International Journal for Housing Science and Its Applications
Publish August 4, 2025. Volume 46, Issue 3 Pages 2518-2527

https://doi.org/10.70517/ijhsa463210

# Application of an athlete injury risk assessment model based on regression analysis method in sports rehabilitation

**Jiayue Yang**[1,*]
[1] Jilin Police College, Changchun, Jilin, 130000, China
Corresponding authors: (e-mail: yang1338430077@163.com).

**Abstract** In recent years, the importance of athletes' injury risk assessment has been increasing in the field of sports training. Through questionnaire survey, expert interviews and mathematical statistics, the risk factors of athletes' sports injuries and illnesses were studied, and the assessment index system was established. Combined with logistic regression analysis, the risk factors were evaluated in four dimensions: athlete's own risk, organizational management risk, equipment and facilities risk, and environmental risk factors, in order to provide references for athletes' sports rehabilitation and training programs. In addition to environmental risk factors, athlete's own risk, organizational management risk, and facility risk were significantly positively correlated with athletes' injuries and illnesses, and athlete's own risk was the main causative factor for the occurrence of injuries and illnesses. The average precision, recall, F2 score and AUC of the logistic regression model were 0.852, 0.9714, 0.9174 and 0.9547, respectively. Compared with other models, the model in this paper improved by 18.56%, 7.63%, 8.24% and 2.96%, respectively, and demonstrated a good precision of assessment.

**Index Terms** logistic regression analysis, indicator system, risk assessment, mathematical statistics, questionnaire, athlete injury

## I. Introduction

In recent years, the heat of sports competitions has remained high, and the physical health of athletes has also attracted much attention. Almost all athletes carry injuries, on the one hand, high-intensity training puts the body in an extreme state for a long time, and the musculoskeletal muscles and bones are repeatedly subjected to stress, which is easy to accumulate damage over time [1], [2]. For example, basketball players frequently jump and land, knee meniscus wear and tear is difficult to avoid. Soccer players high-speed change of direction emergency stop, ankle ligament strain risk is elevated. In addition, unscientific training methods directly bury the hidden danger of injury. Part of the coach excessive pursuit of short-term results, ignoring the rules of physical development of athletes, youth premature strength training may affect bone growth, part of the team rehabilitation means a single means, only simple stretching after training, the lack of cold therapy, pressurized recovery and other modern rehabilitation techniques, fatigue recovery is insufficient to allow the body to enter the next round of training with hidden problems [3]-[6]. On the other hand, the intensity of the tournament exceeds the body's tolerance threshold. Professional league season spanning up to eight or nine months, soccer players average three days a game, muscle fatigue state is more prone to strains, international competition cycle adjustment leads to the compression of the training program, the Winter Olympics athletes in order to adapt to the schedule in advance, had to double the amount of training, the incidence of bone stress fracture increased significantly [7]-[9]. The athlete risk assessment model can identify and evaluate various sports injury risk factors, which ensures the normal training and competition status of athletes, and has attracted the attention of experts and scholars.

Sports injury risk assessment can scientifically evaluate the potential functional problems of the body through a series of tests, and has been widely used at home and abroad in running, skiing, taekwondo and other sports to help athletes recognize their own deficiencies and risk factors, and prepare countermeasures in advance to reduce the risk of sports injuries [10]-[12]. Currently, sports injury assessment is mostly based on functional movement screening (FMS), Y balance test (YBT), and the ability to demonstrate speed, strength, endurance, sensitivity, flexibility and other aspects of physical fitness ability test as the main means. FMS can improve the body's functional movement compensation by evaluating the body's function. However, its scoring criteria are subjectively influenced, the scene of assessment is not in the exercise state, but before and after the exercise, the accuracy of assessment is not more than 60%, and the reliability is limited [13]-[15]. YBT predicts the risk of sports injuries by dynamically assessing the ability of body balance, the ability to control the torso in all directions, and the physical quality predicts the risk of sports injuries by comparing the differences in the quality ability of different subjects. YBT only assesses the athlete's single movement, and it is difficult to comprehensively reflect the risk of injuries. The YBT only assesses a single movement of the athlete, which is difficult to reflect the injury risk in a comprehensive way, and at the same time does not consider the role of other risk factors and is affected by individual differences [16]-[19]. Therefore, it is necessary to develop a new method of injury risk assessment for athletes to improve the basis for rehabilitation and mitigation.

In this study, the Chinese national team members, coaches and domestic experts were taken as the main research objects, and the related questionnaires of “Risk Source Identification Questionnaire” and “Indicator Recognition Questionnaire” were designed to screen out the risk sources and risk factors of athletes' injuries and illnesses, and initially construct a A preliminary index system for assessing the risk of athletes' injuries and diseases was constructed. Athletes' data were collected by Likert 5-level scale, and the results were entered into SPSS-26 for Mac, and multiple regression analysis was used to test the assessment indexes of athletes' risk of sports injuries and illnesses. Precision, recall, F2 score and area under the ROC curve were chosen to measure the performance of the regression model.

