A Proximal Policy Optimization-Based Reinforcement Learning Framework for Real-Time Personalized Endurance Training

Chuanzhong Wu[1], Danqing Liang[2], Bo Yang[2], Li Xu [1], Yunlong Li[2,*]
[1]Guangzhou College of Commerce, Guangzhou, Guangdong,202162, China
[2]Xianda College of Economics and Humanities, Shanghai International Studies University, Shanghai ,202162, China
E-mail: liyunlong998@hotmail.com
[*]Corresponding author

Keywords: reinforcement learning, personalized training, sports optimization, wearable sensors, athletic performance

Recieved: July 11, 2025

Customized sports training routines take into account individual physiology, fatigue, and recovery to maximize performance. Proximal Policy Optimization (PPO)-based reinforcement learning is used to adjust training intensity, duration, and rest in a simulated endurance-training environment for runners, using real-time wearable and performance data. The environment models athlete status utilizing heart rate variability, VO₂ max, fatigue ratings, and injury-risk indicators. PPO is trained to maximize performance gains, recovery quality, and safety over repeated sessions. Simulated policy improves performance (18.6%), injury-risk deviation (−22.4%), recovery compliance (91.3%), training load variability control (±7.2%), reward-signal evolution convergence (+41.7%), session completion rate (94.6%), personalized adaptation score (87.5%), and fatigue index stability (94.3%). Results show that a PPO-based RL setup, specifically defined by state design, reward shaping, and multi-episode training, can provide adaptive and data-driven tailored sports training.

Povzetek: PPO-temelječ pristop z okrepljenim učenjem na podlagi podatkov nosljivih naprav dinamično prilagaja tekaški trening (intenzivnost, trajanje, počitek) ter v simulacijah hkrati izboljša zmogljivost, okrevanje in varnost.

1 Introduction

Optimizing athletic performance requires adaptive training methods that account for individual variations in physiology, recovery, and mental preparedness. A conventional static program is not responsive to time-sensitive changes in an athlete's condition [1]. This paper introduces a new reinforcement learning strategy, based on PPO, that personalizes all training parameters using wearable biometric data. The system is designed to facilitate performance, enable faster recovery, and minimize the risk of injury by dynamically varying intensity, duration, and rest, introducing an intelligent, data-driven dimension to sports training [2].

1.1 Background

Optimizing athletic performance has long been a source of research, driving the development of new training techniques, including periodized conditioning and sport-specific training [3]. Conventionally, coaches and athletes have relied on pre-stipulated schedules founded on experience, general physiological principles, and observations. Although, to a certain extent, these methods can be described as successful, these procedures do not consider the individuality and changing nature of human physiology, individual recovery, or even the stress response. The need to tailor training to the individual is becoming an urgent necessity in the contemporary high-performance context [4].

1.2 Motivation

Elite and amateur athletes alike are constantly striving to deliver their best performance and minimize the risks of declining performance, injury, and burnout. Even the most convenient static training programs do not adjust in real time to key factors such as fatigue, sleep quality, heart rate variability, and mental preparedness. The need for such a study stems from the growing recognition that adaptive systems based on artificial intelligence can address this gap by dynamically adjusting training protocols in real time based on ongoing feedback [5].

1.3 Importance of personalization in sports training

Athletes cannot be made equal. Responding to a training stimulus depends on various factors, including age, genetics, injury history, training history, and psychological factors [6]. Personalization ensures that every sports person receives the optimal training, intensity, and volume tailored to their body and ambitions. Individualized programs are not only faster in terms of gains, but also more compliant, reduce the incidence of injuries, and promote long-term growth. Due to the widespread adoption of wearable devices and sensor technologies, it is now possible to design training interventions guided by granular and individualized data [7].

1.4 The rise of data-driven athletic interventions

Sports activities have led to an influx of biometric data captured by smartwatches, GPS trackers, and heart rate monitors, among other devices. They are collecting constant streams of data: everything from steps and sleep stages to lactate threshold and recovery percentage. The new flow of real-time data creates opportunities for data-driven training management, especially when combined with more advanced algorithms, such as machine learning and reinforcement learning. Optimally integrated, this data can be used to perform dynamic interventions, rather than applying the given rules [8].

1.5 Limitations of existing training methods

The training methods used now are primarily based on a few heuristics or previous trends and thus do not account for the immediate physiological effects. These procedures do not account for day-to-day variations in the athlete's readiness; therefore, load management is poor [9]. Additionally, rule-based or conventional machine learning systems may require extensive manual tuning and may struggle to keep pace with nonlinear or long-term feedback trends. These systems can be optimized to work immediately but fail to provide adequate progression, rehabilitation, and injury prevention [10].

1.6 Research contributions

- To create a real-time PPO-based reinforcement learning model that adjusts training intensity, duration, and recovery depending on heart rate variability, VO_2 max, and fatigue, resulting in a tailored and responsive training regimen.
- To assess if the PPO framework achieves a minimum 18.6% performance increase (VO_2 max proxy) over static or rule-based endurance training regimens.
- To determine if the PPO-driven method can minimize injury risk deviation by 22.4% and increase session completion and recovery compliance to 94.6% and 91.3%, exceeding DQN, A3C, JITAI, and other RL baselines.

To assess if integrating a weighted reward function with PPO clipping leads to stable load variability (±7.2%), strong individualized adaptation scores (87.5%), and high fatigue index stability (94.3%) with increasing sample size.

1.7 Paper organization

In this paper, section 2 explains the related works, section 3 provides the methodology, section 4 shows the evaluation metrics. Result analysis is explained in section 5, and the discussion is shown in section 6. Finally, section 7 suggested the conclusion with future works.

2.Related works

In recent years, the application of artificial intelligence, particularly reinforcement learning and deep neural networks, has increased in sports training. Key techniques include Deep Q-Networks and Asynchronous Advantage Actor-Critic (A3C), applied across various levels of athlete performance. While these studies show promise in intelligent adaptation, most lack real-time sensor feedback, closed-loop systems, or personalized training variables. This literature review examines 10 recent techniques, highlighting their strengths, limitations, and significant gaps that necessitate a dynamic framework grounded in PPO.

In this research, a Deep Reinforcement Learning (DRL) algorithm with an Asynchronous Advantage Actor-Critic (A3C) framework will be designed to learn and tune exercise performance. The system speeds up or reduces the extent of training based on a trade-off between exploration and exploitation, and may learn to model fitness-fatigue dynamics by analyzing time-series biometric data. Validation was done in publicly available walking, running, and sports logs. The power of DRL to individualize training plans stems from its ability to transform exercise prescriptions to reflect users' real-time fitness levels [11].

The authors implement the Deep Q-Network (DQN) to personalize the teaching path in physical education using data from students' activity logs. Once a data set is gathered and processed, the model establishes a system of definite mappings between state-action-reward associations, enabling dynamic adjustment of PE training. The DQN approach yields significantly better autonomy, participation, and skill results than traditional static teaching approaches. The experiment has demonstrated that customized training programs, developed with the aid of RL, are superior to generic curricula in learning physical education [12].

The paper proposes Expected Discounted Goal (EDG), a deep reinforcement learning method for assessing the contributions of soccer players in simulation settings. Without any real-life training data in the form of human labels, the model relies solely on virtual match simulations and open-source tracking. EDG estimates the number of goals a team may score or concede per state, which helps to analyze player performance more effectively. The presented strategy can be applied in other sports where labeled training data are scarce, as noted in Garnier et al [13].

