Luo J & Li N. (2025) REAL-TIME FEEDBACK MECHANISM BASED ON IMAGE ANALYSIS IN MOTOR FUNCTION RECOVERY. Revista Internacional de Medicina y Ciencias de la Actividad Física y el Deporte vol. 25 (100) pp. 281-297.
DOI: https://doi.org/10.15366/rimcafd2025.100.018

ORIGINAL

REAL-TIME FEEDBACK MECHANISM BASED ON IMAGE ANALYSIS IN MOTOR FUNCTION RECOVERY

Jinliang Luo[1], Nianmao Li[2*]

[1]Department of Physical Education, Shandong Jianzhu University, Jinan 250101, China.
[2]School of Physical Education and Health Science, Guang Xi Minzu University, Nanning, Guangxi, 530006, China.
E-mail: lnm262641048@163.com

Recibido 23 de junio de 2024 **Received** June 23, 2024
Aceptado 10 de diciembre de **2024 Accepted** December 10, 2024

ABSTRACT

In the motor function rehabilitation of patients, image analysis technology is being widely used in the rehabilitation treatment of patients. Combined with images, feedback on the patients' rehabilitation status can be provided, and the treatment plan can be adjusted in a timely manner. Existing image analysis technologies mostly use convolutional neural networks (CNN) or long short-term memory (LSTM) network to analyze continuous motion movements. They have good enough performance in terms of accuracy but may lack real-time performance. Based on this, this paper proposes a CNN-TCN architecture that combines the ResNet-50 model of the CNN structure and the temporal convolutional networks (TCN), a variant of the recurrent neural networks (RNN), and uses headphones, videos, etc., for real-time feedback. To verify the effect of the architecture and realize the real-time feedback mechanism of motion function based on image analysis, this experiment selects HMDB51 and KTH datasets as the initial datasets for training, supplemented with common daily action data such as walking, bending, and arm swinging, etc. Then 50 patients who need rehabilitation are recruited as volunteers to verify the results. The results are evaluated using three indicators: accuracy, recall, and feedback time. It is found that the accuracy and recall of CNN alone are 76% and 74%, and the accuracy and recall of LSTM alone are 83% and 84%, while the accuracy and recall of CNN-TCN are 86% and 87%. The feedback time of CNN, LSTM, and CNN-TCN is basically 340 to 400 milliseconds, 360 to 430 milliseconds, and 290-360 milliseconds respectively. CNN-TCN is better than CNN and LSTM in accuracy and also outperforms CNN and LSTM in inference time. Therefore, CNN-TCN is a better choice while ensuring high accuracy and good

effectiveness.

KEYWORDS: Motor Function Recovery; Image Analysis; Real-time Feedback; Recurrent Neural Network; Long Short-Term Memory; CNN-TCN Model.

