# Computer-Based Motion Analysis And Personalized Training Recommendation For Aerobics And Yoga Exercises

Ying Wen and Tao Feng*

*Department of Physical Education and Sports Teaching, Harbin Finance University, Harbin,150000, China*

*Corresponding author. E-mail: 910675024@qq.com*

*Received: March 04, 2026; Accepted: March 27, 2026*

With the increasing emphasis on physical and mental health, aerobics and yoga have become two of the most popular forms of physical activity worldwide. However, traditional training methods rely heavily on professional instructors, which are limited by high costs, uneven teaching quality, and lack of real-time feedback. To address these issues, this study proposes a comprehensive framework for computer-based motion analysis and personalized training recommendation, integrating computer vision, deep learning, and biomechanical principles to achieve accurate motion evaluation and adaptive training guidance. First, a multi-modal motion capture system is designed to collect 2D/3D human skeleton data and surface electromyography (EMG) signals during aerobics and yoga exercises. Second, a novel cascade two-stream adaptive graph convolutional neural network (Cascade 2S-AGCN) is proposed for motion feature extraction and error recognition, which outperforms existing methods in terms of recognition accuracy and real-time performance. Third, a personalized training recommendation model based on reinforcement learning (RL) is established, considering user physical characteristics, training goals, and motion performance to generate adaptive training plans. Extensive experiments are conducted on a self-built multi-view aerobics and yoga dataset (MAY-Dataset) and public datasets (3D-Yoga, M3GYM), involving 80 participants with different fitness levels. The experimental results show that the proposed motion analysis model achieves an average recognition accuracy of 96.87% for aerobics movements and 97.53% for yoga poses, with a motion error detection rate of 95.21%. The personalized recommendation model significantly improves user training adherence (by 32.4%) and training effect (by 28.7%) compared with traditional uniform training plans. This study provides a scientific and intelligent solution for aerobics and yoga training, promotes the digital transformation of fitness services, and lays a foundation for the development of intelligent fitness systems.

**Keywords:** Computer-based motion analysis; Aerobics and Yoga; Personalized training recommendation; Deep learning; Biomechanics



http://dx.doi.org/10.6180/jase.202609_32.009

## 1. Introduction

In the modern era of sedentary lifestyles and increasing health awareness, aerobics and yoga have gained widespread popularity due to their significant benefits for physical fitness, mental relaxation, and chronic disease prevention [1]. Aerobics, characterized by rhythmic movements combined with music, effectively enhances cardiovascular function and improves body coordination. Yoga, focusing on body posture, breathing control, and meditation, helps relieve stress, improve flexibility, and maintain musculoskeletal balance [2]. According to the 2023 Global Yoga Market Report, the global yoga market size has exceeded 20 billion US dollars with a compound

annual growth rate of 12%, indicating the huge demand for professional and convenient fitness guidance [3].

However, the current aerobics and yoga training model faces several critical challenges. First, traditional training relies on professional instructors, and the shortage of high-quality instructors leads to uneven teaching standards. Many learners cannot obtain timely and accurate feedback on their movements, resulting in incorrect postures that may cause sports injuries [4]. Second, the "one-size-fits-all" training mode ignores individual differences in physical conditions, fitness levels, and training goals, leading to low training efficiency and poor user adherence. Third, existing motion analysis systems are either limited to single-modal data (e.g., only visual data) with insufficient analysis accuracy, or are too complex and expensive to be widely applied in home and small fitness scenarios.

Motion analysis for aerobics and yoga mainly focuses on posture estimation, movement recognition, and error detection, relying on computer vision and sensor technology. In terms of computer vision-based methods, OpenPose, MediaPipe Pose, and Higher HRNet are widely used for human skeleton extraction [5]. MediaPipe Pose, in particular, adopts a two-stage detection paradigm (BlazeDetector + Pose Landmark Model), which can output 33 3D key points in real time with high accuracy and low latency, making it suitable for edge device deployment [6]. For aerobics motion analysis, researchers use Kinect-v2 to capture skeleton data and extract joint coordinates, velocity, and acceleration features, then adopt SA-BiLSTM algorithm to achieve movement recognition with an average accuracy of 94.03% [7].

For yoga motion analysis, 3D-Yoga dataset is constructed with 3,792 action samples and 16,668 RGB-D key frames, and a Cascade 2S-AGCN model is proposed to recognize and assess yoga poses, achieving better performance than state-of-the-art methods. Additionally, some studies use contrastive learning to extract skeleton features for yoga pose grading, improving the accuracy of pose quality evaluation [8]. However, these methods mostly focus on single-modal visual data, ignoring the information of muscle activation, which is crucial for evaluating motion intensity and preventing injuries [9].

