SIN-CHN SCIENTIFIC PRESS PTE.LTD.

Molecular & Cellular Biomechanics 2025, 22(2), 653.
https://doi.org/10.62617/mcb653

Article

Sports training injury risk assessment model based on biological mechanisms and complex network analysis

Changyuan Yin, Ting Luo, Zhenping Ye*

Hunan Automotive Engineering Vocational College, Zhuzhou 412002, China
*** Corresponding author:** Zhenping Ye, 13973363090@163.com

CITATION

Yin C, Luo T, Ye Z. Sports training injury risk assessment model based on biological mechanisms and complex network analysis. Molecular & Cellular Biomechanics. 2025; 22(2): 653. https://doi.org/10.62617/mcb653

ARTICLE INFO

Received: 28 October 2024
Accepted: 15 November 2024
Available online: 23 January 2025

Abstract: To improve the accuracy and practicality of sports training injury risk assessment (IRA), this paper constructs a model based on a complex network analysis algorithm and conducts performance comparison experiments across multiple dimensions. The research results demonstrate that the optimized model performs well in terms of risk assessment accuracy, real-time processing, robustness, adaptability, and user satisfaction. Specifically, the Area Under Curve of the Receiver Operating Characteristic Curve (AUC-ROC) of the optimized model reaches 0.928, indicating high accuracy in risk assessment. In addition to these metrics, this study includes a discussion on the biological mechanisms underlying sports injuries, emphasizing how biological signals can be integrated with the complex network analysis to enhance the model's predictive capabilities. This integration allows for a more comprehensive understanding of injury risk factors, such as muscle fatigue, joint stress, and tissue response, which are critical for effective injury prevention strategies. In the real-time experiment, the processing speed score is 4.9. In the robustness experiment, the fault recovery ability score is 4.3. In the adaptive experiment, the diversified data processing ability score is 4.5. In the user satisfaction experiment, the accuracy score of risk assessment is 4.9, and the convenience score is 5.0. These results indicate that the optimized model has significant advantages in handling complex data and adapting to changing environments. Therefore, this paper provides valuable insights for improving injury risk management and decision support in sports training by incorporating biological insights into the assessment model.

Keywords: complex network; sports training injury; machine learning; risk management; model analysis; biological mechanisms; injury prevention

1. Introduction

In modern competitive sports, the intensity and complexity of athletes' training are increasing, and the risk of sports injury is also rising. Injury will not only affect the short-term performance of athletes, but also have a long-term impact on their career [1–3]. Therefore, how to effectively assess and reduce the risk of sports injuries has become the focus of sports and academic circles. As a new data analysis tool, complex network analysis has achieved remarkable results in bioinformatics, social network analysis and other fields [4]. It can help to understand the dynamic interaction and potential risks in complex systems by revealing the internal structure and functional relationship of the system [5]. Therefore, it is of great theoretical significance and practical application value to apply complex network analysis method to sports training injury risk assessment (IRA).

Traditional IRA methods often rely on empirical judgment or simple statistical model, lacking in-depth analysis of multi-dimensional data of athletes [6–8]. By introducing complex network analysis, potential risk factors can be identified more accurately and the accuracy of evaluation can be improved. By analyzing the key nodes and edges in the network structure, people can find the key factors that affect the risk of injury, thus helping coaches and athletes to make more scientific training plans, optimize training strategies and reduce the incidence of injury. By introducing complex network analysis into the field of sports training IRA, this paper aims to promote the cross-integration of sports science and data science and provide new ideas and methods for related research. With the development of competitive sports, sports injuries have increasingly serious impact on the economy and health of sports teams and athletes. Therefore, it is of great social and economic significance to develop an effective risk assessment model to help athletes reduce the possibility of injury in training.

The innovations of this paper are as follows:

Firstly, this paper applies the complex network analysis method to the risk assessment of sports training injury, and constructs a risk assessment model based on network structure to capture the dynamic relationship and potential risk factors of athletes in the training process more comprehensively.

Secondly, this paper integrates athletes' physiological indicators, training data, injury history and other multi-source data, and reveals the complex interaction between these data through complex network analysis, which improves the accuracy and meticulousness of IRA.

These innovations make a significant breakthrough in this paper based on existing literature, and provide a more scientific and efficient method for the risk assessment of sports training injuries. To sum up, this paper aims to explore the application of complex network analysis algorithm in sports training IRA, and provide a new solution to improve the scientific and practical sports IRA.

2. Related works

In the study of IRA in sports training, Schweizer et al. found that in high-intensity sports training, the increase of injury risk was closely related to athletes' physical quality, training load and psychological pressure. Through the comprehensive evaluation of athletes, the incidence of injuries could be effectively reduced [9]. Cui et al. proposed that the use of biomechanical and physiological indicators, combined with data collection of smart wearable devices, could achieve real-time injury risk monitoring of athletes and improve the accuracy and timeliness of early warning [10]. In the application research of complex network analysis algorithm in sports field, Wilke and Groneberg pointed out that complex network analysis algorithm had obvious advantages in identifying athletes' training mode and optimizing competition strategy. Through the network modeling of sports data, the cooperative relationship between athletes and the overall tactical layout could be revealed [11]. The research of Ji et al. showed that complex network analysis could help identify the key nodes and important connections in sports teams, optimize the team structure and improve the overall performance of the team [12]. In the construction of sports training IRA model, Lutter et al. developed a set of IRA model based on complex network analysis, which could predict the possibility of injury according to the training history and physical state of athletes. The model had been proved to be efficient and accurate in many sports [13]. Ageberg et al. found that combining machine learning with complex network analysis could further improve the prediction ability of IRA model. Especially in big data environment, the adaptive ability of the model was significantly enhanced [14].

Although there have been many studies on sports training IRA, there are still some shortcomings. Many studies mainly rely on traditional statistical analysis methods, and the evaluation of athletes' injury risk often ignores the complexity and diversity of data and fails to make full use of multi-source data for comprehensive analysis. In this paper, the complex network analysis algorithm is adopted, which can build an interactive network among athletes based on multidimensional data, reveal the potential injury risk factors and correlation, and provide a more comprehensive risk assessment.