1. INTRODUCTION

Motor function recovery is a significant research area in the field of rehabilitation medicine (Roesner et al., 2024), aiming to help patients with limited motor ability due to disease, trauma, or surgery to restore their normal motor function. As the population ages and lifestyles change, there is already a large base of patients with motor dysfunction, which is a great burden to the patient's family and society, so finding effective methods and technical tools for motor recovery has become especially important. Image analysis-based technology can provide targeted training guidance by acquiring patients' motor data in a variety of ways, including video monitoring and sensor data. This real-time feedback mechanism not only improves patient participation and motivation but also effectively promotes the recovery of motor function. Therefore, it has become an important trend in modern rehabilitation research to apply image analysis technology in motor function recovery (Kidziński et al., 2020). The continuous development of computer vision and deep learning (DL) technology has greatly promoted the research on patient motor rehabilitation and has also performed well in image analysis (Debnath et al., 2022; Swarnakar & Yadav, 2023). Computer vision technology enables real-time analysis of the motion status in the video stream and automatic detection of the patient's movement patterns, posture, and amplitude of movement, which allows doctors to understand the patient's motor performance in real time and make timely adjustments to the rehabilitation plan. Convolutional neural networks (CNN) excel in action recognition tasks and can accurately recognize multiple motion types. This is critical for monitoring the performance of exercise therapy and assessing the progress of patient rehabilitation. Liu K proposed a method based on filter-bank multi-scale convolutional neural network (FBMSNet) to decode motor imagery information in electroencephalogram (EEG) signals. The filter bank technology and multi-scale convolution operations were combined to better extract feature information in different frequency bands, ultimately significantly improving the decoding precision of motor imagery tasks (Liu et al., 2022) Tortora S et al. used long short-term memory (LSTM) neural network to study the decoding of gait patterns in brain signals. The proposed decoding method showed AUC (Area Under Curve)>90% for gait patterns (Tortora et al., 2020). Qiu Y proposed a Siamese convolutional neural network (SCNN) for evaluating the quality of movements in rehabilitation training. It used DL combined with motion posture matching technology to precisely analyze and judge the standardization of rehabilitation training movements, thereby improving the accuracy of rehabilitation training effect evaluation (Qiu et al., 2022). Bijalwan V proposed a heterogeneous computing model for identifying rehabilitation training for recovery of walking patterns and postural stability after injury. The model combined multiple computing resources and algorithms to achieve precise identification and analysis of gait and postural stability during rehabilitation, which helped to evaluate and improve the effectiveness of rehabilitation training (Bijalwan et al., 2022). Xu, Fangzhou proposed a method for EEG data augmentation based on a deep convolutional generative adversarial network (DCGAN) model for motor rehabilitation training after stroke. By generating more EEG data, the decoding performance of the motor imagery task was improved, thereby enhancing the effect of stroke rehabilitation training (Xu et al., 2022). Image analysis technology has demonstrated superior performance in tasks such as motion posture recognition, action classification, and motion trajectory tracking. It can not only objectively quantify the patients' motion status, but also effectively reduce human errors in the rehabilitation process (Chen et al., 2022). By using DL models such as advanced CNN, researchers are able to extract features from complex motion data to improve the personalization of rehabilitation training. Although image analysis technology has made significant progress in improving rehabilitation efficiency (Huo et al., 2021), there are still some limitations. Existing image analysis systems can only passively record and assess the patient's motor performance and lack immediate interactivity and real-time feedback capability, which may lead to difficulties in maintaining the correct posture or rhythm during the rehabilitation process, affecting the rehabilitation effect (Qi et al., 2021). Wang J used a method on the basis of the genetic algorithm-convolutional neural network (GA-CNN) for real-time motion pattern recognition of lower limb exoskeletons. It used an efficient algorithm to accurately identify the user's motion pattern, thereby optimizing the control system of the lower limb exoskeleton and providing more intelligent and adaptive feedback (Wang et al., 2022). Wen Y et al. used deep CNN to accurately identify motor unit activity from high-density electromyography (HD-EMG), demonstrating its feasibility and effectiveness. The latency was less than 80 milliseconds, achieving high-precision identification of motor units and providing an effective tool for EMG signal analysis and motion control (Wen et al., 2021). Tam S proposed an intuitive real-time control strategy that combined DL and transfer learning for the control of a high-density myoelectric prosthetic hand. The DL model was used to process the EMG signals, and transfer learning was used to improve the adaptability between different users, significantly improving the accuracy and response speed of the prosthetic hand control [(Tam et al., 2021). New possibilities can be provided for the personalization and efficiency of motor function recovery by combining image analysis technology with real-time feedback mechanisms (Hashimoto et al., 2020). Therefore, in this paper, a hybrid neural network model CNN-TCN (Ahmadi et al., 2024) that combines CNN and temporal convolutional networks (TCN) is proposed, and multimodal sensors (Ihianle et al., 2020)dedicated to the field of sports rehabilitation are used to collect data and provide real-time feedback. The CNN model selected in this paper is the deep residual network ResNet-50 architecture (Koonce & Koonce, 2021) under the CNN architecture, which is suitable for scenarios with high-precision requirements. The trained ResNet-50 model is used to extract key features such as posture recognition from the motion video frames of the dataset, and then output to TCN to analyze the motion sequence. The time dependency is identified, and finally the results are output (Du et al., 2023). According to the set feedback conditions and the analysis results, real-time feedback is given. By analyzing the experimental data and comparing the accuracy and real-time performance of CNN, LSTM, and CNN-TCN, it is found that CNN-TCN is indeed a better choice in terms of comprehensive accuracy and real-time performance.

2. Experimental Methods

The core of the real-time feedback mechanism is the ability to quickly and accurately evaluate motion performance, which often relies on efficient image analysis technology (Ismail & Malik, 2022). Figure 1 shows the overall structure of the method used in this paper.

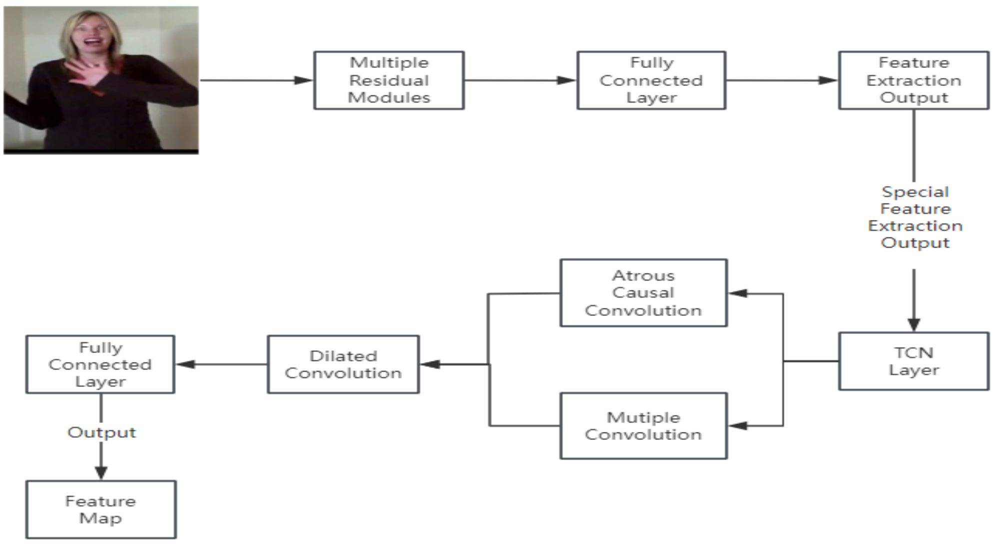

Figure 1: Overall Structure of Method in this Paper

The model first extracts feature such as arm contours and stride lengths in video frames through multiple residual modules, and then integrates these features through fully connected layers. Next, the model further processes the features using dilated convolutions and specific convolutional network layers to capture more complex patterns. Finally, the model outputs feature maps.