## II. Subject and methodology of the study

### *II. A. Subjects of study*

Sports injury is a common problem in athletics. In recent years, athletes' long training time and high body load have led to an increase in the incidence of injuries, which seriously impedes athletes' maintenance and further improvement of athletic performance, and even leads to athletes leaving the field of play prematurely. Therefore, the treatment and prevention of sports injuries have always been the focus of the sports medicine community [20].

In this paper, the risk identification, assessment and response of acute sports injuries in Chinese athletes were studied. Chinese athletes are defined as first- and second-tier athletes of the national team and athletes from various provinces, including the 8-1 team, who participate in national competitions, including national championships, national championships and national games. Due to my limited ability and time, and according to the needs of the research task, the first- and second-line players of the Chinese national team, the coaches of the Chinese national team, and the domestic experts were taken as the research subjects.

### *II. B. Research methodology*

#### II. B. 1) Questionnaire method

The subjects of the survey were the coaches and managers of the national gymnastics team, a total of 10 gymnastics experts, and 22 athletes of the national team, including 13 first-line athletes, 6 men and 7 women, and 9 second-line athletes, 4 men and 5 women. The information of the experts is shown in Table 1. Due to my limited time and ability, I can't investigate all the high level gymnasts and coaches in all provinces and cities in China, so this paper selects some athletes and coaches of the Chinese national gymnastics team and a few gymnastic experts as the target of investigation.

Table 1: Expert statement

| Name | Job title | Management or training years | With men or women |
|---|---|---|---|
| H1 | Deputy director of gymnastics center | Management 5 years | The national gymnastics coach |
| H2 | National coach | Training for 23 years | Men's team |
| H3 | Senior coach | Training for 18 years | Men's team |
| H4 | Senior coach | Training for 20 years | Men's team |
| H5 | Senior coach | Training for 24 years | Men's team |
| H6 | Senior coach | Training for 15 years | Men's team |
| H7 | The national gymnastics team's chief coach | Training for 25 years | Men's team |
| H8 | National coach | Training for 25 years | Men's team |
| H9 | Senior coach | Training for 17 years | Men's team |
| H10 | Senior coach | Training for 14 years | Men's team |
| H11 | National level | Training for 25 years | Men's team |

Athlete information is shown in Table 2.

Table 2: Athlete information

| Name | Gender | Training period | A line or two line |
|---|---|---|---|
| Y1 | Male | 25 | First-line |
| Y2 | Male | 22 | First-line |
| Y3 | Male | 22 | First-line |
| Y4 | Male | 21 | First-line |
| Y5 | Male | 19 | First-line |
| Y6 | Male | 19 | First-line |
| Y7 | Female | 18 | First-line |
| Y8 | Female | 12 | First-line |
| Y9 | Female | 12 | First-line |
| Y10 | Female | 14 | First-line |
| Y11 | Female | 11 | First-line |
| Y12 | Female | 14 | First-line |
| Y13 | Female | 15 | First-line |
| Y14 | Male | 18 | Second line |
| Y15 | Male | 22 | Second line |
| Y16 | Male | 19 | Second line |
| Y17 | Male | 17 | Second line |
| Y18 | Male | 17 | Second line |
| Y19 | Female | 11 | Second line |
| Y20 | Female | 8 | Second line |
| Y21 | Female | 12 | Second line |
| Y22 | Female | 13 | Second line |

According to the content and purpose of this study, and following the basic requirements of the research method on questionnaire design, the "Risk Source Identification Questionnaire", "Indicator Recognition Questionnaire", "Indicator Quantification Table Questionnaire", "Risk Quantity Formula Questionnaire", "Risk Assessment Expert Questionnaire", "Risk Assessment Male Athlete Questionnaire" and "Risk Assessment Female Athlete Questionnaire" were designed.

The reliability test of the questionnaire was carried out using a small-scale "measure-re-measurement" method, and the second questionnaire was distributed 15 days apart, and the reliability test was carried out after collection. The test results of the risk source identification questionnaire (n=4), the indicator recognition questionnaire (n=4), the indicator quantification questionnaire (n=4), the risk quantity formula questionnaire (n=4), the risk assessment expert questionnaire (n=4), the risk assessment male athlete questionnaire (n=5), and the risk assessment female athlete questionnaire (n=5) were R=0.86, R=0.87, R=0.82, R=0.88, R=0.87, R=0.85, and R=0.84, respectively (P <0.05), indicating that the questionnaire was filled out with high reliability.