3. Risk assessment of sports training injury based on complex network analysis algorithm

3.1. IRA in sports training

Sports training IRA is a systematic activity aimed at identifying, analyzing and predicting the injury risks that athletes may face during training. Its goal is to help coaches and athletes formulate effective prevention and management measures, thereby reducing the probability of injury and ensuring the health and training effect of athletes [15–17]. With the development of modern sports science, IRA has gradually become one of the important directions of sports science research.

In risk assessment, many factors affecting the occurrence of injury need to be considered, which can be divided into internal factors and external factors, as shown in **Table 1**.

Table 1. Injury risk factors.

Dimension	Factor	Description
Intrinsic factor	Physical quality	Such as flexibility, muscle strength, endurance and coordination. Lack of physical fitness may lead to deformation of technical movements and increase the risk of injury.
	Physiological characteristic	Include age, gender and individual differences. Different physiological characteristics will affect the adaptability of athletes to training load.
	Psychology	Psychological stress, motivation level and mental state are also important factors affecting the risk of injury.
Extrinsic factor	Training environment	Changes in external conditions such as venues, climate and facilities may pose potential risks to athletes.
	Training load	Unreasonable training plan, too high training intensity and frequency will lead to fatigue accumulation and overuse injury [18].
	Equipment	Improper equipment may increase the probability of injury, such as the cushioning performance of shoes and the safety of sports equipment.

In recent years, with the progress of science and technology and the development of data analysis technology, the methods of sports training IRA have been constantly innovated and improved. Early IRA mostly relies on statistical methods, and common risk factors are identified by analyzing historical data. However, these methods usually assume that the relationship between data is linear, and it is difficult to deal with complex interaction effects. Besides, by measuring and analyzing athletes' postures and action patterns, people can identify abnormal sports behaviors that may lead to injuries. This method needs the help of high-precision motion capture technology and biomechanical model. With the popularization of smart wearable devices and sensor technology, it is possible to monitor athletes' physiological and sports data in real time. These devices can provide real-time feedback and combine with big data analysis technology for personalized risk assessment. In recent years, models based on machine learning have been widely used in IRA. These models can deal with large-scale multidimensional data, and constantly optimize the accuracy and robustness of risk prediction through self-learning algorithms. Effective IRA can help coaches and athletes identify potential injury risks in advance and adjust training plans, thus reducing the incidence of injuries and improving athletes' competitive performance [19–21]. By identifying high-risk factors in time and taking preventive measures, the incidence of injury can be significantly reduced. According to the risk assessment results, the training load and training content are adjusted to ensure that athletes can achieve the best training effect under safe conditions. Through scientific risk management, athletes can be helped to prolong their career and improve their quality of life.

Sports training IRA is an important tool to protect athletes' health and improve sports performance, and its research and application are of great significance to the development of modern sports science [22].

3.2. Application of complex network analysis algorithm in sports field

A complex network consists of many interconnected nodes, which are connected by edges. These nodes and edges can represent different entities and relationships. In the field of sports, nodes can represent athletes, movements, positions or events, while edges represent the interaction or association between these entities [23–25]. Its basic concept is shown in **Table 2**:

Table 2. Basic concepts of complex network analysis algorithm.

Concept	Analysis
Node degree	Represents the number of connections of a node. In sports, node degree can reflect the activity or influence of an athlete in the team.
Clustering coefficient	Measure the degree of interconnection between neighbors of a node. The high clustering coefficient indicates the close cooperation between athletes.
Average path length	Represents the average shortest path length between two nodes in the network, which can be used to evaluate the efficiency of information or goods transmission in the network.
Modularity of network	Used to identify the community or sub-group structure in the network, which is especially important for analyzing team tactics and group behavior.

Complex network analysis has an important application in the interaction analysis between athletes, which can reveal the cooperation mode and key figures between athletes. By constructing and analyzing the passing network or interactive network between athletes, the cooperation mode and strategy in the team can be identified [26]. For example, in a football match, people can determine which players interact most frequently by analyzing the passing network to identify the core players of the team and the main passing paths. Key person identification is a complex network analysis that can help identify the players in the team who have the greatest influence on the game results. These players are usually nodes with high node degree or high intermediary center in the network, which means that they play an important connecting role in the team's passing network [27–29].

The application of complex network analysis in the field of sports has greatly enriched the understanding and analysis methods of team and individual performance [30]. By digging deep into the potential relationships in the data, complex network analysis not only provides strong support for athletes' training and competition, but also provides new tools and methods for coaches and sports science researchers.

3.3. Construction of risk assessment model for sports training injury

It is a complex and systematic process to construct an effective risk assessment model for sports training injuries, and its purpose is to accurately predict the risk of sports injuries by comprehensively analyzing various physiological and environmental factors of athletes [31]. The construction of sports training IRA model needs a comprehensive framework to cover all links from data collection to risk assessment. In complex network analysis, the model uses the importance of nodes and community structure to evaluate risks. The equation is as follows:

$$PR(v) = \frac{1-d}{N} + d * \sum_{u \in \mathrm{In}(v)} \frac{PR(u)}{\mathrm{OutDegree}(u)} \tag{1}$$

In the equation, v and u are nodes. $PR(v)$ is the PageRank value of nodes, d is the damping coefficient, N is the total number of nodes in the network, $\mathrm{In}(v)$ is the set of nodes, and $\mathrm{OutDegree}(u)$ is the number of nodes. After community testing, the equation is as follows:

$$Q = \frac{1}{2m} \sum_{i,j} \left[A_{ij} - \frac{k_i k_j}{2m} \right] \delta(c_i, c_j) \tag{2}$$