2.1 Convolutional Neural Networks

CNN is widely used in many fields such as object detection (X. Wu et al., 2020), image classification (Chen et al., 2021), image analysis (Guan & Liu, 2021), etc. CNN is a feedforward neural network with convolution calculation and deep structure (Cong & Zhou, 2023). Its basic structure mainly includes input layer, convolutional layer, pooling layer, activation function layer, and fully connected layer.

2.1.1 Input Layer

The input layer receives image data related to human body movements from sensors as the basis for analysis, and the data is then input into the network after preprocessing.

2.1.2 Convolutional Layer

The convolutional layer identifies the most basic visual features in the image, such as changes in hand position, tilt of the waist, movement trajectory of the legs, and other details, and provides these feature maps to subsequent layers.

2.1.3 Activation Function

The activation layer helps the model understand nonlinear features such as rapid changes in motion or adjustments to complex postures in the image. The role of the ReLU activation function is to add nonlinearity to the features extracted by the convolutional layer, allowing the network to recognize more complex motion patterns. Common activation functions include Sigmoid, Tanh, Softmax, and ReLU functions (Onwujekwegn & Yoon, 2020; Wang et al., 2020). Figure 2 presents the curves of four commonly used activation functions:

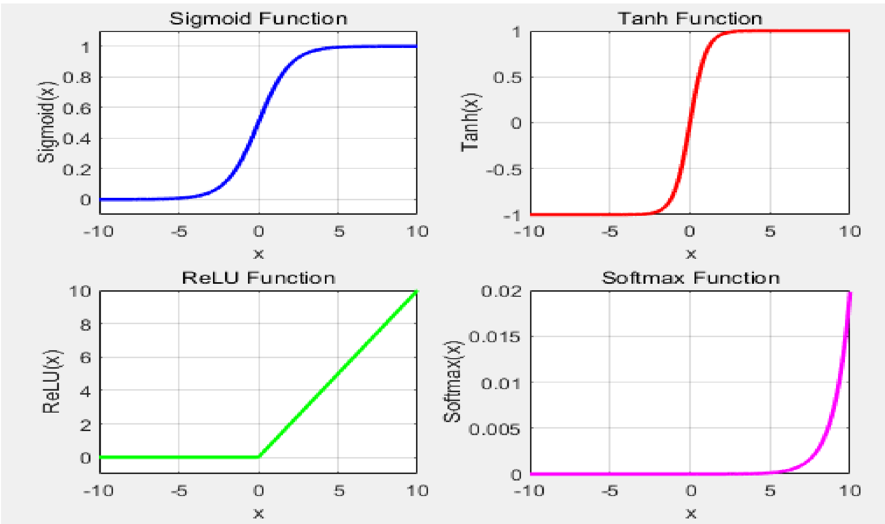

Figure 2: Common Activation Function Curves

Their formulas are as follows:

$$\mathrm{Sigmoid}(x) = \frac{1}{1+e^{-x}} \tag{1}$$

In the Sigmoid function, x is the input value, usually the weighted input of the neuron. e^{-x} controls the shape of the function.

$$\text{Soft max}(x_i) = \frac{e^{x_i}}{\sum_{j=1}^{n} e^{x_j}} \tag{2}$$

x_i represents the i-th element in the input vector. e^{x_i} represents the exponential function of the input element, ensuring that the output is a positive number.

$$\tanh(x) = \frac{e^{x}-e^{-x}}{e^{x}+e^{-x}} \tag{3}$$

x is the input value, usually the sum of the weighted inputs to the neuron, and can be any real number. e^{x} is the positive exponential transform of x. e^{-x} is the negative exponential transform of x.

$$\text{ReLU}(x) = \max(0, x) \tag{4}$$

In the ReLU function, x is the input value, and the function returns the larger value of x and 0. This helps improve computational efficiency in deep networks. The activation function used in this experiment is the ReLU activation function, which performs very well in most neural networks.

2.1.4 Pooling Layer

The pooling layer helps reduce the noise in the image, focusing on the most representative features such as stride, arms, and body contour, ignoring unimportant parts and retaining the most important features, thereby achieving the effect of down sampling. It can usually reduce the amount of data calculation, reduce the parameters of the model, and increase the training speed.

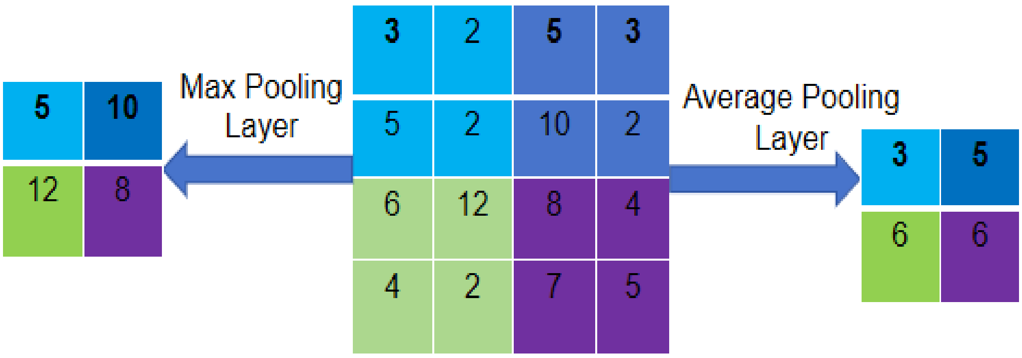

Figure 3: Average Pooling Layer and Max Pooling Layer

Figure 3 displays the specific diagram of average pooling layer and max pooling layer. Max pooling selects the largest value from a region to represent the features of the region. Average pooling calculates the average value within the region.

