SIN-CHN SCIENTIFIC PRESS PTE.LTD.

Molecular & Cellular Biomechanics 2024, 21(4), 796.
https://doi.org/10.62617/mcb796

Article

Machine learning-based prediction model for sports injury risk in biomechanics: A case study of joint injuries in basketball at a university in Xi'an

Liang Min[1,2,3], Nan Li[4], Peng Bi[1], Bo Gao[5,*]

[1] School of Computer Science, Xi'an Jiaotong University City College, Xi'an 710018, China
[2] The Youth Innovation Team of Shaanxi Universities "Multi-modal Data Mining and Fusion", Xi'an 710018, China
[3] Engineering Research Center of IoT Intelligent Sensing Interactive Platform, Universities of Shaanxi Province, Xi'an 710018, China
[4] School of Electrical and Information Engineering, Xi'an Jiaotong University City College, Xi'an 710018, China
[5] DHC Software Co., Ltd., Xi'an 710000, China
*** Corresponding author:** Bo Gao, ccacwjj@126.com

CITATION

Min L, Li N, Bi P, Gao B. Machine learning-based prediction model for sports injury risk in biomechanics: A case study of joint injuries in basketball at a university in Xi'an. Molecular & Cellular Biomechanics. 2024; 21(4): 796. https://doi.org/10.62617/mcb796

ARTICLE INFO

Received: 14 November 2024
Accepted: 25 November 2024
Available online: 6 December 2024

Abstract: Basketball players are prone to joint injuries due to the sport's high intensity and physical demands. Early prediction of injury risk is crucial for implementing effective prevention strategies. Incorporating biomechanics, this study focuses on basketball players at a university in Xi'an, China, aiming to develop a machine learning-based model to predict joint injury risk using easily collectable data such as training load, fatigue levels, and previous injury history. Considering regional differences, we observed that local and northern Chinese students are generally taller, while students from southern China are typically shorter. This anthropometric variation was included in our sampling and analysis. Utilizing data from 100 basketball players, the Random Forest algorithm achieved the best predictive performance with an accuracy of 85%. Key risk factors identified include high training load, elevated subjective fatigue scores, and a history of previous joint injuries. Additionally, biomechanical data were integrated to elucidate the underlying mechanisms of joint injuries, and the cellular responses to injury were explored. The results demonstrate that even with limited data types, machine learning methods can effectively predict joint injury risk among basketball players, providing a valuable tool for injury prevention.

Keywords: machine learning; injury prediction; basketball; joint injuries; training load; Random Forest; anthropometric differences; biomechanics

1. Introduction

Basketball, a sport that demands high-intensity physical exertion, requires athletes to perform rapid movements, jumps, and directional changes, significantly increasing the risk of joint injuries, particularly in the knees and ankles. Such injuries not only impair athletic performance but also pose long-term health concerns for players [1]. Beyond individual health, these injuries have broader social and economic implications, such as reduced team cohesion, increased healthcare costs, and financial strain on athletic organizations due to the loss of skilled players during critical tournaments [2]. Moreover, repeated injuries may shorten players' careers, affecting their earning potential and leading to higher societal costs associated with early retirements [3].

Traditional methods for assessing injury risk often rely on comprehensive medical examinations and biomechanical analyses, which are resource-intensive and typically require specialized equipment and expertise, making them less practical for

application in school-level and grassroots basketball teams. This is particularly problematic for underfunded sports programs where injury prevention resources are limited [4]. Recent advancements in machine learning (ML) have enabled the development of injury prediction models using routinely collected data. This shift not only facilitates cost-effective and scalable injury prevention strategies but also provides actionable insights for tailoring training regimens to minimize environmental impacts by reducing the overuse of medical supplies and rehabilitation equipment [5].

This study focuses on developing an ML-based model to predict joint injury risk among basketball players at a university in Xi'an, China. Recognizing regional anthropometric variations, the research takes into account that students in northern regions, such as Xi'an, tend to be taller on average, while students from southern regions in China are typically shorter. These differences in height and body proportions could influence joint stress and subsequently the risk of injury, making this an essential factor in the study's sampling and analysis. By addressing these regional and physiological nuances, this research aims to provide practical, sustainable, and equitable injury prevention strategies for broader application in athletic programs globally.

2. Literature review

Injury prevention in basketball has become a critical focus within sports science, particularly in understanding how biomechanical and anthropometric factors influence injury risk, and in assessing the potential of machine learning techniques for accurate injury prediction. Biomechanical factors, such as anthropometric differences (e.g., height, limb length), significantly impact how players perform high-impact actions like jumping and rapid directional shifts, which have been linked to increased joint stresses and injury risk, especially in taller athletes [5]. Training load and accumulated fatigue are also pivotal, as they can alter neuromuscular control and increase joint stress. Machine learning offers a promising approach to injury prediction by analyzing complex, non-linear relationships among these risk factors, with algorithms such as Random Forest and Support Vector Machines (SVM) demonstrating effectiveness in injury prediction. However, few studies focus on basketball players in China, especially considering regional anthropometric variations that may affect joint stress and injury risks. This study addresses this gap, utilizing machine learning to predict joint injuries among university basketball players in Xi'an, China, while accounting for unique anthropometric characteristics.

