ingegneria sismica
International Journal of Earthquake Engineering
Anno XLIII - Num. 2 - 2026

Personalized Exercise Training Program Optimization and real-time Feedback System Driven by Reinforcement Learning

Xuexiao Li[1,*]

[1] Henan Judicial Police Vocational College, Zhengzhou, Henan, 450000, China

SUMMARY: *With the development of wearable devices and smart fitness platforms, exercise training is moving from empirical prescription to data-driven personalized decision making. Aiming at the problem that the traditional training scheme is difficult to adjust in real time with the change of individual state, fatigue and action quality, this paper constructs a personalized exercise training scheme optimization and real-time feedback system driven by reinforcement learning. The system fuses multi-source data such as heart rate, IMU signal, action trajectory, training log and subjective feedback, generates training state vector through time alignment, normalization processing, feature fusion and exercise ability portrait, and completes the dynamic decision of training item, load intensity, interval time and feedback type based on PPO policy network. Based on the 8-week training data of 60 subjects, the results show that the state recognition Accuracy of PPO model reaches 94.3%, the F1-score reaches 93.5%, the RMSE and MAE of training load prediction are reduced to 4.2 and 3.2 respectively, and the delay of system complete feedback link is 94 ms. The research provides technical support for real-time optimization and safety control of personalized sports training.*

KEYWORDS: *Reinforcement learning; Personalized exercise training; PPO algorithm; Real-time feedback system*

1 Introduction

1.1 Practical needs of personalized exercise training program optimization

With the popularity of wearable devices, mobile terminals and intelligent fitness platforms, the exercise training process can continuously generate multi-dimensional data such as heart rate, step frequency, speed, joint Angle, energy consumption and training completion rate, which provides a data basis for the optimization of personalized training programs [1]. Traditional sports training programs mostly rely on coach experience or fixed templates, and it is difficult to dynamically adjust training intensity, recovery cycle and action arrangement with individual physical status, fatigue level and target changes, which is prone to problems such as insufficient training stimulus, overload or action execution deviation [2]. Especially in mass fitness, rehabilitation training and competitive auxiliary training scenarios, users' physical conditions, exercise foundation, training preference and adaptability are obviously different, and a single training prescription is difficult to ensure long-term effects [3]. Building an intelligent system that can perceive the training status in real time, automatically

*18569903159@163.com
https://doi.org/10.65102/is2026963 Published: 30 April, 2026

adjust the training content and feedback the execution effect has become an important requirement to improve the scientificity, safety and continuous participation of sports training [4].

1.2 Technical basis of reinforcement learning to drive dynamic decision making in sports training

The optimization of sports training scheme has obvious sequential decision-making characteristics, and the current training load will affect the subsequent physical recovery, movement quality and training benefits. The system needs to maintain a balance between short-term training effect and long-term ability improvement. Through the interaction mechanism of "state-action-reward", reinforcement learning can select the training action, load intensity, group number arrangement and recovery strategy according to the real-time state of the user, and continuously revise the decision-making strategy by using the feedback results [5, 6]. Compared with static recommendation algorithms, reinforcement learning is more suitable to deal with dynamic changes and individual differences during training. The PPO algorithm has the characteristics of stable policy update, high sample utilization efficiency and suitable for continuous interaction scenarios, which can be used to construct a personalized training scheme optimization model [7]. By integrating heart rate changes, fatigue indicators, action completion quality and user subjective feedback, the system can form a closed loop of training decision-making for sustainable updating, thereby improving the adaptability and real-time response ability of training programs [8].

1.3 Main Contributions and research objectives of this paper

Focusing on the optimization of personalized exercise training scheme and real-time feedback requirements, this paper constructs an intelligent training system driven by reinforcement learning, which integrates multi-source exercise data perception, individual ability portrait, PPO policy network and online feedback update mechanism into a unified framework. The goal of this paper is to generate the motion state vector by using wearable sensor data, training behavior data and user feedback information, establish a training scheme optimization decision model for individual differences, and realize the dynamic adjustment of training load, action prompt and recovery suggestion through real-time feedback mechanism. The main contributions of this paper are reflected in three aspects: firstly, the multi-source data fusion and state representation process for motion training scenes is constructed; The generation and optimization mechanism of personalized training scheme based on PPO policy network was designed. The closed-loop of online update driven by training execution feedback is established to improve the training effect, the efficiency of action correction and the ability of fatigue risk warning, which provides technical support for the real-time deployment of intelligent sports training system.

2 Literature Review

The existing research provides a clear technical basis for the optimization of personalized exercise training programs and the construction of real-time feedback systems. Spilz and Munz applied deep learning structure to automatic assessment of functional action screening, proving that action quality recognition based on visual or sensing data can reduce the subjectivity of manual assessment and provide algorithmic support for training action correction [9]. Focusing on the problem of fatigue detection in real running scenarios, Marotta et al. compared the influence of different IMU sensor configurations on the recognition effect of machine learning, indicating that the low-invasiveness wearable acquisition method can obtain stable motion state characteristics while ensuring user experience [10]. Van Hooren et al. verified the influence of wearable real-time feedback on running injury and sports performance through randomized controlled experiments, indicating that real-time feedback mechanism helps to improve training behavior and reduce the risk caused by unreasonable load [11]. Kakhi et al. further summarized the trend of fatigue monitoring from the perspective of the integration of wearable devices and artificial intelligence, and pointed out that fatigue recognition needs to combine physiological signals, motion characteristics and intelligent models for comprehensive judgment [12].

In terms of data acquisition and motion state perception, Kristoffersson and Linden's system reviewed the application of wearable sensors in physical activity monitoring, and emphasized the important value of multimodal data such as heart rate, acceleration, posture and activity intensity for motion process modeling [13]. Migliaccio et al. discussed the influence of wearable technology on the marginal benefit of sports performance from the perspective of competitive performance improvement, indicating that small training differences can also be continuously captured and optimized through data monitoring [14]. Morouco combines wearable technology with movement development and human biomechanical analysis, emphasizing its application potential in movement pattern recognition and exercise ability assessment [15]. Alzahrani and Ullah focus on precision health monitoring in sports performance and point out that wearable devices combined with advanced biomechanical analysis can provide more fine-grained data basis for individualized training load control and risk identification [16].

In terms of human activity recognition and machine learning modeling, Uddin and Soylu combined wearable sensors, discriminant analysis and LSTM neural structure learning to realize human activity recognition, which reflected the advantages of temporal model in motion state discrimination [17]. Papaleonidas et al. obtained high activity recognition accuracy by using raw signals of wearable devices and machine learning methods, indicating that raw signals without complex artificial feature design can also form effective representations through model learning [18]. Sports injury prediction research is also closely related to personalized training optimization. Van Eetvelde et al. conducted a systematic review of machine learning methods in sports injury prediction and prevention, and pointed out that training load, historical injury and body state variables are important inputs for risk modeling [19]. Majumdar et al. applied machine learning to football injury understanding and prediction and showed that the algorithm model could identify potential risk patterns from complex training data [20]. Carey et al. further discussed the practical application and methodological challenges of sports injury prediction, and proposed that model generalization, data quality and clinical interpretability are still key issues in practical deployment [21]. Leckey et al. summarized the evidence of sports injury risk prediction through a scope review and emphasized that the fusion of multi-source training load data and machine learning still needed to improve stability and transferability [22]. Tsilimigkras et al. combined machine learning and training load analysis to optimize football injury risk assessment, which provided reference for real-time load regulation and early warning mechanism design [23]. In general, the existing research has accumulated important results in action recognition, fatigue monitoring, real-time feedback and damage prediction, but most of the work is still focused on state recognition or risk assessment, and insufficient attention is paid to the continuous decision optimization of training programs. Reinforcement learning can incorporate the motion state, training action, feedback reward and long-term effect into a unified decision-making framework, so it is suitable for building an individual difference oriented exercise training program optimization and real-time feedback system.

3 Personalized exercise training program optimization and real-time feedback system construction driven by reinforcement learning

3.1 Overall architecture of personalized exercise training program optimization and real-time feedback system

The personalized training scheme optimization and real-time feedback system is constructed with the main line of "data perception, state modeling, strategy decision, training execution and feedback update". The goal is to dynamically generate training schemes according to individual physical state, action quality, fatigue change and training objectives during continuous training. The bottom layer of the system consists of wearable sensors, mobile terminals and training logs, which collect data such as heart rate, step frequency, acceleration, attitude Angle, action trajectory, training completion rate and subjective fatigue score. The middle layer completed data cleaning, feature extraction, multi-modal fusion and state vector generation. The decision-making layer took the PPO policy network as the core, and output the training action, load intensity, group number, interval time and recovery suggestions according to the current motion state. The feedback layer formed a closed-loop control through the mobile tip, action correction, fatigue warning and the kanban board of the coach. The deployment layer is responsible for experience caching, policy updating, delay monitoring and model service to ensure that the system can continuously optimize training decisions in real-time scenarios. The overall system architecture is shown in Figure 1.

