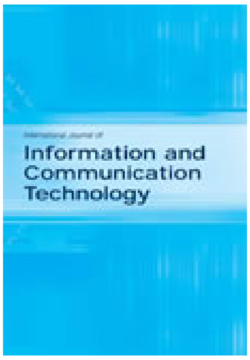
INDERSCIENCE PUBLISHERS
Linking academia, business and industry through research



International Journal of Information and Communication Technology

ISSN online: 1741-8070 - ISSN print: 1466-6642
https://www.inderscience.com/ijict

Application of deep reinforcement learning in sports competitive decision-making

Zijing Wang, Zhijun Li

DOI: 10.1504/IJICT.2026.10075905

Article History:

Received:	19 September 2025
Last revised:	10 October 2025
Accepted:	14 October 2025
Published online:	04 February 2026

Application of deep reinforcement learning in sports competitive decision-making

Zijing Wang

Guangxi Technology and Business Vocational College,
Nanning, 530000, Guangxi, China
Email: wang202102212025@163.com

Zhijun Li*

Guangxi College for Preschool Education,
Nanning, 530000, Guangxi, China
Email: Junzhili860101@163.com
*Corresponding author

Abstract: Demand for tactical optimisation and decision support in sports competitions is growing; traditional rule-based methods suffer from poor adaptability and latency. This study builds a DRL-based decision model for Wushu Sanda, trained on multi-source data and validated in simulation. The agent learns policies via interaction to optimise tactical choices in dynamic contexts. Compared with rule-based and classical RL baselines (Q-learning, SARSA), our model achieves higher decision accuracy, larger cumulative reward, and faster convergence. It adapts to diverse scenarios and supports real-time tactical adjustment. We also identify challenges in data quality, computational cost, and cross-sport generalisation. The findings highlight DRL's practicality for competitive decision-making and outline directions for improving interpretability, sample efficiency, and deployment in live matches.

Keywords: deep reinforcement learning; DRL; sports competition; decision optimisation.

Reference to this paper should be made as follows: Wang, Z. and Li, Z. (2026) 'Application of deep reinforcement learning in sports competitive decision-making', *Int. J. Information and Communication Technology*, Vol. 27, No. 5, pp.61–79.

Biographical notes: Zijing Wang works at Nanyang Normal University from 2005 to 2009 and at Yunnan Normal University from 2009 to 2012. His research interests include development path of ethnic sports special towns in Guangxi.

Zhijun Li works at Yunnan Normal University from 2005 to 2012 and at Eternal University of the Philippines from 2020 to 2023.

1 Introduction

With the rapid development of artificial intelligence technologies, deep reinforcement learning (DRL) has achieved notable applications across various fields. In the realm of sports competition, traditional decision-making methods, which rely on heuristic rules and expert systems, have practical value but fail to generate optimal decisions when confronted with the complex and dynamic nature of competitive environments. In the course of a match, athletes and coaches' judgments are influenced by emotional fluctuations, the strategies of opponents, and environmental conditions, making traditional decision models insufficient for real-time decision-making.

Unlike domains such as finance, manufacturing, or transportation, sports competitions present a set of unique challenges for DRL applications. Decision-making in sports involves continuous high-speed interactions between multiple agents (players), uncertainty from opponents' strategies, and complex spatiotemporal dynamics. Unlike algorithmic trading or production scheduling – where environments can be relatively structured and predictable – sports scenarios require agents to interpret rapidly changing tactical contexts, integrate heterogeneous real-time data (e.g., physiological signals, game states, opponent behaviours), and make split-second decisions that influence outcomes. These characteristics make sports decision-making a particularly demanding environment for DRL, requiring not only accurate policy learning but also high adaptability and robustness in non-stationary, adversarial settings. The theoretical foundation of applying DRL to sports decision-making lies in its ability to bridge the gap between traditional decision models and adaptive, data-driven approaches. Traditional models, such as rule-based systems and expert heuristics, rely on predefined tactical patterns and human judgment. While useful, these models lack the flexibility to adapt to rapidly evolving competitive environments. DRL, in contrast, combines reinforcement learning's sequential decision-making framework with deep neural networks' ability to model high-dimensional states. This allows agents to approximate optimal policies by continuously interacting with dynamic competitive environments. Conceptually, DRL extends Markov decision processes (MDPs) to complex, adversarial settings, enabling strategy optimisation under uncertainty – a fundamental characteristic of competitive sports.

DRL, which simulates human learning and decision-making processes, provides new avenues for decision-making in sports competitions. Unlike traditional models, DRL continuously refines decision-making strategies by interacting with the environment, improving the accuracy and efficiency of decisions. By considering real-time data such as athlete performance, opponent behaviour, and match progress, DRL enables the adjustment of decision strategies to optimise outcomes during competition. For example, in basketball, DRL models can dynamically adjust defensive strategies or identify the player most likely to score, while in soccer, DRL can analyse opponent offensive tactics to adjust defensive formations. These capabilities enhance team performance during matches.

With the advancement of data collection and processing technologies, sports-related data resources have become increasingly abundant. The integration of big data technologies with deep learning provides a solid foundation for DRL applications in sports decision-making. Every stage, from data collection and sample selection to model training and validation, depends on sophisticated data analysis techniques.

This study, taking Wushu Sanda as an example, leverages the advantages of DRL to explore intelligent pathways for sports decision-making, fostering deeper integration between sports science and artificial intelligence. The goal is to provide sports teams with scientific tactical support, enhancing the timeliness and accuracy of decisions during real competitions and advancing the modernisation of competitive sports.

Recent studies have increasingly focused on applying DRL in sports contexts, reflecting growing academic interest in this area. For example, Schütz and Schultheiss (2020) explored implicit motivational patterns in gymnasts using data-driven decision frameworks, while Bae et al. (2024) validated multidimensional assessment tools for combat sports, providing valuable structured data for DRL applications. Shan (2023) examined ethical boundaries in sports competition, highlighting the need for adaptive decision-making frameworks. These recent contributions indicate that DRL in sports is moving beyond conceptual exploration toward practical implementation, underscoring the timeliness and relevance of this study.

At present, the research on the application of DRL in many fields focuses on optimising decision-making and automation. Jaisson (2022) discusses the application of deep differentiable reinforcement learning in financial transactions, and thinks that DRL can improve the optimisation ability of trading strategies, especially when dealing with complex market environment. Guan et al. (2023) proposed that reinforcement learning-driven deep problem generation can deal with semantically rich problems, enhance the semantic depth of generated problems, and improve the quality of information processing and decision-making.

