Copyright ⓒ 2026 KSII

Physical Fitness and Tactical Optimization of Badminton Based on Multimodal Fusion and Swarm Intelligence Algorithm

Yanjing Zheng[1*], Ye Han[1], Wei Wang[1], and Xiao Yu[2]
[1] College of Physical Education, Harbin University, Harbin 150086, China
[2] School of Physical Education and Health Sciences, Mudanjiang Normal University, Mudanjiang 157011, China
[e-mail: zyjssdd2026@163.com, hanye@163.com, liangxingyidu2025@163.com, yuxiao5626@163.com]
[*]Corresponding author: Yanjing Zheng

Received February 24, 2026; revised April 28, 2026; accepted May 30, 2026; published August 31, 2026

Abstract

To address the lack of a unified quantitative relationship among action recognition, physical condition representation, and tactical selection in badminton training, an offline analysis framework based on multimodal fusion and swarm intelligence optimization is proposed. The study utilizes the MultiSenseBadminton dataset, which includes 25 participants and 7,763 stroke records, covering three skill levels: beginners, intermediates, and experts. The experiment extracts three-axis acceleration and three-axis angular velocity sequences from inertial signals closely related to stroke actions and completes model training and evaluation under cross-subject partitioning conditions. The core of the model employs a dual-stream 1D Convolutional Neural Network (CNN) with Multi-Head Self-Attention (MHSA) for jointly modeling local temporal features of different inertial modalities and cross-modal dependencies. To enhance the efficiency of hyperparameter search, the Tent chaos mapping is introduced along with an improved Sparrow Search Algorithm (SSA), which is used to adaptively optimize the learning rate, convolution kernel size, and regularization parameters. Based on this, an offline tactical utility function with physical load proxy terms is constructed to simulate strategy selection under different load conditions. The results indicate that the proposed method achieves a representative optimal accuracy of 96.8% in the cross-subject recognition task, with an average accuracy of 96.72%±0.14% across five repeated experiments. Offline simulation results show that the optimized strategy maintains a high score gain while reducing the load proxy indicator. The findings suggest that multimodal action perception and load-constrained decision modeling have combined value and can provide a foundation for scientific badminton training and subsequent online system development.

Keywords: Multimodal fusion; Sparrow search algorithm; Badminton; Physical fitness and tactical optimization.

1. Introduction

Modern competitive badminton is evolving in the direction of high speed, strong confrontation and fine control, which poses a severe double challenge to athletes' physical reserve and tactical execution ability. On the field, athletes need to complete the transition from high-intensity movement to fine net-front technique in a very short time. This frequent energy metabolism mode switch can easily lead to central and peripheral fatigue, which in turn leads to the deformation of movement techniques and the lag of tactical decision-making [1]. The traditional training system often relies on the coaches' experience observation and subjective judgment to evaluate the athletes' physical condition. This qualitative rather than quantitative guidance mode is difficult to capture the small changes of kinematic parameters and accurately quantify the specific impact of fatigue on the quality of tactical execution [2, 3].

Large-scale sports data may now be obtained thanks to the widespread use of wearable technologies and the Internet of Things (IoT). Single-mode data sources always have limitations. For instance, video images are easily affected by occlusion and illumination, while simple inertial sensor data is difficult to restore complex tactical scenes [4]. How to break the data island and deeply fuse multi-source heterogeneous data to reveal the nonlinear coupling mechanism between physical fitness and tactics has become the bottleneck that sports engineering needs to break through urgently [5].

The core motivation of this study is based on the deep insight into the above pain points, aiming at building an intelligent system that can sense the physical loss in real time and dynamically recommend the best tactics. In the complicated and changeable competition process, the distribution of physical fitness is essentially a multi-objective dynamic optimization problem, that is, how to maximize the scoring probability and minimize the energy consumption. The existing linear programming or traditional logistic regression model is difficult to deal with this kind of high-dimensional and non-convex optimization space, and the Swarm Intelligence Algorithm (SIA) derived from the behavior mechanism of biological communities provides a new perspective to solve this problem with its strong global optimization ability and self-organization characteristics [6]. By simulating the cooperative foraging behavior of biological populations in a resource-limited environment, people can effectively find the best tactical path under physical constraints. However, at present, the application of this kind of algorithm in sports field mostly stays in the offline data analysis stage, lacking adaptive improvement and application for real-time game data.

Existing research often handles action recognition, physical condition assessment, and tactical recommendation separately, with action recognition results seldom being incorporated into subsequent strategy solving. Physical condition is also typically described in isolation. The research goal is to address the unified modeling problem under offline conditions. The processing workflow includes four stages: action recognition, multimodal representation, load proxy estimation, and strategy solving. The focus of the study is not on the mere stacking of individual modules, but rather on the information transfer and constraint modeling across these stages. The related results serve for offline training analysis and method validation, and are not directly equivalent to the real-time decision-making system that has been deployed.

2. Recent Work

With the infiltration of Artificial Intelligence (AI) technology in the competitive sports field, data-driven sports performance analysis has become a key means to improve the competitive level. Ghosh et al. (2023) [7] systematically combed the emerging technologies in sports analysis and pointed out that the introduction of algorithmic perspective was reshaping the traditional training and decision-making model. Deep Learning (DL) occupies a core position in sports video analysis and pattern recognition with its powerful feature extraction ability. Jia et al. (2025) [8] reviewed the application status of DL in sports performance analysis and thought that the current technology could effectively handle complex unstructured data. Zhou et al. (2024) [9] suggested an enhanced Tuna Swarm Optimization (TSO) to optimize Deep Neural Network (DNN) for the image classification challenge. This greatly increased the accuracy of sports image classification and demonstrated the potential of SIA in visual feature processing. Specific to badminton, video stream analysis is an important way to capture the details of the game. See et al. (2025) [10] compared a variety of DL models and gradient calculation methods, and realized the accurate detection of badminton multi-beat rounds. He and Zhang (2024) [11] further constructed a badminton action recognition model aided by DL neural network, and tried to combine technical action recognition with physical training optimization, which provided a preliminary framework for intelligent training.

Tactical analysis and decision support is another research hotspot in this field. Yuan et al. (2024) [12] built a prediction model of women's badminton singles based on Machine Learning (ML), and deeply analyzed the relationship between technology and tactical decision. To dig deeper tactical patterns, Hu et al. (2024) [13] used Knowledge Graph (KG) technology to reason and analyze badminton tactics, and realized the structural expression of unstructured tactical data. To identify probable rules in the competitive process, Wu et al. (2022) [14] created an interactive visual analytic system that focused on the pattern mining of multivariate event sequences in racket sports. On the microscopic index level, Tan and Teoh (2024) [15] introduced the static eye index, combined with neural network to predict the accuracy of hitting, which enriched the dimension of tactical evaluation from the perspective of visual attention. In addition to the analysis of single mode, the application of multimodal data fusion has been paid more and more attention to. Lv et al. (2025) [16] used Long Short-Term Memory (LSTM) to fuse and intervene multimodal data in basketball training, and realized the optimization of personalized exercise, which was of reference significance to badminton physical fitness research. On the algorithm level, Chen et al. (2025) [17] put forward the basketball winning algorithm inspired by basketball tactics, which showed the effectiveness of meta-heuristic optimizer based on sports rules in engineering application and implied that there was isomorphism between sports tactical logic and algorithm optimization mechanism.