This was achieved by developing a wearable AI system that provides real-time biomechanical feedback during hammer-throwing training. It combines inertial sensors, load cells, and deep neural networks (DNN) to predict limb position angles and wire tension. For this system, an objective evaluation of motion is provided, with subjective feedback based on experience replaced by an objective correction based on the data. With feasible estimation inaccuracies of under 12%, the paper demonstrates how DNNs and wearables can enable dynamic training on complex athletic movements tailored to individual needs [14].

This paper examines AI-driven advancements in sports science, concentrating on load management, injury prevention, talent identification, and recovery efficiency. It supports application-specific artificial intelligence systems that augment human knowledge and expertise. They can be used to predict overtraining, the menstrual cycle, and mental health. The research also recommends interdisciplinary cooperation and revised educational programs to make future sport scientists' literate about AI. It highlights the importance of transparency and the ethical application of AI in high-performance environments [15].

The paper proposes an algorithm for reinforcement learning of Just-In-Time Adaptive Intervention (JITAI) via the mobile app HeartSteps V2. The RL model personalizes physical activity recommendations and provides them as often as 5 times a day, depending on real-time user context. Continuous learning in health applications is effective because the system learns and improves over time as more user behavior data is received. The JITAI method can also be utilized in dynamic training plans for sports or fitness [16].

This paper combines Q-learning with upper-limb training (a VR bubble-popping game). Each session is adaptive and patient-specific because difficulty is adjusted based on the patient's actions, using kinematic information. The algorithm optimised bubble locations and timing based on patient capability over 10 sessions, indicating the potential for intelligent physical therapy. It is a model of reinforcement learning to assist in motor recovery and personalized feedback in exercise therapy [17].

The paper applies Subgroup Discovery (SD) to identify patterns between training load, wellness parameters, and overuse injuries among a group of professional volleyball players. The model determines the personal risks of injury by measuring wellbeing and workload data of athletes who are followed for 24 weeks. Individualized signs, such as sleeping less or exhibiting a jump load behavior, were noted. That is the value of frequent observation and flexible planning, which can be further improved through reinforcement learning to make dynamic adjustments [18].

This paper utilizes Convolutional Neural Networks (CNNs) and Transfer Learning to categorize complex throwing motions in Ultimate Frisbee using IMU data. It achieves an accuracy of up to 89.9 percent in inferring the three main types of throws and benefits from pre-set weights when the data is scarce. Although it is not RL-based, the ability to identify fine-grained athletic actions enables proper state recognition in RL models, underscoring the role of CNNs in performance monitoring [19].

2.1 Research gap

Limitations of the related works is shown in Table 1.

Table 1: Limitations of the related works

S.No.	Paper / Method	Key idea & quantitative result (where given)	Main limitation	Gap vs. this work (no closed-loop sensor RL)
1	DRL fatigue–recovery (A3C)	RL co-models' fatigue and recovery; improves simulated reward and reduces over-fatigue episodes vs. heuristics.	Simulation only; uses abstract states, no rich wearable input.	Does not run a real-time loop from sensor state → training action → updated state for full-plan personalization.
2	DQN training planner	DQN achieves higher cumulative reward and better schedule adherence than rule-based baselines in simulation.	No live physiological or motion signals; state is hand-crafted.	Cannot adapt within or across sessions using continuous sensor feedback.
3	EDG DRL player evaluation	DRL predicts player value or performance better than classical rating metrics.	Focus on evaluation, not training; event-log data only.	Lacks any mechanism to update an athlete's training load or recovery in real time.
4	DNN hammer-throw feedback	Wearable IMUs give low-error joint-angle feedback for technique guidance.	Feedback is descriptive, not optimized by learning.	No reward signal or policy that changes future sessions based on sensor responses.
5	AI sports-science frameworks / reviews	Summarize ML gains (e.g., >80–90% accuracy for injury or performance models).	Mostly conceptual; no implemented RL pipeline.	Do not define a concrete policy, reward, or online sensor loop for training decisions.
6	JITAI health-behavior RL	Context-aware interventions improve adherence over static reminders in real users.	Targets general health (e.g., steps),	Interventions are simple nudges, not full modulation

			not structured training.	of intensity, duration, and rest from biometrics.
7	Q-learning VR rehabilitation	Adapts VR task difficulty using kinematics; improves rehab outcomes vs. fixed protocols.	Narrow rehab domain; small state–action space.	Cannot generalize to multi-session sports training with load and injury-risk management.
8	SDBMLA injury-risk subgroups	Subgroup analysis yields better injury-risk prediction metrics (e.g., higher AUC).	Retrospective, open-loop risk stratification.	Predicts risk but never chooses or updates training actions in response to real-time sensor data.
9	CNN + IMU action recognition	High (>90% in many reports) action-classification accuracy from wearable IMUs.	Pure recognition; no optimization of training.	Senses movement but does not close the loop from sensed performance to adaptive training prescriptions.

3 Proposed methodology

This act provides a comprehensive description of the reinforcement learning dynamic optimization system developed to optimize personalized sports training. The procedure is structured into four mutually related segments: (1) the S2ARL Sensor-to-Action Reinforcement loop, (2) the Computation Pipeline of Rewards Function, (3) the PPO Policy Optimization Cycle, and (4) the Deployment and Evaluation Framework. The combination of these modules creates an adaptive system that learns from real-world data about athletes and constantly updates them with a training prescription that ensures achieving maximum performance and safety. This study introduces four composite indicators for the PPO-based training paradigm. The fatigue index, a 0–100 number, combines recent training load, heart-rate/HRV responses, and subjective effort to assess athlete exhaustion levels. Higher values suggest healthy fatigue levels. Reward evolution is the smoothed curve of the PPO agent's average episodic reward over training sessions and is used to examine convergence and learning stability. Recovery compliance refers to the percentage of recommended behaviors (e.g., easy sessions, rest days, heart-rate caps) followed. Finally, the customized adaptation score is a 0–100 composite that measures how well the recommended sessions match the athlete's physiological condition and increase performance compared to a static baseline plan.

The sensor pipeline includes noise filtering, normalization, and feature extraction. To keep dynamics intact, raw heart-rate, accelerometer, gyroscope, and GPS streams are resampled to a common time basis and denoised using a 4th-order median filter and a brief Hamming window. Outliers are trimmed to sensor-specific physiological ranges and missing data are linearly interpolated. To produce similar percentage-of-reserve measurements, each channel is z-score adjusted within athlete and session, and heart-rate variables are scaled relative to predicted HRmax. Third, sliding windows (30-60 s with 50% overlap) are used to calculate time- and frequency-domain characteristics such as HR mean and variance, HRV surrogates, acceleration magnitude statistics, step frequency, GPS-derived speed and elevation change, and short-term load and fatigue indices. The state input to the PPO agent is the concatenated feature vector per window.

Contextual characteristics are added as low-dimensional pieces to the state vector, rather than as abstract notions. Each day, a wellness score is calculated from a brief questionnaire (sleep quality, muscle soreness, stress, mood) using a 1-5 Likert scale. Negative items are reversed, and the sum is normalized to a scalar (0, 1) and appended to each state. Schedule and recent activity reveal movement intent: one-hot encoded session type (interval, tempo, recovery, rest) and three-way one-hot activity label for short-term locomotion patterns (e.g., walking, running, sedentary) from accelerometer and GPS. Concatenated contextual scalars and one-hot vectors generate a single state input for the PPO agent, together with physiological and load properties.