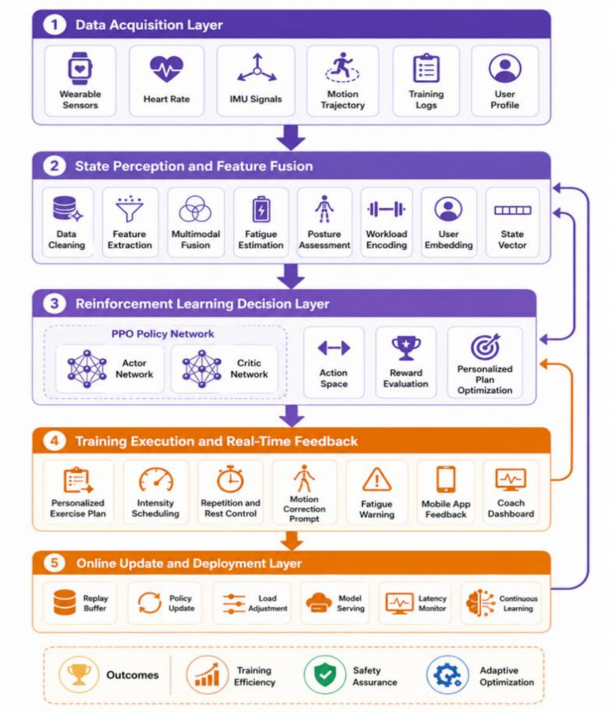

Figure 1: Overall architecture diagram of personalized exercise training scheme optimization and real-time feedback system

In the process of system operation, the original motion sensing data at the TTH training interaction time is set as x_t, the individual static information is u_t, and the historical training record is h_t. The current motion state vector s_t is generated by the feature fusion function, and the calculation process is as follows:

$$s_t = \Phi(x_t, u_t, h_t) \tag{1}$$

where, s_t represents the motion state vector input to the reinforcement learning model at time t, x_t represents the real-time motion data composed of heart rate, IMU signal and action trajectory, u_t represents the individual information such as the user's age, weight, training level and motion goal, and h_t represents the historical training load, completion and recovery records. $\Phi(\cdot)$ represents the multi-source feature fusion mapping function. This state vector encodes individual differences and real-time training performance uniformly, and provides the input basis for subsequent policy network decisions.

According to the state vector, the PPO policy network outputs the training decision action, and the action set includes the selection of training items, the load level, the number of training groups, the number of repetitions, the interval duration and the action correction method. The system selects the current optimal training action under the premise of satisfying the safety constraints, and the decision-making process is as follows:

$$a_t = \arg\max_{a \in \mathcal{A}_t} \pi_\theta(a|s_t) \tag{2}$$

where, at represents the training decision action output by the system at time t, at represents the set of optional training actions in the current state, and $\pi_\theta(a|s_t)$ represents the probability that the PPO policy network with parameter θ selects action A A_t state st. The formula reflects the core decision logic of the system from "state recognition" to "scheme generation", so that the training scheme no longer depends on a fixed template, but dynamically changes according to the individual state.

In order to achieve continuous optimization of the training scheme, the system converts the training completion rate, action quality, heart rate recovery, fatigue change, and user feedback into reward signals, which are used for policy parameter update. The online update goal can be expressed as follows:

$$\theta_{k+1} = \theta_k + \eta \nabla_\theta J(\theta_k) \tag{3}$$

where θ_k represents the policy network parameters before the KTH update round, θ_{k+1} represents the updated policy network parameters, η represents the learning rate, $J(\theta_k)$ represents the expected payoff function formed by the current policy under training feedback, and $\nabla_\theta J(\theta_k)$ represents the gradient of the expected payoff to the policy parameters. Through the update process, the system can continuously modify the strategy according to the user's training performance, so that the subsequent training scheme is more in line with the individual's physical changes and safety control requirements.

3.2 Multi-source motion data acquisition and fusion and real-time perception of training status

Multi-source exercise data collection and fusion is a basic link in the optimization of personalized training scheme. Its role is to transform heterogeneous data scattered in wearable devices, mobile terminals, training equipment and user logs into state information that can be called by reinforcement learning models. The data collected by the system mainly include heart rate, heart rate variability, acceleration, angular velocity, step frequency, action trajectory, joint posture, training duration, number of training groups, interval time, subjective fatigue score and training completion rate. Since the sampling frequency, timestamp accuracy and data scale of different devices are not consistent, the system needs to complete time alignment, outlier removal, missing compensation and scale normalization before entering the state modeling, so that the motion data can express the individual training state in a unified time window.Before the multi-source motion data enters the reinforcement learning decision model, it needs to go through the steps of unified collection, preprocessing, dynamic weighted fusion and state recognition, and its processing flow is shown in Figure 2.

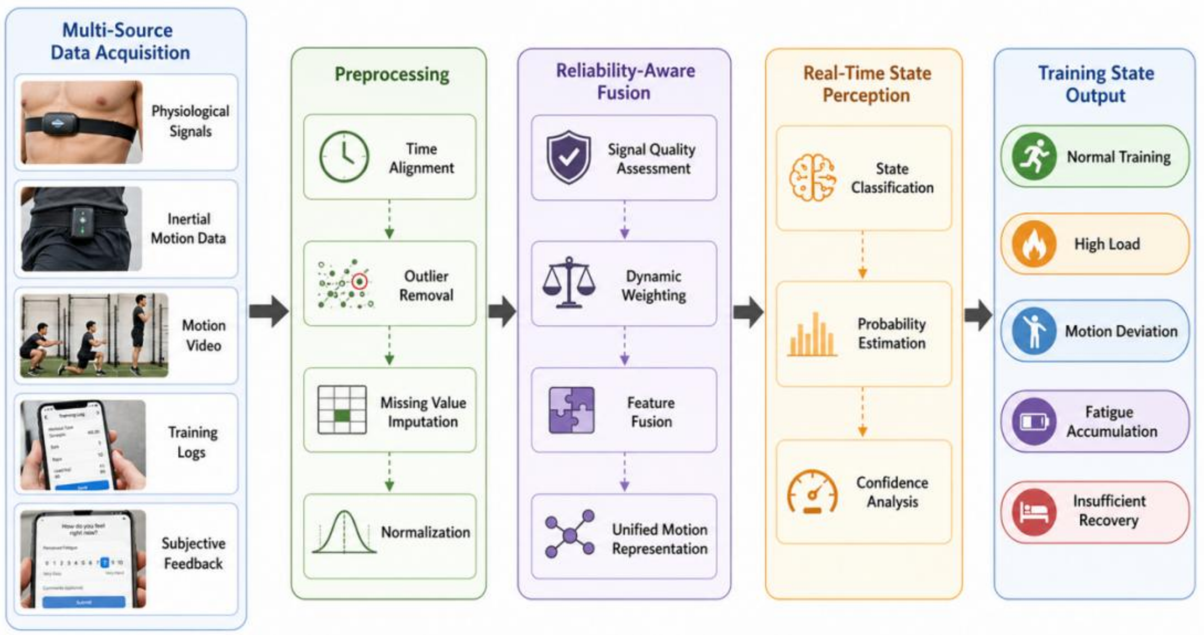

Figure 2: Flow chart of multi-source motion data acquisition fusion and real-time perception of training status

Let the data collected by the m-th motion data source at the original sampling time τ be $z_\tau^{(m)}$, and the feature after time alignment and normalization is expressed as follows.

$$\hat{z}_t^{(m)} = \mathrm{Norm}\left(\mathrm{Align}\left(z_\tau^{(m)}, t\right)\right) \tag{4}$$

where, $\hat{z}_t^{(m)}$ represents the standardized characteristics of the m-th data source at a unified time t, $z_\tau^{(m)}$ represents the original sampled data, $\mathrm{Align}(\cdot)$ represents the time synchronization processing function, and $\mathrm{Norm}(\cdot)$ represents the normalization processing function. Through this processing, heart rate, IMU signal, action trajectory and training log can enter the same calculation scale, reducing the interference of device differences on the state recognition results.

In the actual training process, the reliability of different data sources will change with the motion scene. For example, the IMU signal in high-intensity jump training better reflects the action impact, while the heart rate change in endurance running training better reflects the load bearing capacity. Therefore, the system introduces data source reliability weight to dynamically allocate different modal features:

$$\omega_t^{(m)} = \frac{\exp\left(\kappa_t^{(m)}\right)}{\sum_{j\in\mathcal{M}} \exp\left(\kappa_t^{(j)}\right)} \tag{5}$$

where, $\omega_t^{(m)}$ represents the fusion weight of the m-th data source at time t, $\kappa_t^{(m)}$ represents the reliability score of the data source, M represents the set of all motion data sources, and j represents the index of any data source in the set. The reliability score can be calculated by the signal integrity, noise level, device connection stability and training task correlation, so as to avoid the deviation of training state judgment caused by single sensor anomaly.