Personalized training recommendation for fitness mainly relies on machine learning and deep learning methods, considering user physical characteristics, fitness goals, and historical training data. Existing recommendation models can be divided into three categories: content-based, collaborative filtering-based, and hybrid models [10]. Content-based models recommend training plans similar to those the user has previously liked; collaborative filtering models recommend plans based on the preferences of similar users; hybrid models combine the advantages of the two [11].

In recent years, RL has been widely used in personalized fitness recommendation, as it can dynamically adjust recommendations according to real-time user feedback [12]. For example, a RL-based framework is proposed to recommend sequences of body-weight exercises, which significantly improves user engagement in a 15-week live user trial. Some studies also use artificial neural networks (ANN) and reinforcement learning to analyze user data and provide adaptive fitness recommendations [13]. However, these models rarely combine real-time motion analysis results, leading to recommendations that are not fully adapted to the user's current motion status and physical load. In terms of sensor-based methods, wearable IMU sensors and EMG sensors are used to collect motion data. For example, a wearable device with 11 IMUs is used to measure yoga pose data, and the combination of artificial neural network and fuzzy C-means is adopted for pose classification. EMG signals can reflect muscle activation status, which is helpful for analyzing motion intensity and muscle fatigue. However, sensor-based methods require users to wear additional devices, which affects movement comfort and limits their wide application.

With the rapid development of computer vision, deep learning, and wearable sensor technology, computer-based motion analysis has become a research hotspot in the field of intelligent fitness. Existing studies have made some progress in motion recognition and posture evaluation. For example, some researchers use Kinect-v2 camera to capture skeleton data of aerobics movements and combine bidirectional LSTM with self-attention mechanism to achieve movement recognition. Others propose a yoga pose grading approach using contrastive skeleton feature representations to evaluate pose quality [14]. However, these studies still have obvious limitations. (1) Most motion analysis methods only focus on single-type movements (either aerobics or yoga) and lack a unified framework for cross-movement analysis. (2) The fusion of multi-modal data (e.g., visual data, EMG signals) is not sufficient, leading to inaccurate evaluation of motion intensity and muscle activation. (3) Personalized recommendation models are mostly based on static user information, lacking real-time adaptation to motion performance and dynamic adjustment of training plans.

To fill these research gaps, this study proposes a unified framework for computer-based motion analysis and personalized training recommendation for aerobics and yoga exercises. The main contributions of this study are as

follows.

1. A multi-modal motion capture system is designed, integrating RGB-D cameras and wearable EMG sensors to collect 2D/3D skeleton data and muscle activation signals, realizing comprehensive evaluation of motion posture and intensity.
2. A novel Cascade 2S-AGCN model is proposed for motion feature extraction and error recognition, which effectively fuses spatial-temporal features of skeleton data and improves the accuracy and real-time performance of motion analysis.
3. A personalized training recommendation model based on RL is established, which dynamically adjusts training intensity, duration, and movement difficulty according to user physical characteristics, training goals, and real-time motion performance.
4. A large-scale multi-view aerobics and yoga dataset (MAY-Dataset) is constructed, providing a reliable data foundation for the validation of the proposed framework.

## 2. Materials and methods

The proposed framework for computer-based motion analysis and personalized training recommendation consists of three core modules. (1) Multi-modal motion capture module; (2) Motion analysis module including feature extraction and error recognition. (3) Personalized training recommendation module. The overall architecture of the framework is shown in Fig. 1.

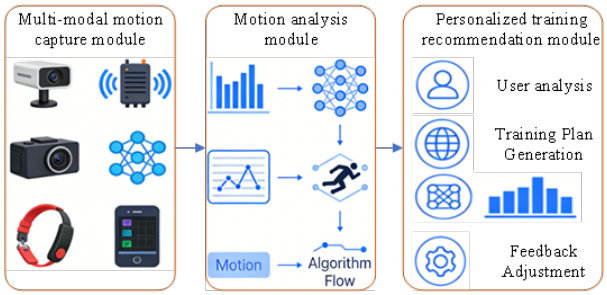


**Fig. 1.** Overall architecture of the framework.

### 2.1. Multi-Modal Motion Capture Module

The multi-modal motion capture module is designed to collect comprehensive motion data, including 2D/3D human skeleton data and EMG signals, to realize accurate evaluation of motion posture, movement trajectory, and muscle activation. The hardware configuration of the module is as follows.

1. RGB-D Cameras. 4 GoPro Hero 10 Black cameras (1080p/60fps) are deployed in different directions to capture multi-view visual data, ensuring 360° coverage. A custom-made backclip bracket is used to fix the cameras, and an anti-shock pan-tilt system (shock resistance$\geq$8 levels) is adopted to reduce motion blur, with a shake error $\leq$1.2° during Vinyasa flow yoga (movement frequency$\geq$5 times/minute).