Most researches tend to analyze tactical decision-making and physical fitness state separately, lacking a quantitative model of the impact on tactical execution quality under the dynamic change of physiological load, that is, failing to fully reveal the nonlinear coupling mechanism of "physical fitness-tactics". The existing tactical recommendation systems are mostly based on offline video data or static statistical indicators, and it is difficult to use multimodal real-time sensing data for dynamic feedback, which leads to the limited significance of real-time guidance of the model in the competition field. When handling high-dimensional and multi-objective physical fitness distribution problems, the conventional optimization approach is prone to falling into a local optimum and does not provide adaptive improvement for sports features. Consequently, this study attempts to construct a closed-loop system that can perceive physical fitness status in real time and dynamically optimize tactical

decision-making through multimodal deep fusion technology and improved SIA to fill the gap in the field of collaborative optimization of physical fitness and tactics.

3. Method

3.1 Dataset and Experimental Protocol

The study uses the publicly available MultiSenseBadminton dataset, which includes data from 25 participants and 7,763 badminton stroke records. The participants consist of 12 beginners, 8 intermediate players, and 5 experts. The data collection covers eye tracking, body tracking, electromyography (EMG), foot pressure, and inertial information. According to publicly available data, the eye tracking data has a sampling rate of 30 Hz, body tracking data has a sampling rate of 96 Hz, foot pressure and inertial data from the Moticon sensors have a maximum sampling rate of 100 Hz, arm EMG has a maximum sampling rate of 1000 Hz, and leg EMG has a sampling rate of 500 Hz. The recognition experiment focuses on the inertial inputs most closely related to the stroke actions, selecting three-axis acceleration and three-axis angular velocity as the main input channels.

To ensure objective evaluation results, the training, validation, and test sets are independently divided along the subject dimension. The experiment adopts a cross-subject grouping scheme, where the 25 participants are split into training, validation, and test sets in a 15:5:5 ratio. This division maintains near-equal balance across skill levels. To test the stability of the results, five repeated grouping experiments are also conducted. In each repetition, a fixed random seed is used, and it is ensured that the same participant's data would not appear in multiple sets simultaneously.

The signal preprocessing consists of four steps. First, the inertial sequences are uniformly resampled to 100 Hz. Second, the motion segments are divided using a sliding window technique, with a window length of 128 samples and a step size of 64 samples. Third, Z-score normalization based on training set statistics is applied to each channel. Fourth, outlier and missing segments are removed, retaining only the time-synchronized and complete samples. This preprocessing pipeline ensures consistency in the input scale across multiple experiments.

To enhance the interpretability of the results, the action category statistics, sample partitioning method, and key preprocessing parameters are provided in **Table 1** and **Table 2**.

Table 1. Summary of datasets and experimental protocols

Item	Content
Dataset	MultiSenseBadminton
Number of subjects	25
Total number of swing records	7,763
Skill level	Beginners 12; Intermediate 8; Expert 5
Identify task input	Triaxial acceleration +triaxial angular velocity
Main inertial source	Plantar wearable inertial unit
Uniform sampling rate	100 Hz
Window length	128
Window step size	64
Data standardization	Z-score based on mean and standard deviation of training set
Division strategy	Cross-subject grouping
Training/validation/testing	15/5/5
Repeated test times	5

Table 1 illustrates the basic protocol of the identification experiment. The number of subjects is 25. Training, verification and testing are divided independently at the subject level. This setting reduces the influence of identity information disclosure on the results. Uniform sampling rate, fixed window length and standardized mode also make the experimental process clearer. After the number of repeated tests is stated, the statistical significance of the identification results is more complete.

Table 2. Summary of datasets and experimental protocols

Action category	Sample number	Proportion (%)
Clear	1,768	22.78
Smash	1,526	19.66
Drop shot	1,384	17.83
Drive	1,657	21.34
Net shot	1,428	18.39
Total	7,763	100.00

Table 2 shows that the sample sizes for the five action categories are in the same order of magnitude. The highest category is Clear, with a sample proportion of 22.78%. The lowest category is Drop shot, with a proportion of 17.83%. There is no extreme imbalance in the category distribution. This balanced data foundation facilitates subsequent comprehensive comparisons of precision, recall, and F1-score.

3.2 Construction of Dual-stream Multimodal Fusion Network Based on MHSA

In view of the high-dimensional, time-varying and heterogeneous characteristics of inertial sensor data in badminton, it is difficult for a single serial network structure to capture the deep correlation between different physical quantities at the same time. In this study, a dual-stream multimodal fusion network architecture based on Multi-Head Self-Attention (MHSA) is designed and constructed. In feature extraction stage, the model adopts dual-stream Convolutional Neural Network (CNN) structure. The triaxial acceleration data contained in MultiSenseBadminton dataset reflects the linear displacement and impact load of limbs, while the triaxial angular velocity data represents the rotation rate and motion trajectory of joints, which are significantly different in physical meaning and numerical distribution. The left-hand input acceleration sequence and the right-hand input angular velocity sequence of the dual-stream architecture are both stacked by multiple one-dimensional convolution layers. Through the sliding window, the convolution layer detects local information on the time axis. This allows it to efficiently extract the rhythmic properties of the pace movement and the explosive qualities of the swing moment. Let the input time series data be $X \in R^{T \times C}$, where T is the time step and C is the number of channels, and the calculation process of the output characteristic graph H^l of the lth convolution layer is shown in (1):

$$H_{i,j}^{l} = f(\sum_{m=1}^{M} H_{i+m-1,k}^{l-1} \cdot W_{m,k,j}^{l} + b_j^l) \tag{1}$$

$H_{i,j}^{l}$ denotes the output value of the j-th convolution kernel in the l-th layer at time step i. W is the convolution kernel weight matrix, M is the convolution kernel size, b is the bias term, and $f(\cdot)$ is the nonlinear activation function. This study uses the Rectified Linear Unit (ReLU) as the activation function [18], whose mathematical expression is displayed in (2), to improve the model's capacity to match nonlinear relationships and mitigate the gradient vanishing problem:

$$f(x) = \max(0, x) \tag{2}$$

After dual-stream CNN extraction, the original signals of acceleration and angular velocity are transformed into high-dimensional feature vectors containing rich semantic information. To explore the dependence of these two types of heterogeneous features in time dimension, the model introduces the MHSA mechanism MHSA module. **Fig. 1** shows the architecture of dual-stream multimodal fusion network based on MHSA.