Before the implementation of the questionnaire, the structure and content of the questionnaire were tested on a 5-point scale, and the average evaluation scores of the experts were 3.64, 3.85, 3.88, 3.79, 3.87, 3.83 and 3.89, respectively. The questionnaire has a high validity.

The questionnaire was distributed and collected between January and March 2025 with the help of the captain of the national men's gymnastics team, and has a high recovery rate. 10 copies of the "Risk Source Identification Questionnaire", "Indicator Recognition Questionnaire", "Indicator Quantification Table Questionnaire", "Risk Quantity Formula Questionnaire" and "Risk Assessment Expert Questionnaire" were distributed respectively, and 10 copies were recovered respectively, with a recovery rate of 100%. The "Risk Assessment Male Athlete Questionnaire" and the "Risk Assessment Female Athlete Questionnaire" were distributed and 11 copies were recovered, with 100% recovered.

#### II. B. 2) Expert interview method

Coaches' experts were consulted on important issues of the thesis, especially in determining the risk sources and risk factors of acute sports injuries and diseases in athletes. Interviews with many experts, such as the head coach of the national gymnastics team, the head coach of the women's gymnastics team, etc., national coaches, senior coaches, as shown in the table, and some of the athletes of the national gymnastics team, provided direction for the progress of the thesis [21].

### II. B. 3) Mathematical and statistical methods

The 72 questionnaires recovered were organized, the data of 72 questionnaires were classified and summarized by EXCLE software, and the function formulas were used to calculate the mean score, sum, product and other operations on the data, and the processed data were used to make tables and bar charts that are valuable for this study [22].

### II. B. 4) Regression analysis

Regression analysis is a statistical method involved in the study of influencing factors, in which the characteristics of data are identified in a large amount of raw data. The mathematical expression of the correlation of each variable is obtained, and the relationship is analyzed with probability statistics to determine the validity of the relationship between variables, and at the same time, through the value of one or more variables (independent variables), the value of another variable (dependent variable) is evaluated or controlled [23]. According to the number of variables involved, it can be categorized into univariate regression and multivariate regression, and according to the different types of dependent variables, it can be categorized into: linear regression, logistic regression, Poisson regression, etc. When the data of the dependent variable is expressed as a continuous value, multivariate linear regression is chosen to establish a linear line of best fit regression between the independent variable $y$ and the dependent variable $x$. Let the evaluated dependent variable be $Y$, and the $k$ independent variables $x_1, x_2, \cdots, x_k$ affecting the dependent variable, the specific mathematical expression is:

$$y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \ldots \beta_k X_k + \varepsilon_a \tag{1}$$

where $\beta_0, \beta_1 \ldots, \beta_k$ are the regression coefficients, $\varepsilon_a$ are the random variables, and for $y$ and $X_1, X_2, \cdots, X_k$, respectively, after n independent observations of $(y_a, x_{1a}, x_{2a}, \cdots, x_{ka}), (a = 1, 2, 3, \cdots, n)$. Obtain n sets of sample data, if $b_1, b_2, \cdots, b_k$, are $\beta_0, \beta_1, \beta_2, \ldots, \beta_k$ for the fitted values, then the regression equation is:

$$\hat{y} = b_0 + b_1 x_1 + b_2 x_2 + \ldots + b_k x_k \tag{2}$$

The parameter estimation is based on the principle of least squares and the equation is expressed as:

$$Q = \sum (Y_i - \hat{Y}_i)^2 = \min \tag{3}$$

Introduce vector and matrix representations:

$$b = \begin{pmatrix} b_0 \\ b_1 \\ b_2 \\ \ldots \\ b_k \end{pmatrix}, Y = \begin{pmatrix} y_1 \\ y_2 \\ \ldots \\ y_n \end{pmatrix}, X = \begin{pmatrix} 1 & x_{11} & x_{21} & \ldots & x_{k1} \\ 1 & x_{12} & x_{22} & \ldots & x_{k2} \\ 1 & x_{13} & x_{23} & \ldots & x_{k3} \\ \ldots & \ldots & \ldots & \ldots & \ldots \\ 1 & x_{1n} & x_{2n} & \ldots & x_{kn} \end{pmatrix}$$

$$A = X^T X = \begin{pmatrix} 1 & 1 & 1 & \ldots & 1 \\ x_{11} & x_{12} & x_{13} & \ldots & x_{1n} \\ x_{21} & x_{22} & x_{23} & \ldots & x_{2n} \\ \ldots & \ldots & \ldots & \ldots & \ldots \\ x_{k1} & x_{k2} & x_{k3} & \ldots & x_{kn} \end{pmatrix} \begin{pmatrix} 1 & x_{11} & x_{21} & \ldots & x_{k1} \\ 1 & x_{12} & x_{22} & \ldots & x_{k2} \\ 1 & x_{13} & x_{23} & \ldots & x_{k3} \\ 1 & \ldots & \ldots & \ldots & \ldots \\ 1 & x_{1n} & x_{2n} & \ldots & x_{kn} \end{pmatrix} \tag{4}$$