After completing the weight calculation, the system compressed the different modal features into a unified motion fusion representation for subsequent training state perception and reinforcement learning decision making. The fusion vector is calculated as follows:

$$e_t = \sum_{m\in\mathcal{M}} \omega_t^{(m)} \hat{z}_t^{(m)} \tag{6}$$

where, e_t represents the multi-source motion fusion feature vector at time t, $\omega_t^{(m)}$ represents the dynamic weight of the m-th data source, and $\hat{z}_t^{(m)}$ represents the modal features after alignment and normalization. The fusion vector not only retains the information of exercise intensity, physiological load and action execution, but also provides compact data input for subsequent training state recognition and reduces the computational overhead of real-time systems.

Real-time perception of training states further maps the fused features into interpretable state categories, including normal training, overloading, action deviation, fatigue accumulation, and insufficient recovery. The system uses a lightweight neural classification layer to complete state probability estimation, and its expression is as follows:

$$\hat{y}_t = \mathrm{Softmax}\left(W_y e_t + b_y\right) \tag{7}$$

where, $\hat{y}_t$ represents the probability distribution of the training state at time t, W_y represents the weight matrix of the state recognition layer, by represents the bias vector, and e_t represents the multi-source motion fusion feature vector. The maximum probability category output by the model is used as the real-time training state judgment result, and the state probability distribution is retained to provide more fine-grained uncertainty information for the PPO policy network.

Through the above process, the system can transform the complex, continuous and heterogeneous motion data into stable training state input, and realize the automatic processing from the original sensing signal to the training state perception. This link not only supports the subsequent generation of motor ability profile, but also provides real-time and credible data basis for the optimization decision of training scheme.

3.3 Motor ability profile and state vector generation for individual differences

Individual differences in exercise training are not only reflected in the static attributes such as age, weight, gender and training goal, but also reflected in the dynamic characteristics such as heart rate recovery speed, action stability, training load tolerance, fatigue accumulation speed and training preference. In order to make the PPO policy network generate a training scheme more suitable for the individual state, the system constructs the motor ability profile based on the multi-source data fusion results, and encodes the profile information, real-time training status and historical training performance into a reinforcement learning state vector. The core goal of this process is to convert "who the user is, what state he is in, how much training stimuli he is suitable for, and what feedback intervention he needs" into a computable vector representation. The core features of motor ability portrait are shown in Table 1.

Table 1: Composition table of core features of motor ability portrait

Feature Category	Collected Indicators	Computational Processing	Profile Meaning	Model Function
Basic Physical Features	Age, height, weight, training years	Structured encoding and standardization	Describes the user’s basic physical conditions	Defines the initial training intensity range
Physiological Response Features	Heart rate, heart rate recovery, heart rate variability	Sliding-window statistics and trend extraction	Reflects cardiopulmonary load tolerance	Determines whether the training stimulus is excessive
Motion Execution Features	Posture angle, joint trajectory, motion completion quality	Trajectory smoothing and motion scoring	Represents motion stability and standardization	Supports the generation of motion correction feedback
Training Load Features	Sets, repetitions, rest interval, training duration	Load accumulation and intensity stratification	Describes recent training pressure level	Controls the increment range of subsequent training load
Fatigue Recovery Features	Subjective fatigue, sleep quality, recovery score	Multi-indicator weighted fusion	Determines whether recovery is sufficient	Triggers load reduction or rest recommendations
Training Preference Features	Preferred training items, completion rate, withdrawal records	Behavior frequency statistics and preference encoding	Reflects the user’s tendency for sustained participation	Improves program acceptance and execution rate

It can be seen from Table 1 that the motor ability portrait does not only record static body information, but comprehensively depict the user's physiological adaptation, action performance and behavior feedback in the training process, so as to provide more stable prior constraints for personalized scheme optimization.After the core features are determined, the system further fuses the basic body features, historical motion representation, stage training performance and real-time state information into a state vector that can be called by the PPO policy network. The generation mechanism is shown in Figure 3.

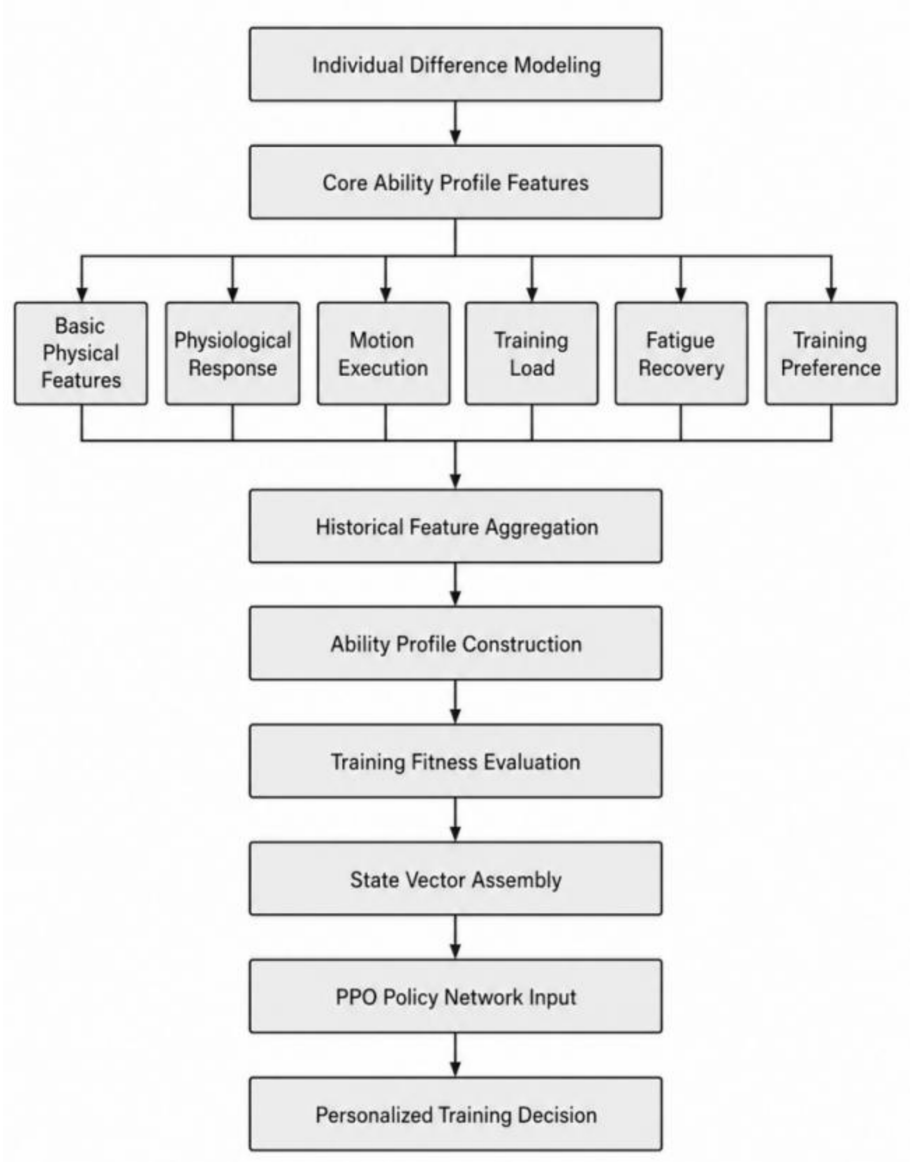

Figure 3: Motor ability portrait and state vector generation mechanism diagram for individual differences

In order to avoid the single training data fluctuation affecting the portrait judgment, the system first aggregates the historical fusion features by a time window. Let the length of the history window be Lw, and the historical motion at time t be characterized as follows:

$$\bar{e}_t = \sum_{\ell=0}^{L_w-1} \alpha_\ell \, e_{t-\ell} \tag{8}$$

where, $\bar{e}_t$ represents the historical motion aggregation feature at time t, $e_{t-\ell}$ represents the multi-source motion fusion feature at the previous ℓ time steps, α_ℓ represents the attenuation weight at different historical moments, and L_w represents the sliding window length. The formula can highlight the recent training state, while retaining a certain historical trend, making the profile update smoother.

On this basis, the system maps individual static information, historical motion representation and phasic training performance into a motor ability portrait vector:

$$g_t = \sigma\left(W_g[u_t \parallel \bar{e}_t \parallel v_t] + b_g\right) \tag{9}$$

where, gt represents the motor ability profile vector at time t, ut represents the user basic attribute vector, $\bar{e}_t$ represents the historical motion aggregation feature, v_t represents the stage training performance vector, ‖ represents vector splicing operation, W_g and b_g represent the weight matrix and bias vector of the portrait generation layer, σ(·) represents the nonlinear activation function. The profile vector provides individual difference information for the policy network to avoid different users being assigned the same training intensity.

To further characterize the current training stimuli that are affordable to the user, the system calculates a training fitness score:

$$\psi_t = \sigma\left(w_\psi^T[g_t \parallel e_t \parallel \hat{y}_t] + b_\psi\right) \tag{10}$$

where ψ_t represents the training fitness score at time t, w_ψ^T represents the transposed weight vector of the fitness evaluation layer, gt represents the motor ability profile vector, e_t represents the current multi-source motion fusion feature, $\hat{y}_t$ represents the probability distribution of the training state, and b_ψ represents the bias term. The higher the score, the more suitable the user is for higher intensity training. When the score is low, the system will tend to reduce the load or increase the recovery proposal.