2.1.5 Fully Connected Layer

The fully connected layer can combine the features extracted by the convolutional layer and the pooling layer, and determine the type of the overall action by integrating the previously extracted features such as the arm movements, waist postures, leg steps, etc.

2.2 ResNet-50

Residual network (ResNet) is a deep CNN architecture. This network has attracted widespread attention due to its deep structure and excellent performance. The core concept of ResNet is to apply residual connections or skip connections (D. Wu et al., 2020). This type of connection enables information to be passed across layers in the network, rather than relying only on layer-by-layer forward propagation. With this structure, the network is able to focus on the difference between the input and the target output, simplifying the learning process and allowing deeper layers of the network to still perform well. The core construction block of ResNet is the residual block, which consists of a convolution path and a skip connection. The skip connection significantly improves the training efficiency and overall performance of the deep network by adding the input directly to the convolution output instead of simply overwriting the input. The residual connection allows the DL module to be passed down in an additive form by connecting the input and output of the module. Formula (5) is the module mathematical expression of this residual connection.

$$y = f(x) + x \tag{5}$$

In Formula (5), y represents the output of the residual connection; $f(x)$ represents the method within the module; x represents the input which is the output of the previous layer. ResNet has a simple structure and a large network depth, and has excellent model training effect. It is very suitable for motion image analysis tasks applied to patients' motor function rehabilitation.

2.3 Temporal Convolutional Networks

The core operation of the TCN when processing time series data is the dilated causal convolution (Sasou, 2021). Compared with ordinary convolution, this convolution method has better temporal causal properties and a larger receptive field, enabling it to effectively capture long-term dependencies. The dilated causal convolution operation is shown in Figure 4. Compared with ordinary convolution, dilated convolution expands the convolution range without increasing the number of parameters. Compared with causal convolution, the output y_t of dilated causal convolution can trace back a longer time interval.

$u_c \in \{u_0, u_1, u_2, \ldots, u_{t-1}\}$ is a one-dimensional time series data, where the dilated causal convolution operation Θ of element c is calculated as follows.

$$\Theta(c) = (u * f)(c) = \sum_{i=0}^{k-1} f(i) u_{c-d \cdot i} \tag{6}$$

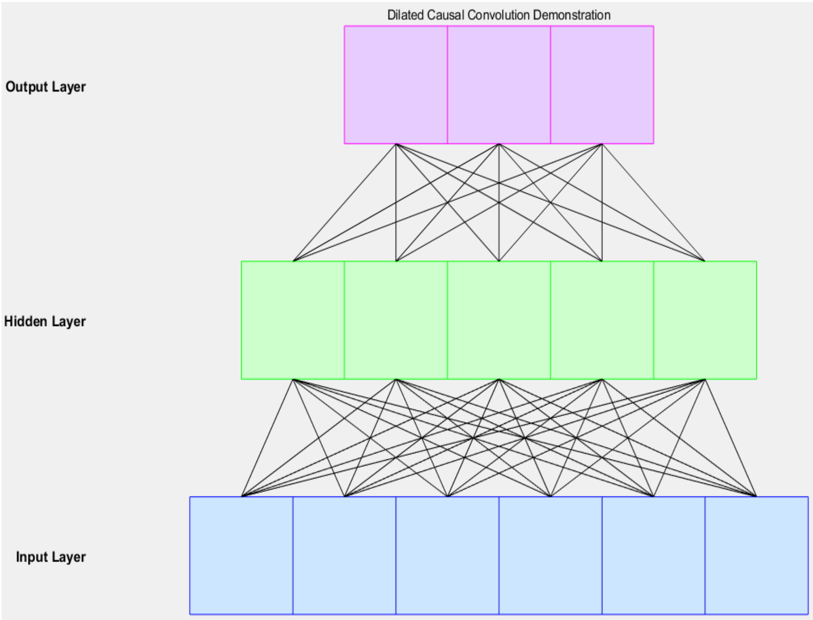

Figure 4: Dilated Causal Convolution

In Formula (6), $*$ is the convolution operator; u is the input time series vector; f represents the convolution kernel; $f(i)$ is the element value of the convolution kernel f; k is the filter size; d is the atrous rate; $u_{c-d \cdot i}$ is the element value of $c - d \cdot i$ in vector u, and $c - d \cdot i$ also considers the past direction. The dilated causal convolution can expand the receptive field by increasing the atrous rate d or the filter k. In the shallow network, a smaller atrous rate is selected to ensure local feature extraction, and in the deep network, the atrous rate is increased to calculate the local information at different times, which ensures that the network is lightweight while ensuring that long-term information is not leaked. The data first enters the ResNet-50 layer, and then goes through multiple residual network modules to extract spatial features to the fully connected layer. The data is then input into the TCN layer, and through dilated convolution and multiple residual connections, long-range dependencies in the sequence are captured, and various time series data problems are handled. The prediction results are output through the fully connected layer and fed back to the multi-modal sensor for real-time feedback.