In manufacturing field, Liang et al. (2022) applied DRL to Lenovo's notebook production scheduling, and found that this method performed well in optimising production scheduling, and the production efficiency and cost control were balanced in the multi-objective and uncertain environment. The research on dynamic scheduling optimisation using DRL provides a new perspective for decision-making in industrial manufacturing, and this technology can cope with highly complex and dynamic production environment. Zeng et al. (2024) studied the application of deep Bayesian inverse reinforcement learning in unmanned driving, and proposed a driving decision-making framework with uncertain knowledge. It simulates the decision-making behaviour of human drivers and shows higher adaptability in the face of uncertain traffic environment. Wang et al. (2023) applied DRL to the problem of sequence goal setting, and found the advantages of this method in long-term decision-making problems, which helped enterprises, optimise resource allocation in multi-stage decision-making. Raziei and Moghaddam (2021) put forward a modular method of DRL, which enables the system to quickly adapt to different tasks through policy transfer technology and improves the flexibility of the automation system. Ahamed et al. (2021) put forward a crowdsourcing urban distribution optimisation model based on DRL, which can solve the problems of path planning and resource scheduling in urban distribution and improve distribution efficiency and service quality.

Wallrafen et al. (2022) discussed the competition between professional sports leagues and fan substitution, analysed the economic characteristics of competition in the sports market, and pointed out that deep learning method can provide a new perspective for the analysis of sports economics. Rossi et al. (2020) analysed the perception of non-profit sports clubs when they compete with commercial sports suppliers, and put forward the factors that affect the competitive strategy of non-profit sports clubs. Furley (2019) explored the relationship between human nature and competition through modern sports competition, and revealed the profound influence of competition on individual psychology. DRL is widely used in industrial manufacturing, transportation, logistics and finance, and provides a new perspective in sports and sociology.

1.1 Research content and purpose

Focusing on the application of DRL in sports decision-making, this study explores how to use DRL to optimise the decision-making process in sports competition and improve the overall performance of athletes and teams. With the continuous accumulation of sports data and the advancement of artificial intelligence technology, DRL has gradually become an important tool to solve complex decision-making problems, and DRL has outstanding advantages in dynamic, changeable and uncertain environments. Decision-making in sports competition involves complex real-time data processing, including the selection and adjustment of various tactical strategies, which has high computational complexity and real-time requirements. Aiming at the practical problems in sports decision-making, this study analyses how DRL is applied to tactical selection, player behaviour prediction and real-time decision-making in the course of competition (Schütz and Schultheiss, 2020). Through data collection and sample selection, combined with the performance and decision-making scenes in the actual competition, a DRL model suitable for sports competition is constructed. This paper discusses the applicability and performance differences of DRL in different sports, compares its application effects in different sports such as football and basketball, and emphatically analyses the promotion effects in tactical decision-making and athlete performance prediction. Compared with traditional decision-making methods, the advantages and limitations of DRL in practical application are evaluated. Aiming at the application effect of DRL in sports competition, a model verification and evaluation mechanism is designed to systematically analyse the influence on the improvement of competitive level.

While these studies demonstrate DRL's effectiveness across various domains, their contexts differ fundamentally from sports decision-making. The reviewed works establish a strong methodological foundation, but a direct translation of these approaches to sports is non-trivial. Sports competitions involve strategic interactions between multiple agents, dynamic tactical adjustments, and real-time physiological responses, all of which require tailored DRL frameworks. Therefore, bridging the methodological advances in other fields with the unique characteristics of sports is essential to develop effective decision-making models.

The purpose of this study is to provide a data-driven method for sports decision-making and explore how DRL can improve the shortcomings in the existing decision-making process. On the one hand, with the help of DRL, the decision-making process of tactical adjustment in the competition can be optimised and self-adjusted in a shorter time; On the other hand, it explores how to use DRL to model the uncertainty in the competition and improve the accuracy and real-time performance of decision-making. Through the research of different sports events, the universality and adaptability of DRL model in various competitive scenes are verified. The research will also pay attention to the data characteristics needed in the model training process, and discuss how to improve the decision-making quality of the model through appropriate data preprocessing, feature engineering and reinforcement learning algorithm design. The ultimate goal of this study is to promote the application of DRL in sports decision-making and provide scientific decision-making support tools for coaches, athletes and event organisers. The research results provide new ideas for improving the level of sports competition, promote the innovative application of artificial intelligence technology in sports field, and provide experience for the development and improvement of intelligent decision-making system.

This study hypothesises that DRL can outperform traditional decision-making approaches in Wushu Sanda by enabling real-time, adaptive tactical strategies in dynamic and adversarial environments. To test this, a DRL-based decision model is constructed and trained using multi-source competition data. The research focuses on evaluating the model's decision accuracy, cumulative reward, and convergence speed compared to traditional methods, while exploring its applicability across different sports. By emphasising this central hypothesis, the study provides a clear framework for analysing DRL's effectiveness and generalisation potential in sports decision-making.

This study specifically focuses on Wushu Sanda, a combat sport that poses additional challenges due to its fast-paced, adversarial nature and the lack of existing DRL research in this field. Unlike team sports such as basketball or football, Sanda involves one-on-one confrontations with highly dynamic tactical shifts, making real-time decision optimisation more complex. Applying DRL to Sanda is novel in that it explores how DRL can model individualised tactical responses and real-time adjustments in a combat environment, providing a foundation for extending DRL beyond traditional team sports.

2 Materials and methods

2.1 Data collection and sample selection

2.1.1 Data sources and collection methods

The data collection of this study depends on modern sensors, video analysis of competitions and athletes' physiological monitoring system (Shan, 2023). The data sources are as follows:

1. Game video and motion capture data: Through high-frequency camera equipment and motion capture system, the athlete's motion trajectory, technical movements and behaviour patterns in the game are collected. Video analysis and computer vision technology transform athletes' performance into data that can be used in deep learning model, forming a multi-dimensional data set of athletes' position, speed, acceleration and passing (Bae et al., 2024).
2. Physiological monitoring data: The physiological data of athletes' heart rate, blood oxygen, body temperature, etc. reflect the athletes' physical condition and their performance in coping with competition pressure. Real-time acquisition is carried out by wearing equipment such as heart rate monitor and physical fitness tracker (Uyar et al., 2022).
3. Game statistics: Get basic statistics such as scores, fouls, assists and rebounds from the post-game statistics of the game, and build a decision-making model based on historical performance (van Everdingen et al., 2019).
4. Opponent behaviour data: Based on the analysis of opponent's behaviour, including opponent's tactical layout, action mode and interaction mode of athletes, the data is

obtained by video playback analysis and research on the behaviour mode of the opposing team.