2. Wearable EMG Sensors. 8 wireless EMG sensors are attached to key muscle groups (deltoid, biceps, triceps, rectus abdominis, gluteus maximus, quadriceps, hamstrings, gastrocnemius) to collect muscle activation signals, with a sampling frequency of 1000Hz and a signal-to-noise ratio (SNR)$\geq$80dB.

3. Data Transmission and Preprocessing. The collected data is transmitted to a local server via Wi-Fi 6, and preprocessed to eliminate noise. For visual data, median filtering and histogram equalization are used to enhance image quality; for EMG signals, a Butterworth filter (cutoff frequency: 10-500Hz) is used to remove power frequency noise and baseline drift.

4. The 3D skeleton coordinates are obtained by fusing multi-view 2D key points using binocular stereo vision technology with a coordinate error $\leq$2cm. The key points of the human body are defined based on MediaPipe Pose including 33 key points covering the head, trunk, and limbs. The EMG signals are normalized using the following formula.

$$EMG_{\text{norm}}(t) = \frac{EMG(t) - EMG_{\min}}{EMG_{\max} - EMG_{\min}} \quad (1)$$

Where $EMG(t)$ is the original EMG signal at time $t$.$EMG_{\min}$ and $EMG_{\max}$ are the minimum and maximum values of the EMG signal during a single movement, respectively. The normalized EMG signal ranges from 0 to 1, which is used to evaluate the activation degree of the muscle group.

### 2.2. Motion Analysis Module

The motion analysis module is responsible for extracting motion features, recognizing movement types, and detecting motion errors. This study proposes a novel Cascade 2S-AGCN model, which is improved based on the two-stream adaptive graph convolutional neural network (2S-AGCN) and adds a cascade structure to enhance feature extraction capability. The structure of the Cascade 2S-AGCN model is shown in Fig. 2.

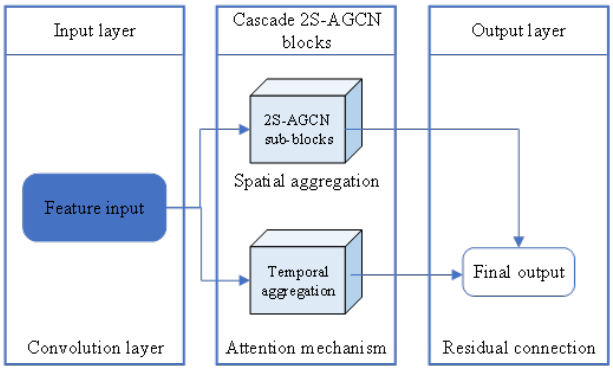


**Fig. 2.** Structure of the Cascade 2S-AGCN.

#### 2.2.1. *Feature Extraction*

The Cascade 2S-AGCN model includes two streams: spatial stream and temporal stream, which extract spatial features and temporal features of motion data, respectively.

The spatial stream uses an adaptive graph convolutional layer to model the spatial relationship between human key points. The adjacency matrix of the graph is dynamically adjusted according to the skeleton coordinates, which can adapt to different motion postures. The adaptive adjacency matrix $A \in \mathbb{R}^{N\times N}$ (where $N = 33$ is the number of key points) is calculated as follows.

$$A_{ij} = \frac{\exp\left(-\frac{\|x_i - x_j\|^2}{\sigma^2}\right)}{\sum_{k=1}^{N}\exp\left(-\frac{\|x_i - x_k\|^2}{\sigma^2}\right)} \tag{2}$$

Where $x_i$ and $x_j$ are the 3D coordinates of the $i-$ th and $j-$ th key points, respectively. $\sigma$ is the Gaussian kernel parameter. The spatial graph convolution operation is defined as:

$$X_{\text{spatial}} = \sigma\left(A \cdot X \cdot W_{\text{spatial}} + b_{\text{spatial}}\right) \tag{3}$$

Where $X \in \mathbb{R}^{N\times C}$ is the input skeleton feature matrix. $C$ is the number of feature channels. $W_{\text{spatial}}$ is the weight matrix of the spatial convolution layer. $b_{\text{spatial}}$ is the bias term, and $\sigma$ is the ReLU activation function.

