# Mid-Air Hand Rehabilitation Evaluation and Virtual Training System

*Rongzhe ZHOU, Peize LAI, Le CHEN, Zhibo HU, Lei QIAN*

School of Electronic and Information Engineering, Tiangong University, 300000, Tianjin, China

qianlei@tiangong.edu.cn

Submitted November 28, 2025 / Accepted February 27, 2026 / Online first April 20, 2026

**Abstract.** *This paper presents the design of a mid-air hand rehabilitation evaluation and virtual training system based on Leap Motion and gesture recognition algorithms. The system aims to provide a scientifically-grounded and accessible home-based rehabilitation solution for patients with hand injuries. It employs a spatio-temporal attention-enhanced multi-scale residual graph convolutional network algorithm for gesture recognition. Following this, specific joint angles are calculated and 14 representative gesture scores are derived. Weights are assigned to each scoring item via ridge regression to achieve a quantitative assessment. The rehabilitation training module comprises two modes: interactive turn-based games and music rhythm games, designed to train hand movement and dexterity. The system was subsequently tested to evaluate the performance of both the assessment function and the two training games. Results show an average gesture recognition accuracy of 94.86%. Furthermore, the reliability scores for both training games exceeded 90%. These findings demonstrate that the system achieves good accuracy in gesture recognition and effective assessment of hand rehabilitation progress.*

## Keywords

Gesture recognition, hand rehabilitation, multi-scale residual graph convolution, attention mechanism, human-computer interaction

## 1. Introduction

The hand is one of the most delicate and complex organs in the human body [1], playing a crucial role in daily activities such as eating, dressing, traveling, and communicating. However, with ongoing industrial development, the incidence of hand injuries is increasing [2], [3]. Major causes include improper machine operation and accidents during activities such as sports or physical labor [3–5]. Furthermore, the age profile of patients with hand injuries is trending younger [2, 6, 7].

Hand function impairment generally refers to a pathological condition in which the movement, sensation, or coordination of the hand is compromised due to injury, disease, or other etiologies [8], [9]. This not only affects physical function but also has psychological and social consequences [10–12]. Therefore, timely and scientifically grounded rehabilitation therapy and training are essential [13]. Although existing hand rehabilitation devices can be effective, most are costly and difficult to implement in home settings.

Given these limitations, developing a low-cost yet effective home-based hand rehabilitation system is of considerable significance. To reduce cost, many existing home systems employ sensors such as cameras for hand tracking and human-computer interaction. The development of such systems primarily involves three components: gesture recognition, hand rehabilitation assessment, and rehabilitation training.

Firstly, regarding the gesture recognition part, as mentioned in [14], Ahmad, K. A., et al. employed Long Short-Term Memory Network (LSTM) for gesture recognition. This network is capable of effectively learning the dependencies of each joint point of the hand in the time series. However, when modeling long sequences, there will be a situation of information forgetting, and the recognition accuracy in complex environments is limited. In [15], Tran, D. S., et al. used 3D Convolutional Neural Network (3D-CNN) for dynamic gesture recognition. Although this model can dynamically recognize spatial-time features, it requires high computational and memory resources. In [16], Liu, Q., et al. used Extended and Supplementary Displacement Multi-scale G3D (ESS MS-G3D) neural network for action recognition. This neural network can effectively capture the subtle patterns of actions through multi-scale spatio-temporal convolution, but this algorithm cannot capture key features, which limits its efficiency in action recognition. In [17], Chen, Z., et al. achieved action recognition through multi-scale spatiotemporal residual graph convolution. They proposed a multi-input branch architecture based on graph convolutional networks, inputting joint points and skeletal information into the network separately and extracting features respectively. This innovation can enhance computational efficiency and feature extraction capability. This algorithm has extremely high

DOI: 10.13164/re.2026.0195

accuracy in action recognition, which inspires us to make improvements to it by introducing an enhanced spatio-temporal attention mechanism, and applying it to high-precision gesture recognition.

The next topic to be discussed is the issue of hand rehabilitation assessment. Currently, many scholars have proposed numerous representative assessment methods. In [18], Yu, H., et al. used exoskeleton equipment to evaluate hand health by recording the torque and angles of the hand joints during bending and extension. In [19], Fang, Q., et al. used a new type of infrared imaging equipment to assess hand health by measuring the angles of multiple hand joints. In Reference [20], Li, C. G., et al. achieved hand health assessment through three consecutive specific actions. However, since the normal operation of daily life largely depends on the coordinated movements of various joints, and the above assessment methods only focus on individual joints without considering the coordination relationship between the overall joints, this limits the accuracy of their assessment results. This prompts us to assess the condition of the hand from both a local and an overall perspective simultaneously.

Finally, there is also hand rehabilitation training. Currently, there are many methods for hand rehabilitation training. In [21], Nath, D., et al. designed variable resistance joystick training. This training can be dynamically adjusted, providing patients with rehabilitation training suitable for their level. However, in the later stages of rehabilitation, higher resistance training will significantly increase the fatigue level of patients and reduce the training efficiency. In [22], Gherman, B., et al. comprehensively studied the development of hand rehabilitation robots. Most of them adopt exoskeleton design to achieve assisted training. Although this can provide patients with high-dose and repetitive training, most of the equipment is large in size or expensive. In [23], Jha, C. K., et al. designed a glove-type virtual hand rehabilitation system. By adding virtual scenes, the training process becomes more interesting. However, this system cannot dynamically adjust the hand rehabilitation plan according to the patient's rehabilitation progress, which affects the rehabilitation effect. This inspires us to adopt a cheap, compact and dynamically adjustable training method.

Although the aforementioned studies have made significant progress in gesture recognition, hand rehabilitation assessment, and rehabilitation training, there are still some shortcomings. Firstly, in the aspect of gesture recognition, most models require a large amount of computing resources and cannot focus on key features. Secondly, in the aspect of hand rehabilitation assessment, most assessment strategies ignore the evaluation of the dynamic integration and coordination mechanism among joints, thereby limiting the authenticity and reference value of the assessment results. Finally, in the aspect of hand rehabilitation training, the existing training schemes have inherent limitations in terms of efficiency, cost, and portability, which severely restrict their popularization in the home environment.

To address the aforementioned issues, this paper has developed an airborne hand rehabilitation assessment and virtual training system. In terms of gesture recognition, a multi-scale residual graph convolutional network with an enhanced spatio-temporal attention mechanism was employed for recognition. Through the enhanced attention mechanism, the efficiency of feature learning was improved, and the residual connection was used to enhance the smooth transmission of feature information. Leap Motion was utilized to collect the hand joint points to enhance the accuracy of collection. 14 representative gestures were selected for quantitative rehabilitation assessment, achieving both local and overall evaluations. During the rehabilitation training phase, two modes were adopted to train the activity and flexibility of both hands, and the training focus was dynamically adjusted based on the assessment results of the hand rehabilitation status.

Our main contributions can be summarized as follows:

- In the gesture recognition algorithm, by introducing an enhanced attention mechanism, it is possible to automatically focus on the key spatiotemporal features, thereby enhancing the sensitivity to important gesture features. Finally, through residual connections, not only has the perception ability of the model been expanded, but also its ability to recognize complex gestures has been enhanced.
- In terms of hand rehabilitation assessment: Through 14 representative movements, a comprehensive evaluation of local joints and overall coordination mechanisms was achieved. Additionally, by using the ridge regression method, weights were added to each movement.
- In terms of hand rehabilitation training, a variety of interesting training methods were adopted to enhance the flexibility and mobility of the hands. Based on the previous assessment results, the training focus will be dynamically adjusted to improve the specificity of the training.
- The innovative adoption of the Leap Motion imaging device enables the collection of gesture joint data. Compared to ordinary devices, its recognition accuracy is higher (reaching millimeter-level), and the latency is lower (less than 10 milliseconds).

The remainder of this paper is structured as follows. Section 2 introduces the data input architecture based on graph convolution and a spatio-temporal attention mechanism. Section 3 presents the multi-scale residual graph convolutional network. Section 4 proposes a quantitative assessment model for hand rehabilitation. Section 5 describes two hand rehabilitation training methods. Section 6 details the system testing and presents the experimental results. Finally, Section 7 provides the conclusion and acknowledgments.