Data collection mainly includes real-time collection and historical data collection. Real-time data collection uses sensors and video technology during the competition to obtain real-time changes during the competition (Hamurcu and Eren, 2020). Historical data collection extracts the competition records of previous years from the competition database, which is a reference for the training of DRL model. As shown in Table 1.

Table 1 Data collection methods and sources

Data type	*Collection method*	*Source*
Athlete behaviour data	Cameras and motion capture devices	Game footage and motion capture system
Physiological data	Heart rate monitors and fitness trackers	Devices worn by athletes
Game Statistics	Post-game statistics system	Post-game data records and statistics platform
Opponent behaviour data	Video analysis and behaviour pattern recognition	Video playback analysis and opponent tactics research

2.1.2 Sample selection

In order to study the validity and representativeness of the results, the selection of samples should consider the individual differences of athletes, and also need a variety of competition environments, different opponent types and various tactical strategies. In this study, different sports, such as basketball, football and tennis, are selected for analysis. These sports have clear tactical characteristics and are highly dependent on decision-making. Sample selection criteria are as follows:

1 Athlete type and ability level: The sample includes athletes of different levels, such as top athletes and intermediate athletes. There are differences in the behaviour and decision-making modes of these athletes in the competition (Tang et al., 2020). According to the differences, the performance of the DRL model at different competitive levels is analysed.

2 Competition venue and opponent type: The types of competition venues, such as indoor and outdoor venues and different competition environments have great influence on tactical decision-making (Ding et al., 2023). The selected samples cover different venues and opponents with different tactical styles and preferences, and the model can cope with complex and changeable confrontation environment.

3 Data integrity: The historical game data of the selected samples have high integrity, and all kinds of data in the course of the game, such as basic statistical data such as scores, assists and fouls, must cover the whole process of each game (Anwar et al., 2023). Physiological data and behavioural data should also be comprehensive, and the model can obtain enough training data for deep learning. As shown in Table 2.

Table 2 Sample selection

Sport	*Athlete type*	*Competition environment*	*Sample size*
Basketball	International-level athletes	Indoor courts	50
Football	National-level athletes	Outdoor fields	60
Tennis	Provincial-level athletes	Indoor courts	40

2.1.3 Data preprocessing

Noise, missing values and other inconsistent problems often exists in the process of data acquisition, which will affect the training effect of the model. Data preprocessing includes data cleaning, missing value processing, abnormal value detection and standardisation processing (Chen et al., 2021). Some sensors in the original data may have missing values due to equipment failure or data transmission errors. These missing values can be processed by interpolation or other completion methods. The abnormal values are identified according to the normal physiological range or historical data of athletes, and they are eliminated from the training data set. The dimension of the original data is not uniform, which will affect the training of the model. All data need to be standardised before entering the model. Especially the physiological data of athletes, such as heart rate and blood oxygen, need to be normalised to ensure that they have the same influence as other data categories in model training. According to the behaviour data of athletes, meaningful features are extracted for training. The athletes' trajectory, acceleration, instantaneous speed and other characteristics are extracted from the competition video, and the athletes' efficiency value, confrontation intensity and other indicators are calculated according to the statistical data. The data preprocessing method is shown in the following formula (1):

$$X_{standardised} = \frac{X - \mu}{\sigma} \tag{1}$$

where $\underline{x}$ is the original data, the mean value of the data, the standard deviation of the data and the standardised data. The comparison after data cleaning is shown in Table 3. μ as the mean of the data, σ is the standard deviation of the data, $X_{standardised}$ for the normalised post-hoc data.

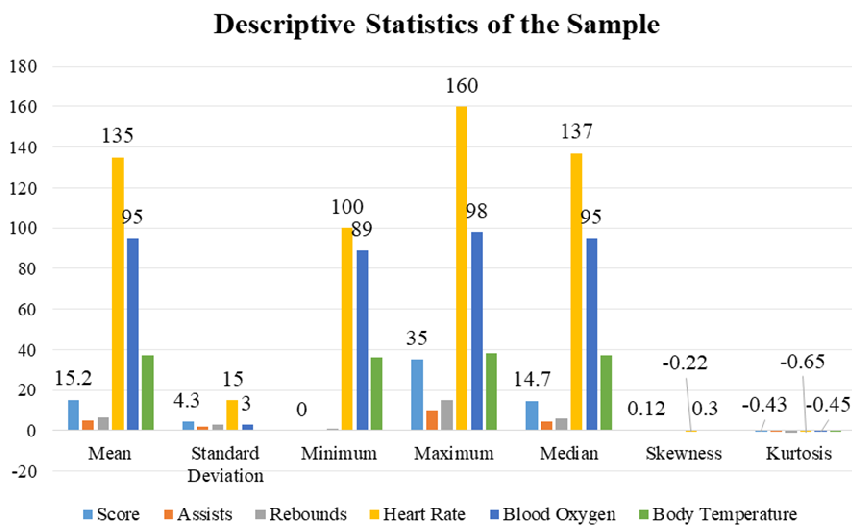
Table 3 Comparison of data cleaning before and after

Data type	*Missing data percentage*	*Missing data percentage after cleaning*	*Outlier treatment*
Heart rate data	10%	0%	No outliers
Behaviour data	5%	0%	Outliers removed
Game score data	0%	0%	No outliers

2.1.4 Descriptive statistics of samples

This study makes descriptive statistics on the analysis of athletes' performance data, physiological data and competition environment data, and reveals the distribution and fluctuation range of various data.