2.4 Mean-Square Error Loss Function

The loss function is a key component in experiments on image analysis, effectively measuring the gap between the predictions of the model and the true labels (Yin et al., 2021). The mean-square error (MSE) loss function is commonly used in model training for regression problems to measure the difference between the model's predicted value and the true value (Miao et al., 2021). The mathematical expression of the MSE loss function is as follows:

$$\mathrm{MSE} = \frac{1}{N}\sum_{i=1}^{N}(y_i - y^{\wedge})^2 \quad (7)$$

N represents the number of samples; y_i represents the true value; $y^{\wedge}$ represents the predicted value. MSE measures the gap between the model's predictions and the actual results at each frame or time point and gradually optimizes the network so that the model's predictions at each moment are closer to the actual action. This is crucial to ensure that the model precisely captures dynamic changes. In the early stages of training, there may be large deviations in model predictions. Using MSE can help the model converge faster, quickly adjust parameters, and reduce large errors, which is especially important when capturing subtle differences in actions such as waving, bending, and walking.

2.5 Motion Performance Evaluation and Feedback Mechanism

After feature extraction is completed, the system performs real-time evaluation based on key motion features, covering posture error and motion coherence (Bu, 2020).

2.5.1 Posture Evaluation

The position information of the key points of the joints is used to evaluate whether the posture of the movement is consistent with the standard posture. The posture quality is judged by calculating the angle deviation between key points, such as the knee angle and the height of the arm raised.

2.5.2 Stability and Consistency Testing

The continuity of joint movement is analyzed, and the stability during the movement is determined, such as the smoothness of gait and the shaking of knee joints. The system detects unstable motion signals through TCN and generates relevant feedback.

2.5.3 Implementation of Feedback Mechanism

According to the results of the motion performance evaluation, the system generates real-time feedback information to help users adjust their movements immediately (Van Hooren et al., 2020). The system displays the user's motion score, posture error reminders, and improvement suggestions on the display screen, and issues a sound warning when the user performs an irregular movement, reminding him to pay attention to his postures.

3. Experimental Settings

3.1 Hardware Configuration

The computer used in the experiment is configured as an NVIDIA RTX 3060 GPU desk computer, Intel i7 CPU, and 32GB memory. The camera used to collect data is configured as an Intel RealSense D435. The feedback equipment is a high-definition display and headphones.

3.2 Data Preprocessing

To realize the real-time feedback mechanism of motor function based on image analysis, this experiment selects HMDB51 and KTH as the initial data sets, supplemented with common daily action data. These data sets have high-resolution action video samples, covering a variety of postures and action changes of human body movements, which are conducive to the model's extraction and learning of motion features.

The actions in the data set are divided into three groups of data related to motor function recovery, such as walking, bending, and arm swinging, with 50 videos in each group. They are divided into a training set and a test set in a ratio of 8:2, and the data of 50 volunteers is used as the validation set. Figure 5 shows some image data in the data set:

Figure 5: Display of Some Image Data

Model training settings: In the experimental training, the backbone feature extraction network selected is ResNet-50. Hyperparameter settings: The initial value of the learning rate of the CNN model is set to 0.0001, and the cosine annealing strategy (Cheng et al., 2023) is adopted to reduce the learning rate to control the training convergence speed. TCN model configuration: The time step is set to 8 frames; the number of TCN layers is set to 2 layers; the batch size is set to 1; the total number of training epochs is 50 for each group of data. Figure 6 displays the training results.

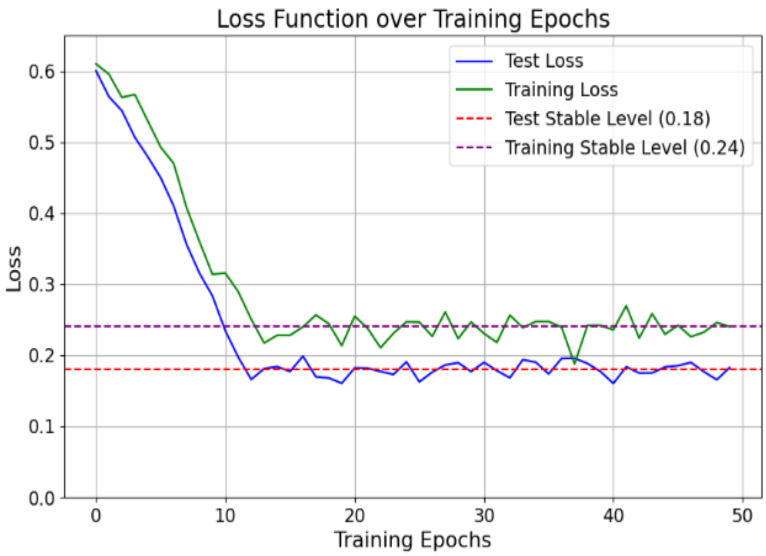

Figure 6: Training Effect Curves

At the beginning of training, both the test loss and the training loss drop rapidly. Around the 10th epoch, both the test loss and the training loss reach a low level. The test loss stabilizes at around 0.18, and the training loss stabilizes at around 0.24. The gap between the test loss and the training loss is not large. The model performs relatively consistently on the test set and the training set.

3.3 Evaluation Criteria

To measure the effectiveness of the model, this paper uses three indicators: accuracy, recall, and feedback time. TP represents the positive sample data with correct detection results; FP is the data that is incorrectly labeled as positive; TN is the number of negative classes correctly classified as negative by the model; FN is the number of positive classes misclassified as negative by the model. The calculation formula of Accuracy is:

$$\text{Accuracy} = \frac{\text{TP+TN}}{\text{TP+TN+FN+FP}} \tag{8}$$

The calculation formula of recall R is:

$$\text{R} = \frac{\text{TP}}{\text{TP+FN}} \tag{9}$$

Feedback time is the average time required for the model to process a single sample, usually expressed in milliseconds. T represents the total time required to process all samples (in milliseconds), and N represents the total number of samples.

$$\text{Inference Time}=\frac{\text{T}}{\text{N}} \tag{10}$$

4. Experimental Results

4.1 Ablation Experiment

Table 1 displays a comparison of the data obtained from ablation experiments on the training set for several mainstream object recognition algorithms.