Finally, the state vector required by the reinforcement learning model is composed of real-time motion fusion features, ability portraits, training state probabilities, and fitness scores:

$$s_t = e_t \parallel g_t \parallel \hat{y}_t \parallel \psi_t \tag{11}$$

where, s_t represents the final state vector of the input PPO policy network, e_t represents the current motion fusion feature, gt represents the motor ability profile vector, $\hat{y}_t$ represents the training state recognition result, and ψ_t represents the training fitness score. Through this state vector, the system can simultaneously perceive the user's immediate exercise performance and long-term ability difference, and provide unified decision-making input for subsequent training item selection, load distribution, action correction and fatigue warning.

3.4 Optimization decision model of training scheme based on PPO policy network

The training scheme optimization decision model based on PPO policy network is the core of the system to realize personalized training scheme generation. The process of exercise training has obvious dynamic interactive characteristics, and the current training action, load intensity and interval arrangement will affect the physical recovery, action quality and fatigue risk in the next stage. Therefore, it cannot rely only on static recommendation or fixed training template. The state vector st generated in Section 3.3 is used as the input of the policy network, and the training item, load level, group number, repetition number, interval duration and feedback type are used as the action output. The reward signal is formed by training completion rate, action quality, heart rate recovery, fatigue change and safety constraints. This enables the model to optimize the training scheme in continuous interaction.

The model adopts the Actor-Critic structure. The Actor network is responsible for outputting the training action probability distribution, and the Critic network is responsible for estimating the current state value. Actor shares the front-end feature extraction layer with Critic, which enables the model to learn the motion state representation and train the decision rule simultaneously. For the TTH training interaction, the policy network outputs the probability distribution of the action at according to the state st as follows:

$$\pi_\theta(a_t|s_t) = \mathrm{Softmax}(W_\theta f_\theta(s_t) + b_\theta) \tag{12}$$

where, $\pi_\theta(a_t|s_t)$ represents the probability of selecting the training action at under the state st, $f_\theta(s_t)$ represents the high-dimensional feature mapping result of the policy network to the state vector, W_θ represents the action output layer weight matrix, b_θ represents the action output layer bias term, θ represents the Actor network parameters. This formulation is used to generate training scheme decisions that enable the system to select differentiated training actions according to different user states instead of adopting a uniform training prescription.

The Critic network is used to evaluate the long-term training gain of the current state and provide a baseline value for policy update. The state value function is defined as follows:

$$V_\varphi(s_t) = W_\varphi f_\varphi(s_t) + b_\varphi \tag{13}$$

where, $V_\varphi(s_t)$ represents the value estimation result of state st, $f_\varphi(s_t)$ represents the feature mapping of the value network to the state vector, W_φ represents the weight matrix of the value output layer, b_φ represents the bias term of the value output layer, and φ represents the network parameters of Critic. The value estimation can judge the potential influence of the current training state on the subsequent training effect, and avoid the model only pursuing short-term completion rate and ignoring fatigue accumulation and exercise safety.

PPO algorithm improves training stability by limiting the change range of new and old strategies. Let the old policy be $\pi_\theta$old, the current policy be π_θ, and the policy probability ratio be $\rho_t(\theta)$. The objective function of PPO trimming is expressed as follows:

$$L_{ppo}(\theta) = \mathbb{E}_t\left[\min(\rho_t(\theta)A_t, \mathrm{clip}(\rho_t(\theta), 1-\epsilon, 1+\epsilon)A_t)\right] \tag{14}$$

where, $L_{ppo}(\theta)$ represents the PPO policy optimization objective, $\rho_t(\theta)$ represents the probability ratio between the current policy and the old policy in the same state action, At represents the advantage estimate, ϵ represents the clipping threshold, clip(·) represents the clipping function, and Et[·] represents the expectation calculation of training interaction samples. This objective function can suppress the training instability caused by excessive policy update, and make the model maintain good convergence and security in personalized training scenarios. Figure 4 shows the structure of the training scheme optimization decision model of the PPO policy network.

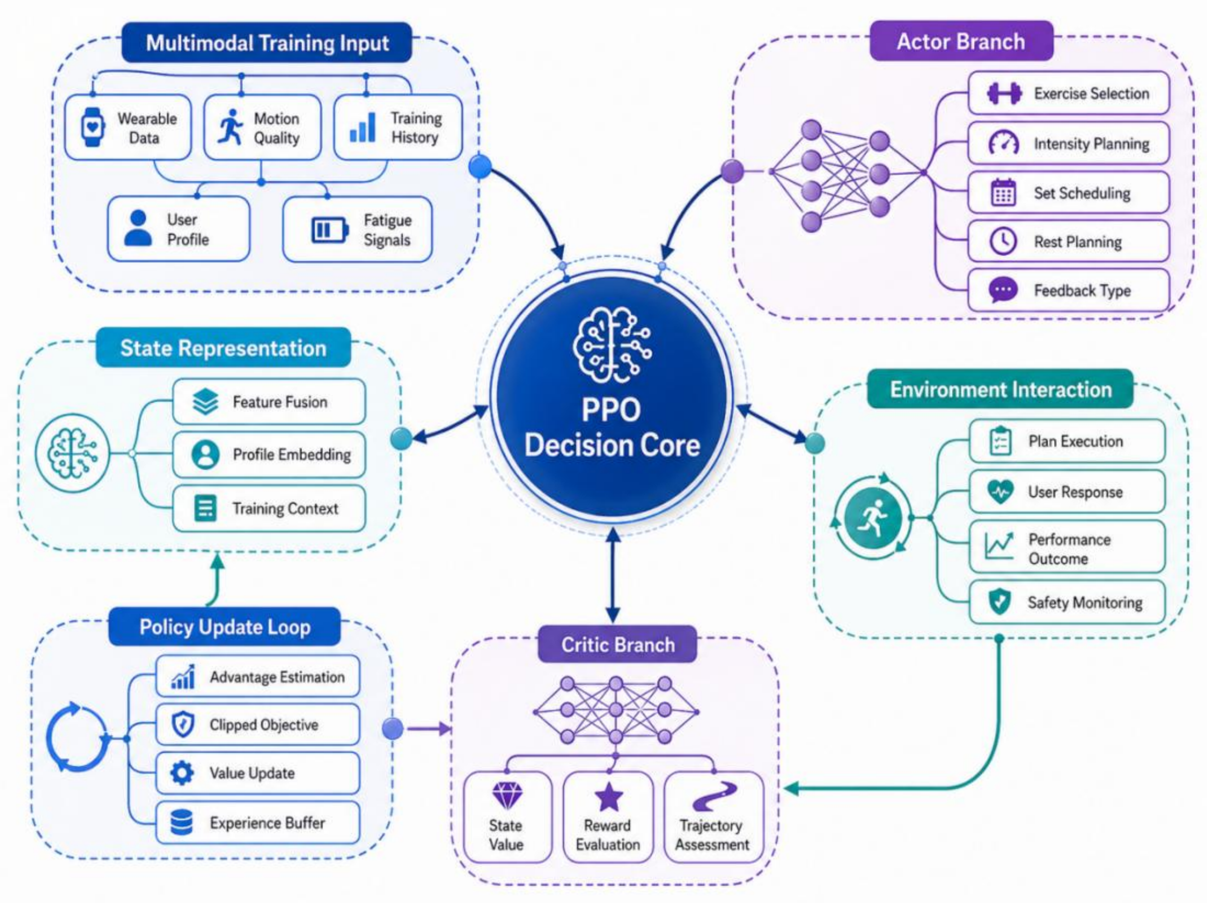

Figure 4: Structure diagram of training scheme optimization decision model of PPO policy network

3.5 Training performs feedback-driven load regulation with an online update closed loop

The actual implementation of the training program will be affected by the fluctuation of physical energy, the quality of action completion, equipment sampling error, user subjective fatigue and the change of training environment. Therefore, the system needs to continuously receive feedback information during the training execution phase, and convert the feedback results into load adjustment basis and strategy update samples. In this paper, training execution feedback is divided into four categories: action feedback, physiological feedback, load feedback and subjective feedback. The motion feedback mainly comes from attitude Angle, motion trajectory and completion quality score. Physiological feedback includes heart rate change, heart rate recovery and abnormal heart rate fluctuation. The load feedback included the number of training groups, the number of repetitions, the interval time and the completion rate. Subjective feedback included fatigue ratings, pain cues, and training satisfaction. The system collects these feedback through the mobile terminal, wearable device and training record interface, and compares them with the training actions output by the PPO policy network, so as to determine whether the current training plan needs to increase load, reduce load, extend rest or trigger action correction.