Figure 1 Descriptive statistics of the sample (see online version for colours)



As shown in Figure 1, the average score of athletes is 15.2, and the standard deviation is 4.3. The score fluctuates in the sample. The skewness of the score is close to 0, and the distribution of the score is symmetrical. The kurtosis of the score is relatively low, and the data is relatively flat, lacking extreme high scores. The average of assists is 4.8, and the standard deviation is 2.1. The performance of athletes' teamwork ability and creating scoring opportunities. Athletes' assists are concentrated, and a few athletes may have a higher number of assists. The average value of rebounding is 6.3, and the standard deviation is high. The rebounding data fluctuates greatly, which may be related to factors such as the role of different athletes and the position on the court. The distribution of rebounds is slightly to the right, and a few athletes have more rebounds. The average heart rate is 135, the standard deviation is 15, and the skewness is negative. The athletes' heart rate distribution is relatively concentrated, and a few athletes have higher heart rates, which may be due to excessive exercise intensity. The mean value of blood oxygen level is 95, the standard deviation is small, and the blood oxygen level fluctuates slightly in the sample. The skewness value is 0.30, and the data distribution tends to be normal. The average body temperature is 37.0, and the standard deviation is 0.6. The athlete's body temperature is kept in a relatively normal range, and the fluctuation of body temperature is small. Descriptive statistical data provide input values for subsequent DRL models, help researchers understand the basic characteristics of data, evaluate the effectiveness of model input, and further design appropriate algorithms for decision optimisation.

2.2 Model selection and construction

2.2.1 Overview of DRL

DRL is a powerful tool that combines deep learning and reinforcement learning. Reinforcement learning is a self-reinforcing learning method, and agents learn how to take actions by interacting with the environment to maximise cumulative rewards. Deep learning technology uses multi-layer neural network to model complex input data and extract effective features from a large number of unstructured data. The combination of the two gives DRL the ability to deal with decision-making problems in high-dimensional and complex environments, and can solve decision-making problems in nonlinear and dynamic environments.

In sports decision-making, DRL simulates the real-time decision-making process of athletes in the competition. By observing the environment (that is, the state in the competition), the agent takes a series of actions to get immediate rewards according to the consequences of each action (such as scores, mistakes, etc.), and finally optimises the overall decision-making strategy. DRL can process high-dimensional input data, such as athletes' historical performance, real-time physiological data, competition environment and other information, and improve the accuracy and real-time decision-making. For example, DRL in basketball matches helps the decision-making system to make the best offensive or defensive strategy in real time according to the players' physical condition, the progress of the game and the tactical choice of the opposing team. Agents constantly interact with the environment and learn how to make reasonable decisions in a complex and changeable environment. Compared with the traditional optimisation method, the advantage of DRL lies in self-adjustment and optimisation of decision-making strategies in the ever-changing environment to adapt to the ever-changing competition dynamics.

2.2.2 Model selection basis

When choosing a model suitable for sports decision-making, we need to consider the nature of the decision-making problem, the computational efficiency of the model and the interpretability. Sports decision-making problems are usually highly complex and dynamic, involving a lot of uncertainties. DRL can deal with this kind of complex dynamic decision-making problem, which has strong adaptability and can adjust its strategy itself. The selection basis should consider the following aspects:

1 Complexity and dimension of the problem: the input of multiple dimensions of sports competitive decision-making, and there is a complex nonlinear relationship between these factors. Traditional machine learning methods, such as support vector machines and decision trees, are difficult to deal with such a high-dimensional state space. DRL can deal with high-dimensional and continuous state space through deep neural network, and extract effective features from it, which is suitable for solving such complex problems.

2 Timing of decision-making: In sports competition, decision-making is dynamic and sequential, and the decision at a certain moment will affect the future results. DRL models the future results by strengthening the reward mechanism in the learning framework, and the decision at every moment is closely related to the long-term goal (such as winning the game). DRL has more advantages than the traditional single decision model (such as classification or regression model).

3 Antagonism and uncertainty: The competitive sports environment is antagonistic and uncertain in nature. Athletes not only need to face the real-time reaction of their opponents, but also deal with unexpected situations (such as injuries and sudden tactical adjustments). DRL can interact with the environment many times, self-adjust

strategies to deal with uncertain factors, and provide more robust decision-making schemes.

4 Model training and optimisation: The training of DRL model requires a lot of time and computing resources, but it can realise self-optimisation through multiple interactions with the environment. The performance data and competition environment data of athletes in sports competition are trained by historical data, and the model can adapt to changing competition situations.

2.2.3 Model construction method

The construction of DRL model includes the following steps: defining state space, action space, reward mechanism, strategy network and optimisation method. The model construction in sports competitive decision-making focuses on how to design the state space and action space adapted to the sports environment, and optimise the decision-making strategy by strengthening the training process of learning.

1 Definition of state space: The state space in sports competition includes the athlete’s physical state (such as heart rate, blood oxygen, body temperature, etc.), the current competition process (such as score and remaining time), the opponent’s tactical arrangement, etc. In order to accurately reflect the real-time situation of the game, the state space covers all factors that may affect the decision. Data usually come from real-time sensor monitoring, game video analysis and historical game data.

2 Action space design: The action space in the decision-making problem of sports competition corresponds to the decisions that athletes can take in the competition. The action space in basketball game includes choosing offensive strategy (such as shooting, passing, breakthrough, etc.), while the action space in football game includes choosing offensive, passing or defending. The choice of each action will have an impact on the result of the competition, and the design of action space needs to be consistent with the actual decision-making needs of sports events.

3 Reward mechanism: The reward mechanism feeds back the advantages and disadvantages of each decision. In sports competition, rewards are usually related to factors such as scores, fouls and mistakes. In basketball games, shooting scores, successful assists and other behaviours can get positive rewards, while mistakes and fouls will get negative rewards. The reward mechanism can reflect the performance of athletes in the competition and guide agents to learn the best decision-making strategy.

4 Strategy network and value function: In DRL, the strategy network is modelled by deep neural network, and the current state is input and the probability distribution of an action is output. The agent chooses actions according to the strategy network and gives feedback according to the reward mechanism. The value function is used to evaluate the value of a state and help the agent to optimise the strategy. In the face of uncertainty, the value function provides a long-term evaluation of the current state.

Through reinforcement learning algorithm formula (2) DRL can self-adjust the strategy, optimise the decision-making process and adapt to different competition environments.