Table 1: Data Comparison of Ablation Experiments

	ACCURACY	RECALL	FEEDBACK TIME (MS)
CNN	76%	74%	340-400
LSTM	83%	84%	360-430
CNN-TCN	86%	87%	290-360

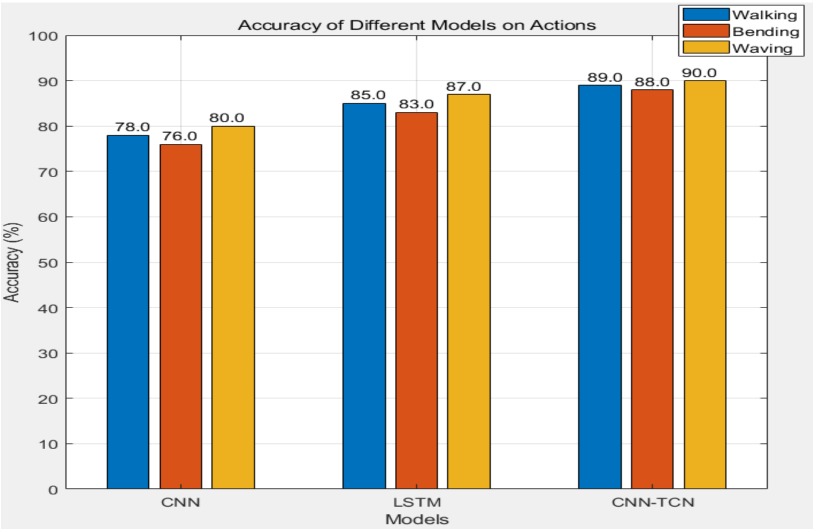

Figure 7: Recognition Accuracy of Different Models for Different Actions on the Validation Set

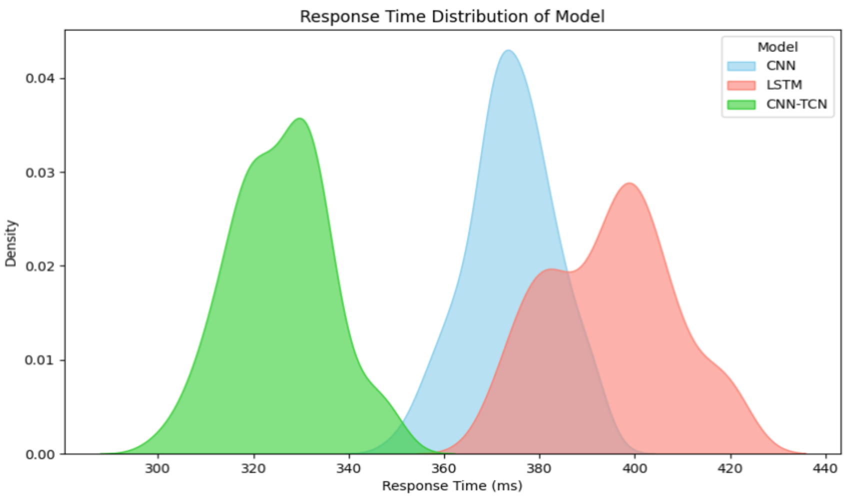

Figure 8: Normal Distribution of Feedback Time of Different Models

Table 1 presents the comparison results of three models in terms of accuracy, recall and feedback time. Figure 7 displays the accuracy comparison of different models in different actions on the validation set. Figure 8 shows the distribution diagram of feedback time of each model. The following results can be obtained from Table 1 and Figures 7 and 8. CNN achieves an accuracy (Acc) of 76% and a recall (R) of 74%, with a feedback time of 340ms-400ms. LSTM achieves an accuracy of 83% and a recall of 84%, with a feedback time of 360ms-430ms. CNN-TCN model shows excellent performance, with an accuracy of 86%, a recall of 87%, and feedback time of 290ms-360ms, which are better than those of CNN and LSTM. On the validation set, the average accuracy of CNN is 78%; that of LSTM is 85%; that of CNN-TCN is 89% which is the highest among the three. Because the posture of the human body changes greatly in the action of bending, the accuracy of bending is the lowest among the three actions(Zhou et al., 2023).

4.2 Test Data Display

A total of 50 rehabilitation patients are recruited as volunteers in this test. This experiment aims to evaluate the functional performance of rehabilitation patients in three actions: walking, bending and waving, so as to understand their rehabilitation progress. Participants are tested in a laboratory environment on the following three actions: walking, bending, and waving, and repeating the training for 5 sets. During the test, sensors and real-time monitoring systems are used to evaluate each action in real time. Data is automatically recorded by a computer system, and real-time feedback is displayed through a visual interface. Participants can see their performance instantly and make adjustments during training. Tables 2, 3, and 4 are the specific data of the three actions in the experiment.