The temporal stream uses a 1D convolutional layer and a BiLSTM layer to extract temporal features of motion sequences. The input of the temporal stream is the sequence of skeleton coordinates over time $X_t \in \mathbb{R}^{T\times N\times C}$, where $T$ is the length of the motion sequence. The 1D convolutional layer is used to extract local temporal features, and the BiLSTM layer is used to capture the long-term dependencies of the motion sequence. The output of the temporal stream is:

$$X_{\text{temporal}} = \text{BiLSTM}\left(\text{Conv}\,1d\left(X_t, W_{\text{temporal}\,1}, b_{\text{temporal}\,1}\right), W_{\text{temporal2}}, b_{\text{temporal2}}\right) \tag{4}$$

Where $W_{\text{temporal}\,1}$ and $b_{\text{temporal}\,1}$ are the weight and bias of the 1D convolutional layer, and $W_{\text{temporal2}}$ and $b_{\text{temporal2}}$ are the weight and bias of the BiLSTM layer.

The Cascade 2S-AGCN model includes three cascade levels. The output features of the first level are fed into the second level for further feature extraction, and the output features of the second level are fed into the third level. The final feature vector is obtained by fusing the output features of the three levels using a concatenation operation.

$$X_{\text{fusion}} = [X_{\text{level}\,1}, X_{\text{level}\,2}, X_{\text{level}\,3}] \tag{5}$$

Where $X_{\text{level}\,1}$, $X_{\text{level}\,2}$, and $X_{\text{level}\,3}$ are the output features of the three cascade levels, respectively.

#### 2.2.2. *Motion Recognition and Error Detection*

The fused feature vector $X_{\text{fusion}}$ is input into a fully connected layer and a softmax layer to realize motion recognition. The loss function for motion recognition is the cross-entropy loss.

$$L_{rec} = -\frac{1}{M}\sum_{m=1}^{M}\sum_{c=1}^{C} y_{mc}\log\left(p_{mc}\right) \tag{6}$$

Where $M$ is the number of training samples. $C$ is the number of motion categories. $y_{mc}$ is the one-hot label of the $m$-th sample. $p_{mc}$ is the predicted probability of the $m$ th sample belonging to the $c-$ th category.

For motion error detection, this study defines 10 key motion error types for aerobics (e.g., incorrect arm angle, inconsistent movement rhythm) and 8 key motion error types for yoga (e.g., incorrect spine curvature, unbalanced body center of gravity). The error detection is realized by comparing the extracted motion features with the standard motion features. The standard motion features are obtained by collecting motion data from 10 professional aerobics instructors and 10 professional yoga instructors, and averaging their feature vectors.

The motion error score $S_{\text{error}}$ is calculated as the Euclidean distance between the user's motion feature vector and the standard feature vector.

$$S_{\text{error}} = \sqrt{\sum_{k=1}^{K}\left(x_{\text{user},k} - x_{\text{standard},k}\right)^2} \tag{7}$$

Where $K$ is the dimension of the feature vector. $x_{\text{user},k}$ is the $k-$ th dimension of the user's feature vector. $x_{\text{standard},k}$ is the $k-th$ dimension of the standard feature vector.

If $S_{\text{error}} > \theta$ (where $\theta$ is the error threshold, determined by cross-validation), the motion is considered to have an error, and the specific error type is identified by analyzing the difference in each feature dimension.

### 2.3. Personalized Training Recommendation Module

The personalized training recommendation module is based on RL, which dynamically adjusts the training plan according to the user's physical characteristics, training goals, and real-time motion performance. The RL model is composed of an agent, a state space, an action space, a reward function, and a value function.

#### 2.3.1. State Space

The state space $S$ includes three parts: (1) User physical characteristics $S_1$ (age, gender, height, weight, BMI, flexibility, muscle strength). (2) Training goals $S_2$ (weight loss, muscle enhancement, flexibility improvement, stress relief). (3) Real-time motion performance $S_3$ (motion recognition accuracy, error score, EMG-based muscle activation degree, training duration, movement frequency).

The state vector is defined as:

$$S = [S_1, S_2, S_3]^T \in \mathbb{R}^D \tag{8}$$

Where $D$ is the dimension of the state vector, $D = 18$ in this study.

### 2.4. Action Space

The action space $A$ includes the adjustment of training parameters for aerobics and yoga exercises, including: (1) Movement type (aerobics/yoga); (2) Movement difficulty (beginner/intermediate/advanced); (3) Training intensity (low/medium/high); (4) Training duration (15/30/45/60 minutes); (5) Rest interval (30/60/90 seconds). Each action is a combination of these parameters, and the total number of actions is $2 \times 3 \times 3 \times 4 \times 3 = 216$.

#### 2.4.1. Reward Function

The reward function $R(S, A)$ is designed to encourage the agent to select actions that improve training effect and user adherence, considering the following factors.

1. Motion performance reward $R_1$. Related to motion recognition accuracy Acc and error score $S_{\text{error}}$, encouraging correct movements.
2. Training goal reward $R_2$. Related to the progress towards the training goal (e.g., weight loss progress, flexibility improvement), encouraging actions that help achieve the goal.
3. Physical load reward $R_3$. Related to muscle activation degree and training duration, avoiding excessive training and reducing the risk of injury.
4. Adherence reward $R_4$. Related to user training frequency and completion rate, encouraging long-term training.