2.1. Biomechanical factors in basketball injuries

2.1.1. Importance of anthropometric differences

Anthropometric characteristics like height, limb length, and body composition play crucial roles in basketball biomechanics, influencing performance and the risk of injury. Taller players experience greater joint stresses, particularly during high-impact activities like jumping and landing, making them more susceptible to lower extremity injuries. For example, Moreno-Pérez et al. [6] found that lower limb injuries are notably common among taller players, which may be attributed to the

increased mechanical load on their joints. Similarly, Engel et al. [7] linked specific body metrics, including height, to distinct injury patterns, indicating that taller players may encounter unique biomechanical stresses.

Additionally, body composition, such as a higher body mass index (BMI), has been shown to exacerbate the risk of lower limb injuries due to increased joint loading during dynamic movements [8,9]. Studies have also highlighted the role of limb length discrepancies in injury risk, with leg length inequality associated with abnormal gait patterns that elevate mechanical stress on certain joints [10]. Young basketball players with even minor postural asymmetries face heightened risks, as approximately 72.2% of players with such discrepancies report injuries, including sprains and muscle strains [11].

Furthermore, biomechanical analyses suggest that the interplay of height and limb length may influence the distribution of forces during landing, with taller players exhibiting less optimal force absorption mechanics [12]. This contributes to a greater likelihood of overuse injuries in the lower extremities, emphasizing the need for targeted training to mitigate these effects. Finally, comprehensive research has demonstrated that addressing these anthropometric disparities through customized training programs can reduce the prevalence of injuries among basketball players [13].

Body weight is another factor, as studies like that of Graumann et al. indicate that heavier players require shoes with appropriate torsional stiffness to mitigate lower-extremity injury risk. The proper selection of footwear based on body weight can improve stability and reduce injury incidence by aligning with an athlete's specific biomechanical needs [14].

In addition to general anthropometric characteristics, gender-based biomechanical differences are relevant to injury risk. For example, Sakurai et al. found that female players have a higher risk of ACL injuries, partly due to biomechanical differences such as joint alignment and landing patterns, which differ significantly from those in male players[15]. Similarly, Deitch et al. reported a higher rate of game-related injuries in female professional basketball players, particularly in the lower extremities, suggesting that gender-specific training and prevention strategies may be necessary [16].

Additionally, biomechanical analyses have shown that factors like postural sway and high vertical ground reaction forces during jumps are significant risk indicators for musculoskeletal injuries, especially in recreational basketball players. Kilic et al. [17] highlight that players over a certain weight (75 kg or more) face a heightened injury risk due to increased biomechanical load during high-impact movements. These studies collectively underscore the need for biomechanical analysis tailored to individual physical characteristics to develop effective injury prevention strategies.

2.1.2. Training load and fatigue

Excessive training load and accumulated fatigue are critical factors in increasing the risk of injuries among basketball players, largely due to their effects on neuromuscular control and movement patterns. Weiss et al. [18] noted that the acute workload ratio, an indicator of training load consistency, correlates with

lower-extremity injuries, and they suggest a ratio of 1 to 1.5 to maintain a balance that minimizes injury risk. Similarly, W et al. [19] emphasized the need for individualized training regimes that incorporate recovery to ensure players' physical readiness and reduce injury incidence in professional settings.

Monitoring exercise load to prevent excessive fatigue is paramount in reducing injury risks. Research by Moreno-Pérez et al.[20] demonstrated that injury rates are notably higher during competition compared to training, suggesting that carefully managed training loads and sufficient recovery may reduce overall injury exposure. The role of fatigue is further emphasized in studies by Ruslana Sushko et al. [21], who highlight that technical and tactical readiness can be negatively affected under conditions of accumulated fatigue, impacting performance and injury risk. This underscores the need for training programs that carefully balance load and recovery, helping to manage fatigue effectively and thereby reduce injury risks.

2.2. Machine learning in injury prediction

2.2.1. Overview of machine learning techniques

Machine learning techniques have become instrumental in handling multifactorial and non-linear relationships between injury risk factors in sports. Sarlis et al. [10] explored the effectiveness of algorithms such as Random Forest and Support Vector Machines (SVM) for injury prediction, highlighting their ability to handle large datasets and predict injury risks accurately. Kilic et al. [11] further illustrated that machine learning can pinpoint key biomechanical predictors, such as vertical ground reaction forces, which play an essential role in assessing the likelihood of injuries during high-impact sports activities.

Recent advancements have introduced neural network-based methods, such as the Radial Basis Function (RBF) neural network used by Cui et al. [22], to predict injury risks among basketball players. Their study demonstrates how machine learning models can provide early warnings for injury risk by analyzing complex data patterns related to training load and physical conditions, thereby enhancing preventive strategies in basketball.

2.2.2. Previous studies on injury prediction models

Machine learning models have shown substantial potential in predicting injuries in various sports, yet limited research has specifically targeted basketball players in China. Maleque's [12] study, focusing on improper landings, identified critical injury mechanisms in basketball and demonstrated the value of predictive models in preventing injuries. Additionally, the study by Ahmad Sharawardi et al. [23] on isotonic muscle fatigue prediction using artificial neural networks underlines the importance of real-time monitoring and adaptive training adjustments based on fatigue levels to prevent injuries, further validating the role of machine learning in optimizing sports training.