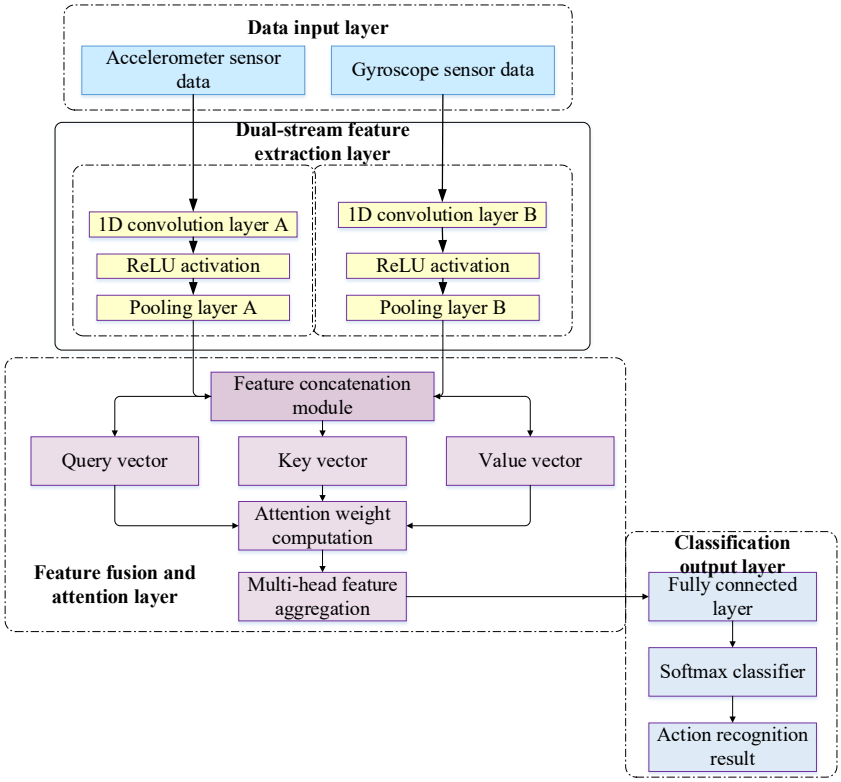

Fig. 1. Dual-stream multimodal fusion network architecture based on MHSA

In MHSA module, the eigenvectors output by the dual-stream network are spliced and fused to form a unified input matrix Z. The input matrix is mapped to different subspaces through three independent linear transformation matrices W^Q, W^K and W^V, and the query matrix, key matrix and value matrix are generated respectively, as shown in (3):

$$Q = \mathrm{Z}W^Q, K = ZW^K, V = ZW^V \tag{3}$$

The attention weight between time steps is calculated by scaling point product attention mechanism. By calculating the dot product of Q and K transposed, the similarity between features in different time steps is measured, and then divided by the scaling factor d_k to prevent the gradient from being too small, and the weight distribution is normalized by the Softmax function, which acts on the V matrix [19], as shown in (4):

$$Attention(Q, K, V) = \mathrm{softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V \tag{4}$$

To capture inherent law of action characteristics from multiple dimensions, MHSA extends the attention mechanism to multiple "heads". Each head pays attention to different subsets of the feature space, splices the output results of h heads, and fuses them through the linear projection matrix W^O to obtain the final multimodal fused feature representation. The calculation process is described by (5):

$$MHSA(Z) = \text{Concat}(head_1, head_2, \cdots, head_h)W^O \quad (5)$$

Each $head_i$ is the result of independently executing the calculation of (4). The features processed by MHSA combine the physical properties of acceleration and angular velocity, and embed the temporal and spatial context information of the action sequence. These features are then sent to the full connection layer for dimensionality reduction, which provides a highly robust input basis for the subsequent physical fitness evaluation and tactical decision optimization based on SIA.

The action recognition experiment uses only the inertial channel as the model input. Eye-tracking, body tracking, and electromyography (EMG) data are not directly fed into the recognition branch. This approach ensures that the model input is consistent with the deployment form and avoids introducing excessive additional variables caused by differences in sampling rates across different sensors. According to the publicly available data, the sampling rates of the different sensors vary significantly: eye-tracking data is sampled at 30 Hz, body tracking at 96 Hz, foot pressure and inertial data at 100 Hz, and EMG data at higher frequencies.

3.3 Improved Sparrow Search Algorithm Based on Chaotic Mapping

Strong global search capabilities are demonstrated by the Sparrow Search Algorithm (SSA), which mimics the sparrow population's foraging and anti-predation behavior [20]. This study introduces Tent chaotic mapping to enhance the standard algorithm and builds an Improved Sparse Search Algorithm (ISSA) that combines chaotic mapping to adaptively optimize the neural network's key parameters to further improve the algorithm's ergodicity in the nonconvex solution space and its ability to jump out of the local optimum.

Let the population size be N, the dimension be D, and the position vector of the i-th sparrow in the t-th iteration be X_i^t. When the early warning value R_2 is lower than the safety threshold ST, it means that the surrounding environment is safe, and the discoverer executes a wide range of global search modes. The populace instantly relocates to the safe region when the early warning value surpasses the threshold. Equation (6) displays the discoverer's location update description:

$$X_{i,j}^{t+1} = \begin{cases} X_{i,j}^t \cdot \exp\left(-\frac{i}{\alpha \cdot T_{max}}\right), if\ R_2 < ST \\ X_{i,j}^t + Q \cdot L, if\ R_2 \geq ST \end{cases} \quad (6)$$

In this equation, t is the current iteration number, T_{max} is the maximum iteration number, $\alpha \in (0,1]$ is a random number, Q is a random number with normal distribution, and L is an all-1 matrix. Joiners follow the discoverer to get food, and their location updating strategy depends on their fitness ranking in the population. The position updating equation of the participant is given by (7):

$$X_{i,j}^{t+1} = \begin{cases} Q \cdot \exp\left(\frac{X_{worst}^t - X_{i,j}^t}{i^2}\right), if\ i > n/2 \\ X_P^{t+1} + \left|X_{i,j}^t - X_P^{t+1}\right| \cdot A^+ \cdot L, otherwise \end{cases} \quad (7)$$

X_P is the optimal position occupied by the discoverer. X_{worst} is the current global worst position. A is a matrix with only 1 or -1 elements, and $A^+ = A^T(AA^T)^{-1}$. Randomly selected scouts in the population are responsible for monitoring the threat of predators. When they perceive the danger, they will quickly approach the global optimal position or conduct anti-predation maneuvers near the current point [21], and their mathematical model is shown in (8):

$$X_{i,j}^{t+1} = \begin{cases} X_{best}^{t} + \beta \cdot \left| X_{i,j}^{t} - X_{best}^{t} \right|, if\ f_i > f_g \\ X_{i,j}^{t} + K \cdot \left(\frac{\left| X_{i,j}^{t} - X_{worst}^{t} \right|}{(f_i - f_w) + \epsilon} \right), if\ f_i = f_g \end{cases} \tag{8}$$