# 2. Data Input Architecture Based on Graph Convolution and Enhanced Spatio-Temporal Attention Mechanism

To balance the accuracy of recognition and the efficiency of computation, this paper employs a data input architecture based on graph convolution and enhanced spatio-temporal attention mechanism. The core of this design lies in first calculating the hand bone data through the input hand joint data, and then using graph convolution and enhanced attention mechanism to enhance and extract the decisive features of the hand joint and hand bone data, and to achieve their fusion. This design enables the model to significantly reduce the computational complexity while still maintaining excellent recognition performance.

## 2.1 Data Architecture Division

This paper divides the input data into two parts: (a) hand joint data and (b) hand bone data. Firstly, the three-dimensional spatial coordinate information of hand joints $v_{i,t} = (x_{i,t}, y_{i,t}, z_{i,t})$, $i = 1,2,\ldots,I$, $t = 1,2,\ldots,T$ is obtained using Leap Motion. Here, $v_{i,t}$ represents the $i$-th hand joint in the $t$-th frame, and $x_{i,t}$, $y_{i,t}$, $z_{i,t}$ respectively represent the horizontal coordinate, vertical coordinate, and depth coordinate.

The hand bone data includes the lengths of hand bones $L = \{l_{(i,i+1),t}\}$, where $l_{(i,i+1),t}$ represents the length of the hand bone between the $i$-th joint and the ($i$+1)-th joint in the $t$-th frame, and the hand joint angle information $\mathrm{Ang} = \{ a_{v_i,t} \}$, indicating the joint angle of the $i$-th joint in the $t$-th frame. The calculation formula for the hand bone length information $l_{(i,i+1),t}$ is given by:

$$l_{(i,i+1),t} = \sqrt{(x_{i+1,t} - x_{i,t})^2 + (y_{i+1,t} - y_{i,t})^2 + (z_{i+1,t} - z_{i,t})^2}. \quad (1)$$

The formula for calculating the joint angle information $a_{v_i,t}$ of the hand is given by:

$$\begin{cases} a_{v_i,t} = \cos^{-1} \dfrac{\mathbf{l}_{(i,i+1),t} \cdot \mathbf{l}_{(i-1,i),t}}{l_{(i,i+1),t} \cdot l_{(i-1,i),t}}, \\ \mathbf{l}_{(i,i+1),t} = (x_{i+1,t} - x_{i,t}, y_{i+1,t} - y_{i,t}, z_{i+1,t} - z_{i,t}) \end{cases} \quad (2)$$

where $\mathbf{l}_{(i,i+1),t}$ represents the bone vector between the $i$-th and ($i$+1)-th joint point in the $t$-th frame.

## 2.2 Graph Convolutional Network

This paper constructs a graph structure $\mathbf{G} = (\mathbf{K}, \mathbf{S})$ based on the connections between hand joint nodes. $\mathbf{K}$ represents the set of hand joint nodes, and the elements within the set are denoted as $k \in \mathbf{K}$; $\mathbf{S}$ represents the set of natural connections of hand joint nodes. The elements within this set are denoted as $s \in \mathbf{S}$. The relationships between hand joint nodes are represented by the adjacency matrix $\mathbf{A} \in \mathbf{R}^m$, where $m$ corresponds to the order. The definition of the adjacency matrix $\mathbf{A}$ is given by:

$$\mathbf{A} = \begin{bmatrix} a_{1,1} & \cdots & a_{1,m} \\ \vdots & \ddots & \vdots \\ a_{m,1} & \cdots & a_{m,m} \end{bmatrix}, a_{ij} = \begin{cases} 1, & k_i \in s_j, \\ 0, & \text{else}. \end{cases} \quad (3)$$

In the spatial-temporal domain, the calculation formulas for the outputs $f_{\text{out}}^{\text{spa}}(\cdot)$ and $f_{\text{out}}^{\text{tim}}(\cdot)$ of the graph convolution are given by:

$$\begin{cases} f_{\text{out}}^{\text{spa}}(v_{i,t}) = \sum_{d=0}^{D} \mathbf{w}_d f_{\text{in}}^{\text{spa}}((\mathbf{\Lambda}_d^{\text{spa}})^{-\frac{1}{2}} \mathbf{A}_d^{\text{spa}} (\mathbf{\Lambda}_d^{\text{spa}})^{-\frac{1}{2}} \odot \mathbf{M}_d), \\ f_{\text{out}}^{\text{tim}}(v_{i,t}) = \sum_{d=0}^{D} \mathbf{w}_d f_{\text{in}}^{\text{tim}}((\mathbf{\Lambda}_d^{\text{tim}})^{-\frac{1}{2}} \mathbf{A}_d^{\text{tim}} (\mathbf{\Lambda}_d^{\text{tim}})^{-\frac{1}{2}} \odot \mathbf{M}_d) \end{cases} \quad (4)$$

where $f_{\text{in}}^{\text{spa}}(\cdot)$ and $f_{\text{in}}^{\text{tim}}(\cdot)$ represent the input features in the spatial domain and the temporal domain, $D$ represents the spatial range in the neighborhood map, $\mathbf{w}_d$ and $\mathbf{M}_d$ respectively represent the convolution operation and weights, $\mathbf{A}_d^{\text{spa}}$ and $\mathbf{A}_d^{\text{tim}}$ represent the $d$-order spatial and temporal adjacency matrixes, $\mathbf{\Lambda}_d^{\text{spa}}$ and $\mathbf{\Lambda}_d^{\text{tim}}$ represent the normalized matrixes, and $\odot$ represents the point-wise multi-plication operation.

## 2.3 Enhanced Spatio-Temporal Attention Mechanism

After completing the aforementioned data processing and graph convolution operations, the formed hand feature maps will be input into the enhanced spatiotemporal attention module to more accurately capture the core features of the actions, ignoring the influence of irrelevant noise and low-importance feature information on gesture recognition, thereby significantly improving the accuracy and robustness of gesture recognition.

The processing flow of the enhanced spatio-temporal attention module is as follows:

Firstly, the input features are respectively subjected to average pooling in the spatial and temporal dimensions for information aggregation. The spatial domain and temporal domain average pooling operations $\mathrm{pool}_{\text{spa}}(\cdot)$ and $\mathrm{pool}_{\text{tim}}(\cdot)$ are given by:

$$\begin{cases} \mathrm{pool}_{\text{spa}}(\cdot) = \mathrm{pool}_{\text{spa}}(f_{\text{in}}), \\ \mathrm{pool}_{\text{tim}}(\cdot) = \mathrm{pool}_{\text{tim}}(f_{\text{in}}). \end{cases} \quad (5)$$

Then, the two obtained feature vectors are concatenated and sent to a fully connected layer to achieve feature dimension reduction. The output $O(\cdot)$ of the fully connected layer is given by:

$$O(\cdot) = \theta(\mathbf{W}_1 \cdot \mathrm{pool}_{\text{spa}}(\cdot) + b_1 \oplus \mathbf{W}_2 \cdot \mathrm{pool}_{\text{tim}}(\cdot) + b_2) \quad (6)$$

where $\mathbf{W}_1$, $\mathbf{W}_2$ represent the learnable weight matrices, $\theta(\cdot)$ represents the Sigmoid activation function, $b_1$, $b_2$ represent

the learnable bias terms, $\oplus$ respectively represents the element-wise addition.

Subsequently, through two independent fully connected layers, the attention weights in the spatial and temporal feature channels are deeply analyzed separately. Then, the spatio-temporal attention matrix $\mathbf{W}_{\text{spa-tim}}$ is given by:

$$\mathbf{W}_{\text{spa-tim}} = \sigma(\mathbf{W}_3 \cdot O(\cdot) + b_3) \otimes \sigma(\mathbf{W}_4 \cdot O(\cdot) + b_4) \quad (7)$$

where $\mathbf{W}_3$, $\mathbf{W}_4$ represent the learnable weight matrices, $\sigma(\cdot)$ represents the HardSwish activation function, $b_3$, $b_4$ represent the learnable bias terms, $\otimes$ represents the outer product.