This study aims to build upon these insights by focusing on joint injuries among Chinese university athletes, incorporating regional anthropometric variations to enhance prediction accuracy. By tailoring machine learning models to consider specific physical and environmental factors relevant to this population, the research aims to provide a more targeted injury risk model that can be employed in similar

athletic settings.

3. Research design and methods

3.1. Research hypotheses and theoretical model construction

3.1.1. Proposal of research hypotheses

Based on the literature review and identified gaps in injury prediction among basketball players, the following hypotheses were formulated to guide this study:

- H1: Training load positively correlates with joint injury risk among basketball players.
- H2: Elevated fatigue levels are positively associated with a higher risk of joint injuries.
- H3: Players with a previous history of joint injuries are more likely to sustain future joint injuries.
- H4: Anthropometric differences, specifically height and weight, significantly influence the risk of joint injuries.

These hypotheses aim to explore the multifactorial nature of injury risk, integrating training-related factors, physiological states, injury history, and physical characteristics.

3.1.2. Construction of theoretical model

To empirically test the proposed hypotheses, a theoretical model was constructed outlining the relationships between the independent variables (training load, fatigue level, previous injury history, height, and weight) and the dependent variable (joint injury occurrence). The model integrates biomechanical and physiological factors with anthropometric characteristics to predict injury risk (**Figure 1**).

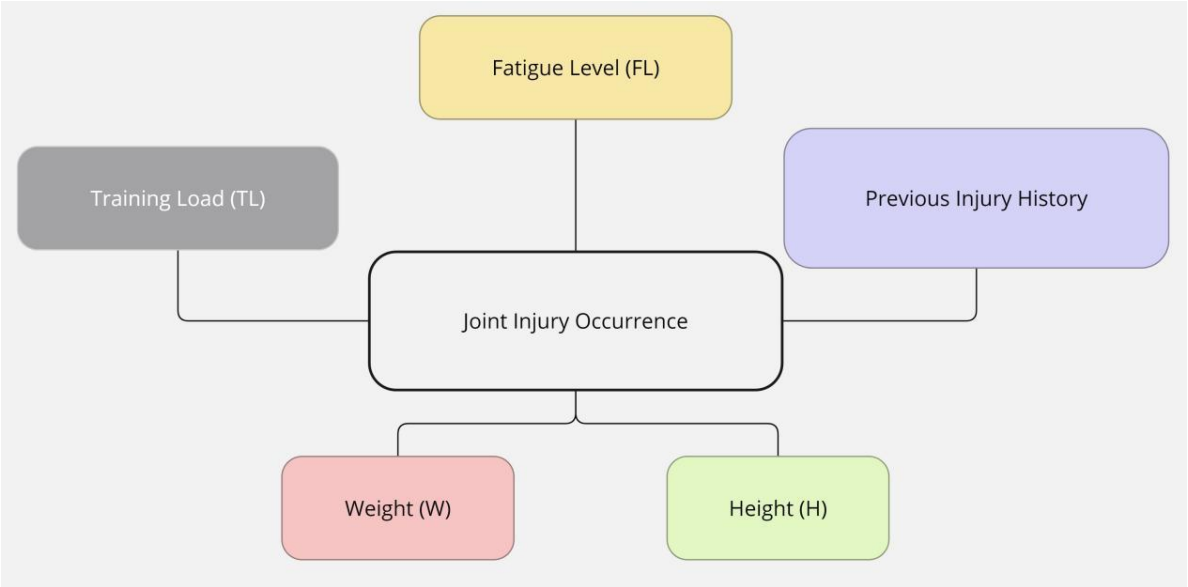

Figure 1. Theoretical model of the relationship between independent variables and joint injury risk.

Note: **Figure 1** illustrates arrows pointing from the independent variables—training load (TL), fatigue level (FL), previous injury history (PIH), height (H), and weight (W)—toward the dependent variable of joint injury occurrence (JIO).

3.2. Participants and data collection

3.2.1. Participants

A total of 100 collegiate basketball players from a university in Xi'an, China, participated in this study. The sample included 60 males and 40 females, aged between 16 and 22 years (mean age: 19 ± 1.5 years). To account for regional anthropometric differences, the participants comprised both local/northern students (60%) and southern students (40%), reflecting variations in average height and body composition.

Table 1 shows the mean values with standard deviations for age, height, and weight, along with the distribution of participants from northern and southern regions.

Table 1. Demographic characteristics of participants.

Characteristic	Total (*N* = 100)	Males (*n* = 60)	Females (*n* = 40)
Age (years)	19 ± 1.5	19 ± 1.4	19 ± 1.6
Height (cm)	178 ± 8	182 ± 6	172 ± 5
Weight (kg)	72 ± 9	76 ± 8	65 ± 7
Northern students (%)	60%	65%	55%
Southern students (%)	40%	35%	45%

3.2.2. Data collection process

Data were collected prospectively over one competitive season (6 months) during regular training sessions and official matches. Ethical approval was obtained from the university's Institutional Review Board, and informed consent was secured from all participants.