X_{best} is the global optimal position. β is the normal distribution random number of the control step. K is a random variable between [1,1]. f_i, f_g and f_w represent the fitness values of the current individual, global optimal and global worst respectively. The initial population of the conventional SSA algorithm is typically generated using pseudo-random numbers, which results in an unequal distribution of initial solutions in the solution space and lowers the algorithm's ergodic efficiency. Chaotic variables have the characteristics of ergodicity, randomness and regularity, which can effectively improve the diversity of population. Tent chaotic mapping is chosen to replace the conventional logistic mapping in this study because its iterative distribution is more uniform in the [0,1] interval, which can greatly increase the algorithm's search efficiency. Equation (9) defines the generating function of the Tent chaotic sequence.

$$z_{k+1} = \begin{cases} 2z_k, 0 \le z_k < 0.5 \\ 2(1 - z_k), 0.5 \le z_k \le 1 \end{cases} \tag{9}$$

Tent mapping is added to initialize the position of sparrow population, and the chaotic sequence formed by it is mapped back to the solution space, which can ensure that the starting population spans a greater range of parameters. In the optimization process of DNN, ISSA maps the weight matrix and bias vector of the network into the spatial coordinates of sparrows, and takes the prediction error on the training set as the fitness function. The algorithm execution stream is shown in **Table 3**.

Table 3. ISSA pseudo-code fused with chaotic mapping

Algorithm step	Execution description
Input	Population size N, maximum number of iterations T_{max}, early warning threshold ST, discoverer ratio PD, and scout ratio SD
Step 1	The tent chaotic mapping of (9) is used to generate chaotic sequences, and the population position X is initialized by inverse mapping
Step 2	Calculate the fitness value f_i of all individuals, and sort out the current optimal position X_{best} and the worst position X_{worst}
Step 3	While ($t < T_{max}$) do
Step 4	Select the individual of $PD \cdot N$ before fitness as the discoverer, and update the position according to (6)
Step 5	The rest of the individuals, as participants, update their positions according to (7)
Step 6	Randomly select $SD \cdot N$ individuals as scouts, and update their positions according to (8)
Step 7	Calculate the fitness value of the new position, and if it is better than the current best, update X_{best}
Step 8	Check whether the boundary conditions are met, and correct the cross-border individuals
Step 9	$t = t + 1$
Step 10	End While
Output	Output the global optimal position X_{best} (corresponding to the optimal neural network parameters) and its fitness value.

3.4 Multi-objective Optimization Model for Tactical Decision-making under Physical Constraints

This study establishes a multi-objective optimization model that integrates the probability of tactical revenue with the physical labor concept. The overall operation logic of the model is shown in **Fig. 2**, which constructs a dynamic Pareto frontier in the decision-making space and adjusts the aggressiveness of tactical recommendation in real time according to the current fatigue reserve of athletes.

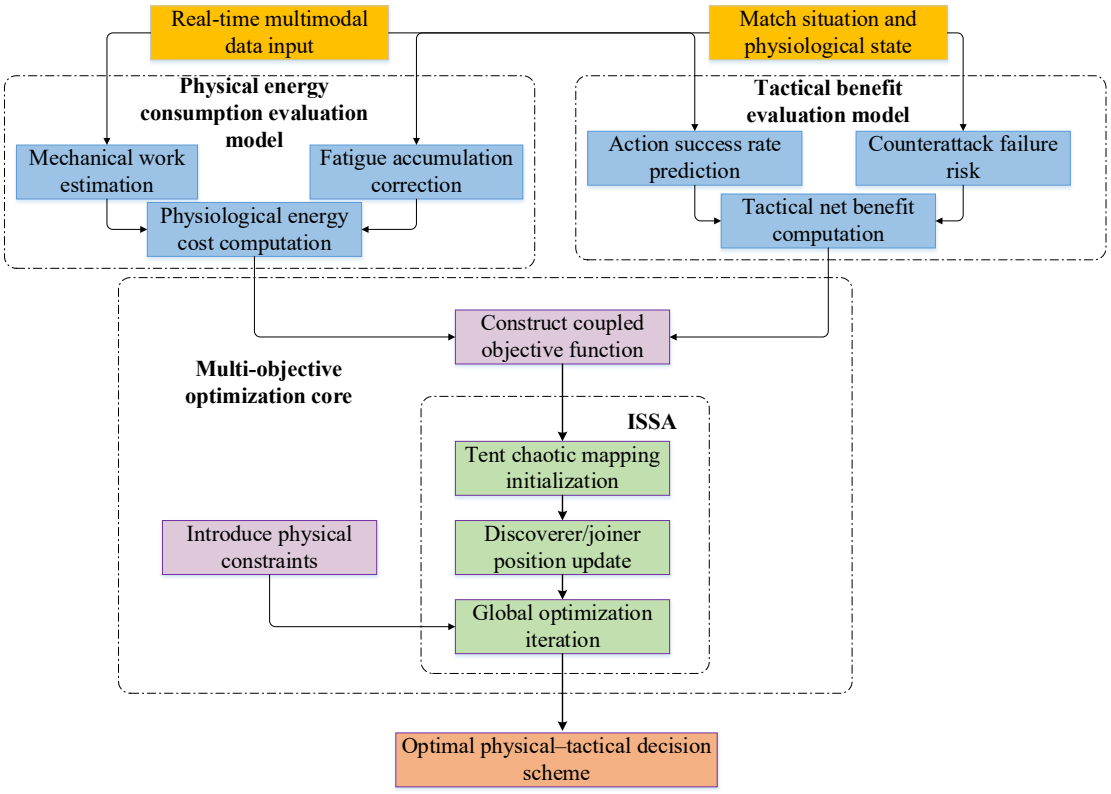

Fig. 2. Logic of multi-objective optimization model of tactical decision under physical constraints

Let the triaxial acceleration in the time window be $a(\tau)$, and after the gravity component is removed, the inertia load proxy in the window is defined as (10):

$$L_t = \frac{1}{\Delta t} \cdot \int_t^{t+\Delta t} \|a(\tau) - g\| \, d\tau \tag{10}$$

Equation (10) represents an inertial intensity-based mechanical load proxy index, which is used to characterize the instantaneous motion load during the execution of actions. Its role is to provide a constraint term under a unified dimension for strategy solving, rather than substituting energy metabolism measurements in exercise physiology experiments. g represents the gravitational acceleration. L_t reflects the average level of action intensity within the window. Let the cumulative load state of individual i be denoted as $H_i(t)$, then the action cost is given by (11):

$$C_i(t) = L_t \cdot (1 + \alpha_i H_i(t) + \beta_i \bar{M}_i(t)) \tag{11}$$

$\bar{M}_i(t)$ represents the normalized mean EMG activation of the subject in the historical window. α_i and β_i are individual parameters. Both of them are obtained by fitting the training set data in the individual calibration stage. The value is determined by minimizing the

cumulative load forecasting error [22].