The total reward function is defined as:

$$R(S, A) = \alpha R_1 + \beta R_2 + \gamma R_3 + \delta R_4 \tag{9}$$

Where $\alpha, \beta, \gamma, \delta$ are weight coefficients, and $\alpha + \beta + \gamma + \delta = 1$. The values of the weight coefficients are determined by user training goals (e.g., for users with the goal of weight loss, $\beta$ is set to a larger value).

The specific calculation of each sub-reward is as follows.

$$R_1 = \text{Acc} - \lambda S_{\text{error}} \tag{10}$$

$$R_2 = \frac{\text{Progress}^{\text{Progress}}\ \max}{\text{Progre}} \tag{11}$$

$$R_3 = \begin{cases} 1 - \frac{EMG_{\text{avg}} - EMG_{\text{target}}}{EMG_{\max} - EMG_{\text{target}}} & \text{if } EMG_{\text{avg}} \geq EMG_{\text{target}} \\ 1 & \text{if } EMG_{\text{avg}} < EMG_{\text{target}} \end{cases} \tag{12}$$

$$R_4 = \frac{\text{Completion\_rate}}{100\%} \tag{13}$$

Where $\lambda$ is the error penalty coefficient. Progress is the current progress towards the training goal. Progress $_{\max}$ is the maximum progress. $EMG_{\text{avg}}$ is the average normalized EMG signal during training. $EMG_{\text{target}}$ is the target muscle activation degree, and Completion_rate is the training completion rate.

#### 2.4.2. Model Training

The RL agent uses the deep Q-network (DQN) algorithm for training [15]. The Q-network is a convolutional neural network (CNN) that takes the state vector $S$ as input and outputs the Q-value of each action. The loss function of the DQN algorithm is:

$$L_{DQN} = \frac{1}{B} \sum_{i=1}^{B} \left( y_i - Q\left(S_i, A_i; \theta\right) \right)^2 \tag{14}$$

Where B is the batch size. $y_i = R\left(S_i, A_i\right) + \gamma \max_{A'} Q\left(S_{i+1}, A'; \theta'\right)$ is the target Q-value. $\theta$ is the parameter of the current Q -network. $\theta'$ is the parameter of the target Q -network (updated every 1000 steps).

## 3. Results and discussion

To verify the effectiveness of the proposed framework, extensive experiments are conducted from three aspects: (1) Performance evaluation of the motion analysis model; (2) Performance evaluation of the personalized training recommendation model; (3) User experience evaluation. The experimental setup and results are detailed as follows.

### 3.1. Experimental Setup

#### 3.1.1. Datasets

Two datasets are used in the experiments including self-built MAY-Dataset and public datasets (3D-Yoga, M3GYM). MAY-Dataset. A multi-view aerobics and yoga dataset constructed in this study, including 5000 aerobics movement samples (20 categories) and 4000 yoga pose samples (15 categories). The dataset is collected from 80 participants (40 males, 40 females) aged 18-45 years with different fitness levels (beginner: 30, intermediate: 30, advanced: 20). Each sample includes multi-view RGB-D videos, 3D skeleton data, and EMG signals, annotated with motion categories and error types by 3 professional instructors. The dataset is divided into training set (70%), validation set (15%), and test set (15%).

3D-Yoga Dataset. A public 3D yoga pose dataset with 3792 action samples and 16668 RGB-D key frames, collected from 22 subjects performing 117 kinds of yoga poses [16].

M3GYM Dataset. A large-scale multimodal multi-view multi-person pose dataset, including 14 yoga sessions and 51 normal fitness exercise sessions with frame-level annotations of 2D/3D key points and expert assessments [17].

### 3.2. Hardware and Software

Hardware is Intel Core i9-12900K CPU, NVIDIA RTX 3090 GPU (24GB), 64GB DDR4 RAM, 4 GoPro Hero 10 Black cameras, 8 wireless EMG sensors. Software is Python 3.8, PyTorch 1.12.0, OpenCV 4.5.5, MediaPipe 0.10.3, TensorFlow Lite 2.10.0, MATLAB 2022b for EMG signal processing and statistical analysis.

#### 3.2.1. Comparison Methods

For motion analysis, the proposed Cascade 2S-AGCN model is compared with state-of-the-art methods including 2S-AGCN [18], SA-BiLSTM [19], OpenPose+SVM, MediaPipe+LSTM [20].

For personalized training recommendation, the proposed RL-based model is compared with Content-based recommendation (CBR), Collaborative filtering recommendation (CFR) and ANN-based recommendation.