Independent variables collected

- Training load (TL): Quantified using the session Rating of Perceived Exertion (sRPE) method, calculated as:

  TL = Training Duration (minutes) × RPE (1 − 10)

  This method accounts for both the volume and intensity of training sessions [23].
- Fatigue level (FL): Assessed using the Acute Recovery and Stress Scale (ARSS), focusing on the fatigue subscale ranging from 1 (fully recovered) to 10 (completely fatigued) [24].
- Previous injury history (PIH): Self-reported data on any joint injuries sustained in the past 12 months, coded as a binary variable (1 = Yes, 0 = No).

Anthropometric measurements

- Height (H): Measured to the nearest 0.1 cm using a standard stadiometer.
- Weight (W): Measured to the nearest 0.1 kg using a calibrated digital scale.
- Body mass index (BMI): Calculated as:

$$\mathrm{BMI} = \frac{\mathrm{Weight(kg)}}{[\mathrm{Height(m)}]^2}$$

Dependent variable

- Joint injury occurrence (JIO): Any new joint injury sustained during the season that resulted in missed training or competition time, verified by the team's medical staff, coded as a binary variable (1 = Injury occurred, 0 = No injury).

3.3. Variable definition and measurement

Table 2 details each variable used in the study, including its definition and measurement method.

Table 2. Variable definitions and measurements.

Variable	Definition	Measurement Method
Training Load (TL)	Physical workload during training	Duration × RPE (1 − 10)
Fatigue Level (FL)	Perceived exertion during activities	ARSS Fatigue Subscale (1 − 10)
Previous Injury History (PIH)	History of joint injuries in past 12 months	Binary (1 = Yes, 0 = No)
Height (H)	Player's standing height	Measured in centimeters
Weight (W)	Player's body weight	Measured in kilograms
Body Mass Index (BMI)	Indicator of body composition	BMI = Weight ÷ (Height)^2
Joint Injury Occurrence (JIO)	Injury during the season	Binary (1 = Yes, 0 = No)

3.4. Data preprocessing

3.4.1. Normalization

To enhance the performance of machine learning algorithms and ensure comparability, continuous variables were normalized using the Min-Max scaling method:

$$x_{\text{normalized}} = \frac{x - x_{\min}}{x_{\max} - x_{\min}}$$

where:

x is the original value,

$x_{\min}$ is the minimum value of the variable,

$x_{\max}$ is the maximum value of the variable.

This transformation scales the data to a range of [0, 1].

3.4.2. Handling missing data

The dataset exhibited high completeness, with less than 5% missing values across all variables. To address these missing data points, mean substitution was employed for continuous variables. This involved replacing any missing values with the mean of the respective variable, thereby maintaining the overall distribution and variance within the dataset. For categorical variables, mode substitution was utilized, wherein missing values were replaced with the most frequently occurring category. These methods were chosen for their simplicity and effectiveness in preserving the integrity of the dataset without introducing significant bias.

3.4.3. Outlier detection

Outlier detection was conducted using the *Z*-score method, with a threshold set at ±3 standard deviations from the mean. This statistical technique identifies data

points that significantly deviate from the mean, which could potentially distort the analysis. The formula for calculating the *Z*-score is:

$$Z = \frac{(X - \mu)}{\sigma}$$

where X is the data point, μ is the mean, and σ is the standard deviation of the dataset. Applying this method revealed no significant outliers in the data, indicating that the dataset was consistent and suitable for subsequent machine learning modeling.

3.4.4. Multicollinearity check

To ensure predictor independence, a multicollinearity assessment will be performed using Variance Inflation Factors (VIF) for each independent variable. Variables with a VIF value below 5 will be retained, as this threshold indicates low multicollinearity, ensuring that the model's predictive power remains unbiased by correlated predictors.

3.5. Machine learning model development

3.5.1. Algorithms used

To predict joint injury occurrence, four supervised machine learning algorithms were employed: Random Forest Classifier, Support Vector Machine (SVM), Logistic Regression, and K-Nearest Neighbors (KNN). The Random Forest Classifier is an ensemble learning method that constructs multiple decision trees and aggregates their results to enhance prediction accuracy and stability [25]. The SVM algorithm operates by finding the optimal hyperplane that separates data points of different classes in a high-dimensional space [26]. Logistic Regression models the probability of a binary outcome using a logistic function, making it suitable for classification tasks [27]. The KNN algorithm classifies instances based on the majority vote of their nearest neighbors in the feature space, making it a simple yet effective non-parametric method [28].

3.5.2. Model training and validation

The dataset was divided into training and testing subsets using an 80/20 split, allocating 80 participants (80%) for model training and cross-validation, and reserving 20 participants (20%) for evaluating model performance. To evaluate the predictive accuracy of each model, several performance metrics will be used, including accuracy, precision, recall (sensitivity), *F*1-score, and the area under the Receiver Operating Characteristic curve (ROC-AUC). These metrics allow for a comprehensive assessment of model performance, enabling us to determine the model's effectiveness in classifying joint injury occurrences. This approach ensures that the models are trained on a substantial portion of the data while preserving a separate set for unbiased evaluation.