In the tactical solution stage, the action benefit is given by the combination of scoring probability P_i^{score} and losing risk P_i^{loss} [23]. The corresponding net utility is written as (12):

$$U_i = \lambda_1 \cdot P_i^{score} - \lambda_2 \cdot P_i^{loss} \tag{12}$$

λ_1 and λ_2 are the weight of scoring reward and penalty for losing points, respectively. Usually, when the score is anxious, the value of λ_2 will increase accordingly to reflect the robust strategy.

To find a balance between the conflicting goals of "minimizing energy consumption" and "maximizing tactical benefits", this study constructs a global multi-objective optimization function $OF(i)$. The strategy optimization objective is defined as (13):

$$OF_i = \gamma(t) \cdot \frac{C_i(t)}{\max(C)} - (1 - \gamma(t)) \cdot U_i \tag{13}$$

$\gamma(t)$ is the load preference factor. This factor changes with the current level of residual physical agent. When the residual load margin is low, the model tends to choose the action with less cost. This module is currently used for off-line strategy simulation and scheme comparison, and does not directly indicate that the actual online deployment has been completed. The key parameters involved in the model and their default settings are displayed in **Table 4**.

Table 4. Key parameter setting of tactical decision optimization model

Parameter symbol	Parameter meaning	Default value/range	Note
α	Fatigue correction coefficient	0.15	Adjust according to the endurance test of athletes
β	Heart rate load index	0.02	Based on the maximum heart rate percentage calculation
w_s	Scoring reward weight	1.0	Key points can be raised to 1.2
w_r	Penalty weight for losing points	1.2	Weight increase under defensive strategy
E_{th}	Physical fitness early warning threshold	40%	Percentage of total energy consumption reserve
k	Policy switching slope	0.5	Control the transition rate between radical and conservative
γ_{safe}	Action safety factor	0.85	Hard constraints to prevent sports injuries

4. Result and Discussion

The overall performance comparison of different motion recognition models is shown in **Fig. 3**. The results demonstrate a clear performance gradient between traditional models and stronger sequential models. Bidirectional LSTM (Bi-LSTM) achieves an accuracy of 91.05%. DeepConvLSTM increases to 93.74%. Temporal Convolutional Network (TCN) and InceptionTime reach 94.11% and 95.02%, respectively. Transformer Encoder achieves 95.36%. The proposed method achieves an accuracy of 96.80% and an F1-score of 96.48%. Support Vector Machine (SVM) performs the worst in this task. 1D CNN and LSTM are able to enhance the temporal feature modeling capability. DeepConvLSTM, TCN, InceptionTime,

and Transformer Encoder further improve performance. The proposed method still achieves the highest accuracy, precision, recall, and F1-score, indicating that the dual-stream structure and Multi-Head Self-Attention (MHSA) provide a stable advantage in multimodal inertial sequence modeling. This result shows that the advantages of the proposed method are not limited to comparisons with basic models, as it maintains superior performance even when faced with stronger temporal modeling architectures.

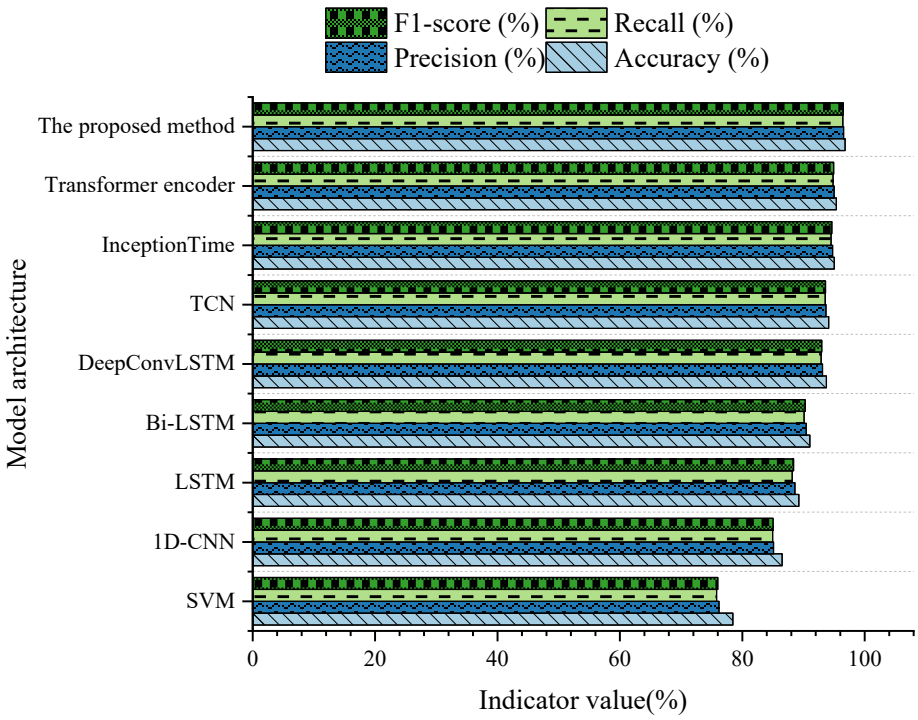

Fig. 3. Comparison of overall performance of different motion recognition models

Fig. 4 displays the suggested model's classification performance on several badminton technical motions. With an average precision of 96.39%, it demonstrates that the model maintains an exceptionally high identification level across the five types of typical technical movements. With a precision of up to 98.45% and an F1-score of 98.17%, the "smash" movement, which is distinguished by a huge feature amplitude and powerful explosive force, achieves the best recognition performance. This suggests that the model's ability to catch high-frequency and high-energy signals is incredibly acute. The "net drop shot" movement is the most difficult to recognize due to its strong concealment and very tiny amplitude. However, the model continues to attain an F1-score of 95.04% and a precision of 95.40%.