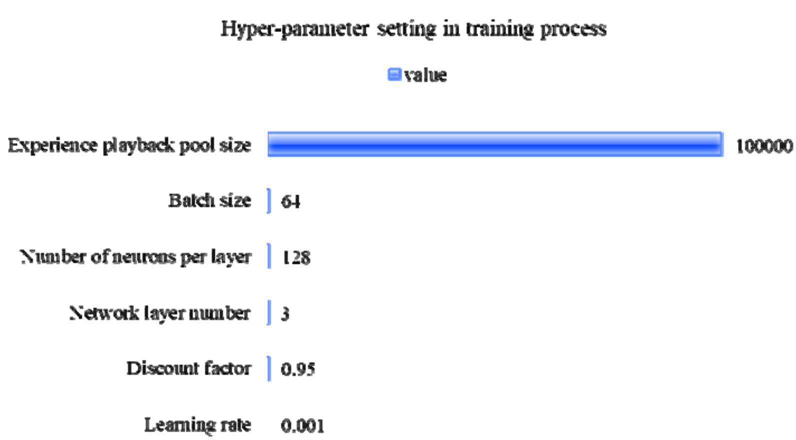
$$Q(s_t, a_t) = Q(s_t, a_t) + \alpha\left(r_t + \gamma \max_{a'} Q(s_{t+1}, a') - Q(s_t, a_t)\right) \tag{2}$$

$Q(s_t, a_t)$ indicates the quality (i.e. value) of the selected action a_t in the state s_t; α is the learning rate; r_t is a reward; γ is a discount factor; $\max_{a'} Q(s_{t+1}, a')$ is the maximum value of the next state.

2.2.4 Model training and optimisation

The training of DRL includes pre-training and online training. In the pre-training stage, historical data is used for preliminary model training, and the interaction between online training and environment constantly optimises the model. The core task of model training in sports competitive decision-making is that agents can learn and adjust strategies efficiently in a limited time and make optimal decisions in actual competitions. The preparation of training data is the premise of model training. The training data comes from historical competition data, athletes' physiological monitoring data and competition videos. In the process of data preparation, in addition to ensuring the integrity of data, it is also necessary to process and clean the data reasonably, remove abnormal data and fill in the missing data. The training of DRL model needs to adjust several super parameters, such as learning rate, discount factor, the number of layers of neural network and the number of nodes in each layer. The choice of super parameter directly affects the efficiency of training and the performance of the model. In the process of training, the best combination of super parameters is selected through experiments to achieve the best learning effect. In the training process, the optimisation method is mainly gradient descent method, and the agent gradually approaches the optimal strategy by calculating the loss function and updating the parameters of the strategy network in each training step. In order to avoid over-fitting and improve the generalisation ability of the model, techniques such as experience playback and target network are adopted to enhance the stability and reliability of training. The super parameter setting in the training process is shown in Figure 2.

Figure 2 Hyper-parameter setting in training process (see online version for colours)



2.3 Model evaluation and verification

2.3.1 Model evaluation indicators

In the training process of DRL model, the evaluation index should aim at the uniqueness of sports decision-making problems and reflect all kinds of decision-making situations that agents may encounter in actual competitions. In this study, the following evaluation indicators are used to consider the performance of the model:

- Cumulative reward: It is the most commonly used evaluation index in reinforcement learning, reflecting the overall reward obtained by agents in the training process. Higher cumulative reward means that agents can maximise their goals (such as winning a game) in the decision-making process. The decision-making effect of the model is displayed intuitively, which is suitable for dynamic and complex decision-making tasks.
- Accuracy of decision-making: A measure of the frequency with which the model chooses the correct action in the actual competition. Accurate decision-making in sports competition is directly related to the outcome of the game.
- Training convergence speed: Refers to the time and iteration times required for the model to reach the optimal performance from the beginning of training. The model with faster convergence rate is more practical and can produce efficient decisions in a short time.
- Model generalisation ability: Refers to the performance of the model on data outside the training set. Due to the differences between different competitions, athletes and opponents, the model must have strong generalisation ability.

The evaluation results of the model in different training stages are shown in Table 4.

Table 4 Evaluation index table

Evaluation index	*Scale*	*Model representation*
Cumulative reward	0–1,000	850
Decision accuracy	0–1	0.87
Training convergence speed	Number of rounds	3,000 rounds
Model generalisation ability	0–1	0.81

2.3.2 Model verification method

This study adopts cross-validation, separation of training and testing, online validation, etc. Cross-validation divides the data set into several subsets, and trains and tests on different subsets respectively. Through repeated cross-validation, the problem of inaccurate model performance evaluation caused by training data distribution deviation is avoided. It can comprehensively evaluate the performance of the model in various environments and reduce the risk of over-fitting. The separation of training and testing is to divide the data set into training set and test set. The training set is used to train the model and the test set is used to verify the generalisation ability of the model. The data of the test set has not been touched in the training process, which helps to ensure that the evaluation of the model is more objective and reliable, and is suitable for model verification in dynamic environment. Online verification method is used to simulate the decision-making process of the model in practical application. The model interacts with the real competition environment, and the agent constantly adjusts the decision-making strategy according to real-time feedback. Through online verification, the adaptability of the model in the face of changing actual environment is evaluated.

To ensure the robustness of the model in the verification process, formula (3) is used to describe the performance evaluation formula in the verification process.

$$V(s) = \mathbb{E}\left[r_t + \gamma \max_{a'} Q\left(s_{t+1}, a'\right) \middle| s_t = s\right] \tag{3}$$

$V(s)$ is the value of state s ,which indicates the expected reward that the agent can get in the current state; r_t is an instant reward for the current moment; γ is a discount factor, indicating the influence of future awards on current decisions; $Q(s_{t+1}, a')$ is the value of action a' in the next state s_{t+1}. This formula calculates the expected return of an agent after taking an action in different states, and evaluates the decision-making process of the model.

3 Results and analysis

3.1 Analysis of results

3.1.1 Performance of reinforcement learning model

1 Compare the results of different models

   After repeated training, the reinforcement learning model is compared with several common decision-making algorithms (such as Q-learning, DQN, SARSA, etc.). The experiment set the same environment and task conditions, and revealed the differences between the DRL model and the traditional reinforcement learning method. Comparing the cumulative reward, decision accuracy and training convergence speed of each model in the same time step, as shown in Figure 3, the DRL model shows higher decision quality and shorter convergence time. Deep Q-network (DQN) has obvious advantages over traditional Q-learning and SARSA when the complexity of tasks increases. DQN is also more outstanding in cumulative awards, and it can make more accurate decisions in a dynamically changing competition environment.

2 Comparison of model training effects

   In order to test the training effect of the model, the training process of different models is tracked and analysed in detail. The change trend of cumulative rewards of different models in the training process is shown in Figure 4. Observe the convergence of the models and the fluctuation in the training process. Compared with the traditional model, the DQN model's reward growth in the initial stage is relatively stable, and its convergence speed is faster. The reward of Q-learning and SARSA models fluctuates greatly, the training process is unstable, and the reward of final convergence is low. Figure 4 clearly shows the advantages of DRL in stability and efficiency. After long-term training, the performance of DQN model is more stable, and the final reward far exceeds the traditional method.