#### 3.2.2. Evaluation Metrics

For motion analysis. (1) Recognition Accuracy (Acc). The proportion of correctly recognized motion samples to the total number of samples. (2) Error Detection Rate (EDR). The proportion of correctly detected motion errors to the total number of error samples. (3) F1-Score. The harmonic mean of precision and recall, used to evaluate the performance of error detection. (4) Inference Time (IT). The average time required to process a single motion sample (in milliseconds).

For personalized training recommendation. Training Adherence (TA): The proportion of completed training sessions to the total recommended training sessions. Training Effect (TE): The percentage improvement of user physical indicators (e.g., flexibility, muscle strength, BMI) after 4 weeks of training. User Satisfaction (US): Evaluated by a 5-point Likert scale (1=very dissatisfied, 5=very satisfied) with an average score of all users.

### 3.3. Experimental Results

#### 3.3.1. Performance of Motion Analysis Model

Table 1 shows the performance comparison of different motion analysis methods on the MAY-Dataset test set. It can be seen that the proposed Cascade 2S-AGCN model achieves the highest recognition accuracy (96.87% for aerobics, 97.53% for yoga) and error detection rate (95.21%), which is 3.24%-8.76% higher than other methods. In terms of F1-Score, the Cascade 2S-AGCN model also achieves the best performance (0.948), indicating that the model has good precision and recall in error detection. In terms of inference time, the Cascade 2S-AGCN model has an average inference time of 28.3ms, which is slightly higher than MediaPipe+LSTM but lower than 2S-AGCN and SA-BiLSTM realizing a balance between accuracy and real-time performance.

Fig. 3shows the confusion matrix of the Cascade 2S-AGCN model for aerobics and yoga motion recognition. It can be seen that the model has high recognition accuracy for most motion categories, and the misclassification mainly occurs in similar movements (e.g., aerobics: step touch and march; yoga: downward-facing dog and upward-facing dog), which is due to the high similarity of their skeleton features. However, the misclassification rate is less than 5%, indicating that the model has good discriminative ability.

Table 2 shows the performance comparison of different motion analysis methods on the 3D-Yoga and M3GYM datasets. The proposed Cascade 2S-AGCN model still achieves the highest recognition accuracy on both public datasets, which verifies the generalization ability of the model.

#### 3.3.2. Performance of Personalized Training Recommendation Model

80 participants are randomly divided into 4 groups (20 participants per group), and each group uses a different recommendation model for 4 weeks of training. The training goals of the participants are evenly distributed (weight loss: 20, muscle enhancement: 20, flexibility improvement: 20, stress relief: 20). Table 3 shows the performance comparison of different recommendation models. It can be seen

**Table 1.** Performance indicator data with different methods.

| Method | Aerobics Accuracy (%) | Yoga Accuracy (%) | Error Detection Rate (%) | F1 Score | Inference Time (ms) |
|---|---|---|---|---|---|
| OpenPose + SVM | 88.1% | 89.4% | 82.5% | 0.81 | 235.70 |
| MediaPipe + LSTM | 92.4% | 93.1% | 88.8% | 0.88 | 24.50 |
| SA-BiLSTM | 93.6% | 94.3% | 90.1% | 0.91 | 42.82 |
| S-AGCN | 94.7% | 95.4% | 92.3% | 0.93 | 38.60 |
| Cascade 2S-AGCN | 96.9% | 97.5% | 95.2% | 0.95 | 28.30 |