$$B = X^T Y = \begin{pmatrix} 1 & 1 & 1 & \ldots & 1 \\ x_{11} & x_{12} & x_{13} & \ldots & x_{1n} \\ x_{21} & x_{22} & x_{23} & \ldots & x_{2n} \\ \ldots & \ldots & \ldots & \ldots & \ldots \\ x_{k1} & x_{k2} & x_{k3} & \ldots & x_{kn} \end{pmatrix} \begin{pmatrix} y_1 \\ y_2 \\ y_3 \\ \ldots \\ y_n \end{pmatrix} = \begin{pmatrix} \sum_{a=1}^{n} y_a \\ \sum_{a=1}^{n} x_{1a} y_a \\ \sum_{a=1}^{n} x_{2a} y_a \\ \sum_{a=1}^{n} x_{2a} y_a \\ \cdots \\ \sum_{a=1}^{n} x_{ka} y_a \end{pmatrix}$$

Using the least squares method, $b$ only needs to satisfy the equation:

$$Ab = B \tag{5}$$

Finally,the least squares solution is obtained as:

$$b = A^{-1}B = (X^T X)^{-1} X^T Y \tag{6}$$

Logistic regression analysis is the most representative linear model in multiple regression, in reality, a variable is often affected by multiple factors, logistic regression can better match the real problem with the model [24]. Secondly, by digging deeper into the quantitative relationship between the variables in the data, not only can we accurately measure the degree of correlation between the factors and the degree of regression fit, but also has the characteristics of automatic diagnosis and risk assessment. After obtaining accurate information about the weights of the independent variables, it helps us to understand the risk factors in the existing information and objectively assess the probability of future risks through the existing risk factor data. Thirdly, the unique results can be obtained through the standard statistical calculation method, which is of great practical significance as a good and accurate guide to the existing risk response and transfer and future risk control.

### *II. C.Athlete Injury Risk Assessment Model Construction*

#### II. C. 1) Sample data collection

The risk assessment index system of athletes' injury and disease is based on the risk identification system established above. The assessment index system covers four major risk sources, namely, athlete's own risk, environmental risk, technical risk and public safety and security risk, and contains 21 types of secondary risk indicators. The Likert 5-level scale was generated using the same method as above, and the second round of questionnaire survey was carried out. 300 questionnaires were distributed and 280 valid questionnaires were collected, with a recovery rate of 93.3%.

#### II. C. 2) Modeling

Athletes' injury risk data were collected through a 5-point Likert scale, which is a typical continuous numerical variable, so multiple linear regression equations were used. The data were entered into SPSS-26 for Mac, and a stepwise multiple regression analysis was used to establish the assessment model, to further illustrate the reasonableness of the established indexes through the overall fit of the model and the path coefficients of each index, to explore the quantitative relationship of the model between multiple variables of the athletes' risk factors for injuries and illnesses and to find out the independent variables that have the most close relationship with the dependent variables, so as to obtain the degree of influence of each risk factor on the overall risk system, and thus to formulate targeted risk avoidance strategies. We will find out the closest relationship with the dependent variable, get the degree of influence of each risk factor on the overall risk system, and then formulate targeted risk avoidance strategies. Athletes' risk of injury and illness involves a wide variety of factors with different degrees of importance and a wide range, first of all, set the first-level indicator as the dependent variable, the subordinate second-level indicator as its independent variable, each first-level indicator and the respective second-level indicator to do the regression equation, that is, to get four regression assessment equations. Similarly, each secondary categorical indicator, as the dependent variable, is regressed with the subordinate tertiary indicators to obtain 10 regression assessment equations.

## III. Results and analysis

### *III. A. Athlete injury risk indicators and regression analysis*

#### III. A. 1) Selection of injury risk indicators

Based on the idea of questionnaire design, this paper sets 4 primary indicators and 21 secondary indicators, as shown in Table 3.

#### III. A. 2) Regression analysis of aggregate risk

In the questionnaire survey on the overall risk, there are two aspects of “whether there has been a sports injury accident” and “the frequency of athletes' injuries and illnesses”, based on the results of the questionnaire, this paper regresses the overall risk by binary choice model and multiple linear regression model. In order to discuss the relationship between the four risks verified above and the occurrence of athletes' injuries and illnesses, a logistic model was established to regress the questionnaire results by using Stata 16.0 software, in which the athlete's own risk (X1), organization and management risk (X2), equipment and facilities risk (X3) and environmental risk (X4) were expressed by the cross-multiplier of the frequency and severity of injuries and illnesses through the combined probability and severity of the risk matrix method. The regression results are shown in Table 4. It was found that the higher the risks of athlete's own risk, organizational management risk, venue facility risk and environmental risk, the higher the probability of athlete's injury and illness.