Figure 3 Comparison of results of different models (see online version for colours)

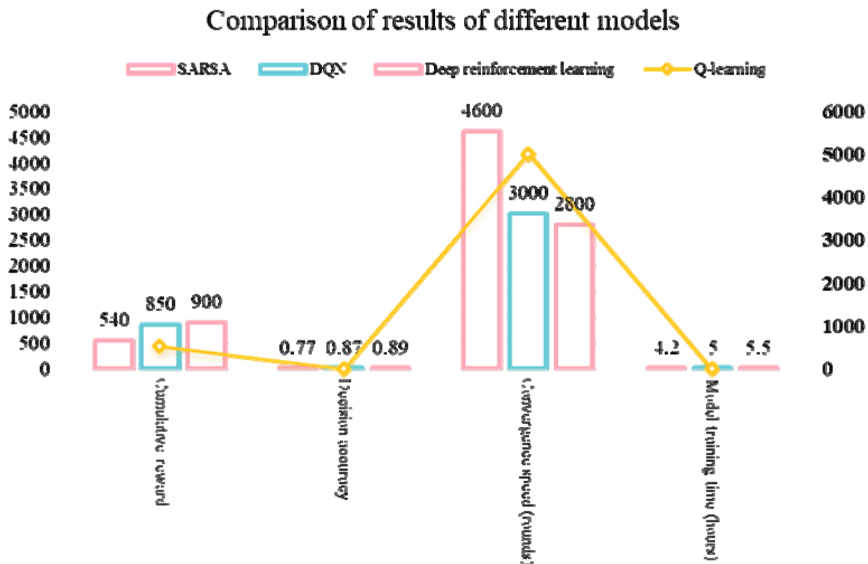
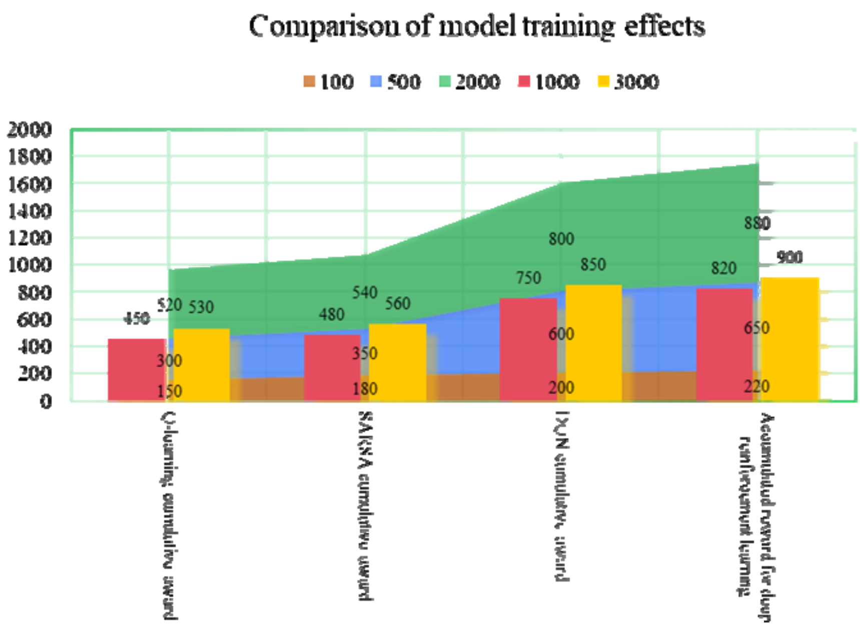


Figure 4 Comparison of model training effects (see online version for colours)

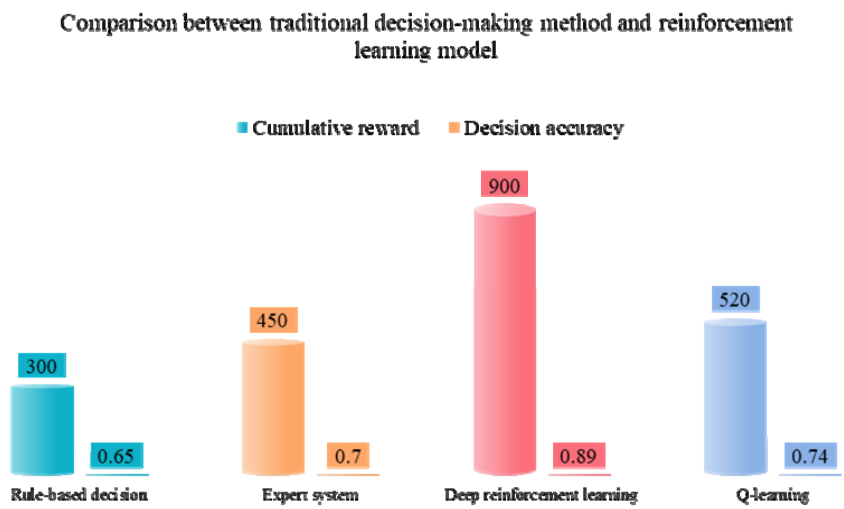


As shown in Figure 4, the cumulative reward trend of different models during the training process. After 1000 rounds, the DRL model shows a relatively stable and continuous increase in reward level, while the traditional model's reward growth is relatively slow. The training effect of the model reflects the convergence speed and the ability of the agent to adapt to the complexity of the task.

3 Comparison of traditional decision-making methods

Measure the advantages of reinforcement learning model, and compare the DRL model with traditional decision-making methods (such as rule-based decision-making, expert system, etc.). Traditional decision-making methods rely on preset rules and manual experience, but lack adaptive ability. Rule-based decision-making in sports competition can not effectively cope with the complex and dynamic competition environment. Figure 5 shows the performance comparison between the traditional decision-making method and the reinforcement learning model in various competition situations. The performance of the traditional decision-making method in the face of complex situations is obviously not as good as that of the DRL model. The traditional method cannot effectively deal with the dynamic competition situations in terms of decision-making accuracy and cumulative rewards, and it is far inferior to the reinforcement learning model in decision-making quality.