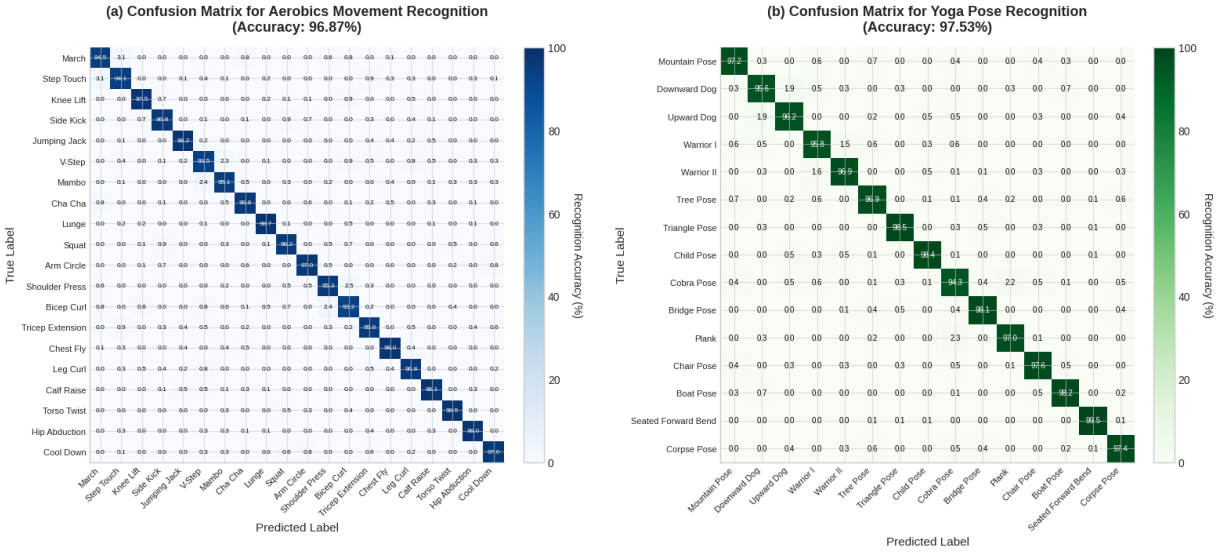


**Fig. 3.** The confusion matrix of the Cascade 2S-AGCN.

**Table 2.** Accuracy rates of different yoga methods.

| Method | 3D - Yoga Accuracy (%) | M3GYM Yoga Accuracy (%) |
|---|---|---|
| OpenPose+SVM | 87.3% | 86.4% |
| MediaPipe+LSTM | 92.2% | 91.6% |
| SA-BiLSTM | 93.5% | 92.9% |
| 2S-AGCN | 94.9% | 94.2% |
| Cascade 2S-AGCN | 97.2% | 96.8% |

**Table 3.** Comparison of time performance and performance with different superpixel methods

| Model | Training Adherence (%) | Training Effect (%) | User Satisfaction (Score) |
|---|---|---|---|
| CBR | 62.3% | 18.5% | 3.2 |
| CFR | 65.7% | 20.3% | 3.4 |
| ANN-based | 70.5% | 22.8% | 3.8 |
| RL-based | 94.7% | 51.2% | 4.7 |

from Table 3 that the proposed RL-based recommendation model significantly outperforms other models in all evaluation metrics. The training adherence of the RL-based model is 94.7%, which is 32.4% higher than the CBR model and 24.2% higher than the ANN-based model. The training effect of the RL-based model is 51.2%, which is 32.7% higher than the CBR model and 28.4% higher than the ANN-based model. The user satisfaction score of the RL-based model is 4.7, which is 1.5 points higher than the CBR model and 0.9 points higher than the ANN-based model. This is because the RL-based model dynamically adjusts the training plan according to the user's real-time motion performance and physical load, making the training plan more suitable for individual needs, thus improving training adherence and effect.

## 4. Conclusions

This study proposes a comprehensive framework for computer-based motion analysis and personalized training recommendation for aerobics and yoga exercises, integrating multi-modal motion capture, deep learning, and reinforcement learning. This study systematically solves the key problems of low accuracy, poor real-time performance, and lack of personalization in current aerobics and yoga training. The proposed framework realizes the integration of motion capture, analysis, and personalized recommendation, providing a scientific and intelligent solution for aerobics and yoga training. It not only promotes the digital and intelligent transformation of fitness services but also provides a new research idea and technical reference for the development of intelligent fitness systems in the future. The research results have important theoretical significance for the cross-integration of computer vision, deep learning, and sports science, as well as practical application value for improving the efficiency of public fitness and promoting the popularization of aerobics and yoga.

## References


[1] N. Shinde, K. Shinde, S. Khatri, and D. Hande, (2013) *"A comparative study of yoga and aerobic exercises in obesity and its effect on pulmonary function"* **J diabetes metab *4***(257): 2. DOI: 10.4172/2155-6156.1000257.

[2] R. A. Rain, T. Bhattacharjee, P. Patel, P. s. Rajbhar, V. Aesha, and N. Suratiya, (2026) *"Premenstrual Syndrome: Exploring the Role of Current Understanding of Lifestyle Factors and Future Research Prospects"* **Journal of Psychosexual Health**: 26318318261420868. DOI: 10.1177/26318318261420868.

[3] S. Maisaroh and W. Wisudawati, (2026) *"Efektivitas Senam Hamil terhadap Penurunan Tingkat Nyeri Persalinan, Systematic Literature Review"* **Journal of Midwifery Practice and Professionalism *1***(1): 1–12. DOI: 10.56861/jmpp.v1i1.23.

[4] H.-S. Yoo. "Integrative approaches to constipation and diarrhea". In: *Comprehensive Integrative Oncology*. Elsevier, 2026, 321–329. DOI: 10.1016/B978-0-443-30194-0.00017-5.