Q in the equation is modularity, which is used to measure the quality of community division. m is the total number of edges. A_{ij} is the edge weight between nodes. i, j are nodes. $k_i k_j$ is the degree of nodes. $\delta(c_i, c_j)$ is that nodes belong to the same community, and the value is 1, otherwise it is 0. In the model, the gradient lifting decision tree is used to predict the risk. The equation is as follows:

$$F(x) = \sum_{m=1}^{M} \gamma_m h_m(x) \tag{3}$$

$F(x)$ is the predicted value. M is the total number of trees. γ_m is the weight of trees. $h_m(x)$ is the prediction of samples by trees, and then the gradient is updated. The equation is as follows:

$$\gamma_m = \mathrm{arg}\min_{\gamma} \sum_{i=1}^{n} L(y_i, F_{m-1}(x_i) + \gamma h_m(x_i)) \tag{4}$$

n is the total number of samples. $L(y_i, F_{m-1}(x_i) + \gamma h_m(x_i))$ is the loss function. $F_{m-1}(x_i)$ is the predicted value after iteration. The model is shown in **Figure 1**:

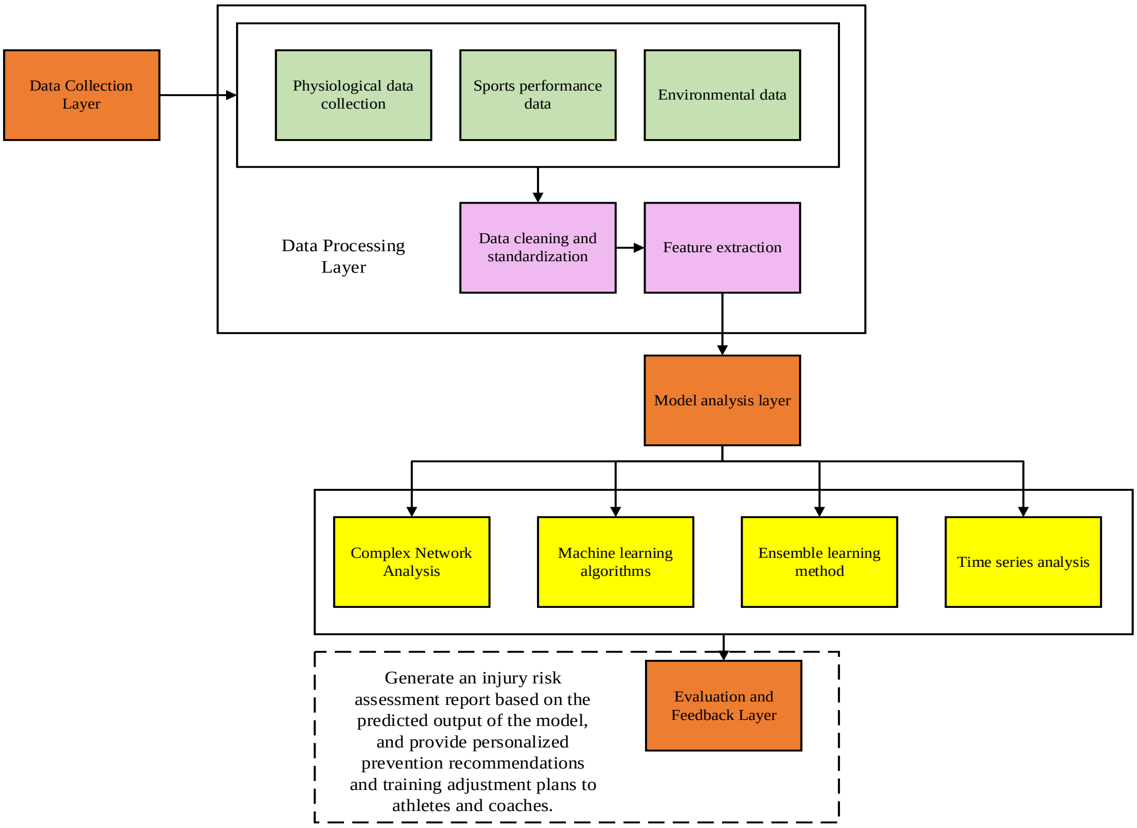

Figure 1. Risk assessment model of sports training injury.

Data is the basis of IRA model. The athletes' heart rate, blood oxygen saturation, muscle fatigue and other physiological indicators are collected by heart rate monitors, exercise bracelets and other equipment. Through video analysis software and motion capture system, the athletes' technical movements, speed, acceleration and posture changes are recorded. The external environmental data such as temperature, humidity and ground conditions of the training ground are recorded, which can affect the performance and injury risk of athletes. Noise and outliers are removed, and data from different sources are standardized to ensure data consistency. Data mining technology is used to extract key features related to injury risk, such as movement pattern change and fatigue accumulation trend. Moreover, the design of IRA model needs to combine a variety of algorithms and technologies to achieve accurate analysis and risk prediction of complex data, build an interactive network model among athletes, and identify high-risk interactions or abnormal sports behaviors that may lead to injuries. The appropriate machine learning algorithms (such as random forest, support vector machine, neural network, etc.) are selected to process multi-dimensional data and improve the accuracy of prediction. Combining the advantages of various models, the robustness and generalization ability of the model are improved by ensemble learning method. The historical data of athletes are analyzed in time series to predict the future injury risk trend.

In this network model, the determination of edge weight is based on several key factors, including physiological data (such as heart rate and muscle fatigue), sports performance data (such as training intensity, speed and reaction time) and environmental data (such as temperature and humidity). By analyzing these data, the correlation or influence degree between them is calculated, and then the weight of edges is determined. The higher the edge weight, the stronger the relationship among nodes and the greater the impact on damage risk. The selection criteria of network features are based on the correlation among features and damage risk, the interpretability of features and the robustness of features in different training scenarios. Specifically, the characteristics that can effectively reflect the athlete's physical condition, training load and environmental pressure are selected. These characteristics include the degree of nodes, aggregation coefficient, the centrality of feature vectors and other complex network indicators, which can reveal the relationship between key nodes and potentially high-risk nodes in the network. In order to deal with the time dependence in training data, the model introduces time series analysis method to model the training data in time series. This includes using sliding window method to segment data to capture trends and fluctuations in continuous time periods. The time decay function is also combined to give higher weight to the recent data to enhance the response ability of the model to the recent state. This approach enables the model to predict the potential injury risk in future training more effectively.