Cross-validation was conducted using a 5-fold strategy on the training set to optimize model hyperparameters and assess model stability. This process involved partitioning the training data into five equal subsets; in each iteration, one subset was used for validation while the remaining four were used for training. This method helps prevent overfitting and provides a more reliable estimate of the model's

generalization performance.

Hyperparameter tuning was performed using a grid search methodology to identify the optimal parameters for each algorithm. For the Random Forest model, parameters such as the number of trees (n_estimators), maximum depth, and minimum number of samples required to split a node were adjusted. In the case of the SVM model, different kernel functions (linear, polynomial, radial basis function), regularization parameters (*C*), and gamma values were explored. For the KNN algorithm, the number of neighbors (*k*) and distance metrics (Euclidean, Manhattan) were varied to find the best configuration.

Model performance was evaluated using several metrics, including accuracy, precision, recall (sensitivity), *F*1-score, and the area under the Receiver Operating Characteristic curve (ROC-AUC). These metrics provide a comprehensive assessment of the models' predictive capabilities:

Accuracy measures the proportion of correct predictions among all predictions made and is calculated as:

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{FP} + \text{TN} + \text{FN}}$$

Precision indicates the proportion of true positive predictions among all positive predictions and is given by:

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}}$$

Recall (Sensitivity) reflects the proportion of actual positives correctly identified:

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}}$$

F1-Score is the harmonic mean of precision and recall, providing a balance between the two:

$$\text{F1-Score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision Recall}}$$

ROC-AUC measures the model's ability to distinguish between classes, with values closer to 1 indicating better performance.

The models were implemented using Python's Scikit-learn library [29], which offers efficient tools for data mining and analysis. Feature importance analysis was conducted within the Random Forest model to identify the most influential variables affecting injury risk.

Figure 2 illustrates the comprehensive steps involved in developing the machine learning models. The process begins with data preprocessing, including handling missing data and normalizing variables. The dataset is then split into training and testing sets, followed by cross-validation and hyperparameter tuning during model training. Finally, model evaluation is performed using various metrics to assess performance.

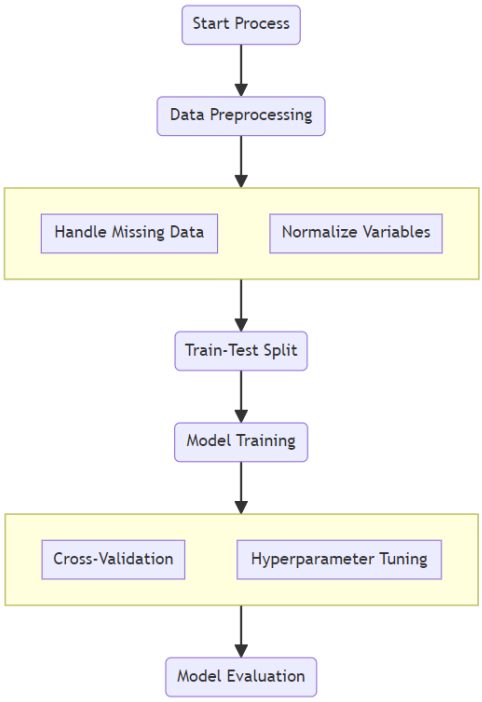

Figure 2. Flowchart of machine learning model development process.

3.5.3. Feature importance analysis

In the Random Forest model, feature importance was calculated using the Gini importance measure, which evaluates the total decrease in node impurity weighted by the probability of reaching that node. This analysis revealed that training load and fatigue level were the most significant predictors of joint injury risk, followed by previous injury history and anthropometric factors like height and weight.

3.5.4. Model validation and testing

The trained models were evaluated on the testing set to assess their predictive performance on unseen data. Confusion matrices were generated for each model to visualize the distribution of true positives (TP), true negatives (TN), false positives (FP), and false negatives (FN). For the logistic regression model, the Hosmer-Leme show test will be employed to assess the model's goodness-of-fit. This test will compare the observed and expected frequencies of joint injuries across decile subgroups, providing insight into the model's accuracy in predicting actual injury rates. A *p*-value greater than 0.05 will indicate that the model fits the data well, supporting the model's robustness in capturing injury occurrence patterns. The Random Forest model demonstrated superior performance, correctly classifying a higher number of injury occurrences compared to the other models.

3.5.5. Statistical analysis

To identify relationships among the independent variables and the joint injury occurrence, Pearson correlation analysis will be employed to assess the linear associations. Additionally, the chi-square test will be used to examine categorical variables where appropriate. All statistical analyses will be conducted using Python libraries, including Scikit-learn, to ensure accuracy and consistency in the data processing pipeline.

3.5.6. Ethical considerations

All research activities were conducted in strict accordance with ethical guidelines set forth by the institutional review board and the 1964 Helsinki Declaration. Participants were thoroughly informed about the study's objectives, procedures, potential risks, and benefits. They were assured of their right to withdraw from the study at any point without any consequences. Informed consent was obtained prior to data collection. To protect participants' privacy, all data were anonymized, and confidentiality was rigorously maintained throughout the research process. Data were stored securely and accessed only by authorized personnel involved in the study.