Table 2: Specific Data of Walking Action

MODEL	ACTION	CORRECT	INCORRECT	AVERAGE FEEDBACK TIME (MS)
CNN	Walking	201	49	377.55
LSTM	Walking	214	36	398.47
CNN-TCN	Walking	227	23	340.67

Table 3: Specific Data of Bending Action

MODEL	ACTION	CORRECT	INCORRECT	AVERAGE FEEDBACK TIME (MS)
CNN	Bending	187	63	385.22
LSTM	Bending	203	47	410.84
CNN-TCN	Bending	219	31	332.91

Table 4: Specific Data of Waving Action

MODEL	ACTION	CORRECT	INCORRECT	AVERAGE FEEDBACK TIME (MS)
CNN	Waving	211	39	373.68
LSTM	Waving	218	32	395.90
CNN-TCN	Waving	233	17	320.12

4.3 Test Results

The experimental results show that when analyzing images of different actions, the accuracy of the CNN model is low, and the feedback time is relatively long. Its real-time feedback capability needs to be strengthened. The accuracy of LSTM is lower than that of CNN-TCN but higher than that of CNN. It processes long sequences slowly because it relies on previous information frame by frame, and the average feedback time is the longest, which limits its potential for real-time application. CNN-TCN has the highest accuracy among the three and the shortest feedback time. After applying TCN to CNN, the model can focus on both spatial and temporal information, thereby showing higher accuracy when processing dynamic action recognition. Its performance in accuracy and feedback time is much better than that of CNN and LSTM.

5. Conclusion

In summary, through the evaluation and research of the performance of different models, the CNN-TCN model is not only superior to CNN and LSTM in accuracy and feedback time, but also has better accuracy and feedback time than CNN and LSTM when performing image analysis on different movements, whether in the waving action with small movements or the bending action with large posture changes. The combination of image analysis technology and real-time feedback mechanism can significantly improve the effect of the patients' rehabilitation process. This shows that in motor function recovery training, the use of image analysis technology can help monitor the movement status of rehabilitation patients in real time and identify and feedback their movement performance in a timely manner, thereby providing strong support for the patients' treatment.

With the continuous advancement of computer vision and DL technologies, intelligent management of motor function recovery is gradually maturing. The real-time feedback mechanism can not only enhance the training efficiency of patients but also provide strong data support for the formulation of personalized rehabilitation plans. Evaluation methods based on image analysis can be more deeply integrated into the rehabilitation process, helping medical workers to better understand and optimize patients' motion performance. Applying these advanced technologies to a wider range of rehabilitation scenarios can help promote the scientific, precise, and personalized development of motor function recovery, and ultimately improve patients' overall rehabilitation experience and effects.

6. Acknowledgment

This work was supported by 2023, Humanities and Social Science Project of Guang Xi Minzu University [number: 2023MDSKYB05];

References

Ahmadi, S. A., Muhammad, F. D., & Al Dawsari, H. (2024). CNN-TCN: Deep Hybrid Model Based on Custom CNN with Temporal CNN to Recognize Sign Language. *Journal of Disability Research*, *3*(5), 20240034.

Bijalwan, V., Semwal, V. B., Singh, G., & Crespo, R. G. (2022). Heterogeneous computing model for post-injury walking pattern restoration and postural stability rehabilitation exercise recognition. *Expert Systems*, *39*(6), e12706.

Bu, X. (2020). Human motion gesture recognition algorithm in video based on convolutional neural features of training images. *Ieee Access*, *8*, 160025-160039.

Chen, L., Li, S., Bai, Q., Yang, J., Jiang, S., & Miao, Y. (2021). Review of image classification algorithms based on convolutional neural networks. *Remote Sensing*, *13*(22), 4712.

Chen, X., Wang, X., Zhang, K., Fung, K.-M., Thai, T. C., Moore, K., Mannel, R. S., Liu, H., Zheng, B., & Qiu, Y. (2022). Recent advances and clinical applications of deep learning in medical image analysis. *Medical image analysis*, *79*, 102444.

Cheng, W.-C., Hsiao, H.-C., & Li, L.-H. (2023). Deep learning mask face recognition with annealing mechanism. *Applied Sciences*, *13*(2), 732.

Cong, S., & Zhou, Y. (2023). A review of convolutional neural network architectures and their optimizations. *Artificial Intelligence Review*, *56*(3), 1905-1969.

Debnath, B., O'brien, M., Yamaguchi, M., & Behera, A. (2022). A review of computer vision-based approaches for physical rehabilitation and assessment. *Multimedia Systems*, *28*(1), 209-239.

Du, Y. H., Li, X., Ren, W., & Zuo, H. (2023). Application of near-infrared spectroscopy and CNN-TCN for the identification of foreign fibers in cotton layers. *Journal of Natural Fibers*, *20*(1), 2172638.

Guan, H., & Liu, M. (2021). Domain adaptation for medical image analysis: a survey. *IEEE Transactions on Biomedical Engineering*, *69*(3), 1173-1185.

Hashimoto, R., Requa, J., Dao, T., Ninh, A., Tran, E., Mai, D., Lugo, M., Chehade, N. E.-H., Chang, K. J., & Karnes, W. E. (2020). Artificial intelligence using convolutional neural networks for real-time detection of early esophageal neoplasia in Barrett's esophagus (with video). *Gastrointestinal endoscopy*, *91*(6), 1264-1271. e1261.

Huo, C.-C., Zheng, Y., Lu, W.-W., Zhang, T.-Y., Wang, D.-F., Xu, D.-S., & Li, Z.-Y. (2021). Prospects for intelligent rehabilitation techniques to treat motor dysfunction. *Neural regeneration research*, *16*(2), 264-269.