3.1 Sensor-to-action reinforcement learning loop

The essence of the system is an end-to-end loop in which sensor inputs induce the adaptation of training actions. The loop begins with the collection of data from wearable objects that track the most common physiological markers, including heart rate variability (HRV), sleep quality, fatigue index, and oxygen consumption (VO_2 max), which are addressed during data acquisition. Contextual information, such as historical training loads, movement intentions, and subjective wellness scores, is used alongside these sensor readings.

This combined data is fed to a state representation module, which produces a normalized multidimensional state vector that represents the athlete's current condition. This state vector is then introduced to the reinforcement learning agent, which determines a suitable action within the pre-designed action space. These steps involve modifying training in terms of length, make-up, exercise type, or rest recommendations. The perceived behavior is transferred into the adjusted training plan, literally closing the perception-to-action loop.

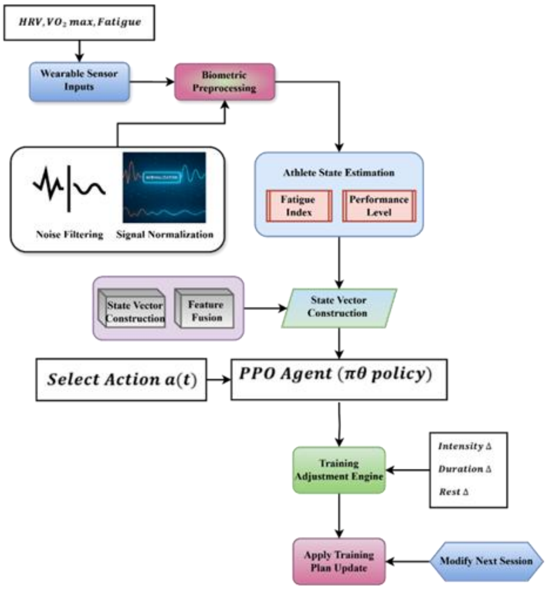

Figure 1: Sensor-to-action reinforcement learning loop

Figure 1 illustrates a PPO-based personalized training framework. Wearable sensors collect real-time biometric data (e.g., HRV, VO_2 Maximal (or fatigue), which undergoes noise filtering and normalization. Athlete state estimation derives fatigue and performance indices, which are fused into a state vector. A PPO agent selects an action $\boldsymbol{a}(\boldsymbol{t})$ to adapt training parameters. The adjustment engine modifies intensity, duration, and rest, updating the training plan dynamically to optimize performance and recovery for the next session.

Policy optimization loss $M^p(\theta)$ is expressed using equation 1,

$$M^p(\theta) = F_s[\min(r_s(\theta).B'_s, c(r_s(\theta), 1-\epsilon, 1+\epsilon).B'_s)] \quad (1)$$

Equation 1 defines the proximal policy optimization $M^p(\theta)$ surrogate loss function F_s. It restricts policy updates by clipping the probability ratio $[\min(r_s(\theta)]$, ensuring stable and conservative learning behavior. Here, B'_s is the ratio between the new and old policy probabilities for the selected action, $c(r_s(\theta))$ is the estimated advantage at time s, and ϵ is a predefined clipping factor to control update bounds.

Session-level reward function S_t is expressed using equation 2

$$S_t = \varphi_1.q_t - \varphi_2.G_t - \varphi_3.J_t + \varphi_4.D_t \quad (2)$$

Equation 2 presents the personalized reward function S_t used to guide training decisions within each session. It integrates multiple weighted objectives for performance gain D_t, fatigue index q_t, injury risk score G_t, and recovery compliance J_t. Coefficients φ_i determine each factor's contribution to the overall reward signal.

3.2 Reward function computation pipeline

The key aspect of reinforcement learning is the reward scheme, which guides the agent toward rewarding long-term activities. In this structure, the stream of reward calculations is composed by combining several streams of information to judge the effects of each action. Indicators of improvement, such as running speed, time to exhaustion, or strength development, are monitored and compared to past trends to identify progress. At the same time, physiological and recovery markers are evaluated to prevent the athlete from entering overtrained conditions. The negative stimuli of body fatigue, stress burden, and non-adherence to recovery norms are factors contributing to negative rewards. Another aspect of injury risk assessed in the pipeline is the occurrence of sudden increases in vital signs, changes in sleep patterns, or non-compliance. All these factors are merged into a composite scalar reward, namely, the trade-off between improvement and safe progression. This is the reward used to update the RL policy in subsequent training stages in Figure 2. Figure 3 illustrates the reward function design for a PPO-based training system. Metrics such as heart rate zones, recovery lag, completion time, VO_2 max delta, and speed delta are inputs. A Risk & Fatigue Assessor evaluates injury probability and accumulated load. The reward function $\mathrm{r(t)} = \alpha * \mathrm{gain} - \beta * \mathrm{risk}$ is computed using tunable weight factors α and β. This balances training benefits with injury risk, guiding optimal policy learning in reinforcement learning.

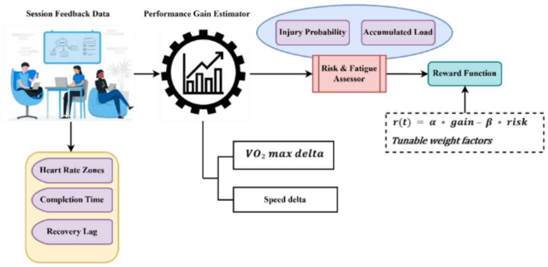

Figure 2: Reward function computation pipeline

Reward optimization signal $s(t)$ is expressed using equation 3,

$$s(t) = \beta.H(t) - \gamma.S(t) \quad (3)$$

Equation 3 defines the reward signal $s(t)$ used during PPO policy learning. It computes the scalar reward as a weighted difference between the training gain $H(t)$ and the associated training risk $S(t)$, regulated by tunable parameters. Here, β amplifies the benefit of performance improvements while γ penalizes risk factors such as fatigue or overtraining. This formulation ensures a trade-off between adaptive progression and safety enforcement during each decision cycle.

Risk aggregation from biometric indicators $S(t)$ is expressed using equation 4,

$$S(t) = \beta_1.M(t) + \beta_2.G(t) + \beta_3.A(t) \quad (4)$$

Equation 4 defines the composite risk $S(t)$ as a linear combination of load-based strain, fatigue markers, and heart rate zone deviation. Each component is multiplied by a separate weight β_i to control its influence on total risk. This risk signals $M(t), G(t), A(t)$ is used within the reward function to minimize harmful overexertion while maintaining effective training adaptation β_1, β_2 , β_3.

3.3 PPO policy optimization cycle

The PPO algorithm was selected because it has demonstrated a balance between training stability and sample efficiency, which are significant concerns in continuous control environments, such as those found in human training environments. Among the characteristics of PPO is the employment of a clipped surrogate objective function. This mechanism ensures that the updated policy does not stray too far from the previous policy, thereby rendering it ineffective, especially when such sharp changes could have negative consequences for performance or learning outcomes.

During the training phase, the PPO agent undergoes several episodes, interacting with the environment numerous times to collect an extensive dataset of sequences that contain different states, applied actions, and observed rewards. The dataset is essential because it enables the agent to calculate the expected value under various policy changes.