Table 3: Selection of risk indicators for sports events

| Primary indicator | Secondary indicator |
|---|---|
| Athlete risk | Physical condition (disease, disability, etc.) |
| | Competition mentality and emotional control |
| | Emergency handling ability |
| | Ability to adapt to physical activity and exercise intensity |
| Organizational risk | Event staff |
| | Security risks such as security security and traffic density control |
| | Leadership ability |
| | Organizer risk prevention awareness |
| | Completeness of contingency plan |
| Facility risk | Traffic congestion, traffic inconvenience, etc |
| | Infrastructure completeness |
| | Temporary safety |
| | Layout rationality |
| | The improvement of medical and rescue equipment |
| Environmental risk | Terrorist behavior |
| | Sabotage risk |
| | Religious and ethnic issues |
| | Demonstration |
| | Outbreak effect |
| | Natural environmental risk |

Table 4: Logistic regression

| | Coefficient | Normal standard error | Z value | P value | Confidence interval of 95% |
|---|---|---|---|---|---|
| X1 | 0.814*** | 0.366 | 2.23 | 0.037 | 0.098,1.532 |
| X2 | 0.422*** | 0.224 | 2..04 | 0.046 | 0.012,0.851 |
| X3 | 0.326* | 0.183 | 1.86 | 0.009 | -0.024,0.685 |
| X4 | 0.284*** | 0.185 | 2.13 | 0.009 | 0.028,0.742 |
| Cons_ | 2.463* | 1.384 | 1.89 | 0.085 | -0.248,0.522 |

Further in order to discuss the association between the frequency of sports injuries and the four aspects of risk, this paper analyzes the issue from three perspectives: overall risk, frequency and severity, and regresses the questionnaire results by establishing a multiple linear regression model through Stata 16.0 software, and the regression results are shown in Table 5. The results found that the frequency of injuries and illnesses was significantly related to the risk of athletes themselves, the risk of organization and management, and the risk of venues and facilities, and the risk of athletes was the main causative factor for the occurrence of injuries and illnesses, and the environmental risk was not significant may be because most of the mass sports events are not sensitive to the changes in the natural environment, and most of the accidents occurred due to the injuries of the athletes.

Table 5: multiple regression results

| | X1 | X2 | X3 | X4 | Cons_ | n | R2 |
|---|---|---|---|---|---|---|---|
| (1) | 0.602***<br>(0.0064) | 0.146***<br>(0.0078) | 0.0704*<br>(0.0415) | 0.0495<br>(0.0072) | 0.425***<br>(0.0652) | 1.168 | 0.594 |

### III. B. Model Performance Evaluation

#### III. B. 1) Assessment of indicators

In the study, the dataset was randomly partitioned into a training set and a validation set at a ratio of 9:1, and 10-fold hierarchical cross-validation was used in the training set for model construction and model internal validity assessment, while the performance of the optimal model was validated using the model validation set to assess the external validity of the model. Given that the cost of missed diagnosis of injuries is much higher than misdiagnosis of injuries in injury detection, it is not appropriate to use only accuracy as a model evaluation metric. Confusion matrix is used to determine the corresponding evaluation criteria as shown in Table 6, in order to better evaluate the classification performance of the model, it is necessary to use the confusion matrix to determine the corresponding evaluation criteria. Precision rate, recall rate, F2 score and area under the ROC curve are chosen as the indicators of model performance.

Table 6: Confusion matrix

| | Forecast undamaged | Prediction injury |
|---|---|---|
| Actual undamaged | True negative (TN) | False positive (FP) |
| Actual injury | False negative (FN) | True positive (TP) |

(1) Accuracy rate: the ratio of injury samples correctly assessed by the model to the samples assessed as injured by the model, calculated as shown in equation (7).

$$Precision = \frac{TP}{TP + FP} \tag{7}$$

(2) Recall rate: the ratio of injury samples correctly evaluated by the model to samples that are actually injured, calculated as shown in (8).

$$Recall = \frac{TP}{TP + FN} \tag{8}$$

(3) F2 score: In practical assessment, precision and recall are mutually constrained. When the precision rate is high, the recall rate is often low. In order to comprehensively weigh the two indicators of precision rate and recall rate, the F2 score indicator was introduced in the study to reconcile Eq. (9). Since the cost of missed diagnosis of injuries and illnesses is much higher than that of misdiagnosis of injuries and illnesses, it is necessary to ensure the precision rate while maximizing the recall rate.The F2 score can fully reflect the overall assessment performance of the assessment model on both majority and minority samples:

$$F2 = \frac{5 \times Precision \times Recall}{4 \times Precision + Recall} \tag{9}$$

(4) Area under the ROC curve (AUC):

The horizontal and vertical axes of the ROC curve consist of the false positive rate and the true positive rate, respectively.The closer the ROC curve is to the upper left corner indicates the better performance of the classifier.However, when comparing different classification models, the direct use of the ROC curve does not allow a quantitative comparison of the classification performance of different classification models. Therefore, AUC was used as a quantitative index.