Figure 5 Comparison between traditional decision-making method and reinforcement learning model (see online version for colours)



3.1.2 Practical significance and application scenarios of the results

The performance of reinforcement learning model in sports decision-making provides a new idea for solving practical problems. In the process of training, the decision-making situation is simulated many times, and the model can adapt to the rapid change and complex environment. In the scene of rapid decision-making and real-time feedback, traditional rule-based decision-making methods and expert systems are difficult to cope with unexpected situations. The training mode of DRL model allows to maintain high decision-making accuracy in the face of unknown environment and gradually optimise decision-making strategies.

Specific to the application scenario, it is of practical significance to strengthen the application of learning in sports competitions. For highly dynamic competition environment, agents can make predictions and adjustments according to real-time data, and provide strategic support for coaches and athletes. For example, in team sports such as football and basketball, the DRL model can evaluate the effectiveness of the current strategy in real time according to every action and event in the competition process, and give improvement suggestions to help formulate more effective competition strategies. The efficient learning and self-optimisation ability of the model can also support athletes' personalised training plan, constantly adjust the training content during long-term training, and improve the overall performance of athletes.

In the future, the application of DRL model is not limited to a single sport, but can also be extended to many fields. For example, in e-sports, agents can provide players with real-time tactical decision support, which surpasses manual decision-making in decision-making accuracy and response speed and helps players get more chances to win. Based on the dynamic analysis of each stage in the competition, the agent can find the opponent's weaknesses in time and make targeted strategies to improve the overall competitive level.

3.2 Discussion

3.2.1 Problems and challenges encountered in the research

One of the key challenges in this study is data acquisition and processing. While simulation environments can provide large amounts of training data, the data in real sports competitions, especially in disciplines like Wushu Sanda, is complex and incomplete, which places higher demands on model training and validation. The complexity and variability of the competition environment in different sports further increase the difficulty of adapting models. In certain decision-making scenarios, data may not accurately reflect the actual situation, leading to discrepancies between the training results and real-world applications. The training process of the model is time-consuming and requires significant computational resources. In DRL, the time and computational costs increase exponentially with the number of model iterations. DRL models depend heavily on numerous experiments and feedback to continuously adjust parameters, creating substantial uncertainty in the training process. Additionally, in specific sports like Wushu Sanda, external factors such as athletes' physical condition and weather conditions may affect the effectiveness of the model, posing challenges to the stability and accuracy of the predictions. Another issue is the generalisation ability of the model. While DRL performs excellently in specific environments, its ability to maintain efficient decision-making in complex, unseen situations still needs to be validated. The sudden changes and complex strategic interactions in real competitions will impact the accuracy of model predictions.

3.2.2 Suggestions and improvement directions for future research

The diversity and quality of data are core factors influencing model performance. To improve the model's generalisation capability, future research should focus on integrating multi-source data. Combining different types of data, such as athletes' physical performance and changing environmental conditions, can enhance the model's adaptability to complex scenarios. By incorporating sensor technology and real-time monitoring, more accurate dynamic changes during the competition can be captured, providing richer and more precise data support for DRL models. Although computational resources are becoming increasingly powerful, the resources and time required for large-scale training remain limiting factors. Future research should explore more efficient algorithms, such as meta-learning-based reinforcement learning, which allows models to learn better decision-making strategies with fewer experiences, thus shortening training time and reducing computational resource consumption. Distributed training and parallel computing techniques can also be adopted to accelerate the DRL training process. Another important future direction is improving the interpretability of DRL models. DRL models are often considered 'black boxes', making it difficult to explain the decision-making process. Understanding why a model made a certain decision is crucial for coaches and athletes in real applications. Research on enhancing model interpretability would make the decision-making process more transparent and provide more actionable guidance for practical use. Lastly, for different types of sports competitions, future research should explore customised reinforcement learning models. In team sports and individual sports, the complexity of decision-making and strategy varies, so customised models will better improve decision accuracy and efficiency. Conducting in-depth studies on specific scenarios will aid in the broader application of DRL in various sports disciplines.

4 Conclusions

This study explores the application of DRL in sports competition decision-making, specifically focusing on Wushu Sanda, with the aim of enhancing decision-making efficiency and precision, optimising athletes' performance, and improving tactical deployment. The results demonstrate that the DRL model exhibits strong adaptability and decision-making capabilities in simulated sports environments. When comparing different models, DRL effectively predicts dynamic changes during competitions and continuously optimises decision-making strategies in complex environments. The model improves decision accuracy by adjusting parameters and learning from experience during the training process, surpassing the limitations of traditional methods in real-time feedback and adaptability.

These challenges are not isolated technical issues but reflect broader themes in sports science, including variability in human performance, contextual decision-making, and the integration of technology into training and competition. Addressing these problems requires interdisciplinary collaboration between AI researchers, sports scientists, and practitioners to ensure that DRL models align with the physiological, tactical, and organisational realities of competitive sports.

The study also reveals several challenges in the practical application of DRL, particularly in data processing, model training, and generalisation capabilities. While the model performs excellently in controlled training environments, external factors and uncertainties in real sports competitions continue to pose challenges to its performance. To meet the differentiated needs of various sports, optimising the model for specific applications and improving its effectiveness in real-world scenarios remains a key direction for future research. DRL presents a new technological framework for sports competition decision-making, with enormous potential for applications in match strategies, athlete training, and team coordination. As data collection technologies and computational power continue to advance, DRL has the potential to be further refined and applied to more sports disciplines, significantly influencing the development of competitive sports. Future research should focus on improving the model's interpretability, generalisation ability, and training efficiency, driving the widespread application of DRL in practical sports competition scenarios. From a practical perspective, integrating DRL into sports decision-making can offer coaches and organisations data-driven tools for tactical analysis, training optimisation, and real-time match support. For example, DRL-based systems can provide scenario simulations for pre-match planning, recommend tactical adjustments during games, and analyse post-match data for continuous improvement. By embedding DRL tools into daily coaching workflows, sports organisations can enhance decision quality, accelerate feedback loops, and strengthen competitive advantages.