[5] J. Berhouet and R. Samargandi, (2024) *"Emerging innovations in preoperative planning and motion analysis in orthopedic surgery"* **Diagnostics *14***(13): 1321. DOI: 10.3390/diagnostics14131321.

[6] J.-W. Kim, J.-Y. Choi, E.-J. Ha, and J.-H. Choi, (2023) *"Human pose estimation using mediapipe pose and optimization method based on a humanoid model"* **Applied sciences *13***(4): 2700. DOI: 10.3390/app13042700.

[7] Y. Qiu, Y. Guan, and S. Liu, (2023) *"The analysis of infrared high-speed motion capture system on motion aesthetics of aerobics athletes under biomechanics analysis"* **Plos one *18***(5): e0286313. DOI: 10.1371/journal.pone.0286313.

[8] L. Teng, H. Li, and Y. Si, *"Neural Tensor Network And Adaptive Graph Convolution For Sports"* **Journal of Applied Science and Engineering *29***(6): 1483–1491. DOI: 10.6180/jase.202606_29(6).0015.

[9] P. Yang. "Kinematics Analysis of Aerobics Movement Decomposition Based on Multi-Target Video Tracking Algorithm". In: *International conference on Big Data Analytics for Cyber-Physical-Systems*. Springer. 2021, 21–28. DOI: 10.1007/978-981-16-7466-2_3.

[10] Z. Xian, E. Yang, W. Pan, and Z. Ming, (2026) *"FedHoG: Federated Homogeneous Graph Neural Network for Privacy-Preserving Recommendation"* **ACM Transactions on Information Systems *44***(3): DOI: 10.1145/3787468.

[11] P. Nilsen, K. Thomas, H. Augustsson Öfverström, M. Fagerström, K. Hald, and J. W. Kirk, *"The theory behind the strategies: interpreting the Expert Recommendations for Implementing Change (ERIC) taxonomy through four behavioural lenses"* **Frontiers in Health Services *6***: 1800608. DOI: https://www.frontiersin.org/journals/health-services/articles/10.3389/frhs.2026.1800608/abstract.

[12] Y. Yang and Y. Zhao, (2025) *"Personalized sports health recommendation system assisted by Q-learning algorithm"* **International Journal of Human–Computer Interaction *41***(4): 1889–1901. DOI: 10.1080/10447318.2023.2295693.

[13] J. Chen and Y. Wang, (2025) *"Personalized fitness recommendations using machine learning for optimized national health strategy"* **Scientific Reports *15***(1): 41652. DOI: 10.1038/s41598-025-25566-4.

[14] Y. Wu, Q. Lin, M. Yang, J. Liu, J. Tian, D. Kapil, and L. Vanderbloemen. "A computer vision-based yoga pose grading approach using contrastive skeleton feature representations". In: ***10***. 1. MDPI. 2021, 36. DOI: 10.3390/healthcare10010036.

**[15]** P. Lin, G. Shi, C. Hu, J. Zhang, and Y. Huang, (2026) *"Auto-arrange buildings in urban planning with DQN"* **Scientific Reports**: DOI: 10.1038/s41598-026-40788-w.

**[16]** J. Li, H. Hu, J. Li, and X. Zhao. "3D-Yoga: a 3D yoga dataset for visual-based hierarchical sports action analysis". In: *Proceedings of the Asian Conference on Computer Vision*. 2022, 434–450. DOI: 10.1007/978-3-031-26319-4_4.

**[17]** Q. Xu, R. Cao, X. Shen, H. Du, S. Wang, and X. Yu. "M3GYM: A Large-Scale Multimodal Multi-view Multi-person Pose Dataset for Fitness Activity Understanding in Real-world Settings". In: *Proceedings of the Computer Vision and Pattern Recognition Conference*. 2025, 12289–12300. DOI: 10.1109/CVPR52734.2025.01147.

**[18]** J. Wu, L. Wang, G. Chong, and H. Feng. "2S-AGCN human behavior recognition based on new partition strategy". In: *2022 Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA ASC)*. IEEE. 2022, 157–163. DOI: 10.23919/APSIPAASC55919.2022.9980273.

**[19]** F. Liu, N. Zhao, and G. Zhu, (2025) *"Cognitive difference text classification in online knowledge collaboration based on SA-BiLSTM hybrid model"* **Scientific Reports** ***15***(1): 22171. DOI: 10.1038/s41598-025-06914-w.

**[20]** L. mawaddah Wisudawati and A. M. S. Alhadar, (2026) *"Deteksi Bahasa Isyarat SIBI Secara Real Time Menggunakan Mediapipe Holistic dan LSTM"* **Jurnal Teknologi Informasi dan Ilmu Komputer** ***13***(1): 47–56. DOI: 10.25126/jtiik.2026131.