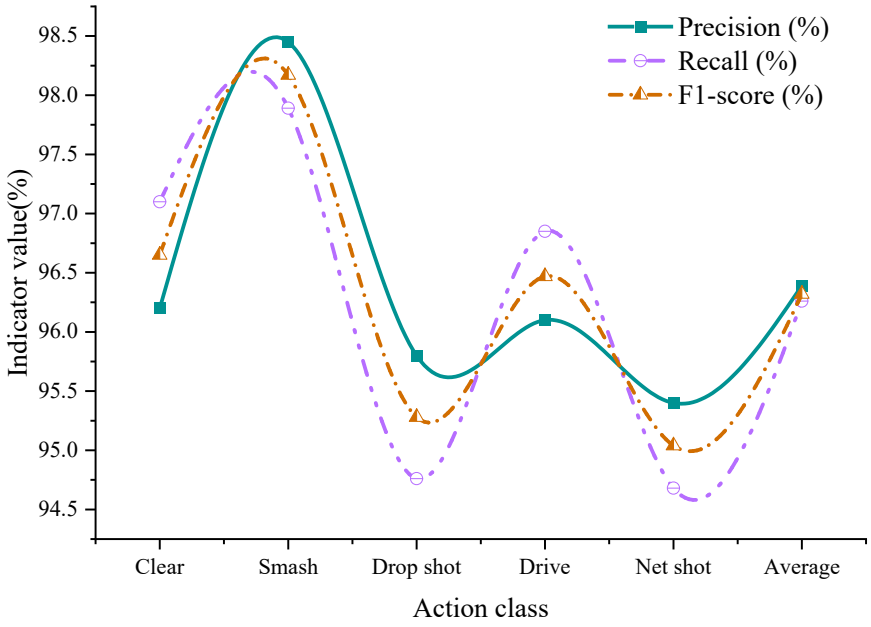

Fig. 4. The classification performance of the proposed model in different badminton technical movements

Table 5 displays the ablation experiment results that each model module provided. It demonstrates that the accuracy of the full dual-stream multimodal fusion model (Model E) reaches 96.80%, which is 8.05% greater than that of Model A (88.75%), which solely employs acceleration data. This indicates that the addition of an angular velocity data stream is required. Although the accuracy of Model C, which only carries out simple data splicing without introducing attention mechanism, is improved to 93.10%, it is still lower than the complete model, which shows that MHSA module plays a key role in feature fusion layer.

Table 5. Ablation experimental results contributed by each module of the model

Experimental setup	Acceleration stream	Angular velocity stream	MHSA module	Accuracy (%)	Lifting range
Model A	√	×	×	88.75	-
Model B	×	√	×	85.40	-3.35%
Model C	√	√	×	93.10	+4.35%
Model D	√	×	√	91.50	+2.75%
Model E (in this study)	√	√	√	96.80	+8.05%

The independent contribution comparison of hyperparameter search strategies is shown in **Table 6**. **Table 6** compares manual tuning, standard SSA, and the improved SSA under the same network structure. The results show that with manual tuning, the accuracy is 95.41%. Standard SSA improves the accuracy to 96.12%. The improved SSA further reaches 96.80%. In terms of loss metrics, the best loss decreased from 0.0189 to 0.0105, the mean loss dropped

from 0.0247 to 0.0118, and the standard deviation reduced from 0.0038 to 0.0009. This result indicates that manual tuning achieves relatively high results, standard SSA can further improve recognition performance, and the improved SSA achieves the best results in terms of accuracy, F1-score, and loss convergence stability.

Table 6. Comparison of independent contributions of hyperparametric search strategies

Hyperparameter strategy	Accuracy (%)	Precision (%)	Recall (%)	F1-score (%)	Best loss	Mean loss	Std. dev
Manual tuning	95.41	95.08	94.93	95.00	0.0189	0.0247	0.0038
SSA	96.12	95.95	95.76	95.85	0.0210	0.0255	0.0032
ISSA	96.80	96.55	96.42	96.48	0.0105	0.0118	0.0009

Fig. 5 demonstrates the comparison of performance indexes of several optimization strategies in model parameter optimization. It shows that ISSA with Tent chaotic mapping is significantly superior to the traditional algorithm in optimization efficiency. ISSA only needs 16 generations to converge, much faster than the 42nd generation of genetic algorithm and the 35th generation of particle swarm optimization, and its optimal fitness value is as low as 0.0105, which is obviously better than the standard SSA's 0.0210. The standard deviation data is only 0.0009, which shows that the algorithm has high stability in many runs and is not easy to fall into local extremum.

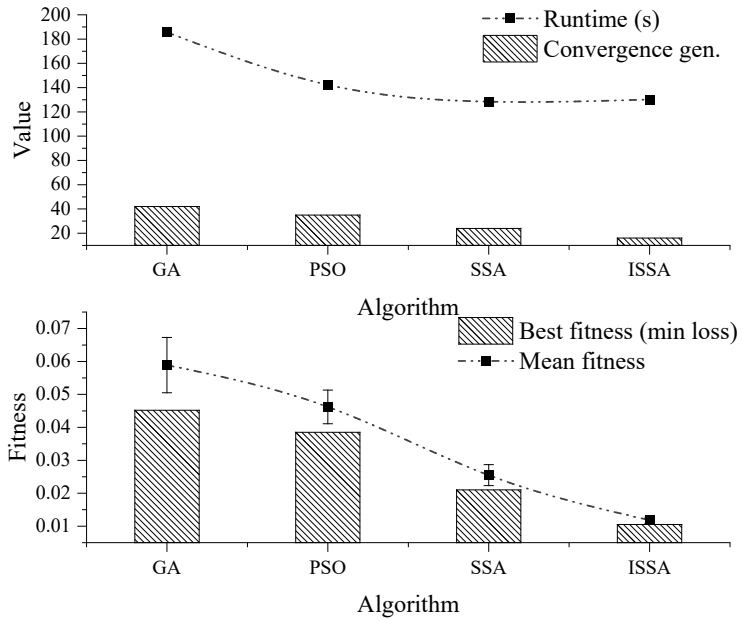

Fig. 5. Comparison of performance indexes of different optimization algorithms in model parameter optimization

The recognition results of the 5 repeated cross-subject experiments are shown in **Table 7**. **Table 7** indicates that the accuracy fluctuates between 96.54% and 96.91%, with a small range of fluctuation. The standard deviations of precision, recall, and F1-score are all controlled within approximately 0.15. This result suggests that the model is not sensitive to changes in cross-subject partitioning, and the single best result is not a random occurrence.

Table 7. Identification results of 5 repeated cross-subject tests

Split ID	Accuracy (%)	Precision (%)	Recall (%)	F1-score (%)
Split-1	96.54	96.28	96.15	96.21
Split-2	96.73	96.49	96.31	96.40
Split-3	96.80	96.55	96.42	96.48
Split-4	96.91	96.68	96.55	96.61
Split-5	96.62	96.37	96.24	96.30
Mean ± Std	96.72 ± 0.14	96.47 ± 0.15	96.33 ± 0.15	96.40 ± 0.15

Table 8 displays the important hyperparameter results of the neural network that ISSA optimized. It demonstrates how the method automatically locks the set of parameters that maximizes the model's performance within the predetermined search space. To guarantee gradient descent's effectiveness and prevent shock, the initial learning rate is set at 0.00142. The first layer has 64 convolution kernels, whereas the second layer has 128. To extract semantic characteristics from low level to high level, it presents a feature pyramid structure with an increasing number of channels layer by layer. The weight attenuation coefficient is set to 4.5e-5, and with the Dropout ratio of 0.35, the regularization ability of the model is effectively enhanced and the over-fitting phenomenon is prevented. These precise parameter settings are the inherent guarantee for the model to achieve a high accuracy of 96.8%.