#### III. B. 2) Risk assessment performance analysis

A total of five athletes were included in the individualized injury risk assessment model based on the exclusion criteria, and a separate dataset was created for each athlete. Each data set contained 256.4±42.4 valid data samples per training session. The five athletes had a total of 25 non-contact injuries during the monitoring period. The basic data are shown in Table 7.

The variables in the optimal feature subset were incorporated into the individualized injury risk assessment model construction for each athlete, and the results of the 10-fold hierarchical cross-validation showed that, as shown in Table 8, the average precision, recall, F2 score, and AUC of the individualized injury risk assessment models for the five athletes were 0.852, 0.9714, 0.9174, and 0.9547, respectively.

Table 7: Data base

| Variable | Athlete 1 | Athlete 2 | Athlete 3 | Athlete 4 | Athlete 5 |
|---|---|---|---|---|---|
| Behavior | 45<br>(23.94) | 65<br>(23.12) | 35<br>(12.25) | 43<br>(15.26) | 55<br>(20.46) |
| sRPE | 512.2±460.2 | 578.2±450 | 460.3±422.8 | 598.3±490.5 | 600.8±442.8 |
| Training time | 82.5±74.8 | 95.2±74.1 | 90.4±77.5 | 97.3±75.4 | 104.5±73.8 |
| fatigue | 3.05±0.45 | 3.20±0.40 | 2.63±0.58 | 2.89±0.40 | 3.07±0.25 |
| Quality of sleep | 2.68±0.69 | 3.05±0.09 | 2.67±0.54 | 3.07±0.69 | 3.18±0.39 |
| Muscle soreness | 3.22±0.58 | 3.89±0.45 | 3.86±0.52 | 2.69±0.62 | 3.26±0.61 |
| Pressure level | 2.98±0.36 | 3.12±0.03 | 2.94±0.30 | 2.94±0.30 | 3.28±0.56 |
| Training initiative | 2.96±0.36 | 3.02±0.02 | 3.02±0.02 | 3.00±0.09 | 3.25±0.52 |
| Deep squat relative strength | 0.92±0.33 | 0.89±0.07 | 0.95±0.32 | 1.72±0.12 | 1.56±0.07 |
| Lie relative strength | 0.82±0.06 | 0.89±0.05 | 0.64±0.07 | 0.64±0.07 | 0.89±0.08 |
| The 5.8 meters six go back | 9.56±0.2 | 10.3±0.31 | 9.8±0.4 | 9.4±0.32 | 9.0±0.5 |
| 15 meters 17 to return | 66.3±0.9 | 72.5±2.2 | 67.5±0.9 | 67.5±0.5 | 67.2±0.4 |
| Running and jumping | 120.3±0.2 | 112.3±1.6 | 112.3±0.2 | 105.6±2.2 | 117.2±0.3 |

Table 8: Times hierarchical cross-validation of individual models

| Numbering | Accuracy rate | Recall rate | F2 score | AUC |
|---|---|---|---|---|
| Athlete 1 | 0.9425±0.1423 | 0.9982±0.0001 | 0.9820±0.0405 | 0.9924±0.0170 |
| Athlete 2 | 0.9256±0.2384 | 1.0000±0.0000 | 0.9782±0.1194 | 0.9945±0.0223 |
| Athlete 3 | 0.7284±0.3115 | 0.9748±0.1063 | 0.8625±0.1628 | 0.9184±0.1194 |
| Athlete 4 | 0.8849±0.2784 | 0.9994±0.0002 | 0.9587±0.1198 | 0.9958±0.0256 |
| Athlete 5 | 0.7786±0.3279 | 0.8846±0.2248 | 0.8057±0.1958 | 0.8725±0.1267 |

In order to illustrate the effectiveness of regression analysis modeling strategy in assessing individual injury risk, this study compared the 10-fold stratified cross-validation results of the individual model with the internal validity assessment knot of the optimal model. The results showed that the average precision, recall, F2 score and AUC of the model constructed based on the overall modeling strategy were 0.6664, 0.8951, 0.8350 and 0.9251, respectively.Compared with the model constructed based on the overall modeling strategy, the average precision, recall, F2 score and AUC of the model constructed based on the regression analysis constructed strategy were improved by 18.56%, 7.63%, 8.24% and 2.96%, respectively. This indicates that the model constructed based on the regression analysis modeling strategy has good precision and low leakage rate for the identification of injury and disease risk, and can effectively assess the risk of injury and disease of individual athletes.