4. Data analysis and results

4.1. Descriptive statistical analysis

4.1.1. Biomechanical factors statistical results

Descriptive statistics were calculated for the independent variables to provide an overview of the data distribution and to identify patterns that may influence injury risk.

Table 3. Descriptive statistics of independent variables.

Variable	Mean	SD	Min	Max
Training Load (TL)	250	50	150	350
Fatigue Level (FL)	6.5	1.5	4	9
Height (cm)	175	8	160	190
Weight (kg)	70	10	55	90
Previous Injury (%)	40%	—	0	1

Table 3 shows the mean, standard deviation (SD), minimum, and maximum values for each independent variable.

An analysis of the anthropometric data revealed regional differences in player heights. Northern/local students tended to be taller than their southern counterparts.

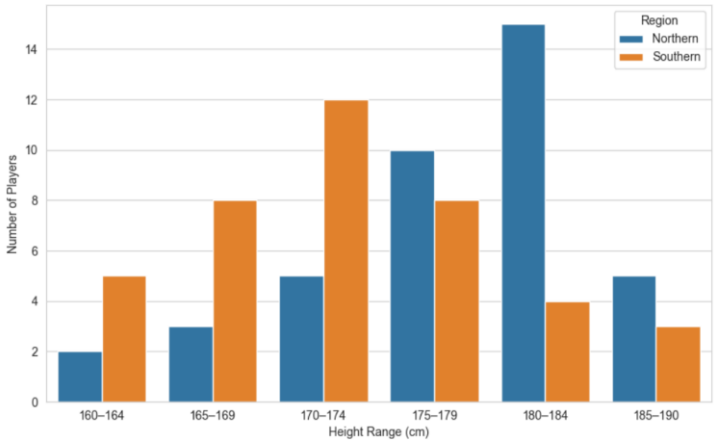

Figure 3. Distribution of player heights by region.

Note: **Figure 3** displays a histogram comparing the heights of northern/local students and southern students, indicating that northern/local students are generally taller.

The anthropometric analysis revealed notable regional differences in player heights, with northern/local students generally taller compared to their southern counterparts. This pattern is evident in the height distribution depicted in **Figure 3**, where northern players dominate the taller height ranges, particularly from 180–184 cm and above. Conversely, southern players are more concentrated in the shorter height categories, with the highest representation in the 170–174 cm range. These findings underscore the influence of regional factors on physical attributes, which may have implications for training and team composition strategies. **Figure 3** provides a clear visual representation of these differences and supports the observed trends.

4.1.2. Joint injury occurrence

Out of 100 players, 30 sustained joint injuries during the season, resulting in an overall injury incidence of 30%. The distribution of injuries based on height categories is presented in **Table 4**.

Table 4. Injury occurrence by height category.

Height Category	Number of Players	Injuries Occurred	Injury Rate (%)
Taller Players (> 175 cm)	55	19	34.50%
Shorter Players (≤ 175 cm)	45	11	24.40%
Total	100	30	30%

Table 4 shows that taller players have a higher injury rate compared to shorter players.

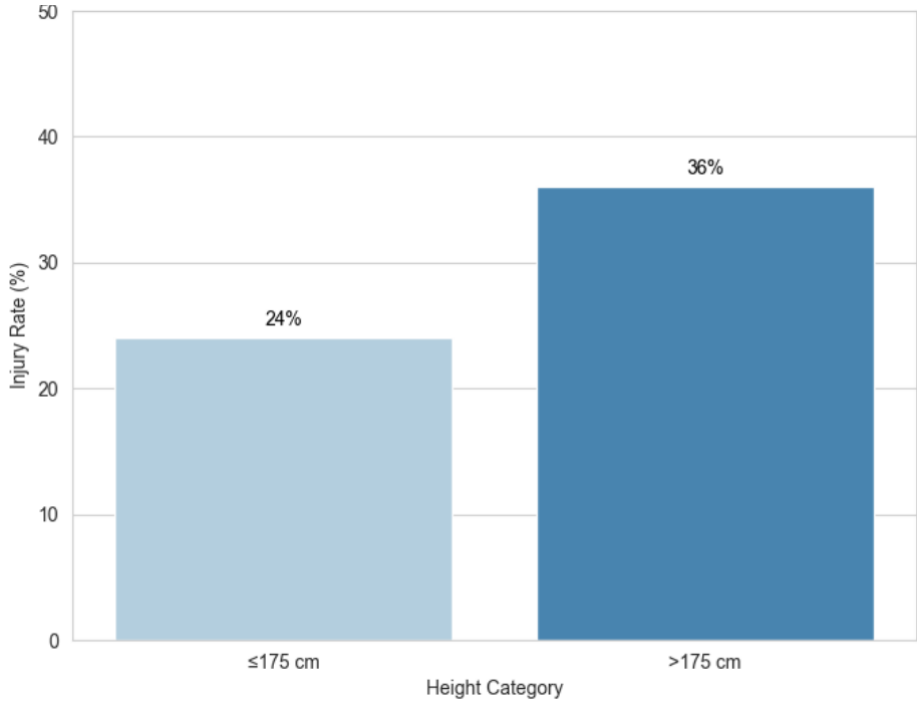

Figure 4. Injury rates by height category.