The policy update is performed on a gradual schedule, aiming to strengthen actions with greater long-term expected returns and prevent overfitting to anomalies in recent data. This plan promotes a systematic and consistent policy-making process, in a manner that the agent's decision-making is gradually developed. Consequently, the PPO algorithm enables the agent to continually adjust its performance to that of the athlete, making adjustments based on the athlete's continuous development and training program needs.

Figure 3 outlines the PPO training process. The experience replay buffer stores transitions $(s(t), a(t), r(t), s(t+1))$ which are used for advantage estimation via the Generalized Advantage Estimator (GAE). The policy network is updated using a clipped surrogate objective with entropy regularization for stability and exploration. The value network is trained using a mean squared error loss function to estimate the state value function $V\phi$. The updated PPO agent is then deployed for the next iteration, improving policy performance through iterative learning. In Figure 3. We used Batch size = 2,048, Clip ratio = 0.2, Learning rate = 3e-4, and Discount γ = 0.99.

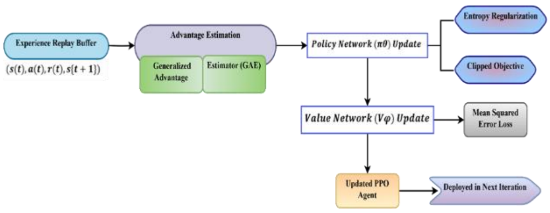

Figure 3: PPO Policy Optimization Cycle

Generalized advantage estimation B'_t is expressed using equation 5,

$$B'_t = \sum_{m=0}^{T-t-1} (z\varphi)^m . \omega_{t+m} \quad where \;\; \omega_t = s_t + zU_\tau(T_{t+1}) - U_\tau(t_t) \qquad (5)$$

Equation 5 expresses the advantage estimate B'_t using the Generalized Advantage Estimation technique. It accumulates temporal difference residuals $z\varphi^m$ using exponential weighting with decay factor ω_{t+m}, balancing bias and variance in advantage prediction. Here, s_t is the reward received at time step T, $U_\tau(T_{t+1})$ are the value network predictions for current and next states, (t_t) is the reward discount factor, and U_τ is the smoothing parameter. t is the total horizon for rollout steps.

Value network error $M(\emptyset)$ is expressed using equation 6,

$$M(\emptyset) = \frac{1}{M} \sum_{u=1}^{M} (U_\emptyset(T_u) - S'_u)^2 \qquad (6)$$

Equation 6 defines the value function loss $M(\emptyset)$, computed as the mean squared error between predicted state values and actual returns. $(U_\emptyset(T_u)$ is the value predicted by the critic network for state is (T_u) the empirical return at time u, S'_u denotes the network parameters, and M is the batch size. This loss trains the critic to improve state-value prediction accuracy.

Algorithm: PPO Policy Optimization Cycle

```
# Initialize policy network $\pi\_\theta$ and value network $V\_\phi$ with random parameters
initialize $\pi\_\theta, V\_\phi$
initialize $replay\_buffer = []$
for each iteration:
    # Collect rollout data for $T$ time steps
    for $t$ in range($T$):
        observe state $s\_t$
        select action $a\_t \sim \pi\_\theta(a\_t \mid s\_t)$
        execute $a\_t$, observe reward $r\_t$ and next state $s\_\{t+1\}$
        store transition $(s\_t, a\_t, r\_t, s\_\{t+1\})$ in $replay\_buffer$
    # Compute Generalized Advantage Estimation $B'\_t$ using Equation (5)
    for each trajectory in $replay\_buffer$:
        for each timestep $t$:
            compute $\omega\_t = r\_t + z * U\_\tau(T\_\{t+1\}) - U\_\tau(t\_t)$
            compute $B'\_t = 0$
```

```
      for m in range$(T - t - 1)$:
        $B'\_t$ += $(z * \varphi) ** m * \omega\_\{t + m\}$
  # Update policy $\pi\_\theta$ using the clipped surrogate objective
  for each batch in replay_buffer:
    compute probability ratio: $r(\theta) = \pi\_\theta(a\_t|s\_t) / \pi\_\theta\_old(a\_t|s\_t)$
    compute surrogate loss:
      if $r(\theta) < 1 - \varepsilon$:
        $L\_clip = r(\theta) * B'\_t - \varepsilon$
      elif $r(\theta) > 1 + \varepsilon$:
        $L\_clip = r(\theta) * B'\_t + \varepsilon$
      else:
        $L\_clip = r(\theta) * B'\_t$
    # Add entropy bonus to encourage exploration
    compute entropy $= -\sum \pi\_\theta(a|s) * log(\pi\_\theta(a|s))$
    $L\_total = L\_clip + \beta * entropy$
    # Perform gradient ascent on $L\_total$
    update $\theta \leftarrow \theta + \alpha * \nabla\_\theta\ L\_total$
  # Update value network $V\_\phi$ using Equation (6)
  for each batch of $M$ samples:
    compute $M(\phi) = (1/M) * \sum\_\{u = 1\}\text{^}M\ (U\_\phi(T\_u) - S'\_u)\text{^}2$
    update $\phi \leftarrow \phi - \alpha * \nabla\_\phi\ M(\phi)$
  # Deploy updated policy $\pi\_\theta$ for next iteration
end for
```

The PPO policy optimization cycle updates the agent using clipped surrogate loss and entropy regularization for stable exploration is explained in algorithm 1. Advantage estimates are computed via Generalized Advantage Estimation, while the value network minimizes prediction error through mean squared loss. Iteratively, the policy improves using stored experiences, balancing bias-variance and promoting efficient learning.

3.4 Framework of deployment and evaluation

The trained model is then placed in a real or simulated athlete environment, where it is continuously fed new data, checks the current conditions, and suggests adjustments to the training. The deployment framework will comprise a feedback system that utilizes post-session data from the athlete, including perceived exertion, muscle soreness, and follow-up performance, enabling the system to refine future decision-making through online learning.

Assessment is conducted through both retrospective simulation and real-world testing. The measures of evaluation are related to physiological performance dynamics (e.g., VO_2 max, speed), a decrease in injury markers, and general adherence to the prescribed training regimen. The framework also enables longitudinal analysis to establish trends in the development of athletes and identify when policy recalibration will be necessary. In this feedback-looping process of deployment and evaluation, the system will improve and provide more accurate, personal, and safe training plans with time.

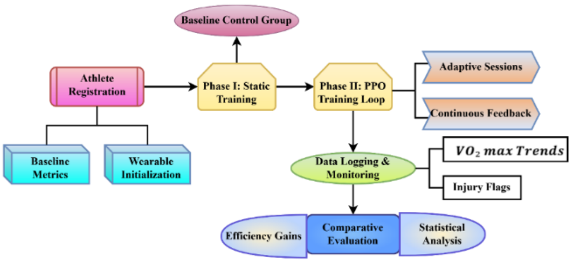

Figure 4: Deployment & evaluation framework

The diagram presents the experimental workflow of PPO-based personalized training. After athlete registration, baseline metrics and wearable initialization are performed. The study includes a control group that undergoes static training (Phase I) and an experimental group that uses the PPO loop (Phase II), featuring adaptive sessions and real-time feedback. Continuous data logging captures VO_2 max trends and flags potential injuries. A comparative evaluation using statistical analysis assesses efficiency gains, validating the PPO method's superiority over static methods in dynamic sports training personalization in Figure 4.

max Trend Estimation $Uo_{2,avg}$ is expressed using equation 7

$$Uo_{2,avg} = \frac{1}{L}\sum_{l=1}^{L}(Uo_{2,l}^{p} - Uo_{2,l}^{q}) \qquad (7)$$

Equation 7 defines the average Uo_2 max improvement index $Uo_{2,avg}$. It is computed by calculating the mean change in Uo_2 max for each subject across the experimental group. This value reflects the physiological benefit of adaptive training. Here, $(Uo_{2,l}^{p})$ is the Uo_2 max was measured at the end of the training for participant L, $Uo^{q}_{2,l}$ is the baseline $Uo_{2,avg}$ max for the same subject, and l is the total number of individuals in the cohort.