Through the above steps, an accurate, real-time and personalized sports training IRA model can effectively reduce the injury risk of athletes and improve the training effect and competitive level of athletes. This model provides strong technical support and decision-making basis for coaches and athletes under the background of modern sports science.

4. Experimental analysis of IRA model in physical training

4.1. Model performance comparison

The dataset selected in the experiment is the Sport Vu National Basketball Association Player Movement Dataset, which contains the movement data of players in National Basketball Association (NBA) matches. Although it is mainly used for basketball game analysis, it can be used to simulate the movement patterns of athletes in different environments. By analyzing the players' running and movements, we can infer the possible risk of injury. The specific contents of the data set are shown in **Table 3**:

Table 3. Dataset content.

Dimension	Content
Time stamp	The specific time point of each record.
Position coordinates	The player's position on the court.
Speed	The speed at which players move.
Acceleration	Acceleration of the player's movement.
Competition event	For example, shooting, passing, foul, etc.

In the experiment of sports training IRA model, a suitable experimental environment is studied, in which the computer processor is Intel Core i7, the memory is 16 GB, the storage device is 500 GB, the graphics card is NVIDIA GTX 1660, the operating system is Windows 10 (64-bit), and the Python library is Pandas. The parameters of the model are also set, in which the Damping Factor is 0.75, the Resolution is 0.5, the Random State is 20, and the Number of Estimators is 100. The performance of the model is compared in the experiment, and the comparison indexes are accuracy, recall, specificity and Area Under Curve of The Receiver Operating Characteristic Curve (AUC-ROC). The comparison models are eXtreme Gradient Boosting (XGBoost) and Light Gradient Boosting Machine (LightGBM). XGBoost has performed well in many machine learning competitions and is famous for its excellent performance and stability. By enhancing the combination of trees, it is excellent in dealing with complex features and large-scale datasets. As a widely used machine learning algorithm, XGBoost has proved its effectiveness in many fields, including classification, regression and sorting tasks, which makes it an ideal choice for benchmarking. XGBoost supports regularization (L1 and L2), which can effectively prevent over-fitting. In addition, its built-in parallel processing ability and efficient memory use also make the model more advantageous when dealing with large-scale data. Developed by Microsoft, LightGBM is famous for its fast-training speed and low memory consumption, and is especially suitable for processing large-scale data sets. LightGBM adopts decision tree algorithm based on histogram, which can quickly process high-dimensional sparse data, which makes it have significant advantages in the case of huge data and feature quantity. LightGBM supports distributed training and Graphic Processing Unit (GPU) acceleration, and can be easily extended to larger data sets and complex models, thus improving its applicability in practical applications. The performance comparison results are shown in **Figure 2**:

In **Figure 2**, in terms of accuracy, the accuracy of the proposed optimized model is 0.883 when the data volume is 1000. When the data volume is 2000, the accuracy is significantly improved to 0.923, which is the best performance. When the data is 3000, the accuracy is 0.910, which is slightly lower, but it is still better than other models. XGBoost and the optimized model show better adaptability and improvement when the data volume increases, while the accuracy of LightGBM does not change much when the data volume increases, so it may be necessary to further optimize its parameters to adapt to the large data volume. In terms of recall rate, the XGBoost model has a recall rate of 0.835 when there are 1000 data. In 2000, the recall rate dropped to 0.801. When the data volume is 3000, the recall rate rises to 0.848. The recall rate of LightGBM model is 0.862 when the data volume is 1000. In 2000, the recall rate dropped to 0.831. When the data volume is 3000, the recall rate drops slightly to 0.829. When the proposed optimized model has 1000 data, the recall rate is 0.849. When the data volume is 2000, the recall rate is 0.855. When the data volume is 3000, the recall rate is increased to 0.866, which is the best performance. The optimized model is stable under all data volumes, especially when the data volume is 3000, which shows its good recognition ability for complex patterns. The recall rate of XGBoost is significantly improved under a large amount of data, which shows its effectiveness under certain conditions. In terms of specificity, the specificity of the proposed optimized model reaches 0.883, 0.899 and 0.87 under different data volumes, and the optimized model has high specificity under all data volumes, especially when the data volume is 2000, showing its strong non-destructive identification ability. However, LightGBM is stable when the amount of data increases, and its specificity is gradually improved. XGBoost also improved significantly when the data volume increased to 3000, showing good adaptability. In terms of AUC-ROC, the AUC-ROC of the proposed optimized model reaches 0.927 when the data volume is 1000. When the data volume is 2000, AUC-ROC is 0.928, which is the best performance. When the data volume is 3000, AUC-ROC is 0.920. The AUC-ROC of the optimized model is significantly higher than other models in all data volumes, especially in 1000 and 2000 data volumes, which shows that it has strong classification ability.