Future research should focus on enhancing the interpretability of DRL models to increase trust and adoption in real-world sports contexts. Methods such as attention visualisation, policy saliency mapping, and post-hoc explanation techniques can make DRL decision processes more transparent to coaches and analysts. Additionally, developing sport-specific DRL architectures that account for the unique tactical and temporal structures of different sports (e.g., team-based vs. individual combat) will be crucial. Customisation will allow models to capture domain-specific nuances while maintaining generalisable decision-making capabilities. Although this study focuses on Wushu Sanda, its findings have broader implications for other sports. The underlying DRL framework is not limited to combat scenarios; it can be adapted to team sports such as basketball and football, where dynamic tactical decision-making also plays a critical role. By adjusting the state and action spaces to reflect different tactical contexts, the proposed model can support diverse competitive settings. This cross-sport adaptability demonstrates the generalisation potential of DRL in sports science and underscores its role as a versatile decision-support tool.

Declarations

The data used to support the findings of this study are all in the manuscript.

The authors declare no competing interests.

References

Ahamed, T., Zou, B., Farazi, N.P. and Tulabandhula, T. (2021) 'Deep reinforcement learning for crowdsourced urban delivery', *Transp. Res.*, October, Part B, Vol. 152, pp.227–257, DOI: 10.1016/j.trb.2021.08.015.

Anwar, G.A., Hussain, M., Akber, M.Z., Khan, M.A. and Khan, A.A. (2023) 'Sustainability-oriented optimization and decision making of community buildings under seismic hazard', *Sustainability*, March, Vol. 15, No. 5, p.4385, DOI: 10.3390/su15054385.

Bae, J., Sim, Y., Seo, E., Seo, G. and Nam, Y. (2024) 'Multidimensional assessment of preparation for sport competition (MAPSC): psychometric properties and initial validation among Korean student-athletes in combat sport', *Meas. Phys. Educ. Exerc. Sci.*, 1 October, Vol. 28, No. 4, pp.360–376, DOI: 10.1080/1091367X.2024.2365213.

Chen, D.Q., Preston, D.S. and Swink, M. (2021) 'How big data analytics affects supply chain decision-making: an empirical analysis', *J. Assoc. Inf. Syst.*, Vol. 22, No. 5, pp.1224–1244, DOI: 10.17705/1jais.00713.

Ding, C.D.Y., Fung, H.G. and Zhang, Y. (2023) 'Yin-yang framework of decision-making: development and assessment of Yin-yang decision-making styles questionnaire', *Pers. Individ. Differ.*, April, Vol. 205, p.112090, DOI: 10.1016/j.paid.2023.112090.

Furley, P. (2019) 'What modern sports competitions can tell us about human nature', *Perspect. Psychol. Sci.*, March, Vol. 14, No. 2, pp.138–155, DOI: 10.1177/1745691618794912.

Guan, M.H., Mondal, S.K., Dai, H.N. and Bao, H.Y. (2023) 'Reinforcement learning-driven deep question generation with rich semantics', *Inf. Process Manag.*, March, Vol. 60, No. 2, p.103232, DOI: 10.1016/j.ipm.2022.103 232.

Hamurcu, M. and Eren, T. (2020) 'Strategic planning based on sustainability for urban transportation: an application to decision-making', *Sustainability*, May, Vol. 12, No. 9, p.3589, DOI: 10.3390/su12093589.

Jaisson, T. (2022) 'Deep differentiable reinforcement learning and optimal trading', *Quant. Finance*, 3 August, Vol. 22, No. 8, pp.1429–1443, DOI: 10.1080/14697688.2022.2062431.

Liang, Y., Sun, Z., Song, T.H., Chou, Q., Fan, W., Fan, J.P., Rui, Y., Zhou, Q.P., Bai, J., Yang, C. and Bai, P. (2022) 'Lenovo schedules laptop manufacturing using deep reinforcement learning', *INFORMS J. Appl. Anal.*, January–February, Vol. 52, No. 1, pp.56–68, DOI: 10.1287/inte.2021.1109.

Raziei, Z. and Moghaddam, M. (2021) 'Adaptable automation with modular deep reinforcement learning and policy transfer', *Eng. Appl. Artif. Intell.*, August, Vol. 103, p.104296, DOI: 10.1016/j.engappai.2021.104296.

Rossi, L., Breuer, C. and Feiler, S. (2020) 'Determinants of non-profit sports clubs' perceived competition from commercial sports providers', *Sport Manag. Rev.*, August, Vol. 23, No. 4, pp.736–749, DOI: 10.1016/j.smr.2019.10.001.

Schütz, L.M. and Schultheiss, O.C. (2020) 'Implicit motives, laterality, sports participation and competition in gymnasts', *Front. Psychol.*, 26 May, Vol. 11, p.900, DOI: 10.3389 /fpsyg.2020.00900.

Shan, F.J. (2023) 'Limits of sports ethics and sports competition: the matter of sports humanity', *Int. J. Sport Psychol.*, March–April, Vol. 54, No. 2, pp.91–108, DOI: 10.7352/IJSP .2023.54.091.

Tang, J., Meng, F.Y., Cabrerizo, F.J. and Herrera-Viedma, E. (2020) 'Group decision making with interval-valued intuitionistic multiplicative linguistic preference relations', *Group Decis. Negoti.*, February, Vol. 29, No. 1, pp.169–206, DOI: 10.1007/s10726-019-09646-1.

Uyar, Y., Gentile, A., Uyar, H., Erdeveciler, O., Sunay, H., Mindrescu, V., Mujkic, D. and Bianco, A. (2022) 'Competition, gender equality, and doping in sports in the red queen effect perspective', *Sustainability*, March, Vol. 14, No. 5, p.2490, DOI: 10.3390/su14052490.

van Everdingen, Y., Hariharan, V.G. and Stremersch, S. (2019) 'Gear manufacturers as contestants in sports competitions: breeding and branding returns', *J. Mark.*, May, Vol. 83, No. 3, pp.126–144, DOI: 10.1177/0022242919831996.

Wallrafen, T., Nalbantis, G. and Pawlowski, T. (2022) 'Competition and fan substitution between professional sports leagues', *Rev. Ind. Organ.*, August, Vol. 61, No. 1, pp.21–43, DOI: 10.1007/s11151-022-09860-3.

Wang, W., Li, B.B., Luo, X.M. and Wang, X.Y. (2023) 'Deep reinforcement learning for sequential targeting', *Manag. Sci.*, September, Vol. 69, No. 9, pp.5439–5460, DOI: 10.1287/mnsc.2022.4621.

Zeng, D., Zheng, L., Yang, X.T. and Li, Y.N. (2024) 'Uncertainty-aware human-like driving policy learning with deep Bayesian inverse reinforcement learning', *Transp. Res.*, February, Part A, p.15, DOI: 10.1080/23249935.2024.2318621.