Finally, these two sets of weights are integrated through the outer product operation and applied to the node channels. The resulting output is the attention score representing the entire sequence information. The resulting output $f_{\text{out}}$ is given by:

$$f_{\text{out}} = f_{\text{in}} \odot \mathbf{W}_{\text{spa-tim}}. \quad (8)$$

# 3. Multi-Scale Residual Graph Convolutional Neural Network

The main body of this network is composed of five cascaded multi-scale residual graph convolutional modules. Each module performs alternating spatial and temporal graph convolution and extracts and fuses temporal and spatial features from multiple scales. Subsequently, the network integrates the aggregated spatiotemporal information using a global average pooling layer, and finally outputs the result of action recognition through a fully connected layer and a Soft-max layer.

Specifically, the spatio-temporal features input to this part are $\mathbf{X} \in \mathbf{R}^{C \times T \times K}$, where $C$ represents the number of channels, $T$ represents the length of the time series, and $K$ represents the number of hand joint points. This paper divides the gesture completion process into five stages: the starting stage, the acceleration stage, the peak stage, the stable stage, and the ending stage. The input spatio-temporal features are divided into 5 parts, with each part's input feature dimension being $C/5 \times T \times K$, and feature learning is completed successively. Each part consists of a spatial graph convolution and a time graph convolution, and is called a sub-multi-scale residual graph convolution module. In the part of spatio-temporal graph convolution, formula (4) is still used for calculation. In addition, the network also introduces cross-stage residual connections. This design not only expands the model's receptive field but also enhances the modeling ability for the local and overall dependency relationships between joints. The output of the $i$-th sub-multiplescale residual graph convolution module $f_1^{\text{st}}$ is given by:

$$f_i^{\text{st}} = \begin{cases} \text{conv}_{\text{st}}(x_i), & i = 1, \\ \text{conv}_{\text{st}}(x_i + f_{i-1}^{\text{st}}), & i > 1 \end{cases} \quad (9)$$

where $\text{conv}_{\text{st}}(\cdot)$ denotes a single sub-multiplescale residual graph convolution module, and $x_i \in \mathbf{X}$ represents the input features.

In order to better study the characteristics of the hand at different time scales (long-term and short-term), this paper uses 1×1, 5×1, and 7×1 convolution kernels to process and extract the input hand features. It separately extracts the hand motion features at single frame, 5 frames, and 7 frames. After each extraction, a 3×1 maximum pooling operation is performed on the extracted features to refine and compress them. Finally, the extracted results of the features at different time scales after processing are combined to obtain the output. This enables the model to combine different time scales for recognition and improve the recognition accuracy.

In this module, the global average pooling operation is first used to aggregate information from $f_1^{\text{st}}$, receiving feature information from a 5-hop neighborhood, thereby establishing a connection from fine details to the macroscopic level between finger joint points, palm and wrist. This mechanism enables the equivalent receptive field of the terminal segments to be significantly expanded. Subsequently, all segments are concatenated and stabilized during training and promoted to converge by means of a cross-module residual connection. The final output of this module $f$ is given by:

$$f = \theta([f_1^{\text{st}}, f_2^{\text{st}}, \cdots, f_5^{\text{st}}] + \mathbf{X}) \quad (10)$$

where the output features $f$ are processed through the Sigmoid activation function $\theta(\cdot)$, which can capture the spatio-temporal features between nodes of different distances. This representation enables the model to simultaneously learn the short-term dynamics between joints and the long-term spatiotemporal dependencies, and furthermore, to improve the recognition accuracy of the model.

Figure 1 shows the structure diagram of the gesture recognition model.

# 4. Quantitative Assessment Model for Hand Rehabilitation

In this model, this paper achieves the quantitative assessment of hand rehabilitation through two methods, namely the hand joint angle calculation model and the gesture completion degree assessment model.

## 4.1 Hand Joint Angle Calculation Model

In order to enable the hand to function properly for daily life and work, the functional range of motion of the wrist, metacarpal joints, interphalangeal joints, and the thumb should reach their basic requirements as follows: flexion of 45°, proximal flexion of 90°, interphalangeal extension, etc. [24].

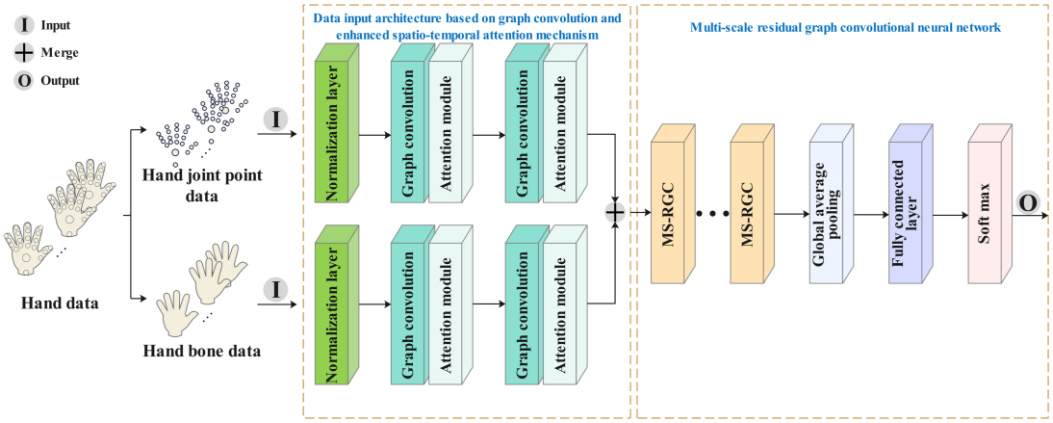



**Fig. 1.** Hand gesture recognition model structure diagram.

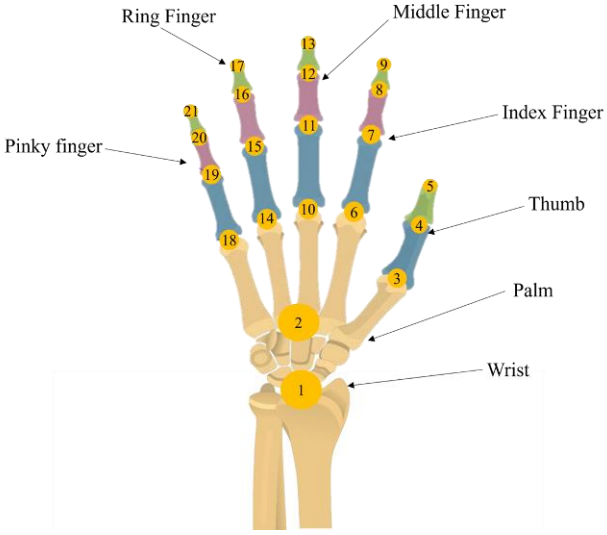
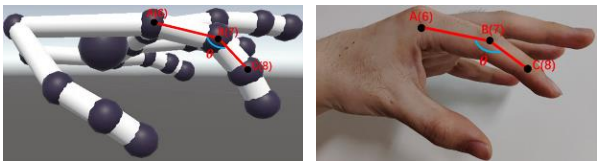
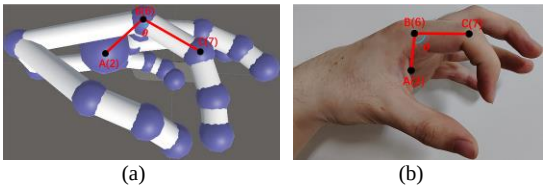



**Fig. 2.** Hand joint marker diagram.



(a) (b)

**Fig. 3.** Diagram showing the flexion angle of the proximal interphalangeal joint.



(a) (b)

**Fig. 4.** Diagram showing the flexion angle of the metacarpophalangeal joints.

Take the left hand as an example. The joint labels of the hand are shown in Fig. 2.

a) Measurement of The Flexion Angle of The Proximal Interphalangeal Joint: Taking the index finger as an example, the three-dimensional positions of the metacarpophalangeal joint ⑥, the proximal phalangeal joint ⑦, and the distal phalangeal joint ⑧ of the index finger is required, as shown in Fig. 3.