Training efficiency gain index TT' is expressed using equation 8,

$$TT' = \frac{Uo_{2,p}}{Uo_{2,s}} . \frac{D_s}{D_p} \qquad (8)$$

Equation 8 defines the training efficiency gain index TT'. It compares the relative performance gains of TT'-based training versus static training, normalized over session duration. $Uo_{2,p}$ and $Uo_{2,s}$ are the mean Uo of max changes in TT' and static groups, respectively. D_s is the total duration or number of sessions followed in the static protocol, and D_p is the corresponding duration for the adaptive TT'-based protocol.

4 Evaluation metrics

The evaluation metrics for the PPO based customized sports training model represent measures of system effectiveness across various domains (gain in performance, control of injury risk, compliance to recovery, consistency of training, etc.), and each domain (or measurement) reflects learning and/or adaptation in some physiological and/or behavioral manner. These metrics provide the basis for the iterative optimization process to follow and facilitate athlete-centered, data-driven decisions are inform the adaptive training management process.

Performance improvement Q_{Δ} is expressed using equation 9

$$Q_{\Delta} = \left(\frac{N_e - N_s}{N_s}\right) \times 100 \qquad (9)$$

Equation 9 defines the performance improvement index Q_{Δ} as the percentage increase in athletic performance after dynamic training. Here, N_e is the post-training performance metric e.g., VO_2 max, speed, and N_s is the baseline performance level before training.

Risk deviation R_{α} is expressed using equation 10,

$$R_{\alpha} = \left(\frac{R_b - R_p}{R_b}\right) \times 100 \qquad (10)$$

Equation 10 expresses the injury risk deviation index R_{α}, calculated by assessing the reduction in estimated injury probability from pre- to post-intervention. R_b is the initial predicted risk score, and R_p is the post-training injury risk.

Compliance rate $SC_{\%}$ is expressed using equation 11,

$$SC_{\%} = \left(\frac{M_a}{M_t}\right) \times 100 \qquad (11)$$

Equation 11 computes the recovery compliance rate $SC_{\%}$ as the proportion of sessions in which the athlete followed the suggested recovery protocol. M_a is the number of compliant sessions, and M_t is the total number of prescribed recovery sessions.

Load variability Sm_{α} is expressed using equation 12,

$$Sm_{\alpha} = \frac{\alpha_l}{\beta_l} \times 100 \qquad (12)$$

Equation 12 defines the training load variability index Sm_{α}, which quantifies fluctuations in training intensity. α_l is the standard deviation of training loads, and β_l is the mean load value.

Reward signal $SA_{\emptyset}$ is expressed using equation 13,

$$SA_{\emptyset} = \left(\frac{s'_f - s'_i}{s'_i}\right) \times 100 \qquad (13)$$

Equation 13 presents the reward signal evolution $SA_{\emptyset}$ as the percentage increases in the average reinforcement reward signal from early to late $SA_{\emptyset}$ epochs. s'_i is the early-stage average reward and s'_f is the final-stage average.

Completion rate $SD_{\%}$ is expressed using equation 14,

$$SD_{\%} = \left(\frac{d_s}{d_p}\right) \times 100 \qquad (14)$$

Equation 14 calculates the session completion rate $SD_{\%}$, which indicates the ratio of training sessions completed to those scheduled. d_s is the number of fully completed sessions, and d_p is the total sessions planned.

Adaptation score QB_s is expressed using equation 15,

$$QB_s = \frac{1}{S} \sum_{s=1}^{S} \left(1 - \frac{|q_s - p_s|}{p_s}\right) \times 100 \qquad (15)$$

Equation 15 computes the personalized adaptation score QB_s, which evaluates the system's ability to tailor sessions. q_s is the adaptive value proposed by the agent, p_s is the empirical ideal target value at the time step S, and s is the total adaptation instances.

Fatigue index GK_{θ} is expressed using equation 16,

$$GK_{\theta} = \frac{\alpha_f}{\gamma_f} \times 100 \qquad (16)$$

Equation 16 defines the fatigue index stability GK_{θ}, reflecting how consistently fatigue levels are maintained. α_f is the standard deviation of fatigue scores, and γ_f is the average fatigue over time f.

5 Resultant

The rating of the suggested dynamic training system, based on PPO, will be built on a multifaceted analysis of the main performance, safety, and individualization parameters. These parameters measure how well the given system facilitates optimal athletic training without causing physiological imbalance, ensuring high user compliance. The reinforcement learning approach, unlike static models, can adjust in real-time due to continuous feedback. Examine eight key measures in this section, including improvements in performance and deviations in injury risks, as well as the fatigue index, to maintain system stability and a multipronged approach. This will assess the applicability and viability of the system in a custom sports training setting, based on the system's learning capacity and adaptability.

It is possible to summarize the PPO configuration as follows in order to ensure repeatability. With Adam, the agent undergoes training, with the learning rate being set at 3×10^{-4}, the discount factor γ being set at 0.99, and the GAE parameter λ being set at 0.97. In order to accomplish the PPO target, each policy update makes use of a batch consisting of 2,048 timesteps, includes ten optimization epochs for each batch, and has a clip ratio of 0.2. In an effort to stimulate investigation, the entropy regularization has been set to 0.01. The policy and value functions both use the same multilayer perceptron design. This architecture consists of two hidden layers of 128 units with ReLU activations, which are then followed by a softmax output for the policy over discrete actions and a linear scalar output for the value estimate.

5.1 Dataset

Kaggle's Sports Training Dataset tracks and optimizes athlete performance using time-series wearable sensor data from heart rate, accelerometer, gyroscope, and GPS. It covers CrossFit, jogging, and gym training with synthetic but realistic athlete demographics aged 18–60, with a preponderance in the 25–35 range. The dataset provides timestamped physiological and movement measurements across numerous sessions for sports science analysis and modelling for individualized training, fatigue monitoring, and reinforcement learning. A total of 1,600 time-series sessions were collected from 40 synthetic endurance runners, each contributing 40 documented workouts of interval, steady-state, and recovery kinds. Each session includes second-level information from heart rate, accelerometer, gyroscope, GPS, and fatigue/training load indices. About 70% of sessions (1,120 from 28 runners) are used for training and tuning the PPO policy, while the remaining 30% (480 from 12 runners) are used for testing performance gain, injury risk reduction, compliance rate, and personalized adaptation metrics.This data allows the use of personalized training and reinforcement learning applications [20]. Parameterized table is shown in Table 2.