In order to identify training risks in real-time decision-making, the system constructs a fatigue risk index, which maps heart rate overload, fatigue score and recovery status into a unified risk probability:

$$\zeta_t = \sigma\left(\lambda_1 H_t^{hr} + \lambda_2 D_t^{fat} - \lambda_3 R_t^{rec} + b_\zeta\right) \tag{15}$$

where, ζ_t represents the fatigue risk index at time t, H_t^{hr} represents the heart rate overload score, D_t^{fat} represents the user fatigue score, R_t^{rec} represents the recovery status score, λ_1, λ_2, λ_3 represent the risk weight of different feedback indicators, b_ζ represents the bias term of fatigue risk calculation, and $\sigma(\cdot)$ represents the Sigmoid function. A higher index indicates that the current training load is more stressful to the user and the system needs to prioritize safety constraints.

In the reinforcement learning closed loop, the execution feedback also needs to be transformed into a reward signal, which is used to evaluate the rationality of the current training action. In this paper, training completion rate, action quality, recovery state, fatigue risk and plan deviation are jointly incorporated into the reward function:

$$r_t = \chi_1 C_t + \chi_2 Q_t + \chi_3 R_t^{rec} - \chi_4 \zeta_t - \chi_5 P_t^{dev} \tag{16}$$

where, r_t represents the training feedback reward at time t, C_t represents the training completion rate, Q_t represents the action quality score, and R_t^{rec} denotes the recovery status score, ζ_t denotes the fatigue risk index, P_t^{dev} denotes the deviation between the actual execution result and the planned goal, and χ_1 to χ_5 denote the weight of each reward component. The reward function not only encourages users to complete the training and maintain the action specification, but also punifies the risk of fatigue and execution deviation, avoiding the model from simply pursuing the improvement of training intensity.

Load regulation is done synchronously according to reward outcome and fatigue risk. Let the current training load be Λ_t and the proposed load in the next round be Λ_{t+1}. The system adopts the dynamic adjustment mode with lower limit constraints:

$$\Lambda_{t+1} = \mathrm{clip}\left(\Lambda_t + \mu r_t - \nu \zeta_t, \Lambda_{low}, \Lambda_{high}\right) \tag{17}$$

where Λ_t represents the training load level at time t, Λ_{t+1} represents the adjusted training load at the next time, μ represents the reward-driven load gain coefficient, ν represents the fatigue risk suppression coefficient, Λ_{low} and Λ_{high} represent the lowest and highest allowed training load, respectively, and clip(·) represents the boundary clipping function. This formulation is able to moderately increase stimulation when training performance is good and reduce load in time when fatigue risk is elevated, making personalized training programs progressive and safe.

The online update closed loop is used to write the interaction samples during the training execution to the experience cache and decide whether to trigger the policy update based on the recent reward and risk changes:

$$\mathcal{D}_{k+1} = \mathcal{D}_k \cup \{(s_t, a_t, r_t, s_{t+1}, \zeta_t)\}, \quad \mathcal{U}_k = 1\left(\bar{r}_k \geq \delta_r \wedge \bar{\zeta}_k \leq \delta_\zeta\right) \tag{18}$$

where, D_k represents the experience cache before the KTH update, D_{k+1} represents the experience cache after writing the new sample, s_t represents the current state vector, a_t represents the training action output by the system, r_t represents the feedback reward, s_{t+1} represents the next state after the execution of the action, ζ_t represents the fatigue risk index, U_k represents the trigger mark of the policy update.$\bar{r}_k$ represents the recent average reward, $\bar{\zeta}_k$ represents the recent average fatigue risk, and δ_r and δ_ζ represent the reward threshold and risk threshold, respectively. This update rule can avoid frequently updating the policy when the user's state is unstable or the risk is too high, and improve the security of the online learning process.

Through the above closed-loop design, the system can feed back the training execution results to the load adjustment and PPO policy update process in real time. When training performs well and the risk is low, the system ramps up the training stimulus and accumulates positive samples. When movement quality decreases, fatigue risk rises, or recovery is insufficient, the system reduces training load, extends intervals, or pushes corrective prompts. This mechanism makes the personalized sports training scheme have the continuous adaptive ability, and also provides a technical guarantee for the stable deployment of the system in the real sports scene.

4 Analysis of experimental results

4.1 Configuration of experimental environment and construction of sports training data set

In order to verify the effectiveness of the reinforcement learning driven personalized exercise training scheme optimization and real-time feedback system, the experiment was completed in a unified software and hardware environment. The server is configured with Intel Xeon Silver 4214R processor, NVIDIA RTX 3090 GPU, 64 GB memory, Ubuntu 22.04 operating system, Python 3.10 algorithm development environment, and PyTorch 2.1 deep learning framework. The mobile terminal is deployed on the Android 12 smart terminal, and the wearable acquisition device includes a heart rate belt, a six-axis IMU sensor and an intelligent exercise bracelet to obtain the physiological response, movement posture and exercise load data during training. The experimental subjects were 60 subjects with basic exercise experience, aged from 19 to 35 years old. The training period was 8 weeks, and the training tasks were completed 3 times a week, covering running, squat, lunge, plank, jump training and core strength training. The system divided the training set, validation set and test set according to the user dimension, and the ratio was 7：1.5：1.5, to avoid the high results caused by the same user data appearing in the training and testing phase at the same time. During data sampling, the heart rate signal sampling frequency is 1 Hz, the IMU signal sampling frequency is 50 Hz, the action video is recorded at 30 fps, and the training log is written by the mobile terminal in real time. In order to ensure that multi-source data can enter a unified state vector, the system performed timestamp alignment, outlier removal, missing imputation, sliding window segmentation and normalization on different source data, and finally formed a sports training data set that could be used for training state recognition, load prediction and PPO strategy optimization. The multi-source motion data collection content and preprocessing method are shown in Table 2.

Table 2: Multi-source motion data collection content and preprocessing method table

Data Type	Collected Content	Sampling Scale or Frequency	Preprocessing Method	Technical Function
Physiological Response Data	Heart rate, heart rate recovery, heart rate variability	1 Hz, approximately 103,680 records	Abnormal peak removal, sliding mean filtering, interval normalization	Determines training load tolerance and recovery status
Inertial Motion Data	Acceleration, angular velocity, posture changes	50 Hz, approximately 5,184,000 records	Timestamp alignment, low-pass filtering, window slicing	Extracts motion rhythm, impact intensity, and posture stability
Motion Video Data	Motion sequences such as squats, lunges, and jumps	30 fps, 2,880 video clips	Keyframe extraction, pose keypoint detection, trajectory smoothing	Evaluates motion standardization and motion deviation
Training Load Data	Sets, repetitions, rest interval, training duration	1,440 training segments or task windows	Structured encoding, load accumulation, intensity stratification	Supports training load prediction and program adjustment
Subjective Feedback Data	Fatigue score, pain alert, training satisfaction	4,320 questionnaire records	Missing value compensation, grade encoding, outlier correction	Corrects the model's judgment of fatigue and training experience
System Interaction Data	Completion rate, withdrawal records, feedback response time	1,440 interaction logs	Log cleaning, event serialization, behavior label generation	Evaluates real-time feedback effectiveness and user execution stability

It can be seen from Table 2 that the experimental data cover multiple dimensions such as physiology, action, load, feedback and system interaction, which can reflect the state changes of users in the training process more completely. After unified preprocessing, different modal data are converted into structured feature input, which provides a reliable data basis for subsequent individual motion state recognition, training load prediction and reinforcement learning strategy optimization.

4.2 Performance analysis of individual motion state recognition and training load prediction

In order to verify the state perception and load prediction ability of the proposed PPO strategy model, Rule-based, SVM, LSTM and DDPG are selected as the baseline models and evaluated from two aspects of motion state recognition and training load prediction respectively. The motion state recognition task mainly distinguished five kinds of states:

normal training, action deviation, high load, fatigue accumulation and insufficient recovery. The evaluation indexes were Accuracy and F1-score. The training load prediction task is the prediction target of the load level in the next training window, and the evaluation indexes are RMSE and MAE. The performance comparison between individual motion state recognition and training load prediction is shown in Figure 5.