Table 8. Key hyperparametric results of neural network optimized by ISSA

Hyperparameter name	Search space range	Value selection after optimization
Initial learning rate	[0.0001, 0.01]	0.00142
Batch size	[16, 128]	32
Number of Convolution Kernels-Layer 1	[16, 128]	64
Number of Convolution Kernels-Layer 2	[32, 256]	128
Dropout ratio	[0.1, 0.5]	0.35
Weight attenuation coefficient	[1e-5, 1e-3]	4.5e-5

The mean and standard deviation of the results from the 5 offline tactical simulations are shown in **Table 9**. Compared with the empirical strategy, the optimized strategy demonstrates lower load proxy indicators and higher overall utility across 50 offline scenarios. The average load proxy value decreased from 48.65 to 41.58. The non-forced error rate decreased from 18.50% to 14.20%. The score probability slightly increased from 62.40% to 63.15%. This indicates that, after adding the physical load constraint, the strategy optimization did not simply pursue a conservative approach but instead formed a more robust balance between load and reward. The overall standard deviation is small, suggesting that the offline simulation results are stable across multiple runs.

Table 9. Load agent and utility comparison in off-line tactical simulation scenario

Metrics	Baseline strategy	Optimized strategy
Total rallies	50	50
Avg. load proxy/min	48.65 ± 2.11	41.58 ± 1.74
Avg. power proxy (W)	810.5 ± 31.4	692.4 ± 28.6
Exp. scoring prob. (%)	62.40 ± 1.80	63.15 ± 1.60
Unforced error rate (%)	18.50 ± 1.20	14.20 ± 1.00
Efficiency ratio	1.28 ± 0.07	1.52 ± 0.06

The change of policy distribution under different load agent states is shown in **Fig. 6**. It shows that with the loss of athletes' physical reserves, the tactical recommendation strategy of the model shows an obvious adaptive adjustment trend. In the stage of abundant physical fitness, the model tends to recommend the "kill the ball/assault" tactics with high energy consumption and high income, accounting for 45.5%. However, in the state of severe fatigue (residual physical strength < 40%), the recommended proportion of this kind of tactics dropped sharply to 12.4%, and it is replaced by low-energy "pull-and-hoist control" tactics, and its proportion rises to 65.2%. This distribution changes the average energy consumption coefficient of tactics from 0.88 to 0.48. It objectively reflects that the model can dynamically reshape the tactical structure according to physiological constraints, and ensure that athletes will not be exhausted or make mistakes when they are at the critical point of physical fitness.

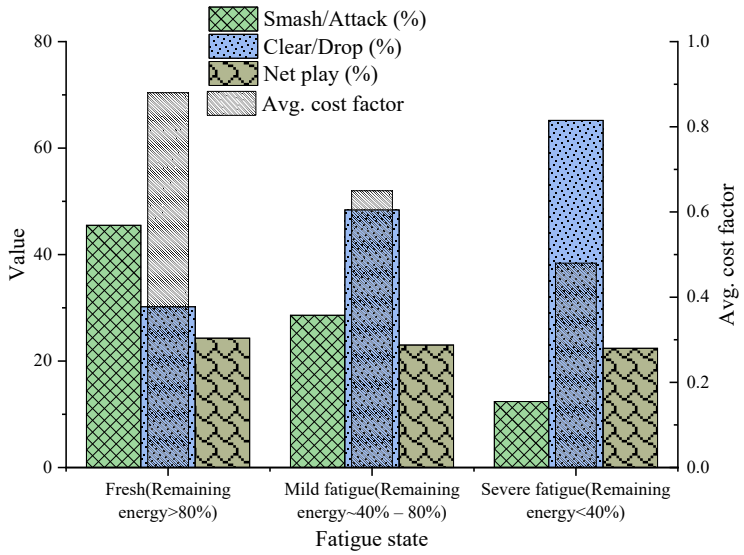

Fig. 6. Policy distribution changes under different load agent states

The current study has not yet completed online validation at the full-match level, nor has it undergone blind evaluations by professional coaching teams or practical field tests with athletes. The load metric used is the mechanical load proxy, which is suitable for engineering modeling but cannot directly replace metabolic energy expenditure measurements. Future work should incorporate more comprehensive physiological indicators, along with real-time latency testing and edge deployment experiments, to further validate the practical value of the system.

5. Conclusion

This study has constructed a multimodal recognition and load-constrained strategy optimization framework for badminton training analysis. The recognition component uses triaxial acceleration and triaxial angular velocity as core inputs, and features are fused through a dual-stream 1D convolutional neural network with MHSA. The improved SSA is applied for hyperparameter optimization, achieving better convergence quality and recognition performance in independent comparisons. Cross-subject experiments show that the model can stably recognize multiple badminton technical movements. Offline tactical simulation results indicate that after incorporating load constraints, the strategy optimization can balance the

distribution between movement rewards and load proxies. The current conclusions are primarily based on offline recognition and offline simulation, with online deployment, system latency, coaching assessments, and real-world tests still needing further validation. This framework can serve as the foundation for the development of future intelligent training systems, though it should not be equated with a fully deployed real-time closed-loop assistance platform.

Funding

This paper was supported by the 2023 Harbin Institute Youth Doctoral Research Launch Fund Project: Research on the Integration of "Big Health" Concept and Ice and Snow Sports Development, Project Number: HUDF2023109; Research Project on Higher Education Teaching Reform in Heilongjiang Province: Internal and External Cultivation+Bacon Casting Soul: Research on the Construction of Chinese Style Yoga Curriculum Project Number: SJGZB2024251; Basic Research Business Project of Heilongjiang Provincial Undergraduate Universities in 2023: Value and Implementation Path of Inner Development of Rural Sports in Heilongjiang Province from the Perspective of Life Education Project Number: 1453YB006.