# IV. Discussion

## *IV. A. Advantages of Modeling in Sports Rehabilitation*

The athlete risk assessment model based on regression analysis constructed in this study can explore the causal relationship between the risk of injury and disease, the cumulative amount of indicators and the amount of change, and utilize such causal relationship to further guide the training practice. It can provide useful reference for sports performance analysis, sports rehabilitation and other related researches to fully explore the scientific research value and practical value of big data, provide assistance for the implementation of "one person, one case" key scientific and technological protection for key athletes, and ultimately achieve the purpose of individualization of injury and disease early warning and personalization of sports training.

In addition, in the problem of injury risk assessment, the potentially ill population is often the focus of attention. If the constructed assessment model has high misdiagnosis and omission rates, the model is unable to recognize the potentially ill population and cannot meet the needs of practical applications. The individualized non-contact injury risk assessment model proposed in this study can effectively assess the risk of non-contact injuries in individual athletes, which has potential application in training practice and sports injury prevention.

## *IV. B. Model Decision Analysis*

The ultimate goal of injury risk assessment research in athletes is not simply to assess the occurrence of sports injuries, but to identify injury risk at the individual level and to adapt interventions to reduce injury risk. However, under what circumstances are athletes susceptible to non-contact injuries? How to make adjustments for specific situations? These are the two main questions facing the practice of sports injury prevention. The shift from empirical to data-driven injury prevention strategies requires not only screening individual risk factors for non-contact injuries, but also providing specific decision-making information to assist coaches, researchers, and team physicians in making quantitative adjustments to athletes. Since the present study is to assess whether an injury occurs in the sliding window, when the model evaluates a positive value, it suggests that the athlete has a potential risk of injury or illness. At this time, the analysis of the decision-making process of the model can help coaches, researchers and team doctors to understand the specific injury risk factors and the decision-making basis of the model, and then make timely adjustments to the athletes to avoid injuries and illnesses.

## V. Conclusion

In this paper, we summarize the risk factors affecting athletes' injuries and illnesses through questionnaire survey, expert interview, questionnaire survey, mathematical statistics, etc., and establish the risk indicator system of athletes' injuries and illnesses. Regression analysis was used to comprehensively identify and accurately assess the risk indicators, and logistic regression analysis showed that, based on the results of the survey and analysis, the indicator system included four primary indicators, namely, athletes' own risk, organization and management risk, equipment and facilities risk, and environmental risk, as well as 21 secondary indicators. The frequency of athletes' injuries and illnesses was significantly correlated with athletes' own risk, organizational management risk and facility risk, and athletes' own risk was the main cause of injuries and illnesses. In the performance analysis of the model risk assessment, the regression-based assessment model had good accuracy and a low leakage rate.

## References


[1] Barranco-Ruiz, Y., Villa-González, E., Martínez-Amat, A., & Da Silva-Grigoletto, M. E. (2020). Prevalence of injuries in exercise programs based on Crossfit®, cross training and high-intensity functional training methodologies: a systematic review. Journal of human kinetics, 73, 251.

[2] Kalkhoven, J. T., Watsford, M. L., & Impellizzeri, F. M. (2020). A conceptual model and detailed framework for stress-related, strain-related, and overuse athletic injury. Journal of science and medicine in sport, 23(8), 726-734.

[3] Jayanthi, N. A., LaBella, C. R., Fischer, D., Pasulka, J., & Dugas, L. R. (2015). Sports-specialized intensive training and the risk of injury in young athletes: a clinical case-control study. The American journal of sports medicine, 43(4), 794-801.

[4] Dudley, C., Johnston, R., Jones, B., Till, K., Westbrook, H., & Weakley, J. (2023). Methods of monitoring internal and external loads and their relationships with physical qualities, injury, or illness in adolescent athletes: A systematic review and best-evidence synthesis. Sports Medicine, 53(8), 1559-1593.

[5] Parish, M. R. (2020). On determining factors affecting injury and recovery in athletes. Health, sport, rehabilitation, 6(3), 26-33.

[6] Bestwick-Stevenson, T., Toone, R., Neupert, E., Edwards, K., & Kluzek, S. (2022). Assessment of fatigue and recovery in sport: narrative review. International journal of sports medicine, 43(14), 1151-1162.

[7] Kelly, S., Pollock, N., Polglass, G., & Clarsen, B. (2022). Injury and illness in elite athletics: a prospective cohort study over three seasons. International journal of sports physical therapy, 17(3), 420.

[8] Soligard, T., Palmer, D., Steffen, K., Lopes, A. D., Grant, M. E., Kim, D., ... & Engebretsen, L. (2019). Sports injury and illness incidence in the PyeongChang 2018 Olympic Winter Games: a prospective study of 2914 athletes from 92 countries. British Journal of Sports Medicine, 53(17), 1085-1092.