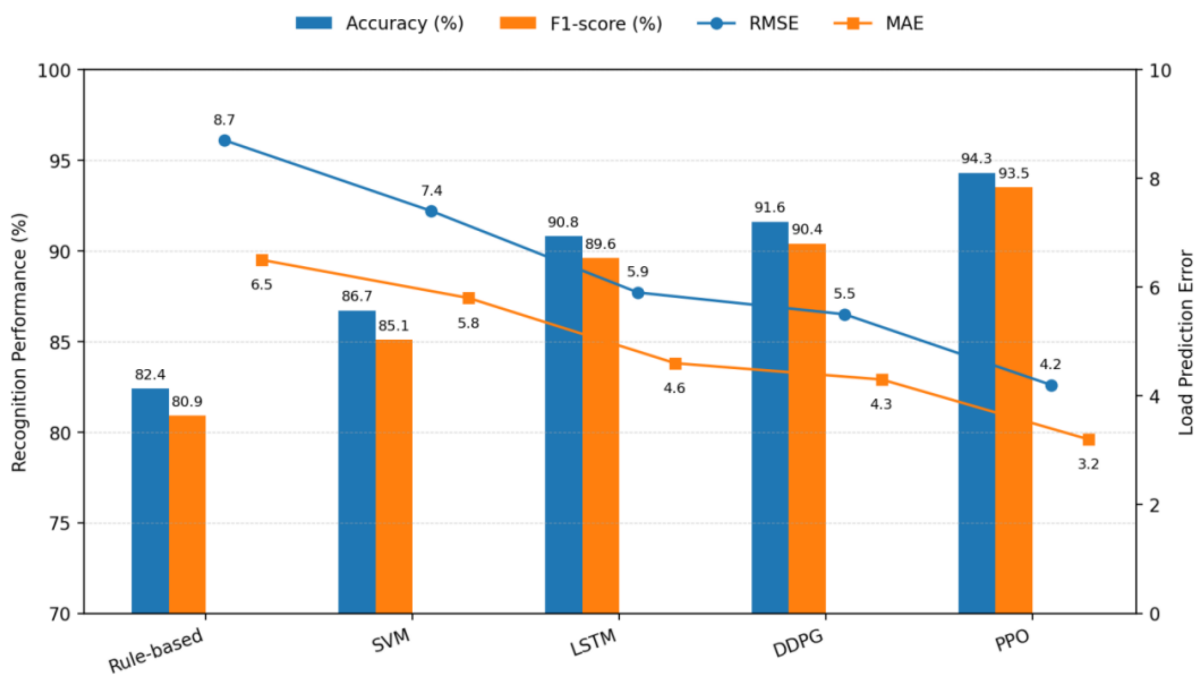

Figure 5: Comparison of individual motion state recognition and training load prediction performance

It can be seen from Figure 5 that the PPO model performs best in both tasks of state recognition and load prediction, with an Accuracy of 94.3% and an F1-score of 93.5%, which are 3.5 and 3.9 percentage points higher than those of the LSTM model respectively, indicating that the reinforcement learning strategy network performs best after fusing motion state, ability portrait and feedback reward. It can distinguish different training states more accurately. In terms of load forecasting, the RMSE and MAE of PPO model are reduced to 4.2 and 3.2, which are significantly lower than 8.7 and 6.5 of Rule-based model, and better than 5.5 and 4.3 of DDPG model. The results show that PPO can not only improve the accuracy of individual motion state recognition, but also predict subsequent training load changes more stably, which provides a reliable basis for the generation of personalized training programs and real-time load adjustment.

4.3 Analysis of personalized training action selection mechanism under PPO policy network

In order to further analyze the decision-making mechanism of PPO policy network in the generation of personalized training scheme, this paper statistics the training action selection probability output by the model in the test phase, and focuses on the distribution characteristics between action types such as action correction, load reduction, extended interval, intensity planning and load maintenance. Different from the fixed rule recommendation method, the PPO policy network does not directly output a single training template, but dynamically adjusts the action selection probability according to the exercise performance, fatigue risk, action quality and fitness score in the state vector. When the user has action deviation or the risk of fatigue increases, the model tends to choose the safety control actions such as action correction, load reduction and interval extension. When the user state is stable and the fitness score is high, the model improves the selection probability of the maintenance load and intensity planning action. Figure 6 shows the probability distribution of action selection for PPO policy.

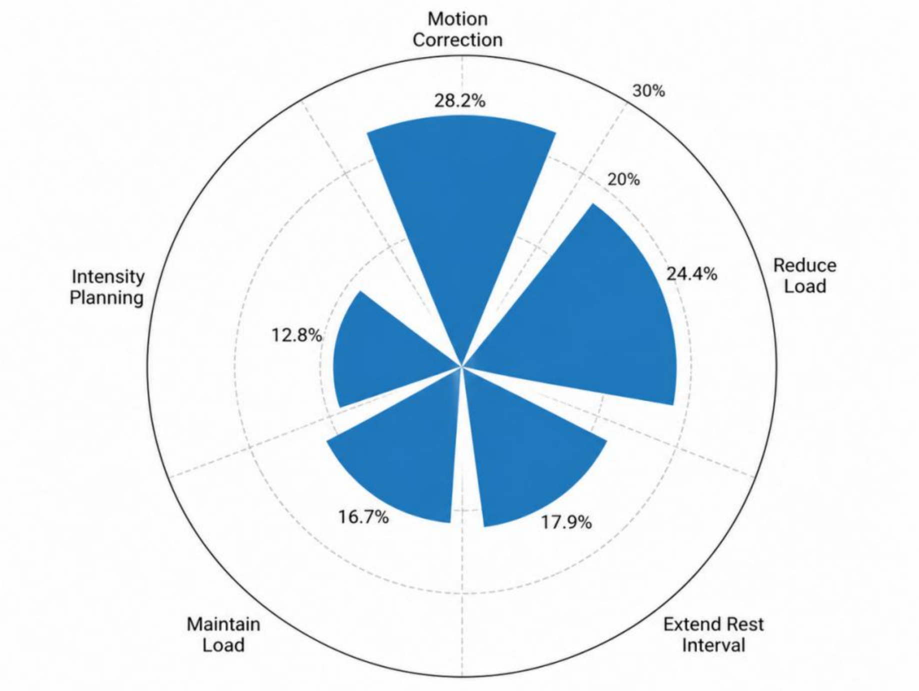

Figure 6: Radar rose plot of probability distribution for action selection of PPO policy

It can be seen from Figure 6 that the PPO strategy does not centrally select a single action in the training scheme generation, but forms distributed decisions among action correction, load reduction, interval extension, intensity planning and load maintenance according to the user state. Among them, the probability of action correction is the highest, which is 28.2%. The probability of load reduction is 24.4%. The probability of extended interval is 17.9%, indicating that the model is more inclined to give priority to ensuring training safety when it finds action deviation or fatigue risk. The maintenance load and intensity planning probabilities are 16.7% and 12.8%, respectively, indicating that the system will maintain the training stimulus and moderately advance the training intensity when the state is stable. The results show that PPO can integrate exercise performance, fatigue risk and training goals into the decision-making process, making the personalized training program more adaptive and sustainable.

4.4 Influence analysis of real-time feedback mechanism on training action correction and fatigue risk early warning

Real-time feedback mechanism is an important part of the online optimization of personalized sports training system. Its role is to transmit the action deviation, physiological load change, recovery state and subjective fatigue information back to the decision-making model in time, so that the training program can be dynamically modified with the change of individual state. In the actual training process, the user's action completion quality and fatigue level will continue to fluctuate, and a single training plan is difficult to cover all the execution changes.

To this end, this paper integrates action trajectory, heart rate fluctuation, action quality score, training completion rate and fatigue risk index into the feedback closed loop, and adjusts the training load, interval time, action tips and recovery suggestions in real time through the PPO strategy network. This mechanism can push corrective feedback in time when movement deviation occurs, reduce training stimulus or extend recovery time when fatigue risk increases, so as to improve the safety and continuity of the training process.

The process of training load regulation can reflect the dynamic balance between the maintenance of training stimulus and the control of fatigue risk. In order to observe the influence of the real-time feedback closed loop on the direction of load adjustment, this paper uses the fatigue risk index and training load fitness as coordinate variables to draw the training load adjustment trajectory, as shown in Figure 7.

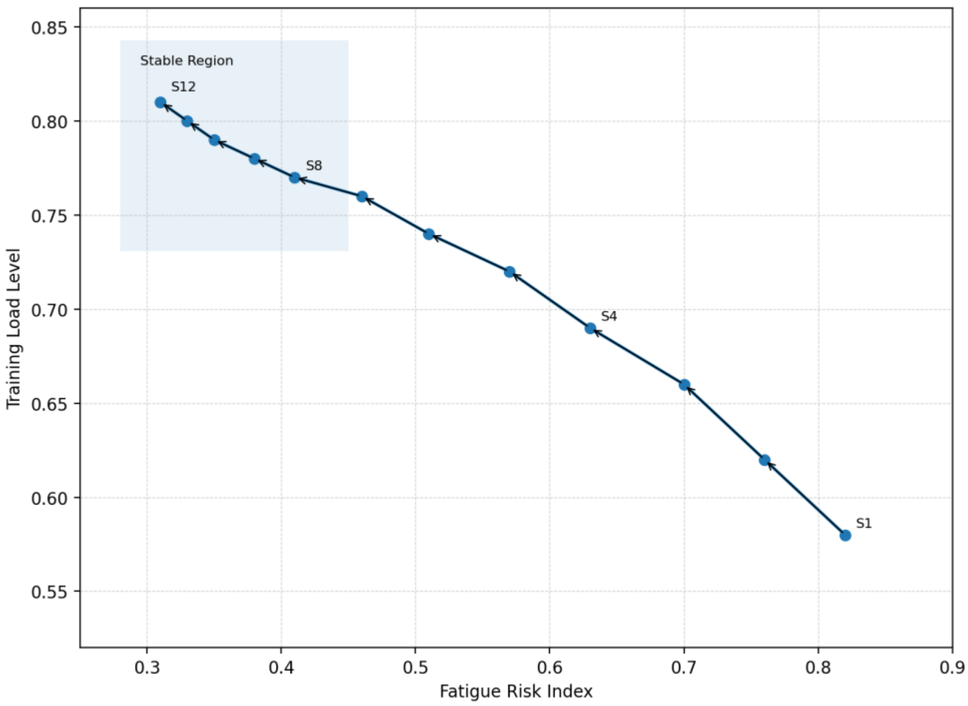

Figure 7: Phase diagram of training load regulation trajectory

It can be seen from Figure 7 that the training load adjustment trajectory gradually migrates from "high fatigue risk, low training load adaptability" to "low fatigue risk, high training load adaptability" region. In the initial stage, the fatigue risk index was 0.82, and the training load level was 0.58, indicating that it was not suitable for users to continue to increase the training intensity under high fatigue pressure, and the system gave priority to the implementation of risk reduction and action modification strategies. With the continuous intervention of real-time feedback, the fatigue risk index gradually decreased to 0.31, the training load level increased to 0.81, and the trajectory finally entered the stable training region. The results show that the real-time feedback closed loop can adjust the training stimulus according to the change of user state, make the load change smoother, and avoid the deterioration of action quality or the increase of risk due to the accumulation of fatigue in the training process.