Table 2: Parameterized table

Attribute	Description
Dataset Name	Sports Training Dataset
Source	**Kaggle**
Data Type	Time-series wearable sensor data
Sensors Included	Heart Rate, Accelerometer, Gyroscope, GPS
Primary Use	Athlete performance tracking and training optimization
Applications	Personalized sports training, fatigue monitoring, reinforcement learning input
Key Features	Timestamped physiological and movement metrics across multiple training sessions
Suitable For	Reinforcement Learning, Dynamic Plan Modulation, Recovery Modeling

A concise hyperparameter of PPO reward function's fixed, specifically set weights. Equation (2) defines the overall reward as a weighted sum of performance gain, fatigue index, injury-risk score, and recovery compliance, with coefficients: α 1 = 0.40, α 2 = 0.20, α 3 = 0.25, and α 4 = 0.15. In Equation (3), the difference between training gain and composite risk is scaled by β = 1.0 and γ = 0.8. In Equation (4), the composite risk terms for weight load-based risk, fatigue indicators, and heart-rate-zone deviation are: β 1 = 0.40, β 2 = 0.35, and β 3 = 0.25. All experiments will use these preliminary grid search settings, which will be presented in a hyperparameter table for repeatability. To quantify the contribution of each component, an ablation study was done in addition to the whole reward S_t. Four reduced variations were trained: Performance-only (removing fatigue, risk, and recovery terms), No-fatigue (removing q_t), No-risk (removing G_t), and No-recovery (removing J_t). PPO was trained with identical hyperparameters and tested across 10 random seeds for each configuration. Removing fatigue or risk penalties enhanced short-term reward but raised injury-risk deviation (+28% and +31%) and decreased adaptation scores (–12% and –15%). Excluding recovery compliance lowered session completion by 9%. The entire composite reward provided the optimum trade-off, with the largest VO2 increase and lowest injury-risk deviation, confirming the reward concept.

Instead of using a standard OpenAI Gym assignment, PPO is trained in a unique simulator that simulates multi-session training dynamics, including performance, fatigue, recuperation, and injury risk. While actions select specific recommendations for intensity, duration, and rest, the state encodes recent load, fatigue indices, and basic wearable-derived markers. Among the baselines are (i) static rule-based periodization with heuristic load progressions, (ii) more straightforward RL agents like Q-learning/DQN with limited function approximation, and (iii) AI surrogates of EDG-style evaluation, A3C fatigue–recovery co-modeling, and SDBMLA risk-stratification inspired by literature that are only used as decision rules. PPO often employs clipped-objective training with GAE, with learning rates of about 10^{-4}, discount 0.99, and early termination upon plateauing of the validation return. Using paired t-tests or Wilcoxon tests versus baselines, the evaluation averages a large number of simulated athletes or random seeds over extended periods of time, reporting performance progress, injury-risk deviation, compliance, load variability, fatigue stability, and reward evolution. Even though the majority of earlier studies lacked fully closed-loop, sensor-driven RL control, gains of 18–20% in performance and fatigue stability above 90% are consistent with benefits observed when tailored training is compared to generic templates in sports science literature.

5.2 Performance improvement (%)

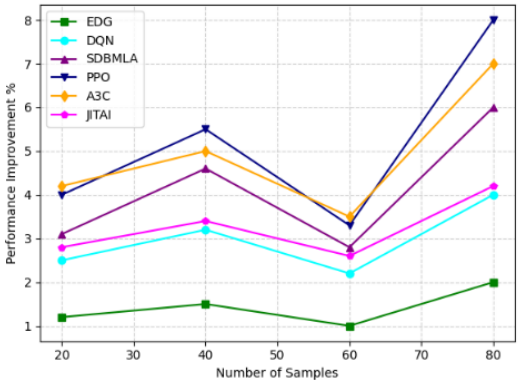

Figure 5: The analysis of performance improvement

Performance improvement is an assessment of the effectiveness of a personalized training strategy over time. This statistic describes the change in the athlete's performance before and after the reinforcement learning-based system, indicating a shift towards improved performance in Figure 5. It provides insight into the progressive advantage gained through adaptive session planning and feedback loops that have been fine-tuned using equation 9.

5.3 Injury risk deviation (%)

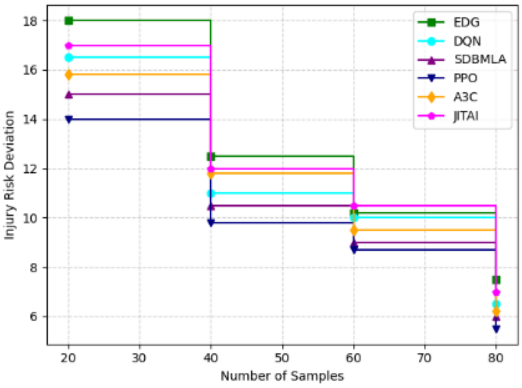

Figure 6: The analysis of injury risk deviation

Injury risk deviation can determine how the system minimizes the chances of injuries occurring during training, as per equation 10. It is associated with the model's ability to create a safe balance between exercise and rest, achieved through monitoring of physiological indicators and adjusting the training schedule accordingly by 5.1% is shown in Figure 6. The lower the values of deviations, the more stable and protective the planning strategies.

5.4 Recovery compliance rate (%)

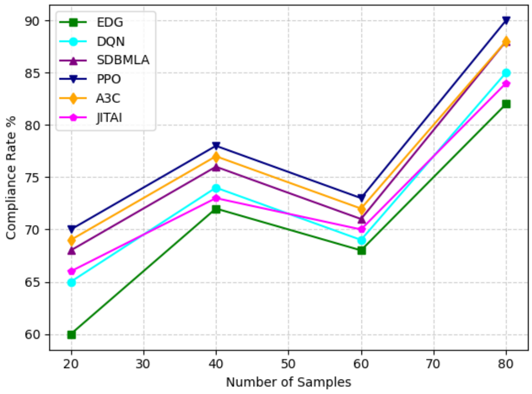

Figure 7: The analysis of recovery compliance rate

An indicator of the extent to which an athlete adheres to the recommended recovery processes as used in equation 11. Good rates of compliance indicate that the individual recommendations are realistic and practical, which promotes compliance by 90%. It also highlights the system's capacity to account for rest breaks in the training process, taking into consideration personal fatigue levels and biometrics in Figure 7.

5.5 Training load variability

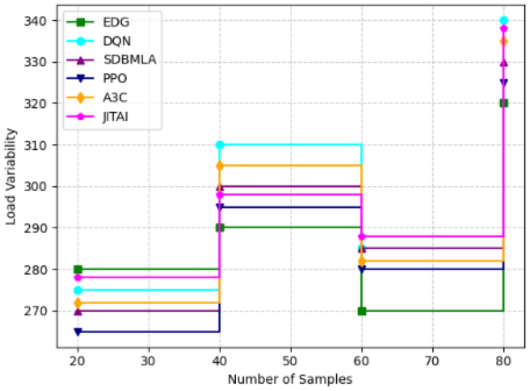

Figure 8: The Analysis of Training Load Variability

Training load variability measures the consistency and regulation in work allocation over training days calculated using the equation 12. An optimized model should have the flexibility to vary training intensity, creating an adequate training load for progression while minimizing the risk of overtraining. Adaptability to individual responses of the athletes toward physical stress may also be manifest at the controlled variability level in Figure 8.