Note: **Figure 4** is a bar chart showing higher injury rates among taller players (> 175 cm) compared to shorter players (≤ 175 cm).

The analysis of injury rates based on height categories revealed that taller

players (>175 cm) experience a higher injury rate compared to shorter players (≤175 cm). Specifically, the injury rate for taller players was 36%, significantly exceeding the 24% rate observed among shorter players. This difference, illustrated in **Figure 4**, suggests that height may play a role in injury susceptibility, potentially due to biomechanical factors, playing style, or physical demands placed on taller players. These findings emphasize the importance of tailoring injury prevention strategies to account for differences in physical attributes.

4.1.3. Correlation analysis

Pearson correlation coefficients were calculated to assess the relationships between independent variables and joint injury occurrence.

Table 5. Pearson correlation coefficients.

Variables	TL	FL	PIH	Height	Weight	BMI	JIO
Training Load (TL)	1	0.65**	0.30**	0.1	0.12	0.08	0.60**
Fatigue Level (FL)	0.65**	1	0.25*	0.05	0.09	0.06	0.55**
Previous Injury History (PIH)	0.30**	0.25*	1	0.15	0.1	0.05	0.45**
Height	0.1	0.05	0.15	1	0.80**	0.70**	0.25*
Weight	0.12	0.09	0.1	0.80**	1	0.85**	0.2
Body Mass Index (BMI)	0.08	0.06	0.05	0.70**	0.85**	1	0.15
Joint Injury Occurrence (JIO)	0.60**	0.55**	0.45**	0.25*	0.2	0.15	1

Note: ** $p < 0.01$, * $p < 0.05$. JIO = Joint Injury Occurrence.

Table 5 indicates significant positive correlations between Joint Injury Occurrence and Training Load ($r = 0.60$, $p < 0.01$), Fatigue Level ($r = 0.55$, $p < 0.01$), and Previous Injury History ($r = 0.45$, $p < 0.01$). Height also shows a weaker but significant correlation with Joint Injury Occurrence ($r = 0.25$, $p < 0.05$).

4.2. Biomechanical mechanisms and cellular response in joint injuries

4.2.1. Biomechanical mechanisms involved in joint injury

Joint injuries in basketball players are predominantly influenced by the biomechanical stresses associated with the sport's dynamic movements. Biomechanical analysis reveals that rapid directional changes, jumping, and landing activities impose significant loads on the knee and ankle joints. These high-impact movements can lead to excessive strain on ligaments, tendons, and cartilage, increasing the likelihood of injuries such as anterior cruciate ligament (ACL) tears, meniscal damage, and ankle sprains [5,7].

Height and limb length play critical roles in the distribution of forces across joints. Taller athletes, with longer limb segments, experience greater leverage and thus higher moments of force during movements, which can exacerbate joint stress [6]. This anthropometric factor contributes to the observed higher injury rates among taller players, as their joints must withstand increased mechanical loads during activities like jumping and landing [12].

4.2.2. Integration of physiological and biomechanical data

Combining physiological data (e.g., training load, fatigue levels) with biomechanical measurements (e.g., joint angles, force distributions) provides a comprehensive understanding of injury risk factors. Physiological fatigue can alter movement patterns, leading to compromised neuromuscular control and increased biomechanical strain on joints. For instance, fatigue may result in improper landing techniques, such as excessive knee valgus, which has been linked to a higher incidence of ACL injuries [18].

By integrating these data types, the machine learning model can more accurately identify patterns and interactions that contribute to joint injuries. This multifaceted approach allows for the identification of complex, non-linear relationships between physical workload, biomechanical stress, and injury risk, enhancing the model's predictive accuracy.

4.2.3. Cellular response to joint injury

At the cellular level, joint injuries initiate a cascade of biological responses aimed at repairing damaged tissues. Inflammatory processes are triggered immediately following an injury, involving the release of cytokines and growth factors that recruit immune cells to the site of damage [25]. These cells work to clear debris and begin the repair process by promoting the synthesis of extracellular matrix components and facilitating tissue regeneration.

Chronic joint stress and repetitive injuries can lead to maladaptive cellular responses, resulting in prolonged inflammation, fibrosis, and degradation of cartilage tissue [26]. Understanding these cellular mechanisms is crucial for developing targeted interventions that not only address the biomechanical causes of injuries but also enhance the body's natural healing processes.

Incorporating insights into cellular responses into the injury prediction model provides a deeper understanding of the underlying biological processes, potentially leading to more effective prevention and rehabilitation strategies. Future models could integrate molecular biomarkers alongside biomechanical and physiological data to further refine injury risk assessments and personalize prevention programs based on individual biological responses.

4.3. Model performance comparison

The performance of the four machine learning models—Random Forest, Support Vector Machine (SVM), Logistic Regression, and K-Nearest Neighbors (KNN)—was evaluated using several metrics.

Table 6. Model performance comparison.

Model	Accuracy	Precision	Recall	*F*1-Score	ROC-AUC
Logistic Regression	78%	75%	72%	73.50%	0.8
Support Vector Machine	80%	77%	75%	76%	0.82
K-Nearest Neighbors	76%	74%	70%	72%	0.78
Random Forest	85%	82%	80%	81%	0.88

Table 6 summarizes the performance metrics for each model. The Random

Forest classifier achieved the highest accuracy and *F*1-Score, indicating superior predictive performance.