The effectiveness of fatigue risk warning depends not only on whether the model can identify high-risk states, but also on the consistency between the predicted probability and the real frequency of fatigue events. In order to further verify the credibility of the risk early warning results, the fatigue risk calibration curve of the PPO early warning model is drawn and compared with the ideal calibration line, as shown in Figure 8.

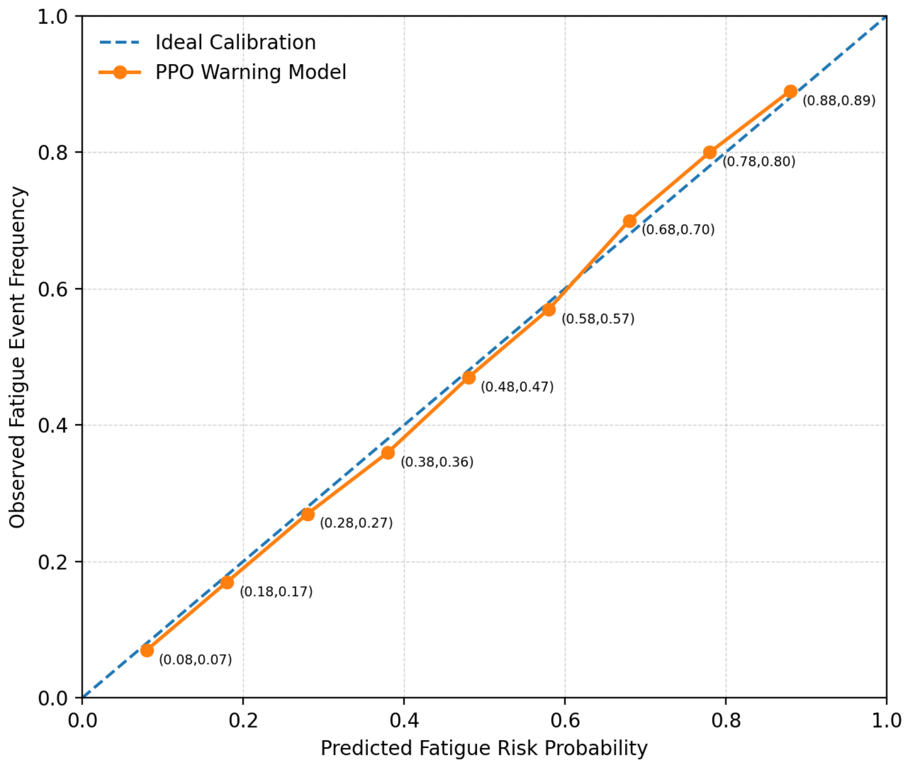

Figure 8: Fatigue risk warning calibration curve

It can be seen from Figure 8 that the calibration curve of PPO early warning model is close to the ideal calibration line, and the observed fatigue event frequency and the predicted fatigue risk probability maintain a high consistency within the predicted probability interval of 0.08 to 0.88. Among them, when the predicted risk probability is 0.58, the real fatigue event frequency is 0.57. When the predicted risk probability is 0.78, the true frequency is 0.80, and the deviations are controlled within a small range. The results show that the fatigue risk probability output by the model has good interpretability and credibility, and can provide a basis for the system to trigger load reduction, prolong the interval and recovery recommendations. In summary, the real-time feedback mechanism can transform the training execution results into the basis of load adjustment and risk warning in time, so that the personalized training scheme has stronger online adaptability.

4.5 System response time delay and feasibility analysis of real-time deployment

The system response delay directly determines whether the real-time feedback mechanism can play a role in the exercise training process, so this paper starts from the end-to-end feedback link to test the feasibility of system deployment. In the experimental environment, the processing time of data acquisition, signal cleaning, status recognition, PPO decision, feedback rendering and online log writing were counted respectively, and a single complete feedback process was used as the delay evaluation object. The waterfall delay of the system real-time feedback link is shown in Figure 9.

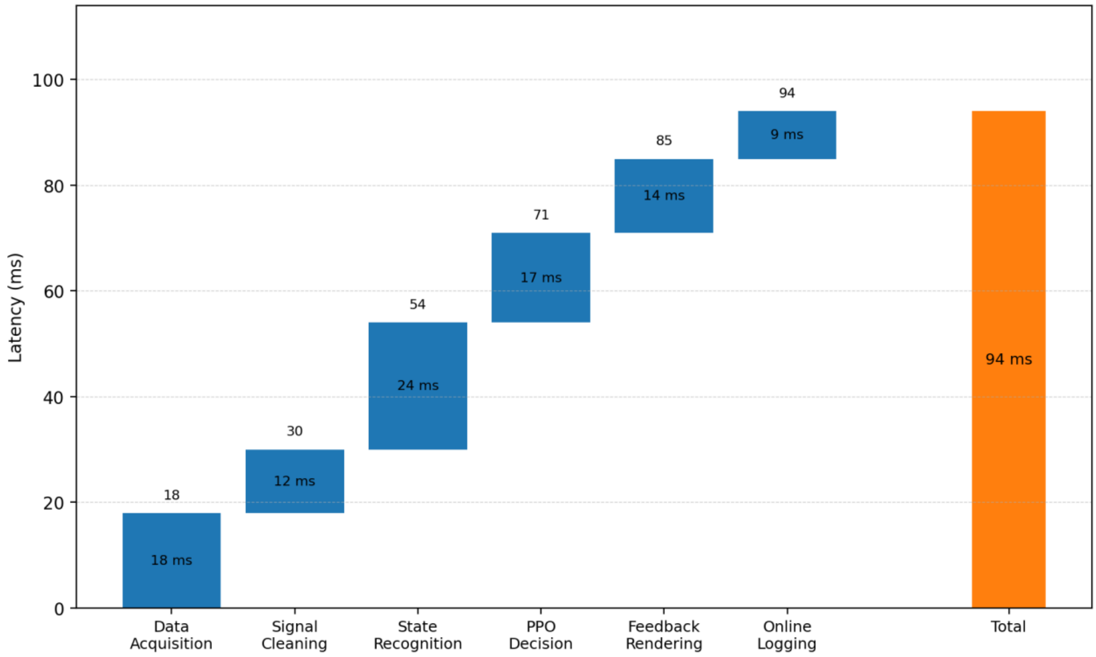

Figure 9: Waterfall delay diagram of the real-time feedback link of the system

It can be seen from Figure 9 that the total delay of the single complete real-time feedback link of the system is 94 ms, which meets the requirements of low delay response in personalized sports training scenarios. Among them, data acquisition took 18 ms, signal cleaning took 12 ms, and state recognition took 24 ms, which accounted for the highest proportion of the stage in the whole link, indicating that multi-source data feature extraction and state discrimination were still the main computational overhead. PPO decision took 17 ms, indicating that the policy network can complete the generation of training actions in a short time. Feedback rendering and online log writing took 14 ms and 9 ms respectively, which had relatively little impact on the overall latency. In summary, the whole process of the system from state perception to scheme generation and then to feedback output can be stably controlled within 100 ms, indicating that the constructed reinforcement learning driven system has good real-time performance and engineering deployment feasibility, and can support online feedback and dynamic load adjustment in actual training scenarios.

5 Discussion

The personalized exercise training scheme optimization and real-time feedback system constructed in this paper combines multi-source motion data perception, exercise ability portrait, PPO policy decision and online feedback update, so that the training scheme can be dynamically adjusted according to the change of individual state. Compared with the fixed training template, the system can take heart rate change, action trajectory, fatigue score, training completion rate and recovery status into consideration to determine whether the user is currently suitable to continue to increase training stimuli, and give corrective tips in time when movement deviation or fatigue risk increases. Such a closed-loop structure makes the training process have the ability of continuous perception and continuous correction, which helps to improve the matching degree between the training scheme and the individual state.