The calculation formula for angle $\theta$ is given by:

$$\theta = \cos^{-1}\frac{\overrightarrow{\mathbf{BA}}\bullet\overrightarrow{\mathbf{BC}}}{\left|\overrightarrow{\mathbf{BA}}\right|\bullet\left|\overrightarrow{\mathbf{BC}}\right|} \tag{11}$$

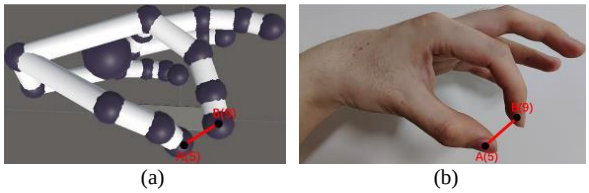
where **A**, **B**, and **C** respectively represent the three-dimensional coordinate information of joints ⑥, ⑦, and ⑨. The scoring formula for the patient's proximal interphalangeal joint flexion of $\text{Score}_\theta$ is given by:

$$\text{Score}_\theta = (\frac{\theta}{\theta_{\max}})\times 100 \tag{12}$$

where $\theta$ represents the actual flexion angle of the proximal interphalangeal joint of the patient, while $\theta_{\max}$ represents the maximum theoretical flexion angle of the proximal interphalangeal joint.

b) Measurement of The Flexion Angle of The Metacarpophalangeal Joints: Taking the index finger as an example, the three-dimensional positions of the palm ②, the metacarpophalangeal joint of the index finger ⑥, and the proximal phalangeal joint of the index finger ⑦ need to be obtained, as shown in Fig. 4.

The three-dimensional positions of the palm ②, the metacarpophalangeal joint of the index finger ⑥, and the proximal phalangeal joint of the index finger ⑦ need to be respectively taken as **A**, **B**, and **C** and substituted into (11) to calculate $\theta$. The score for the flexion of the metacarpophalangeal joint can be calculated through (12).



(a) (b)

**Fig. 5.** Illustration of the feature points for the thumb-to-thumb interlocking function.

c) Measurement of The Distance Between The Thumb and The Index Finger: It is necessary to obtain the three-dimensional position information of the thumb fingertip joint ⑤ and the index finger fingertip joint ⑨, as shown in Fig. 5.

The Euclidean distance is selected as the similarity calculation method for detecting the interphalangeal function of the thumb. Assuming that the three-dimensional position information of the thumb fingertip joint ⑤ is $\mathbf{A}(x_1, y_1, z_1)$, and the position information of the index finger fingertip joint is also $\mathbf{B}(x_2, y_2, z_2)$, then the distance between the fingertip joints of the thumb and the index finger can be calculated through the Euclidean distance calculation formula. The calculating formula for the distance $d$ is given by:

$$d = \sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2 + (z_1 - z_2)^2}\ . \tag{13}$$

The scoring formula for the patient's thumb apposition function $\text{Score}_\text{d}$ is given by:

$$\text{Score}_\text{d} = 100 - (\frac{d}{d_{\max}}) \times 100 \tag{14}$$

where $d$ represents the actual thumb-to-thumb distance measured for the patient, while $d_{\max}$ represents the maximum theoretical thumb-to-thumb distance.

Finally, the total score of the patient's hand joint angles can be obtained by integrating the scores of the proximal interphalangeal joint flexion of the hand, the metacarpophalangeal joint flexion, and the thumb opposition function. The calculation formula for the total score $\text{Score}_\text{t}$ is given by:

$$\text{Score}_\text{t} = \frac{\text{Score}_{\theta_1} + \text{Score}_{\theta_2} + \text{Score}_\text{d}}{3} \tag{15}$$

where $\text{Score}_{\theta_1}$ and $\text{Score}_{\theta_2}$ represent the scores for flexion of the proximal interphalangeal joints of the hand and the flexion of the metacarpophalangeal joints, respectively.

## 4.2 Gesture Completion Degree Evaluation Model

This paper investigates the range, accuracy and speed of hand movements by observing the process of performing specific actions with the hands. This helps to assess the overall coordination mechanism of hand movements. Referring to the research on hand rehabilitation assessment gestures [25], this paper selected 14 actions as the basis for the study, including the detection of finger flexion and extension functions as well as wrist movement detection, etc. The specific actions are shown in Tab. 1.

This article applies these quantitative data to conduct a quantitative assessment of the patients' recovery progress. Based on the data-driven quantitative assessment method, it can optimize the rehabilitation process, guide patients to recover hand function more effectively, and provide a database path for patients by quantitatively feedback their recovery status. This helps patients maximize their reconstruction of hand mobility. The specific calculation formulas for each indicator score are as follows.

a) Hand Movement Range: By utilizing Leap Motion technology to track the key points of the hand, the maximum and minimum ranges of hand movements can be calculated. At the same time, the flexibility and movement ability of the hand joints can also be evaluated. The calculation formula for this range $R$ is given by:

$$R = \max(L_i - L_{i-1}) - \min(L_i - L_{i-1}) \tag{16}$$

where $L_i$ and $L_{i-1}$ represent the positions of the first joint at a certain moment and its position at the previous moment.

The scoring formula for the hand's range of motion $\text{Score}_\text{R}$ is given by:

$$\text{Score}_\text{R} = (\frac{R}{R_{\max}}) \times 100 \tag{17}$$

where $R_{\max}$ represents the theoretically maximum movement range of the hand.

b) Accuracy of Gestures: The positions of the detected key points are compared with the standard key point positions of the preset action, and the average deviation is calculated. The precision is used to represent the average value of the difference between the actual positions of all key points and the standard positions. The specific calculation formula for precision $A$ is given by:

$$A = \frac{1}{n}\sum_{i=1}^{n} \left| P_i - S_i \right| \tag{18}$$

where $P_i$ represents the actual position of the $i$-th joint point, while $S_i$ is the standard position of that joint point and $n$ represents the total number of key points on the hand.

The scoring formula for the accuracy of $\text{Score}_\text{A}$ is given by:

$$\text{Score}_\text{A} = \begin{cases} 0, & A_{\max} < A, \\ 100 - (\frac{A}{A_{\max}}) \times 100, & A_{\max} \geq A \end{cases} \tag{19}$$

where $A_{\max}$ represents the ideal maximum error range.

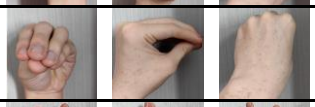
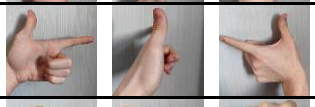
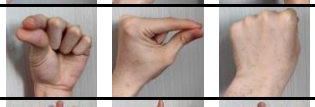
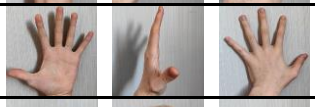
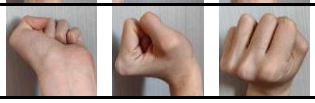
c) Speed of Gestures: Measure the movement rate of key points between adjacent frames to evaluate the speed and smoothness of hand movements. The formula for calculating speed $V$ is given by:

$$V = \frac{L_i - L_{i-1}}{\Delta t} \tag{20}$$

where $\Delta t$ represents the time difference between consecutive frames. The calculation formula for the accuracy score $\text{Score}_\text{V}$ is given by:

$$\text{Score}_\text{V} = (\frac{V}{V_{\max}}) \times 100 \tag{21}$$

where $V_{\max}$ represents the ideal average maximum speed.

| Serial Number | Gesture | Name |
|---|---|---|
| 1 |  | Fist Clench |
| 2 |  | Confirmation Gesture |
| 3 |  | Hook Hand Gesture |
| 4 |  | Pinch Gesture |
| 5 |  | Scissors Gesture |
| 6 |  | Palm Gesture |
| 7 |  | Eight Gesture |
| 8 |  | Two-Finger Pinch Gesture |
| 9 |  | Five-Finger Spread Gesture |
| 10 |  | Thumb Left Extension Gesture |
| 11 |  | Six Gesture |
| 12 |  | One Gesture |
| 13 |  | Wrist Flexion |
| 14 |  | Wrist Extension |

**Tab. 1.** Evaluation of gesture indication.

The average score of these three scores is the score for this gesture. The calculation formula for the comprehensive score $\text{Score}_f$ of a single gesture is given by:

$$\text{Score}_f = \frac{\text{Score}_R + \text{Score}_A + \text{Score}_V}{3}. \tag{22}$$

## 4.3 Weight Calculation Model

This paper assigns weights to each gesture and sums them up to obtain the final hand rehabilitation score. The calculation formula for the comprehensive score Score of the patient's hand is given by:

$$\text{Score} = \sum_{i=1}^{14} w_i \cdot (\text{Score}_f + \text{Score}_t) \tag{23}$$

where $w_1,\ldots,w_{14}$ represent the weights corresponding to the gestures.