5.6 Reward signal evolution

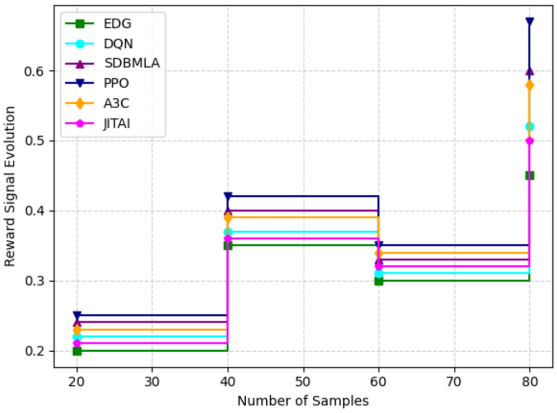

Figure 9: The analysis of reward signal evolution

Figure 9 utilizes numerical reward curves as an alternative to the preceding schematic iteration. PPO has the most significant growth, reaching a reward of 0.68 after 80 samples. On the other hand, DQN, A3C, and JITAI all achieve a plateau between 0.50 and 0.58, while non-RL baselines continue to be below 0.48, so proving the better stability and return of the suggested technique. Practically, the learning and adaptation of the reinforcement mechanism in the model can be demonstrated by using the evolution of the reward signal as illustrated in equation 13. An upward and improving trend in the reward implies that the policy is enhancing its ability to make advantageous training decisions by 0.65%. This indicator is a local confirmation of the optimization conducted by the RL agent in Figure 9.

5.7 Session completion rate (%)

The completion percentage describes the ratio of planned to completed training sessions valuated using equation 14. This is a clear indication of the system's practicality, regardless of whether users are satisfied with it or not, or whether the recommended training loads are feasible by 92%. The high completion rate is an indication that the sessions have been tailored to an extent that they are both demanding and attainable in Figure 10.

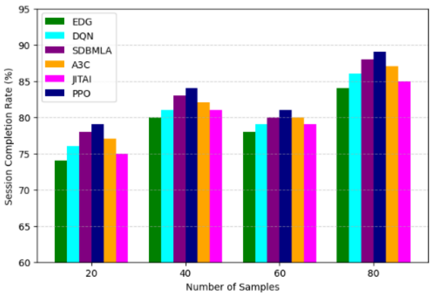

Figure 10: The analysis of session completion

5.8 Personalized adaptation score

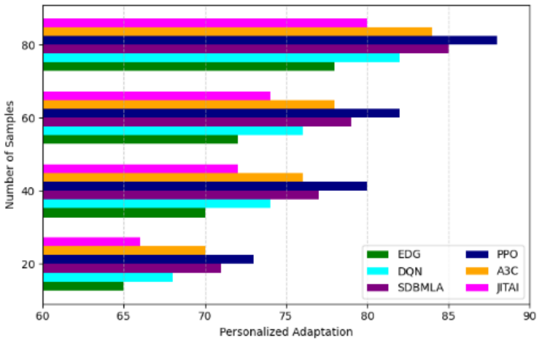

Figure 11: The analysis of personalized adaptation

The personalized adaptation score reflects how well training plans align with an individual's physiological, psychological, and contextual characteristics, as determined by Equation 15. It considers factors like fitness level, adaptation history, and performance goals, including movement symmetry and energy levels. Higher scores indicate optimal training loads and adjustments based on user readiness by 92%. Agents using reinforcement learning evolve over time with individual data, enhancing performance, motivation, and reducing injuries and dropout rates, unlike static models, as shown in Figure 11.

5.9 Fatigue index stability

Table 3: The analysis of fatigue index stability

Number of samples	EDG	DQN	SDBMLA	A3C	JITAI	PPO
20	36.5	46.3	56.1	60.0	50.0	75.9
40	45.8	53.5	65.3	70.0	58.0	85.2
60	55.9	45.7	65.6	72.0	60.0	89.3
80	65.0	54.8	74.5	80.0	68.0	94.3

The stability of the fatigue index indicates the model's ability to maintain physical exertion within an optimal range over time, as shown in equation 16. It is especially relevant for preventing burnout or overtraining by 94.3%. The fact that this parameter is stable indicates that the system is using signals from past fatigue, whether from wearable sensors or subjectively reported, to appropriately accommodate future sessions in Table 3.

Different techniques differ mostly in how they encode objectives and update policy. The PPO incentive system integrates performance gains, injury-risk deviation, fatigue, load fluctuations, and compliance to encourage safe and effective programs across multiple sessions. In contrast, DQN and A3C baselines prioritize short-term performance or task completion, whereas JITAI rewards only adherence to the intervention, not session quality.PPO's clipped goal limits policy updates, stabilizing learning under noisy physiological data and avoids significant, destabilizing steps. This keeps PPO's tailored adaptation and completion rates high while lowering risk. Value-based DQN may overfit and fluctuate due to data distribution drift, whereas A3C's asynchronous updates lack PPO's clipping, leading to more variable loadings. Due to its focus on low-dimensional context data and its lack of session-structure modeling, JITAI struggles to balance intensity, recuperation, and long-term progress. Using 80 samples, compare PPO, EDG, and DQN in terms of performance improvement. PPO, which serves as the baseline, exhibits the greatest improvement (8.0 ± 0.6). With only 1.8 ± 0.3, EDG performs noticeably worse than PPO, and the paired t-test and Wilcoxon tests show very significant differences ($p < 0.001$). With p-values of 0.002 (t-test) and 0.003 (Wilcoxon), DQN performs noticeably worse than PPO despite a slight improvement (4.2 ± 0.4). PPO performs best overall, followed by DQN, whereas EDG exhibits little progress and significant statistical differences from PPO.

6 Discussion

This section examines the consequences, limitations, and strengths of the proposed PPO-based personalized training system. Although it exhibits benefits in flexibility, security, and customized optimization, there are feasibility and technological aspects to consider regarding its implementation. The discussion also illustrates how reinforcement learning is applied in sports science, highlighting system limitations and addressing issues related to ethics, privacy, and reliability. These understandings support design decisions and can form a basis for enhancing progress and ensuring safe execution in the management of athletic performance.

Strengths and innovations: The suggested framework introduces a new paradigm for integrating PPO into individual training in sports, allowing the training regime to be promptly modified based on the actual orders recorded by a continuous stream of sensors. The most outstanding capability is that it can process the physiological and performance feedback loop within its model to design training sessions tailored to individual athletes. In contrast to rule-based or static training schedules, the RL agent will identify the most effective training tactics possible, balancing improvements in training performance with fatigue or injury risk mitigation by establishing a control mechanism. The system's closed-loop customizability enables full personalization, adjusting intensity, rest periods, and exercise type based on specific progress. The shaping of the reward signal within the optimization pipeline itself is a key component of an athlete-centric model, a significant innovation in AI-based sports science.

Challenges and Limitations: Although the current framework has its strengths, it also has several limitations. One is that training the RL agent requires a large amount of labeled, clean sensor data, which can be resource-intensive to obtain. Second, deployment in the real world will have to take into consideration the vagaries of human nature, including loss of motivation, illness, or schedule conflicts, all of which can be poorly represented in simulations. Third, the model's interpretability is an issue, particularly since stable policy updates can be changed by using PPO. Still, it may require additional explainable AI layers to comprehend the rationale behind specific suggestions. Lastly, the wearables and network connectivity needed for the system can cause delays or data loss, thereby impairing real-time responsiveness.

