Neural Systems and Rehabilitation Engineering, 27(3), 552-561.

[10] Toledo-Pérez, D. C., Martínez-Prado, M. A., Gómez-Loenzo, R. A., Paredes-García, W. J., & Rodríguez-Reséndiz, J. (2019). A study of movement classification of the lower limb based on up to 4-EMG channels. Electronics, 8(3), 259.
[11] Gu, F., Chung, M.-H., Chignell, M., Valaee, S., Zhou, B., & Liu, X. (2021). A survey on deep learning for human activity recognition. ACM Computing Surveys (CSUR), 54(8), 1-34.
[12] Song, Q., Ma, X., & Liu, Y. (2023). Continuous online prediction of lower limb joints angles based on sEMG signals by deep learning approach. Computers in Biology and Medicine, 163, 107124.
[13] Han, Y., & Tao, Q. (2025). Lower Limb Movement Recognition Based on a Hybrid Deep Learning Model Using Surface Electromyography. IEEE Access.
[14] Chaobankoh, N., Jumphoo, T., Uthansakul, M., Phapatanaburi, K., Sindhupakorn, B., Rooppakhun, S., & Uthansakul, P. (2022). Lower-Limb Motion-Based Ankle-Foot Movement Classification Using 2D-CNN. Comput Mater Continua, 73, 1269-1282.
[15] Zhou, C., Yang, L., Liao, H., Liang, B., & Ye, X. (2021). Ankle foot motion recognition based on wireless wearable sEMG and acceleration sensors for smart AFO. Sensors and Actuators A: Physical, 331, 113025.
[16] Li, M., Wang, J., Yang, S., Xie, J., Xu, G., & Luo, S. (2023). A CNN-LSTM model for six human ankle movements classification on different loads. Frontiers in Human Neuroscience, 17, 1101938.
[17] Vijayvargiya, A., Prakash, C., Kumar, R., Bansal, S., & Tavares, J. M. R. (2021). Human knee abnormality detection from imbalanced sEMG data. Biomedical Signal Processing and Control, 66, 102406.
[18] Castañeda, T. S., Connan, M., Capsi-Morales, P., Beckerle, P., Castellini, C., & Piazza, C. (2024). Experimental evaluation of the impact of sEMG interfaces in enhancing embodiment of virtual myoelectric prostheses. Journal of neuroengineering and rehabilitation, 21(1), 57.
[19] Merletti, R., Disselhorst-Klug, C., Rymer, W. Z., & Campanini, I. (2021). Surface Electromyography: Barriers Limiting Widespread Use of sEMG in Clinical Assessment and Neurorehabilitation. Frontiers in neurology, 2021, 12: 642257.
[20] Rainoldi, A., Melchiorri, G., & Caruso, I. (2004). A method for positioning electrodes during surface EMG recordings in lower limb muscles. Journal of neuroscience methods, 134(1), 37-43.
[21] Huenaerts, C., Dewit, T., Soentjens, D., Monari, D., Van Campenhout, A., & Desloovere, K. (2022). Differences in surface electrode placement and its effect on duration of EMG activity. Gait & Posture, 97, S7-S8.
[22] Kurihara, C., Kerpel-Fronius, S., Becker, S., Chan, A., Nagaty, Y., & Naseem, S. (2024). Declaration of Helsinki: ethical norm in pursuit of common global goals. Front Med. 2024; 11: 1360653.
[23] Vijayvargiya, A., Singh, B., Kumar, R., Desai, U., & Hemanth, J. (2022). Hybrid Deep Learning Approaches for sEMG Signal‐Based Lower Limb Activity Recognition. Mathematical Problems in Engineering, 2022(1), 3321810.
[24] John, C. T., Anderson, F. C., Higginson, J. S., & Delp, S. L. (2013). Stabilisation of walking by intrinsic muscle properties revealed in a three-dimensional muscle-driven simulation. Computer methods in biomechanics and biomedical engineering, 16(4), 451-462.