In order to initially make the quantitative evaluation scores obtained in this paper have medical reference value and certain feasibility in practical application, the ridge

regression method was adopted in this study. The weight coefficients of each indicator in (23) were calculated.

a) Expression Definition:

$$\begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1m} \\ \vdots & \vdots & \ddots & \vdots \\ x_{n1} & x_{n2} & \cdots & x_{nm} \end{bmatrix} \times \begin{bmatrix} \beta_1 \\ \vdots \\ \beta_n \end{bmatrix} + \begin{bmatrix} \xi_1 \\ \vdots \\ \xi_n \end{bmatrix} \quad (24)$$

where $y_n$ represents the $n$-th comprehensive score in the actual assessment. $x_{nm}$ is the $n$-th scoring result of the $m$-th indicator in the actual assessment. $\xi_n$ is the random error term. $\beta_n$ is the regression coefficient. The matrix containing the independent variable $x_{nm}$ is defined as the design matrix $\mathbf{X}$.

b) Loss Function Definition: The loss function of this model is the sum of the squared residuals between the predicted value and the actual value, plus the regularization term. The regularization term is the L2 norm (sum of squares) of the regression coefficients. Therefore, the formula for the ridge regression loss function $L(\beta)$ of this model is given by:

$$L(\beta) = \sum_{i=1}^{n} (y_i - \sum_{j=1}^{m} \beta_j x_{ij})^2 + \lambda \sum_{j=1}^{m} \beta_j^2 \quad (25)$$

where $n$ represents the number of samples, $m$ represents the number of features, $y_i$ represents the $i$-th comprehensive score in the actual assessment, $x_{ij}$ represents the $j$-th feature value of the $i$-th sample, $\beta_j$ represents the weight coefficient of the $j$-th feature, and $\lambda$ represents the regularization parameter.

c) Weight Coefficient Estimation: The calculation formula for the weight coefficient matrix $\boldsymbol{\beta}$ is given by:

$$\boldsymbol{\beta} = (\mathbf{X}^{\mathrm{T}}\mathbf{X} + \lambda \mathbf{I})^{-1}\mathbf{X}^{\mathrm{T}}\mathbf{y} \quad (26)$$

where $\mathbf{y}$ represents the target value vector and $\mathbf{I}$ is the identity matrix.

## 5. Hand Rehabilitation Training

The training games designed in this paper are all developed based on the Unity3D game engine. The non-contact camera device Leap Motion is used to collect the joint information of the user's hands, in order to enable the user to control the virtual hand in the training game through actual hand movements. Figure 6 shows the hardware equipment required for this training system.

Before undergoing hand rehabilitation training, patients need to first complete the assessment of their hand rehabilitation status. In this system, to reduce the complexity of hand rehabilitation assessment, the hand joint angle calculation system is integrated into the hand activity assessment system. The finger function assessment is incorporated into gestures such as Thumb-Index Opposition Pinch. The two joint angle flexion assessment is integrated into gestures such as Finger Hook Posture, Full Hand Grip, and Full Palm Extension. Users only need to complete 14

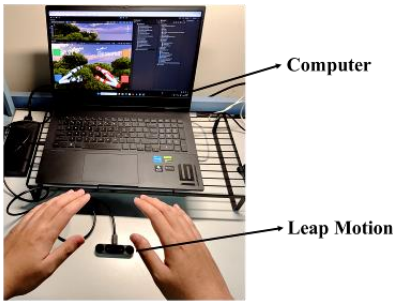
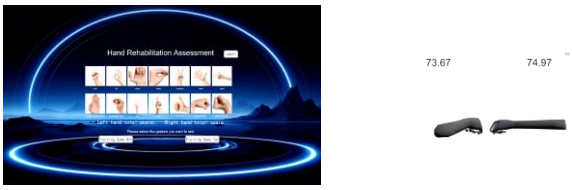
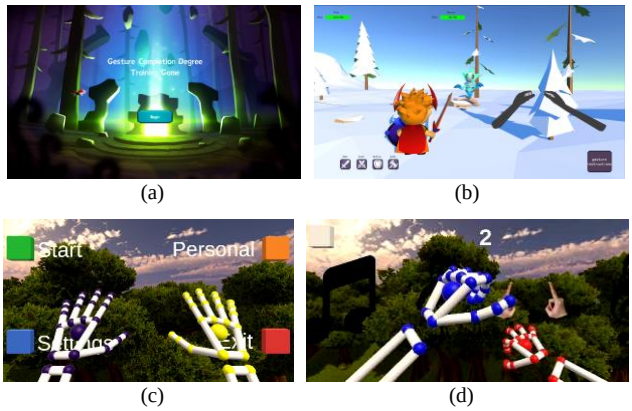



**Fig. 6.** Required hardware equipment.




(a) (b)

**Fig. 7.** Schematic diagram of the hand rehabilitation assessment system.




(a) (b) (c) (d)

**Fig. 8.** Schematic diagram of the hand rehabilitation training system.

gestures respectively to obtain the hand rehabilitation assessment score. Figure 7 shows the schematic diagram of the operation interface for hand rehabilitation assessment.

There are a total of 14 gestures on this interface. Patients need to complete the assessment of these 14 gestures in sequence to obtain the quantitative comprehensive score of the current hand rehabilitation status.

After completing the hand rehabilitation score, patients can click the two training game buttons below Fig. 7a and enter the hand rehabilitation training interface. Figure 8 shows the operation interfaces of two hand rehabilitation training games.

These games are divided into two types. Figure 8a and Figure 8b are training games for hand activity, while Figure 8c and Figure 8d are training games for hand flexibility. Both of these training games are based on 14 gestures for training, and according to the previous quantitative assessment results of hand rehabilitation, the training frequency for each gesture is dynamically adjusted to achieve targeted training.

### 5.1 Hand Mobility Training

This training game allows users to control the character by performing specific gestures to activate skills, thereby engaging in battles against enemies. The game has a total of four skills, each corresponding to three gestures. The gestures and skills are dynamically adjusted based on the results of hand rehabilitation assessment. The default gesture-to-skill correspondence is as shown in Tab. 2.

In this game scenario, each skill is associated with a button. Users complete three specific actions in sequence instead of clicking the buttons. During the game training process, the interactive attributes of each skill and its activation attributes at different levels are first evaluated to determine if the skill is available. Once a user completes the first action of an available skill, all other skills will be set to unavailable status to avoid the influence of other skills. For each completed action, an IEnumerator loop mode will be initiated, generating a 5-second interval to provide users with sufficient time to complete the next action. When users complete these three actions in sequence, the system will combine the hand movement assessment system and calculate the fluctuation of the skill effect based on the average score of these three actions. Then, the corresponding Invoke method of the skill button will be called to activate the corresponding skill logic. If users want to improve the skill effect, they need to make more standardized gesture actions to achieve the purpose of training hand movement.

### 5.2 Hand Dexterity Training

This training game comprises three scenes: the title screen, the training screen, and the evaluation screen.

On the title screen, four buttons are configured with the PhysicalHandsButton script from the Ultraleap plugin. These buttons utilize Leap Motion to capture hand data, thereby controlling the CapsuleHands hand model prefab from the same plugin. Clicking a button triggers its corresponding ButtonPressedEvent.

In the training scene, once a countdown—implemented using an IEnumerator coroutine—concludes, a musical staff containing notes and gesture icons scrolls from right to left. As a gesture icon approaches the center of the screen, its corresponding gesture detector is activated; it is deactivated as the icon moves away. During this active window, the user must perform the correct gesture to achieve recognition. A combo counter at the top of the screen tracks consecutive successful recognitions. The Update method continuously polls the AudioSource state; when the audio finishes, the scene transitions to the evaluation screen.