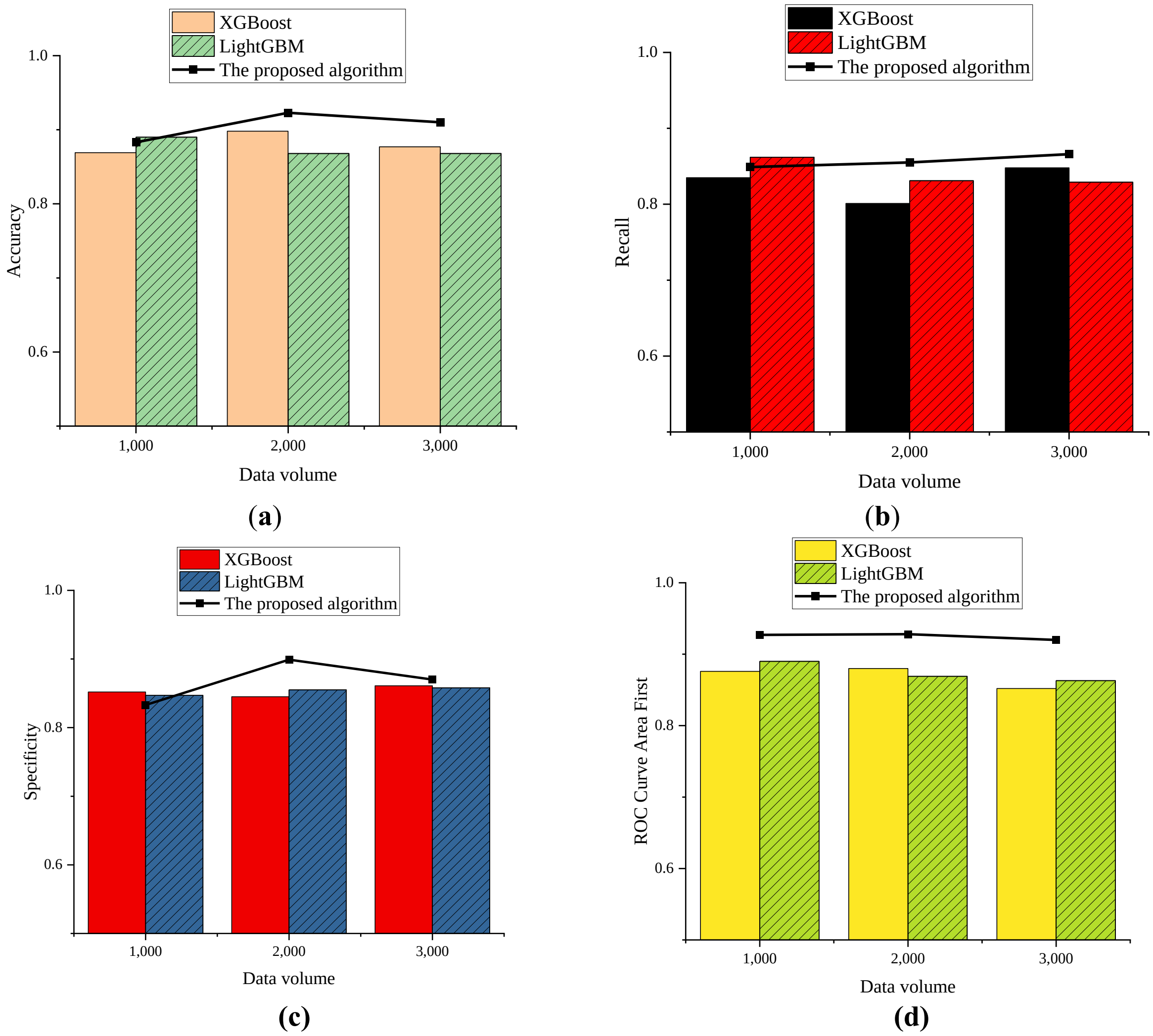

Figure 2. Performance comparison experiment **(a)** accuracy; **(b)** recall; **(c)** specificity; **(d)** ROC curve area first.

4.2. Comparison of simulation experiments

To further verify the effectiveness of the sports training IRA model, a simulation experiment is set up, which compared the indicators with real-time, robustness, adaptability and user satisfaction. The experiment is conducted by scoring, with a score of 1–5. The higher the score, the better the result of the model. The research divides each index into three dimensions, and the experimental results are shown in **Figure 3**:

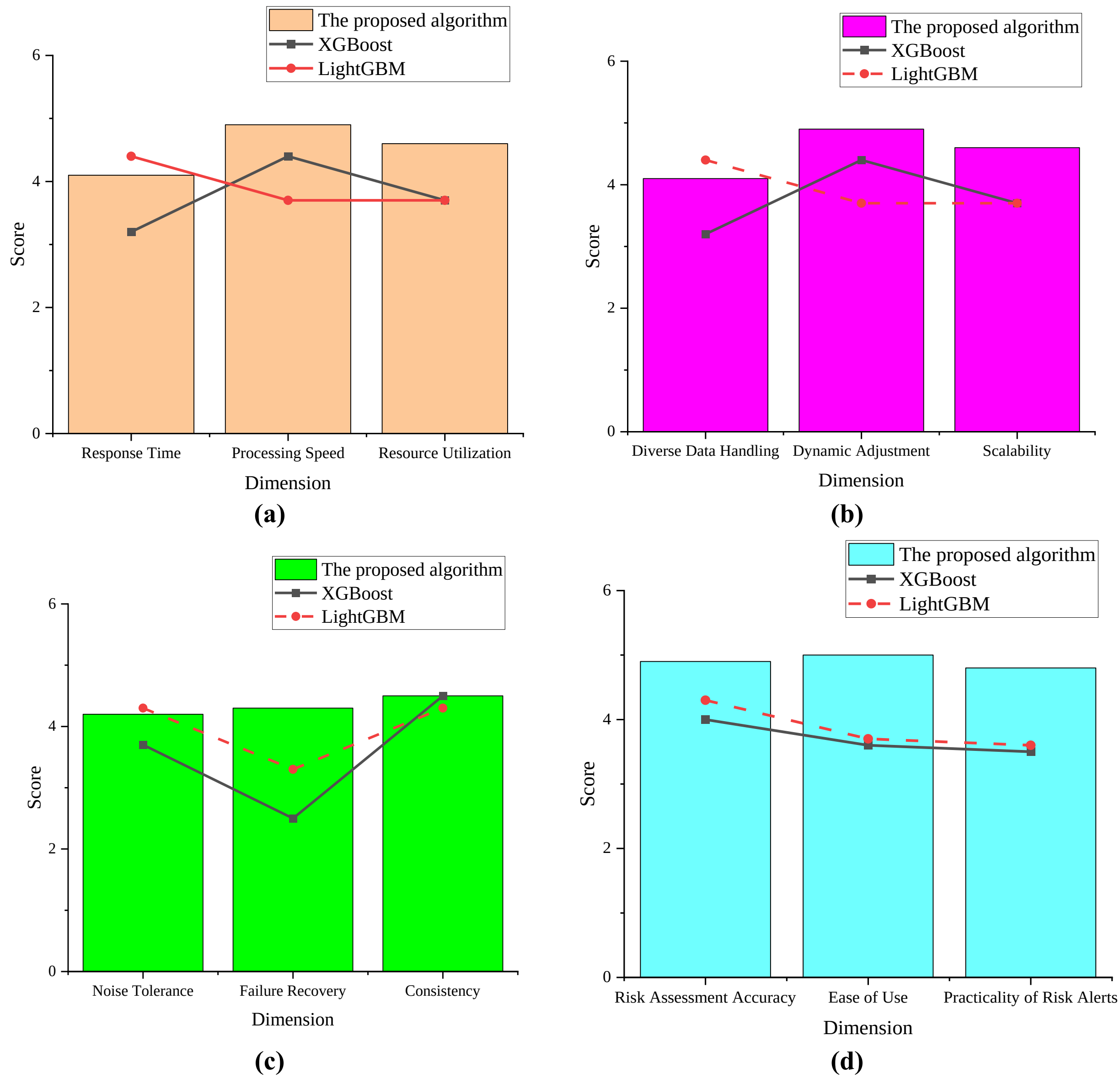

Figure 3. Simulation experiment results **(a)** real-time performance; **(b)** robustness; **(c)** adaptability; **(d)** user satisfaction.

The results in **Figure 3** show that the response time score of the proposed optimized model is 4.1, which is excellent in real-time comparison. The processing speed score is 4.9, which is the best. The score of computing resource usage is 4.6, which is excellent. XGBoost performs well in processing speed, but it is slightly insufficient in response time, so it may be necessary to further optimize the model to improve response efficiency. LightGBM is outstanding in response time, but it is slightly lower than the optimization model in processing speed and computing resource usage, showing its potential in efficient task processing. The optimized model performs well in all dimensions, especially in processing speed, which shows its superiority in application scenarios with high real-time requirements. In the comparison of robustness, the noise tolerance score of XGBoost model is 3.7, and the performance is moderate. The score of fault recovery ability is 2.5, which is weak and needs to be improved. The consistency score is 4.5, which shows excellent performance and high stability of the results. The noise tolerance score of LightGBM model is 4.3, which shows excellent performance and handles the noise influence well. The recovery ability score is 3.3, and the performance is acceptable. The consistency score is 3.7, and the performance is relatively stable. The noise tolerance score of the proposed optimized model is 4.2, which shows excellent performance. The score of fault recovery ability is 4.3, which shows the best performance and has good fault handling ability. The consistency score is 4.5, which is equivalent to XGBoost consistency performance, indicating that the results are highly stable. In the comparison of adaptability, the score of diversified data processing ability of the proposed optimized model is 4.5, which shows excellent performance and strong ability to process different types of data. The score of dynamic adjustment ability is 4.8, which shows the best performance and can adapt to changes quickly. The expansibility score is 4.2, which shows excellent performance and good expansibility. XGBoost performs well in diversified data processing and expansibility, but it is slightly insufficient in dynamic adjustment ability, and may need to be optimized in adapting to the rapidly changing environment. LightGBM performs well in dynamic adjustment and shows some adaptability, but it is like XGBoost in diversified data processing and scalability. The proposed optimized model performs well in all dimensions, especially in dynamic adjustment ability, which shows that it has high adaptability and can flexibly cope with various data and environmental changes. In the comparison of user satisfaction, the accuracy score of the risk assessment of the proposed optimized model is 4.9, which is excellent. The score of ease of use is 5.0, which shows excellent performance, indicating that the user experience is very good. The practical score of risk early warning is 4.8, which is excellent. The optimized model performs best in the comparative experiment of user satisfaction, especially in terms of ease of use and accuracy of risk assessment, which is obviously superior to other models, and is suitable for scenarios that require high user experience and accurate risk assessment. LightGBM performs well in the accuracy of risk assessment, but needs to further improve the user experience in other aspects. XGBoost is relatively average in all aspects, but there is room for improvement in user experience and early warning practicability.

5. Discussion

From the perspective of computing efficiency, the optimized model performs well in processing speed and computing resource usage, especially in real-time and processing speed. This shows that the model has obvious advantages in processing data efficiently. However, it should be noted that the improvement of computing efficiency is usually accompanied by the consumption of computing resources. For example, although the optimized model is superior in processing speed, it may rely heavily on hardware resources. Thus, in the environment with limited resources, it may be necessary to simplify the model to balance the relationship between computing efficiency and resource use. Secondly, according to the performance of each model, this paper analyzes the advantages of different models in applicable scenarios. The optimized model has the best performance in terms of risk assessment accuracy, adaptability and user satisfaction. Moreover, the model is especially suitable for scenes that require high accuracy and user experience, such as personalized training plans of professional athletes and real-time risk monitoring of high-intensity training. In addition, because of its good dynamic adjustment ability, the optimized model can provide effective support in the case of rapid changes in the training environment, so it also has advantages in complex and changeable scenes.

XGBoost shows good adaptability when the amount of data increases, especially in recall and specificity. This means that XGBoost model is more applicable in scenes that need high recall rate, such as in the early warning stage, to ensure that potential risks will not be ignored. In addition, although the computational efficiency of XGBoost is not as good as that of the optimization model, it can still provide stable performance in the case of limited resources, which makes it suitable for training environments with few resources. LightGBM model performs well in response time and specificity, but it is not as good as other models in precision and recall. Therefore, LightGBM may be more suitable for scenarios with low real-time requirements, such as data batch processing or post-event risk assessment and analysis. In these scenarios, the LightGBM model can provide stable performance, and can maintain low computational overhead in the case of limited resources.