Ihianle, I. K., Nwajana, A. O., Ebenuwa, S. H., Otuka, R. I., Owa, K., & Orisatoki, M. O. (2020). A deep learning approach for human activities recognition from multimodal sensing devices. *Ieee Access*, *8*, 179028-179038.

Ismail, N., & Malik, O. A. (2022). Real-time visual inspection system for grading fruits using computer vision and deep learning techniques. *Information*

Processing in Agriculture, *9*(1), 24-37.

Kidziński, Ł., Yang, B., Hicks, J. L., Rajagopal, A., Delp, S. L., & Schwartz, M. H. (2020). Deep neural networks enable quantitative movement analysis using single-camera videos. *Nature communications*, *11*(1), 4054.

Koonce, B., & Koonce, B. E. (2021). *Convolutional neural networks with swift for tensorflow: Image recognition and dataset categorization*. Springer.

Liu, K., Yang, M., Yu, Z., Wang, G., & Wu, W. (2022). FBMSNet: A filter-bank multi-scale convolutional neural network for EEG-based motor imagery decoding. *IEEE Transactions on Biomedical Engineering*, *70*(2), 436-445.

Miao, M., Cai, W., & Li, X. (2021). Parameter Estimation of Gamma-Gamma Fading with Generalized Pointing Errors in FSO Systems. *Wireless Communications and Mobile Computing*, *2021*(1), 1301878.

Onwujekwegn, G., & Yoon, V. Y. (2020). Analyzing the Impacts of Activation Functions on the Performance of Convolutional Neural Network Models.

Qi, Y., Yang, Z., Sun, W., Lou, M., Lian, J., Zhao, W., Deng, X., & Ma, Y. (2021). A comprehensive overview of image enhancement techniques. *Archives of Computational Methods in Engineering*, 1-25.

Qiu, Y., Wang, J., Jin, Z., Chen, H., Zhang, M., & Guo, L. (2022). Pose-guided matching based on deep learning for assessing quality of action on rehabilitation training. *Biomedical Signal Processing and Control*, *72*, 103323.

Roesner, K., Scheffler, B., Kaehler, M., Schmidt-Maciejewski, B., Boettger, T., & Saal, S. (2024). Effects of physical therapy modalities for motor function, functional recovery, and post-stroke complications in patients with severe stroke: a systematic review update. *Systematic Reviews*, *13*(1), 270.

Sasou, A. (2021). Deep residual learning with dilated causal convolution extreme learning machine. *Ieee Access*, *9*, 165708-165718.

Swarnakar, R., & Yadav, S. L. (2023). Artificial intelligence and machine learning in motor recovery: A rehabilitation medicine perspective. *World Journal of Clinical Cases*, *11*(29), 7258.

Tam, S., Boukadoum, M., Campeau-Lecours, A., & Gosselin, B. (2021). Intuitive real-time control strategy for high-density myoelectric hand prosthesis using deep and transfer learning. *Scientific Reports*, *11*(1), 11275.

Tortora, S., Ghidoni, S., Chisari, C., Micera, S., & Artoni, F. (2020). Deep learning-based BCI for gait decoding from EEG with LSTM recurrent neural network. *Journal of neural engineering*, *17*(4), 046011.

Van Hooren, B., Goudsmit, J., Restrepo, J., & Vos, S. (2020). Real-time feedback by wearables in running: Current approaches, challenges and suggestions for improvements. *Journal of Sports Sciences*, *38*(2), 214-230.

Wang, J., Wu, D., Gao, Y., Wang, X., Li, X., Xu, G., & Dong, W. (2022). Integral real-time locomotion mode recognition based on GA-CNN for lower limb exoskeleton. *Journal of Bionic Engineering*, *19*(5), 1359-1373.

Wang, Q., Ma, Y., Zhao, K., & Tian, Y. (2020). A comprehensive survey of loss functions in machine learning. *Annals of Data Science*, 1-26.

Wen, Y., Avrillon, S., Hernandez-Pavon, J. C., Kim, S. J., Hug, F., & Pons, J. L. (2021). A convolutional neural network to identify motor units from high-density surface electromyography signals in real time. *Journal of neural engineering*, *18*(5), 056003.

Wu, D., Wang, Y., Xia, S.-T., Bailey, J., & Ma, X. (2020). Skip connections matter: On the transferability of adversarial examples generated with resnets. *arXiv preprint arXiv:2002.05990*.

Wu, X., Sahoo, D., & Hoi, S. C. (2020). Recent advances in deep learning for object detection. *Neurocomputing*, *396*, 39-64.

Xu, F., Dong, G., Li, J., Yang, Q., Wang, L., Zhao, Y., Yan, Y., Zhao, J., Pang, S., & Guo, D. (2022). Deep convolution generative adversarial network-based electroencephalogram data augmentation for post-stroke rehabilitation with motor imagery. *International journal of neural systems*, *32*(09), 2250039.

Yin, T., Liu, N., & Sun, H. (2021). Self-paced active learning for deep CNNs via effective loss function. *Neurocomputing*, *424*, 1-8.

Zhou, Q., Chen, X., Wu, H., Ma, S., & Wang, Q. (2023). Diffuse uerine leiomyomatosis complicated with cerebral thrombosis after oral contraceptive treatment: a case report of athletic patients. *Revista multidisciplinar de las Ciencias del Deporte*, *23*(90).