The evaluation screen displays data preserved from the training scene via the DontDestroyOnLoad method. The highest historical score and the last three training records are saved locally using PlayerPrefs. This interface effectively visualizes the user's training progress. To achieve a higher score, users must perform more precise gestures within the allotted time, thereby practicing and improving their gestural agility.

| | Gesture One | Gesture Two | Gesture Three |
|---|---|---|---|
| Skill One | Full Hand Grip | Finger Hook Posture | Thumb Radial Extension |
| Skill Two | Full Finger Extension & Abduction | Thumb-Little Finger Extension | Thumb-Single Finger Fine Pinch |
| Skill Three | Thumb-Index Opposition Pinch | Multi-Digit Pinch | Full Palm Extension |
| Skill Four | Isolated Index Finger Extension | Index-Middle Finger Extension & Abduction | Thumb-Index Right-Angle Opposition |

**Tab. 2.** Default skill gesture.

## 6. System Testing and Result Analysis

### 6.1 Test Details

This paper uses the PyTorch framework for training. The input data adopts the intra-frame normalization strategy. During the training process, the batch size is set to 64, the weight decay is set to 0.0005 to prevent overfitting of the model. The number of training epochs is set to 25, and the learning rate is reduced to one-tenth of the original in the 6th and 8th training stages to avoid over-training. The optimizer used for training is Adam (Adaptive Moment Estimation), with an initial learning rate of 0.0008. The loss function uses the cross-entropy loss function with label smoothing (label smoothing = 0.05), and the global average pooling layer uses the average pooling function. The structure of some convolutional blocks in the network is shown in Tab. 3.

### 6.2 Gesture Recognition System Test

The data used for model training was initially collected from actual patients with different hand health conditions. For each gesture, a total of 1,000 sets of data were collected, with each set containing the three-dimensional

| Number | Module name | Input channels | Output channels | Step length |
|---|---|---|---|---|
| 1 | Input projection convolution | 3 | 32 | 1 |
| 2 | Normalization layer | 32 | 32 | 1 |
| 3 | Graph convolution layer | 32 | 32 | 1 |
| 4 | Attention layer | 32 | 32 | 1 |
| 5 | Multi-scale residual graph convolutional | 32 | 32 | 1 |
| 6 | Global adaptive average pooling | 32 | 32 | 1 |
| 7 | Fully connected classifier | 32 | 14 | 1 |

**Tab. 3.** Some convolutional block structures in the network.

spatial coordinates of the hand joint points from the beginning to the end of the gesture. In total, there were 14,000 sets of data. 70% of all the data was used as the training set to train the gesture recognition algorithm. The performance of the algorithm was reflected by the accuracy of gesture recognition.

To ensure the rigor of the model evaluation, we adopted a subject-level data partitioning strategy to conduct subject-independent tests. All patients were randomly divided into two groups: 70% of the patients and all 9,800 sets of data they generated formed the training set, which was used for the training of the gesture recognition algorithm; The remaining 30% of patients and all 4,200 sets of data they generated constitute an independent test set, specifically designed to evaluate the performance of the algorithm. This method completely eliminates the possibility of data leakage of the same patient between the training and test sets, ensuring that the accuracy metric can truly reflect the effectiveness of the algorithm when applied to new patients.

Figure 9 shows the accuracy of each gesture recognition, with an average accuracy of up to 94.86%. The recognition accuracy for the selected gestures is very high. Figure 10 shows the confusion matrix for gesture recognition.

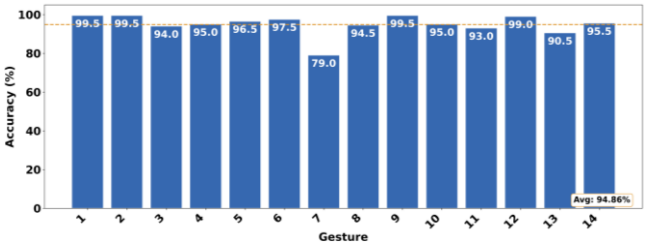



**Fig. 9.** Graph showing the accuracy of each gesture recognition method.

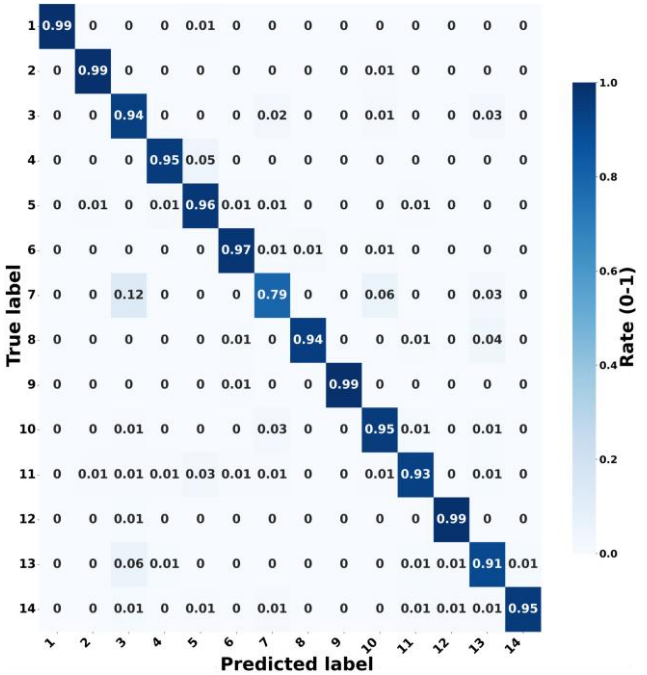



**Fig. 10.** Gesture recognition confusion matrix.

## 6.3 Ablation Experiments

a) Data architecture ablation experiment

The data architecture adopted in this paper includes hand joint data and hand bone data. This experiment compares the recognition accuracy of the results presented in this paper with that of using only hand joint data or hand bone data separately. Based on the gesture recognition algorithm designed in this paper, only the input data type was modified, and the experiment was conducted on the previously collected 14,000 gesture data set. The experimental results are shown in Tab. 4.

As can be seen from Tab. 4, compared with using either hand joint data or hand bone data alone, the recognition accuracy is the highest when both types of data are used simultaneously, increasing by 1.93% and 3.54% respectively. By simultaneously using the data of hand joints and hand bones as the input data for the model, it is possible to effectively overcome the shortcomings that exist when using a single data input, enhance the accuracy and robustness of gesture recognition, and ensure that this algorithm still has a high recognition accuracy for patients with poor hand conditions.

b) Attention mechanism ablation experiment

Based on the original Multi-Scale Residual Graph Convolutional Neural Network, this paper introduces the spatio-temporal attention mechanism. In this experiment, on the basis of the previously collected 14,000 sets of gesture data, the gesture recognition accuracy of the Multi-Scale Residual Graph Convolutional Neural Network with and without the attention mechanism is compared. The experimental results are shown in Tab. 5.

As can be seen from Tab. 5, introducing an enhanced attention mechanism in the traditional multi-scale residual graph convolution can effectively improve the recognition accuracy of gestures by 2.57%. By introducing the enhanced attention mechanism, it can effectively enhance the model's focus on important feature information, improving the accuracy and robustness of gesture recognition, enabling the model to maintain high recognition accuracy even when some secondary input features are missing.

| Input | Recognition accuracy |
|---|---|
| Hand joint data | 92.93% |
| Hand bone data | 91.32% |
| Hand joint data and hand bone data | 94.86% |

**Tab. 4.** Results of the data architecture ablation experiment.

| Attention mechanism configuration | Recognition accuracy |
|---|---|
| Not included | 92.29% |
| Included | 94.86% |

**Tab. 5.** Results of the attention mechanism ablation experiment.

| Method | Recognition accuracy |
|---|---|
| Ours | 94.86% |
| LSTM | 89.54% |
| 3D-CNN | 91.28% |
| ESS MS-G3D | 95.32% |

**Tab. 6.** Experimental results of different algorithms on the same dataset.

## 6.4 Comparative Experiments with Existing Methods

To verify the effectiveness of this algorithm, this experiment was conducted on the previously collected 14,000 gesture data set. The algorithm was compared with LSTM, 3D-CNN, and ESS MS-G3D in terms of recognition accuracy. The experimental results are presented in Tab. 6.