Integrating this model into the existing training management system needs to consider data collection, data processing, and results display. Firstly, the real-time collection of physiological data, sports performance data and environmental data can be realized through seamless docking with existing physiological monitoring equipment, training data acquisition system and environmental data sensor. Then, the data is imported into the model for preprocessing steps such as cleaning, standardization and feature extraction to ensure the standardization and consistency of model input. For the convenience of coaches and managers, the output of the model should be integrated into the user interface of the training management system to provide concise risk assessment results and related suggestions to help coaches monitor the athletes' status in real time. In addition, an alarm mechanism is set up. When the risk assessment value reaches a certain threshold, the system will give an alarm to the coach, prompting the possible injury risk. In the highly dynamic sports training environment, real-time processing is very important to ensure the safety of athletes. The real-time processing requirements of this model include fast data preprocessing, calculation and feedback to provide real-time risk assessment results during training. Specifically, the system should complete data cleaning, feature extraction and model calculation within a few seconds to ensure that the output results of the model can be updated in real time during the training process for coaches' timely reference. At the same time, the model needs to have the ability to dynamically adjust to changes in the data in real time to improve the accuracy and robustness of the prediction. This real-time processing ability can not only help the coach to adjust the training intensity and content at any time during the training, but also provide retrospective analysis after the training, thus providing a basis for future training strategies. In order to make the prediction of the model operable, the coach needs to be able to understand the risk assessment results output by the model and the meaning behind it. The risk assessment results output by the model can be quantified as a score or rating, indicating the level of damage risk. According to this score, the coach can divide the risk into different levels, such as low, medium and high, and take corresponding preventive measures according to different risk levels. For example, when the risk is at a low level, athlete can continue the current training intensity. When the risk reaches the middle or high level, the coach can choose to reduce the training load appropriately, increase the recovery time, or adjust the training content. In addition, the prevention suggestions and training adjustment programs provided by the model can help coaches to take more scientific intervention measures and reduce the possibility of injury. By regularly checking the risk assessment report generated by the model, the coach can also make a personalized training plan for each athlete, and continuously optimize the strategy according to the training feedback to ensure that the athletes train in a safe state.

Compared with the research of Zhou et al., this paper mainly evaluates the injury risk of sports training based on statistical regression model, focusing on analyzing the relationship between simple physiological data and injury risk. This method has certain accuracy when dealing with a small amount of data, but the adaptability and prediction accuracy of its model are limited in complex environment and large amount of data [32]. In contrast, this paper uses complex network analysis algorithm, combined with multi-dimensional data (physiological data, sports performance data, environmental data) for risk assessment, which significantly improves the model's ability to deal with diverse data and adapt to complex environment. Additionally, the time dependence of data is dealt with by time series analysis method, and the optimized model is superior in real-time performance and dynamic adjustment ability, which makes it have better prediction effect in high dynamic training scene. Secondly, compared with the research of Zhan et al., they used the traditional machine learning algorithm to predict the risk of sports training injuries. Although there was some performance in precision, there were limitations in computational efficiency and robustness, especially in the case of increasing data and noise, the accuracy of the model had declined [33]. The optimized model is excellent in precision and robustness, especially in dealing with different data volumes and noise tolerance. In addition, by combining complex network analysis and ensemble learning methods, the optimized model not only improves the identification accuracy of damage risk, but also is significantly better than their proposed model in terms of user satisfaction and practicality. This makes the proposed model have greater application value in scenes that need high user experience and complex data processing.

To sum up, this paper has significantly improved the data processing ability, real-time performance and user experience compared with the traditional research, indicating that the optimization model has stronger application potential and adaptability in the changeable environment.

6. Conclusion

In this paper, by constructing and optimizing the complex network analysis algorithm model, the injury risk in sports training is evaluated, and multi-dimensional model performance comparison experiments are carried out, including accuracy, real-time, robustness, adaptability and user satisfaction. The optimized model performs well in many indicators, especially in the accuracy of risk assessment and user satisfaction, which is significantly better than the traditional model. This shows that the model has higher prediction accuracy and good user experience. The optimized model shows superior performance in processing speed and fault recovery ability, and is suitable for real-time systems that need high efficiency and stability. This advantage enables it to quickly adapt to the changing training environment and maintain efficient damage risk assessment ability. The optimization model performs well in diversified data processing ability, dynamic adjustment ability and expansibility, and can flexibly cope with different types of data and changing training conditions, so it has high application potential. By comprehensively evaluating the accuracy of risk assessment, ease of use and practicability of risk early warning, the optimization model gets the highest score on user satisfaction, which reflects its high recognition and use value in practical application.

This paper also has some shortcomings. Although the model adopts complex network analysis method, some potential important factors may be ignored in feature selection, which affects the comprehensiveness and accuracy of the model. Moreover, although the optimized model has obvious advantages in performance, its computational complexity is high, and its requirements for computing resources are great, which may limit its application in resource-limited environments. Future research will further improve the accuracy and robustness of the model by introducing richer features and multimodal data (such as physiological signals and video data), and explore the application of automatic feature extraction technology (such as deep learning) in risk assessment. Meanwhile, it studies how to simplify the model structure and algorithm flow, reduce the computational complexity and resource consumption, and improve the efficiency and operability of the model in practical application.

Author contributions: Conceptualization, CY; methodology, CY; software, TL; validation, CY, ZY and TL; formal analysis, ZY; investigation, ZY; resources, TL; data curation, TL; writing—original draft preparation, CY; writing—review and editing, CY; visualization, ZY; supervision, TL; project administration, CY; funding acquisition, CY. All authors have read and agreed to the published version of the manuscript.

Ethical approval: Not applicable.

Conflict of interest: The authors declare no conflict of interest.

References

1. Bullock G S, Mylott J, Hughes T, et al. Just how confident can we be in predicting sports injuries? A systematic review of the methodological conduct and performance of existing musculoskeletal injury prediction models in sport. Sports medicine, 2022, 52(10): 2469-2482.
2. Meng L, Qiao E. Analysis and design of dual-feature fusion neural network for sports injury estimation model. Neural Computing and Applications, 2023, 35(20): 14627-14639.
3. Yung K K, Ardern C L, Serpiello F R, et al. Characteristics of complex systems in sports injury rehabilitation: examples and implications for practice. Sports medicine-open, 2022, 8(1): 24.
4. Nassis G, Verhagen E, Brito J, et al. A review of machine learning applications in soccer with an emphasis on injury risk. Biology of sport, 2023, 40(1): 233-239.
5. Martins F, Przednowek K, França C, et al. Predictive modeling of injury risk based on body composition and selected physical fitness tests for elite football players. Journal of Clinical Medicine, 2022, 11(16): 4923.
6. McDevitt S, Hernandez H, Hicks J, et al. Wearables for biomechanical performance optimization and risk assessment in