The core value of the system is to transform the continuous changes in the training process into an updatable decision process. The state vector integrates real-time motion performance, ability profile, training state probability and fitness score, so that the policy network can select training actions in a relatively complete individual background. The PPO policy network generates the training scheme through the Actor-Critic structure, and uses the cropped policy update to control the adjustment amplitude of the model, so as to reduce the unstable impact caused by policy fluctuations. Experimental results show that PPO performs well in state recognition, load prediction and action selection, indicating that the mechanism can adapt to the scene of rapid state change, frequent feedback and high safety requirements in sports training.

Real-time feedback mechanism plays an important role in training safety and continuity. Exercise training needs to maintain a balance between effective stimulation and fatigue control. Excessive pursuit of load lifting may lead to action deformation, insufficient recovery and increased risk of injury. Through fatigue risk index and feedback reward function, the training completion rate, action quality, recovery status and risk factors are incorporated into the judgment of the system, so that the training load can be adjusted with the change of user status. When the user performance was stable, the system moderately increased the training stimulus. When the risk of fatigue is elevated, the system reduces the load, extends the interval, or pushes motion correction suggestions, thereby enhancing the safety boundary of the training process.

There is still room for further improvement of the system in this paper. The experimental data covers multiple types of basic training items, but the samples of different age groups, different sports foundations and special training groups need to be expanded, and the adaptability of the model in long-term high-intensity training needs to be further verified. In the real training environment, wearable devices may be affected by wearing position, signal drift and network delay. In the future, anomaly detection, edge-end reasoning and feedback presentation can be further optimized. At the same time, the system can also enhance the explanation ability of training suggestions, so that users can more clearly understand the basis of load adjustment, action correction and fatigue warning, thereby improving the use trust and long-term participation intention.

6 Conclusion

In this paper, an intelligent training system framework based on reinforcement learning is proposed around the optimization of personalized sports training programs and the requirements of real-time feedback. Based on multi-source motion data acquisition, the system encodes physiological response, inertial motion, action video, training load and subjective feedback uniformly, improves individual differential expression ability through exercise ability portrait and state vector generation mechanism, and uses PPO policy network to realize the dynamic selection of training actions, load levels, recovery suggestions and feedback methods. The experimental results show that the proposed system performs well in individual motion state recognition, load prediction and feedback response. The Accuracy of PPO model reaches 94.3% and F1-score reaches 93.5%, which are 3.5 and 3.9 percentage points higher than those of LSTM respectively. In the real-time feedback link, the fatigue risk index was reduced from 0.82 to 0.31, the training load level was increased from 0.58 to 0.81, and the end-to-end delay was controlled at 94 ms. The results show that the system can take into account the requirements of training effect, security warning and real-time deployment, and the application value can be further improved by combining edge computing and explanatory feedback in the future.

Author's Profile

Xuexiao Li (born November 1989), male, Han ethnicity, from Anyang, Henan Province. He holds a Master’s degree and is a lecturer. His research interests include physical education and training, as well as Wushu Sanda (martial art free combat).

References

[1] Gönül S, Namlı T, Coşar A, Toroslu I H. A reinforcement learning based algorithm for personalization of digital, just-in-time, adaptive interventions[J]. Artificial Intelligence in Medicine, 2021, 115: 102062. DOI: 10.1016/j.artmed.2021.102062.

[2] Mair J L, Hayes L D, Campbell A K, Buchan D S, Easton C, Sculthorpe N. A personalized smartphone-delivered just-in-time adaptive intervention to increase physical activity in older adults: Mixed methods feasibility study[J]. JMIR Formative Research, 2022, 6(4): e34662. DOI: 10.2196/34662.

[3] Gabarron E, Larbi D, Rivera-Romero O, Denecke K. Human factors in AI-driven digital solutions for increasing physical activity: Scoping review[J]. JMIR Human Factors, 2024, 11: e55964. DOI: 10.2196/55964.

[4] Vandelanotte C, Trost S, Hodgetts D, To Q G, Williams S L, Alley S. MoveMentor—examining the effectiveness of a machine learning and app-based digital assistant to increase physical activity in adults: Protocol for a randomised controlled trial[J]. Trials, 2025, 26: 233. DOI: 10.1186/s13063-025-08926-3.

[5] Carey R L, Sheats J L, Ha Le A, Hiremath S V, Nahum-Shani I. mHealth-based just-in-time adaptive intervention to improve the physical activity levels of individuals with spinal cord injury: Protocol for a randomized controlled trial[J]. JMIR Research Protocols, 2024, 13: e57699. DOI: 10.2196/57699.

[6] Nahum-Shani I, Dziak J J, Wetter D W. MCMTC: A pragmatic framework for selecting an experimental design to inform the development of digital interventions[J]. Frontiers in Digital Health, 2022, 4: 798025. DOI: 10.3389/fdgth.2022.798025.

[7] Nahum-Shani I, Dziak J J, Walton M A, Dempsey W, Guastaferro K. Hybrid experimental designs for intervention development: What, why, and how[J]. Advances in Methods and Practices in Psychological Science, 2022, 5(3): 25152459221114279. DOI: 10.1177/25152459221114279.

[8] Mottaghi E, Akbarzadeh-T M R. Automatic evaluation of motor rehabilitation exercises based on deep mixture density neural networks[J]. Journal of Biomedical Informatics, 2022, 130: 104077. DOI: 10.1016/j.jbi.2022.104077.

[9] Spilz A, Munz M. Automatic assessment of functional movement screening exercises with deep learning architectures[J]. Sensors, 2023, 23(1): 5. DOI: 10.3390/s23010005.

[10] Marotta L, Buurke J H, van Beijnum B J F, Reenalda J. Towards machine learning-based detection of running-induced fatigue in real-world scenarios: Evaluation of IMU sensor configurations to reduce intrusiveness[J]. Sensors, 2021, 21(10): 3451.

DOI: 10.3390/s21103451.

[11] Van Hooren B, Plasqui G, Meijer K. The effect of wearable-based real-time feedback on running injuries and running performance: A randomized controlled trial[J]. The American Journal of Sports Medicine, 2024, 52(3): 750-765. DOI: 10.1177/03635465231222464.

[12] Kakhi K, Jagatheesaperumal S K, Khosravi A, Alizadehsani R, Acharya U R. Fatigue monitoring using wearables and AI: Trends, challenges, and future opportunities[J]. Computers in Biology and Medicine, 2025, 195: 110461. DOI: 10.1016/j.compbiomed.2025.110461.

[13] Kristoffersson A, Lindén M. A systematic review of wearable sensors for monitoring physical activity[J]. Sensors, 2022, 22(2): 573. DOI: 10.3390/s22020573.

[14] Migliaccio G M, Padulo J, Russo L. The impact of wearable technologies on marginal gains in sports performance: An integrative overview on advances in sports, exercise, and health[J]. Applied Sciences, 2024, 14(15): 6649. DOI: 10.3390/app14156649.

[15] Morouço P. Wearable technology and its influence on motor development and biomechanical analysis[J]. International Journal of Environmental Research and Public Health, 2024, 21(9): 1126. DOI: 10.3390/ijerph21091126.

[16] Alzahrani A, Ullah A. Advanced biomechanical analytics: Wearable technologies for precision health monitoring in sports performance[J]. Digital Health, 2024, 10: 20552076241256745. DOI: 10.1177/20552076241256745.

[17] Uddin M Z, Soylu A. Human activity recognition using wearable sensors, discriminant analysis, and long short-term memory-based neural structured learning[J]. Scientific Reports, 2021, 11: 16455. DOI: 10.1038/s41598-021-95947-y.

[18] Papaleonidas A, Psathas A P, Iliadis L. High accuracy human activity recognition using machine learning and wearable devices’ raw signals[J]. Journal of Information and Telecommunication, 2022, 6(3): 237-253. DOI: 10.1080/24751839.2021.1987706.

[19] Van Eetvelde H, Mendonça L D, Ley C, Seil R, Tischer T. Machine learning methods in sport injury prediction and prevention: A systematic review[J]. Journal of Experimental Orthopaedics, 2021, 8(1): 27. DOI: 10.1186/s40634-021-00346-x.

[20] Majumdar A, Bakirov R, Hodges D, Scott D, Rees T. Machine learning for understanding and predicting injuries in football[J]. Sports Medicine - Open, 2022, 8: 73. DOI: 10.1186/s40798-022-00465-4.

[21] Carey D L, Crossley K M, Whiteley R. Machine learning approaches to sports injury prediction: Practical applications and methodological challenges[J]. Sports Medicine - Open, 2023, 9: 52. DOI: 10.1186/s40798-023-00589-1.

[22] Leckey C, van Dyk N, Doherty C, Lawlor G, Delahunt E. Machine learning approaches to injury risk prediction in sport: A scoping review with evidence synthesis[J]. British Journal of Sports Medicine, 2025, 59(7): 491-500. DOI: 10.1136/bjsports-2024-108576.

[23] Tsilimigkras T, Kakkos I, Matsopoulos G K, Bogdanis G C. Enhancing sports injury risk assessment in soccer through machine learning and training load analysis[J]. Journal of Sports Science and Medicine, 2024, 23: 537-547. DOI: 10.52082/jssm.2024.537.