From Tab. 6, it can be seen that on the same gesture dataset, the recognition accuracy of the algorithm selected in this paper is higher than that of LSMT and 3D-CNN, achieving an accuracy improvement of 5.32% and 3.58% respectively. However, although its accuracy is slightly lower than ESS MS-G3D, the algorithm adopted in this paper has better training efficiency than ESS MS-G3D. The training of this algorithm only takes 5–10 minutes, while the training of ESS MS-G3D requires approximately 15 minutes. Therefore, although the accuracy of this algorithm is slightly lower than ESS MS-G3D, it has higher algorithm efficiency and lower computing power requirements, making it more suitable for home environments with limited computing power.

## 6.5 Hand Rehabilitation Assessment System Test

We synthesized the scale assessment method to obtain the data. This assessment met the criteria of blind assessment in previous studies and ensured the reliability among raters [26]. Therefore, our weight calculation model initially has a certain degree of reliability. We obtained the information of each hand joint point of each patient at each time frame through Leap Motion, and calculated 1000 sets of comprehensive scores and quantified scores of each gesture based on this.

Based on the 1000 sets of comprehensive scoring results and the quantitative evaluation results of each gesture designed by the system in this study, weighted processing was carried out for each score using ridge regression. Conduct feasibility studies and preliminary verification of the hand rehabilitation assessment system, Figure 11 and Figure 12 show the results of the ridge regression.

From Fig. 12, it can be seen that under this regression coefficient, the mean square error is very small. Thus the evaluation has accuracy and reference value.

To preliminarily verify the evaluation effect of the above system, 30 volunteers were selected. Among them, 10 had normal hand functions, 10 had incomplete hand functions, and 10 had poor hand functions. Each person completed 10 evaluations for both the left and right hands, totaling 600 evaluations. Table 7 shows the test results of the system under different hand function conditions.

Based on the test results, it can be seen that this hand rehabilitation assessment system, for users with different hand function conditions, has no overlapping in assessment

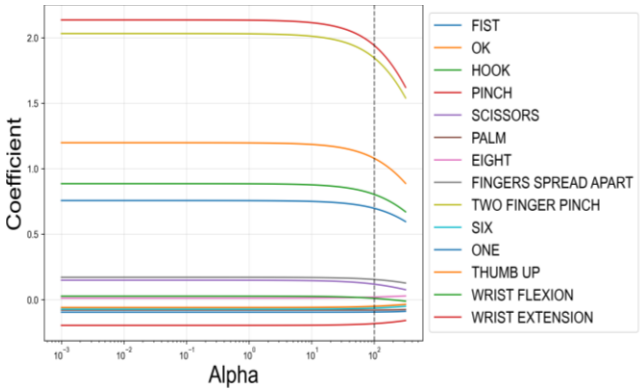



**Fig. 11.** Constrained ridge path.

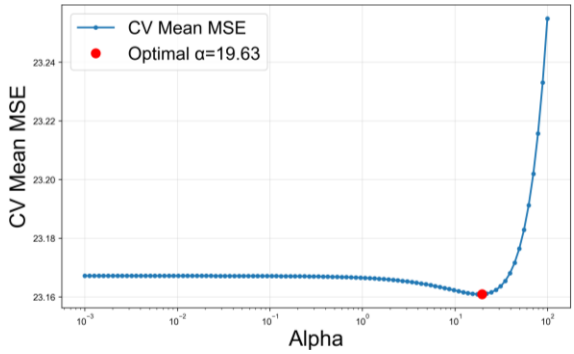



**Fig. 12.** Cross-validation curve.

| Hand function status | Left-hand overall score | Right-hand overall score |
|---|---|---|
| Normal | [91.62,83.43] | [90.71,83.69] |
| Incomplete | [79.21,72.56] | [79.34,73.62] |
| Poor | [70.31,62.87] | [69.68,60.98] |

**Tab. 7.** Test results of the hand rehabilitation assessment system.

scores, and has a good distinguishing effect. It can obtain the user's hand rehabilitation status in real time and provide feedback through quantitative scoring to the user, serving as a reference for the user.

## 6.6 Hand Rehabilitation Training System Test

To preliminarily verify the reliability and sensitivity of the aforementioned hand rehabilitation training system, 30 volunteers were selected. Among them, 10 had normal hand functions, 10 had incomplete hand functions, and 10 had poor hand functions. Each person completed 10 full training sessions. Table 8 shows the test results of the hand rehabilitation training system.

Compared with previously used hand rehabilitation training and clinical practice [27], [28], the above test results indicate that for users with different hand function conditions, there is a good reliability rate. It is possible to achieve the goal of training the hand functions while making the training process entertaining, in order to better assist users in restoring their normal hand functions.

| Hand position | Training type | Number of tests | Number of Successes | Reliability rate |
|---|---|---|---|---|
| Normal | Training Game 1 | 100 | 97 | 97% |
| | Training Game 2 | 100 | 95 | 95% |
| Incomplete | Training Game 1 | 100 | 93 | 93% |
| | Training Game 2 | 100 | 95 | 95% |
| Poor | Training Game 1 | 100 | 91 | 91% |
| | Training Game 2 | 100 | 90 | 90% |

**Tab. 8.** Test results of the hand rehabilitation training system.

# 7. Summary

This paper presents a mid-air hand rehabilitation evaluation and virtual training system, employing a spatio-temporal attention-enhanced multi-scale residual graph convolutional network for gesture recognition. First, users are guided by the assessment module to perform a sequence of 14 specific gestures, enabling a quantitative evaluation of both local joint function and overall hand performance. Subsequently, based on the initial assessment results, the system dynamically adjusts the training focus and formulates a personalized training plan for each user. Experimental validation shows that the proposed gesture recognition algorithm achieves high accuracy across different gestures and health conditions. Furthermore, both the rehabilitation assessment and training modules demonstrate effectiveness and reliability.

Although the hand rehabilitation system developed in this study has achieved satisfactory preliminary verification results in terms of accuracy and system effectiveness during the experimental stage, there is still a long way to go before it can be transformed into a true clinical tool. This research currently mainly focuses on technical feasibility and algorithm verification. Future work will be deepened along the path of from laboratory to clinical practice.

# Acknowledgments

This work was funded by Tiangong University's 2025-level College Students Innovation and Entrepreneurship Training Plan Project (Project No. 202510058089).

# References

[1] MOUNTCASTLE, V. B. *The Sensory Hand: Neural Mechanisms of Somatic Sensation*. 1st ed., Cambridge (Mass, USA): Harvard University Press, 2005. ISBN: 9780674019744

[2] MAYRHOFER-SCHMID, M., AMAN, M., STOLLE, A., et al. Analysis of epidemiology, etiology and injury patterns in 2,179 digit amputations. *Scientific Reports*, 2025, vol. 15, no. 1, p. 1–8. DOI: 10.1038/s41598-025-18983-y

[3] TAMULEVICIUS, M., BUCHER, F., DASTAGIR, N., et al. Demographic shifts reshaping the landscape of hand trauma: A comprehensive single-center analysis of changing trends in hand injuries from 2007 to 2022. *Injury Epidemiology*, 2024, vol. 11, no. 1, p. 1–10. DOI: 10.1186/s40621-024-00510-8

[4] SAVAS, S., ALTUNTAS, S. H., USLUSOY, F., et al. Epidemiology and injury characteristics of children with acute traumatic hand injuries undergoing surgery: A 10-year retrospective cohort study. *Ulusal Travma Ve Acil Cerrahi Dergis-Turkish Journal of Trauma & Emergency Surgery*, 2025, vol. 31, no. 8, p. 729–738. DOI: 10.14744/tjtes.2025.07892

[5] MOELLHOFF, N., THRONER, V., FRANK, K., et al. Epidemiology of hand injuries that presented to a tertiary care facility in Germany: A study including 435 patients. *Archives of Orthopaedic and Trauma Surgery*, 2023, vol. 143, no. 3, p. 1715 to 1724. DOI: 10.1007/s00402-022-04617-9

[6] KANKAM, H. K., IBRAHIM, H., LIEW, M. S., et al. Epidemiology of adult hand injuries presenting to a tertiary hand surgery unit: A review of 4216 cases. *Journal of Hand Surgery-European Volume*, 2024, vol. 49, no. 1, p. 48–53. DOI: 10.1177/17531934231195499