[9] Mosler, A. B., Weir, A., Eirale, C., Farooq, A., Thorborg, K., Whiteley, R. J., ... & Crossley, K. M. (2018). Epidemiology of time loss groin injuries in a men’s professional football league: a 2-year prospective study of 17 clubs and 606 players. British journal of sports medicine, 52(5), 292-297.

[10] Ruddy, J. D., Cormack, S. J., Whiteley, R., Williams, M. D., Timmins, R. G., & Opar, D. A. (2019). Modeling the risk of team sport injuries: a narrative review of different statistical approaches. Frontiers in physiology, 10, 829.

[11] Tenforde, A. S., Carlson, J. L., Chang, A., Sainani, K. L., Shultz, R., Kim, J. H., ... & Fredericson, M. (2017). Association of the female athlete triad risk assessment stratification to the development of bone stress injuries in collegiate athletes. The American journal of sports medicine, 45(2), 302-310.

[12] Shanley, E., Thigpen, C. A., Collins, G. S., Arden, N. K., Noonan, T. J., Wyland, D. J., ... & Bullock, G. S. (2022). Including modifiable and nonmodifiable factors improves injury risk assessment in professional baseball pitchers. journal of orthopaedic & sports physical therapy, 52(9), 630-640.

[13] Dorrel, B., Long, T., Shaffer, S., & Myer, G. D. (2018). The functional movement screen as a predictor of injury in national collegiate athletic association division II athletes. Journal of athletic training, 53(1), 29-34.

[14] Moore, E., Chalmers, S., Milanese, S., & Fuller, J. T. (2019). Factors influencing the relationship between the functional movement screen and injury risk in sporting populations: a systematic review and meta-analysis. Sports Medicine, 49, 1449-1463.

[15] Bond, C. W., Dorman, J. C., Odney, T. O., Roggenbuck, S. J., Young, S. W., & Munce, T. A. (2019). Evaluation of the functional movement screen and a novel basketball mobility test as an injury prediction tool for collegiate basketball players. The Journal of Strength & Conditioning Research, 33(6), 1589-1600.

[16] Bennett, H., Chalmers, S., Milanese, S., & Fuller, J. (2022). The association between Y-balance test scores, injury, and physical performance in elite adolescent Australian footballers. Journal of science and medicine in sport, 25(4), 306-311.

[17] Plisky, P., Schwartkopf-Phifer, K., Huebner, B., Garner, M. B., & Bullock, G. (2021). Systematic review and meta-analysis of the Y-balance test lower quarter: reliability, discriminant validity, and predictive validity. International journal of sports physical therapy, 16(5), 1190.

[18] Greenberg, E. T., Barle, M., Glassman, E., Jacob, L., Jaafar, H., Johnson, A., ... & Jung, M. K. (2019). Reliability and stability of the Y Balance Test in healthy early adolescent female athletes. Orthopaedic Journal of Sports Medicine, 7(3_suppl), 2325967119S00051.

[19] Bauer, J., Panzer, S., Gruber, M., & Muehlbauer, T. (2023). Associations between upper quarter Y-balance test performance and sport-related injuries in adolescent handball players. Frontiers in sports and active living, 5, 1076373.

[20] Alexander Havertz,David Uebis,Rudolph Schifflers,Frank Hildebrand & Christian David Weber. (2025) .Sports injury risk assessment based on a training and functional movement analysis of young elite equestrian athletes– an exploratory cross-sectional study. BMC Sports Science, Medicine and Rehabilitation,17(1),83-83.

[21] Gu Chengzhen,Reefke Hendrik & Yates Nicola. (2025) .Autonomous vehicle adoption and supply chain social sustainability: Delphi study and expert interviews. International Journal of Physical Distribution & Logistics Management,55(3),275-306.

[22] Shuai Changhao & Zhang Hongtao. (2025) .The Path Exploration of Ideological and Political Education in Probability Theory and Mathematical Statistics Courses under the Background of “All-staff Education”. Education Reform and Development,7(3),277-282.

[23] Ahmad Ameri,Zohreh Azma,Khashayar Fattah,Fereshteh Talebi,Pooya Ameri,Nazanin Rahnama... & Farzad Taghizadeh Hesary. (2025) .Clinical and Dosimetric Predictors of Early Onset Postradiation Hypothyroidism in Patients with Head and Neck Malignancies: A Logistic Regression Analysis. Oncology and Therapy,(prepublish),1-17.

[24] Hiwot Altaye Asebe,Beminate Lemma Seifu,Kusse Urmale Mare,Bizunesh Fantahun Kase,Tsion Mulat Tebeje,Yordanose Sisay Asgedom & Zufan Alamrie Asmare. (2024) .The magnitude of stunting and its determinants among late adolescent girls in East Africa: Multilevel binary logistics regression analysis. PloS one,19(5),e0298062-e0298062.