References

[1] Y. Sheng, C. Liu, Q. Yi, W. Ouyang, R. Wang, and P. Chen, "Predicting badminton outcomes through machine learning and technical action frequencies," *Scientific Reports*, vol.15, no.1, Mar. 2025. Article (CrossRef Link)

[2] W. Liu, Y. Zhu, W. Guo, X. Wang, and S. Yu, "Gaming Tree Based Evaluation Model for Badminton Tactic Benefit Analysis and Prediction," *Applied Sciences*, vol.13, no.13, Jul. 2023. Article (CrossRef Link)

[3] F. Xu and W. Zhu, "Evalution of neurodiagnostic insights for enhanced evaluation and optimization of badminton players' physical function via data mining technique," *SLAS Technology*, vol.29, no.4, Aug. 2024. Article (CrossRef Link)

[4] J. An and Y. Zhao, "Research on the Application of Apriori Algorithm in the Teaching of Ball Sports Techniques and Tactics," *Scalable Computing: Practice and Experience*, vol.25, no.1, pp.383-394, Jan. 2024. Article (CrossRef Link)

[5] Y. khan, A. Abdalrahman and R. Ullah, "Predicting Badminton Player Performance: Integrating Physical, Psychological, and Technical Factors Using Machine Learning," *Spectrum of Engineering Sciences*, vol.3, no.1, pp.1-20, Jan. 2025. Article (CrossRef Link)

[6] Y. Xie, X. Xie, H. Xia, and Z. Zhao, "Algorithmic Study on Position and Movement Method of Badminton Doubles," *Scientific Programming*, vol.2022, Jan. 2022. Article (CrossRef Link)

[7] I. Ghosh, S. R. Ramamurthy, A. Chakma, and N. Roy, "Sports analytics review: Artificial intelligence applications, emerging technologies, and algorithmic perspective," *WIREs Data Mining and Knowledge Discovery*, vol.13, no.5, Sep./Oct. 2023. Article (CrossRef Link)

[8] Y. Jia, N. A. Abdullah, H. Eliza, Q. Lu, D. Si, H. Guo, and W. Wang, "A narrative review of deep learning applications in sports performance analysis: current practices, challenges, and future directions," *BMC Sports Science, Medicine and Rehabilitation*, vol.17, no.1, Aug. 2025. Article (CrossRef Link)

[9] Z. Zhou, H. Zhang, and M. Effatparvar, "Improved sports image classification using deep neural network and novel tuna swarm optimization," *Scientific Reports*, vol.14, no.1, Jun. 2024. Article (CrossRef Link)

[10] S. Y. See, R. Paramesran, M. F. Hassan, G. Krishnasamy, "A Comparative Analysis of Deep Learning Models and Gradient Computation for Rally Detection in Badminton Videos," *SN Computer Science*, vol.6, no.5, 2025. Article (CrossRef Link)

[11] C. He and M. Zhang, "Deep learning neural network-assisted badminton movement recognition and physical fitness training optimization," *Heliyon*, vol.10, no.20, pp.1-24, Oct. 2024. Article (CrossRef Link)
[12] H. Yuan, Y. Wang, K. Yang, Y. Bin, "Prediction model and technical and tactical decision analysis of women's badminton singles based on machine learning," *PLoS ONE*, vol.19, no.11, Nov. 2024. Article (CrossRef Link)
[13] X. Hu, J. Li, and R. Cai, "Knowledge Graph-Based Badminton Tactics Mining and Reasoning for Badminton Player Training Pattern Analysis and Optimization," *International Journal of Advanced Computer Science and Applications (IJACSA)*, vol.15, no.10, Oct. 2024. Article (CrossRef Link)
[14] J. Wu, D. Liu, Z. Guo, and Y. Wu, "RASIPAM: Interactive Pattern Mining of Multivariate Event Sequences in Racket Sports," *IEEE Transactions on Visualization and Computer Graphics*, vol.29, no.1, pp.940-950, Jan. 2023. Article (CrossRef Link)
[15] S. Tan and T. T. Teoh, "Predicting Shot Accuracy in Badminton Using Quiet Eye Metrics and Neural Networks," *Applied Sciences*, vol.14, no.21, Nov. 2024. Article (CrossRef Link)
[16] X. Lv, Y. Tao, and Y. Xue, "Research on Personalized Exercise Volume Optimization in College Basketball Training Based on LSTM Neural Network with Multi-Modal Data Fusion Intervention," *Applied Sciences*, vol.15, no.16, Aug. 2025. Article (CrossRef Link)
[17] Y. Chen, G. Wang, B. Yin, C. Ma, Z. Wu, and M. Gao, "Basketball team optimization algorithm (BTOA): a novel sport-inspired meta-heuristic optimizer for engineering applications," *Scientific Reports*, vol.15, no.1, Jul. 2025. Article (CrossRef Link)
[18] K. Liu, X. Li, and C. A. Rochester, "Relationship between Physical Training and Tactical Training in Sports Training Relying on Boosting and Bagging Algorithms," *Scientific Programming*, vol.2022, Apr. 2022. Article (CrossRef Link)
[19] Z. Ding, "Personalized Optimization of Sports Training Plans Based on Big Data and Intelligent Computing," *Scalable Computing: Practice and Experience*, vol.26, no.3, pp.1395-1402, Apr. 2025. Article (CrossRef Link)
[20] Y. Zhang, Z. Ning, and Y. Song, "Design and optimization of intelligent sports IoT system based on big data and quantum computing," *Peer-to-Peer Networking and Applications*, vol.18, no.4, Jun. 2025. Article (CrossRef Link)
[21] Q. Wu, Y. Sun, L. Gao, and W. Yin, "Evaluation of sports training methods and technical characteristics based on multi-dimensional driving fuzzy intelligent computing," *PLoS One*, vol.20, no.6, Jun. 2025. Article (CrossRef Link)
[22] Y. Li and M. N. Jawis, "Construction and effect of interactive teaching methods in the college badminton classroom using Fuzzy-Assisted virtual reality modeling: A case study in Chinese universities," *Journal of Fuzzy Extension and Applications*, vol.5, no.2, pp.199-222, 2024. Article (CrossRef Link)
[23] P. Wang, "Modeling of Badminton Intelligent Teaching System Based on Neural Network," *Wireless Communications and Mobile Computing*, vol.2021, May 2021. Article (CrossRef Link)

Yanjing Zheng received her bachelor's degree in Social Sports and master's degree in Sports Humanities and Sociology from Harbin Institute of Physical Education in 2007 and 2011, respectively. She commenced her doctoral studies in Sports Science at the College of Physical Education, Woosuk University, South Korea, in September 2019. She currently works at Harbin University, where her research interests focus on physical education teaching and winter sports.

Ye Han is a Professor who received her master's degree in Athletic Training from Harbin Institute of Physical Education in 2021. Her primary research interests are yoga courses and public health and wellness, and she currently works at Harbin University.

Wei Wang received her bachelor's degree in Sports Performance from Harbin Institute of Physical Education in 2014, and her master's degree in Physical Education and Training from the same university in 2017. She currently works at Harbin University, where her research primarily focuses on physical education teaching.

Xiao Yu received her master's degree in Physical Education and Training from Harbin Institute of Physical Education in 2010. He currently works at Mudanjiang Normal University, where his primary research focuses on physical education.