[7] LIM, J. X., LE, L. A. T., YEH, J. Z. Y., et al. The epidemiology and distribution of hand fractures in Singapore. *Singapore Medical Journal*, 2025, vol. 66, no. 9, p. 476–480. DOI: 10.4103/singaporemedj.SMJ-2021-334

[8] AKSAN, A., DURUSOY, R., ADA, S., et al. Epidemiology of injuries treated at a hand and microsurgery hospital. *Acta Orthopaedica et Traumatologica Turcica*, 2010, vol. 44, no. 5, p. 352–360. DOI: 10.3944/AOTT.2010.2372

[9] SLADE, J. F., MILEWSKI, M. D. Management of carpal instability in athletes. *Hand Clinics*, 2009, vol. 25, no. 3, p. 395 to 408. DOI: 10.1016/j.hcl.2009.05.002

[10] CHANG, J. H., SHIEH, S. J., KUO, L. C., et al. The initial anatomical severity in patients with hand injuries predicts future health-related quality of life. *Journal of Trauma and Acute Care Surgery*, 2011, vol. 71, no. 5, p. 1352–1358. DOI: 10.1097/TA.0b013e318216a56e

[11] EISELE, A., DERESKEWITZ, C., KUS, S., et al. Factors affecting time off work in patients with traumatic hand injuries - A bio-psycho-social perspective. *Injury*, 2018, vol. 49, no. 10, p. 1822 to 1829. DOI: 10.1016/j.injury.2018.07.012

[12] KOLOVICH, G. P., HEIFNER, J. J. Proximal interphalangeal joint dislocations and fracture-dislocations. *Journal of Hand Surgery-European Volume*, 2023, vol. 48, no. 2_SUPPL, p. 27S–34S. DOI: 10.1177/17531934231183259

[13] CANTERO-TÉLLEZ, R. A global proprioception concept after hand injury - Patient report. *Journal of Hand Therapy*, 2024, vol. 37, no. 2, p. 293–295. DOI: 10.1016/j.jht.2023.10.002

[14] AHMAD, K. A., HIGASHI, T., YOSHIDA, K. Dynamic hand gesture recognition by hand landmark classification using long short-term memory. *Pertanika Journal of Science and Technology*, 2025, vol. 33, no. S2, p. 73–84. DOI: 10.47836/pjst.33.S2.05

[15] TRAN, D. S., HO, N. H., YANG, H. J., et al. Real-time hand gesture spotting and recognition using RGB-D camera and 3D convolutional neural network. *Applied Sciences-Basel*, 2020, vol. 10, no. 2, p. 1–15. DOI: 10.3390/app10020722

[16] LIU, Q., CAI, M., LIU, D., et al. ESS MS-G3D: Extension and supplement shift MS-G3D network for the assessment of severe mental retardation. *Complex & Intelligent Systems*, 2024, vol. 10, no. 2, p. 2401–2419. DOI: 10.1007/s40747-023-01275-1

[17] CHEN, Z., LAN, X., ZENG, Y., et al. Multiscale spatiotemporal

residual graph convolutional neural network for action recognition. (in Chinese) *Journal of Naval Aviation University*, 2025, vol. 40, no. 2, p. 292–302. DOI: 10.7682/j.issn.2097-1427.2025.02.007

[18] YU, H., NELSON, A., ERDEN, M. S. A multi-layered quantitative assessment approach for hand spasticity based on a cable-actuated hand exoskeleton. *IEEE Robotics and Automation Letters*, 2025, vol. 10, no. 3, p. 2335–2342. DOI: 10.1109/LRA.2025.3530320

[19] FANG, Q., MAHMOUD, S. S., GU, X., et al. A novel multistandard compliant hand function assessment method using an infrared imaging device. *IEEE Journal of Biomedical and Health Informatics*, 2019, vol. 23, no. 2, p. 758–765. DOI: 10.1109/JBHI.2018.2837380

[20] LI, C. G., CHENG, L., YANG, H., et al. An automatic rehabilitation assessment system for hand function based on leap motion and ensemble learning. *Cybernetics and Systems*, 2020, vol. 52, no. 1, p. 3–25. DOI: 10.1080/01969722.2020.1827798

[21] NATH, D., SINGH, N., BANDUNI, O., et al. Variable handle-resistance based joystick for post-stroke neurorehabilitation training of hand and wrist in upper extremities. *IEEE Transactions on Human-Machine Systems*, 2024, vol. 55, no. 1, p. 93–101. DOI: 10.1109/THMS.2024.3486123

[22] GHERMAN, B., ZIMA, I., VAIDA, C., et al. Robotic systems for hand rehabilitation- past, present and future. *Technologies*, 2025, vol. 13, no. 1, p. 1–51. DOI: 10.3390/technologies13010037

[23] JHA, C. K., SHUKLA, Y., MUKHERJEE, R., et al. A glove-based virtual hand rehabilitation system for patients with post-traumatic hand injuries. *IEEE Transactions on Biomedical Engineering*, 2024, vol. 71, no. 7, p. 2033–2041. DOI: 10.1109/TBME.2024.3360888

[24] BAIN, G. I., POLITES, N., HIGGS, B. G., et al. The functional range of motion of the finger joints. *Journal of Hand Surgery-European Volume*, 2015, vol. 40, no. 4, p. 406–411. DOI: 10.1177/1753193414533754

[25] CIENKI, A. *The Cambridge Handbook of Gesture Studies*. Cambridge (UK): Cambridge University Press, 2024. ISBN: 9781108486316

[26] BOBOS, P., MACDERMID, J. C., BOUTSIKARI, E. C., et al. Evaluation of the content validity index of the Australian/Canadian osteoarthritis hand index, the patient-rated wrist/hand evaluation and the thumb disability exam in people with hand arthritis. *Health and Quality of Life Outcomes*, 2020, vol. 18, no. 1, p. 1–9. DOI: 10.1186/s12955-020-01556-0

[27] PENG, L., ZHANG, C., ZHOU, L., et al. Traditional manual acupuncture combined with rehabilitation therapy for shoulder hand syndrome after stroke within the Chinese healthcare system: A systematic review and meta-analysis. *Clinical Rehabilitation*, 2018, vol. 32, no. 4, p. 429–439. DOI: 10.1177/0269215517729528

[28] SERBAN, A., GROSU-BULARDA, A., BORDEANU-DIACONESCU, E. M., et al. Early rehabilitation versus conventional approaches in post-traumatic hand injuries with multiple lesions: Clinical outcomes and future directions. *Medicina*, 2025, vol. 61, no. 11, p. 1–24. DOI: 10.3390/medicina61112063

## About the Authors ...

**Rongzhe ZHOU** was born in Henan Province, China, in January 2005. He is an undergraduate student majoring in Communication Engineering at Tiangong University. During his studies, he has participated in his teachers' scientific research projects. His research interests include intelligent algorithms and action recognition.

**Peize LAI** was born in HeBei Province, China, in April 2005. He is an undergraduate student majoring in Communication Engineering at Tiangong University. During his studies, he has participated in his teachers' scientific research projects. His research interests include intelligent algorithms and action recognition.

**Le CHEN** was born in Anhui Province, China, in March 2005. He is an undergraduate student majoring in Communication Engineering at Tiangong University. During his studies, he has participated in his teachers' scientific research projects. His research interests include intelligent algorithms and action recognition.

**Zhibo HU** was born in Heilongjiang Province, China, in July 2005. He is an undergraduate student majoring in Communication Engineering at Tiangong University. During his studies, he has participated in his teachers' scientific research projects. His research interests include intelligent algorithms and action recognition.

**Lei QIAN** (corresponding author) received the B.Eng. and Ph.D. degrees from the College of Communications Engineering, Jilin University, Changchun, China, in 2016 and 2021, respectively. Now she is with the Tianjin Key Laboratory of Optoelectronic Detection Technology and System, School of Electronic and Information Engineering, Tiangong University, as a Lecturer. She was a visiting Ph.D. student with the School of Engineering, University of British Columbia, Canada, from 2019 to 2020, sponsored by the Chinese Scholarship Council. Her current research interests include delay QoS guarantee, effective capacity, visible light communications, resource allocation, physical layer security and channel estimation.