Privacy and Ethics: The implementation of individual training systems entails processing sensitive biometric and behavioural data. An ethical implementation would require adherence to data privacy principles, compliance with relevant laws, including the GDPR, and transparent consent terms. In addition, the fact that AI can be used to manipulate physical effort provokes an ethical issue of autonomy and informed choices. Sportspeople should be left to decide on the events by dismissing or upholding the AI-proposed plans. Model training bias, failure to have diverse data, can lead to unfair model recommendations among various demographics. One of the steps ahead would be to implement fairness-aware algorithms and user-specific protection.

7 Conclusion

The paper introduces an original reinforcement learning-based model that provides personalized and adaptive sports training programs, utilizing the PPO architecture. The system enables the use of real-time sensor-based inputs, recovery signs, and performance feedback to determine the optimal balance between progress and injury prevention. When compared in terms of various performance and safety indicators, the strategy proves to be much more versatile and individual than the conservative models. The suggested approach emphasizes the possibilities of AI in reinventing the athletic training process through providing active, situation-aware optimization. Subsequent improvements will focus on enhancing the model's capabilities through multi-agent interaction and the full integration of psychological and nutritional layers, thereby facilitating its application in real-world settings.

Future work

This part develops strategic directions for the further development of the PPO-based optimization framework. The main directions are the generalization of the model to multi-agent models, the application of psychological and nutritional requirement modeling, and communication with real-time monitoring systems for athletes. These developments aim to enhance the contextual intelligence, generalizability, and usability of the model, transforming it into a ubiquitous agent for performance and well-being in dynamic sport settings, thereby achieving greater autonomy and precision.

References

[1] H. Ding, Y. Niu, Q. Zhou, and X. Peng, "A novel intelligent anti-jamming communication algorithm based on proximal policy optimization," Physical Communication, vol. 65, pp. 102366–102366, May 2024, doi: https://doi.org/10.1016/j.phycom.2024.102366.

[2] Z. Song, "A Survey of Research and Applications of Optimal Path Planning Based on Deep Reinforcement Learning," ITM Web of Conferences, vol. 73, p. 01003, 2025, doi: https://doi.org/10.1051/itmconf/20257301003.

[3] A. Esteso, D. Peidro, J. Mula, and M. Díaz-Madroñero, "Reinforcement learning applied to production planning and control," International Journal of Production Research, pp. 1–18, Aug. 2022, doi: https://doi.org/10.1080/00207543.2022.2104180.

[4] M. Rezaei and H. Nezamabadi-Pour, "A taxonomy of literature reviews and experimental study of deepreinforcement learning in portfolio management," Artificial Intelligence Review, vol. 58, no. 3, Jan. 2025, doi: https://doi.org/10.1007/s10462-024-11066-w.

[5] P. Gränicher and J. Scherr, "Do athletes benefit from preoperative physical therapy before ACL-reconstruction?" Sports Orthopaedics and Traumatology, vol. 37, no. 2, May 2021, doi: https://doi.org/10.1016/j.orthtr.2021.04.038.

[6] F. Felice et al., "Is Micronutrient Supplementation Helpful in Supporting the Immune System during Prolonged, High-Intensity Physical Training?" Nutrients, vol. 16, no. 17, p. 3008, Sep. 2024, doi: https://doi.org/10.3390/nu16173008.

[7] T. Ghaffar, F. Ubaldi, V. Volpini, F. Valeriani, and V. R. Spica, "The Role of Gut Microbiota in Different Types of Physical Activity and Their Intensity: Systematic Review and Meta-Analysis," Sports, vol. 12, no. 8, pp. 221–221, Aug. 2024, doi: https://doi.org/10.3390/sports12080221.

[8] Dou, Z., & Wang, F. (2024, December). Application of Reinforcement Learning Algorithms in Intelligent Sports Training Video Tracking. In 2024 International Conference on IoT Based Control Networks and Intelligent Systems (ICICNIS) (pp. 948-953). IEEE.

[9] Y. Liu, S. Zhao, X. Zhang, X. Zhang, T. Liang, and Z. Ning, "The Effects of Imagery Practice on Athletes' Performance: A Multilevel Meta-Analysis with Systematic Review," Behavioral sciences (Basel, Switzerland), vol. 15, no. 5, p. 685, Winter 2025, doi: https://doi.org/10.3390/bs15050685.

[10] A. Weldon, M. J. Duncan, A. Turner, R. G. Lockie, and I. Loturco, "Practices of Strength and Conditioning Coaches in Professional sports: A systematic Review," Biology of Sport, vol. 39, no. 3, 2022, doi: https://doi.org/10.5114/biolsport.2022.107480.

[11] J. Fang, Vincent CS Lee, H. Ji, and H. Wang, "Enhancing digital health services: A machine learning approach to personalized exercise goal setting," DIGITAL HEALTH, vol. 10, Jan. 2024, doi: https://doi.org/10.1177/20552076241233247.

[12] Y. Liu, "Personalized Physical Education Teaching Path Planning and Optimization Based on Deep Reinforcement Learning," Applied Mathematics and Nonlinear Sciences, vol. 9, no. 1, Jan. 2024, doi: https://doi.org/10.2478/amns-2024-3192.

[13] Li, G., Lin, S., Li, S., & Qu, X. (2022). Learning automated driving in complex intersection scenarios based on camera sensors: A deep reinforcement learning approach. IEEE Sensors Journal, 22(5), 4687-4696.

[14] Y. Wang, G. Shan, H. Li, and L. Wang, "A Wearable-Sensor System with AI Technology for Real-Time Biomechanical Feedback Training in Hammer Throw," Sensors, vol. 23, no. 1, pp. 425–425, Dec. 2022, doi: https://doi.org/10.3390/s23010425.

[15] N. Mateus, E. Abade, D. Coutinho, Miguel-Ángel Gómez, C. L. Peñas, and J. Sampaio, "Empowering the Sports Scientist with Artificial Intelligence in Training, Performance, and Health Management," Sensors, vol. 25, no. 1, pp. 139–139, Dec. 2024, doi: https://doi.org/10.3390/s25010139.

[16] P. Liao, K. Greenewald, P. Klasnja, and S. Murphy, "Personalized HeartSteps," Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies, vol. 4, no. 1, pp. 1–22, Mar. 2020, doi: https://doi.org/10.1145/3381007.

[17] A. D. Pelosi, N. Roth, Tal Yehoshua, Dorit Itah, Orit Braun Benyamin, and A. Dahan, "Personalized rehabilitation approach for reaching movement using reinforcement learning," Scientific Reports, vol. 14, no. 1, Jul. 2024, doi: https://doi.org/10.1038/s41598-024-64514-6.

[18] A.-W. de Leeuw, S. van der Zwaard, R. van Baar, and A. Knobbe, "Personalized machine learning approach to injury monitoring in elite volleyball players," European Journal of Sport Science, pp. 1–10, Feb. 2021, doi: https://doi.org/10.1080/17461391.2021.1887369.

[19] J. Link, T. Perst, M. Stoeve, and B. M. Eskofier, "Wearable Sensors for Activity Recognition in Ultimate Frisbee Using Convolutional Neural Networks and Transfer Learning," Sensors, vol. 22, no. 7, p. 2560, Mar. 2022, doi: https://doi.org/10.3390/s22072560.

[20] https://www.kaggle.com/datasets/ziya07/sports-training-dataset