industrial and sports applications. Bioengineering, 2022, 9(1): 33.
7. Jayanthi N, Schley S, Cumming S P, et al. Developmental training model for the sport specialized youth athlete: a dynamic strategy for individualizing load-response during maturation. Sports health, 2022, 14(1): 142-153.
8. Dhanke J A, Maurya R K, Navaneethan S, et al. Recurrent neural model to analyze the effect of physical training and treatment in relation to sports injuries. Computational Intelligence and Neuroscience, 2022, 2022(1): 1359714.
9. Schweizer N, Strutzenberger G, Franchi M V, et al. Screening tests for assessing athletes at risk of acl injury or reinjury—a scoping review. International Journal of Environmental Research and Public Health, 2022, 19(5): 2864.
10. Cui J, Du H, Wu X. Data analysis of physical recovery and injury prevention in sports teaching based on wearable devices. Preventive medicine, 2023, 173(56): 107589.
11. Wilke J, Groneberg D A. Neurocognitive function and musculoskeletal injury risk in sports: A systematic review. Journal of science and medicine in sport, 2022, 25(1): 41-45.
12. Ji S, Ghajari M, Mao H, et al. Use of brain biomechanical models for monitoring impact exposure in contact sports. Annals of Biomedical Engineering, 2022, 50(11): 1389-1408.
13. Lutter C, Jacquet C, Verhagen E, et al. Does prevention pay off? Economic aspects of sports injury prevention: a systematic review. British journal of sports medicine, 2022, 56(8): 470-476.
14. Ageberg E, Brodin E M, Linnéll J, et al. Cocreating injury prevention training for youth team handball: bridging theory and practice. BMJ Open Sport & Exercise Medicine, 2022, 8(2): e001263.
15. Jauhiainen S, Kauppi J P, Krosshaug T, et al. Predicting ACL injury using machine learning on data from an extensive screening test battery of 880 female elite athletes. The American Journal of Sports Medicine, 2022, 50(11): 2917-2924.
16. Richter C, O'Reilly M, Delahunt E. Machine learning in sports science: challenges and opportunities. Sports Biomechanics, 2024, 23(8): 961-967.
17. Bullock G S, Hughes T, Arundale A H, et al. Black box prediction methods in sports medicine deserve a red card for reckless practice: a change of tactics is needed to advance athlete care. Sports Medicine, 2022, 52(8): 1729-1735.
18. Li N, Zhu X. Design and application of blockchain and IoT-enabled sports injury rehabilitation monitoring system using neural network. Soft Computing, 2023, 27(16): 11815-11832.
19. Chidambaram S, Maheswaran Y, Patel K, et al. Using artificial intelligence-enhanced sensing and wearable technology in sports medicine and performance optimisation. Sensors, 2022, 22(18): 6920.
20. Cabre H E, Moore S R, Smith-Ryan A E, et al. Relative energy deficiency in sport (RED-S): scientific, clinical, and practical implications for the female athlete. Deutsche Zeitschrift fur Sportmedizin, 2022, 73(7): 225.
21. Mason J, Rahlf A L, Groll A, et al. The interval between matches significantly influences injury risk in field hockey. International journal of sports medicine, 2022, 43(03): 262-268.
22. Collings T J, Diamond L E, Barrett R S, et al. Strength and biomechanical risk factors for noncontact ACL injury in elite female footballers: a prospective study. Medicine & Science in Sports & Exercise, 2022, 54(8): 1242-1251.
23. Yung K K, Ardern C L, Serpiello F R, et al. A framework for clinicians to improve the decision-making process in return to sport. Sports medicine-open, 2022, 8(1): 52.
24. Ramkumar P N, Luu B C, Haeberle H S, et al. Sports medicine and artificial intelligence: a primer. The American Journal of Sports Medicine, 2022, 50(4): 1166-1174.
25. Davis G A, Echemendia R J, Ahmed O H, et al. Introducing the child sport concussion assessment tool 6 (child Scat6). British Journal of Sports Medicine, 2023, 57(11): 632-635.
26. Costa E Silva L, Teles J, Fragoso I. Sports injuries patterns in children and adolescents according to their sports participation level, age and maturation. BMC sports science, medicine and rehabilitation, 2022, 14(1): 35.
27. Nilstad A, Petushek E, Mok K M, et al. Kiss goodbye to the 'kissing knees': No association between frontal plane inward knee motion and risk of future non-contact ACL injury in elite female athletes. Sports biomechanics, 2023, 22(1): 65-79.
28. Tan T, Gatti A A, Fan B, et al. A scoping review of portable sensing for out-of-lab anterior cruciate ligament injury prevention and rehabilitation. NPJ Digital Medicine, 2023, 6(1): 46.
29. Broglio S P, McAllister T, Katz B P, et al. The natural history of sport-related concussion in collegiate athletes: findings from the NCAA-DoD CARE Consortium. Sports medicine, 2022, 52(2): 403-415.
30. Torres-Ronda L, Beanland E, Whitehead S, et al. Tracking systems in team sports: a narrative review of applications of the data and sport specific analysis. Sports Medicine-Open, 2022, 8(1): 15.
31. Bird M B, Koltun K J, Mi Q, et al. Predictive utility of commercial grade technologies for assessing musculoskeletal injury

risk in US Marine Corps Officer candidates. Frontiers in Physiology, 2023, 14(2): 1088813.
32. Zhou H, Nau C, Xie F, et al. A machine-learning prediction model to identify risk of firearm injury using electronic health records data. Journal of the American Medical Informatics Association, 2024, 31(10): 2173-2180.
33. Zhan Z, Pan L, Zhu Y, et al. Moderate-intensity treadmill exercise promotes mtor-dependent motor cortical neurotrophic factor expression and functional recovery in a murine model of crush spinal cord injury (SCI). Molecular Neurobiology, 2023, 60(2): 960-978.