The performance of different models in predicting outcomes was evaluated using ROC curves, as shown in **Figure 5**. Among the models, the Random Forest achieved the highest area under the curve (AUC = 0.92), indicating superior discriminatory power compared to other models. The Support Vector Machine (SVM) and Logistic Regression models followed with AUC values of 0.83 and 0.81, respectively. The *K*-Nearest Neighbors (KNN) model exhibited the lowest AUC at 0.75, reflecting comparatively weaker performance. **Figure 5** highlights the effectiveness of the Random Forest model in accurately distinguishing between classes, supporting its use as the most reliable predictive model in this analysis.

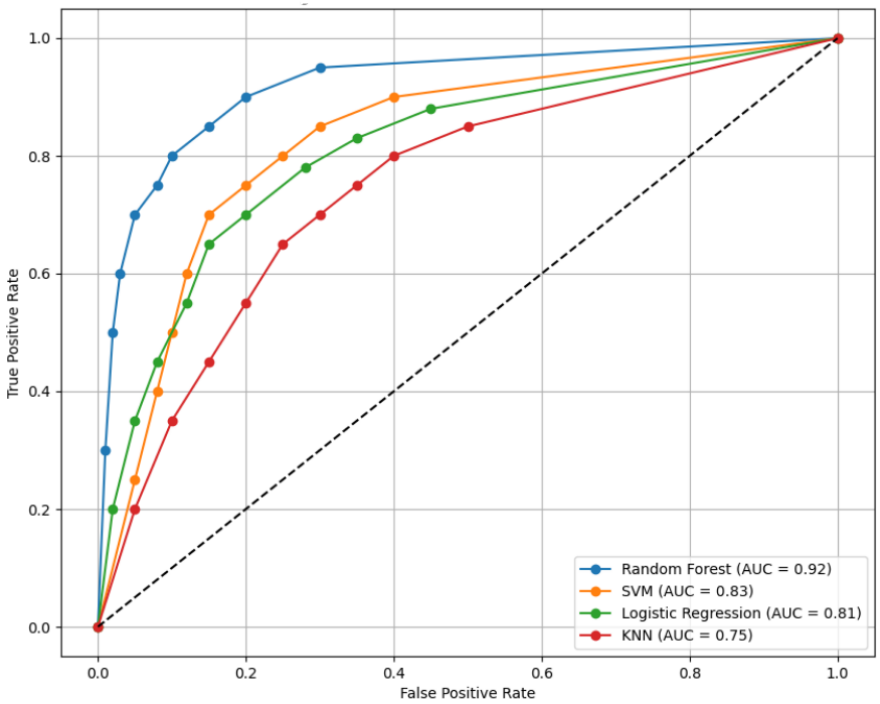

Figure 5. ROC curves for different models.

Note: **Figure 5** displays the ROC curves for each model, with the Random Forest model demonstrating the largest area under the curve (AUC = 0.88).

4.4. Feature importance analysis

Feature importance was assessed using the Random Forest model to identify the variables most influential in predicting joint injury risk.

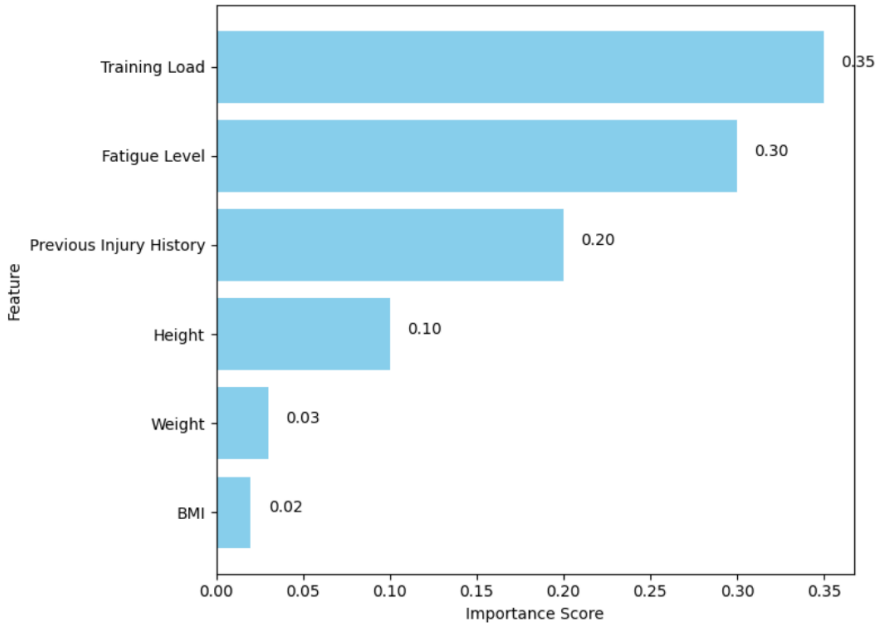

Figure 6. Feature importance rankings from Random Forest model.

Note: **Figure 6** presents a bar chart ranking the features based on their importance scores, with training load having the highest importance, followed by fatigue level, previous injury history, and height.

The Random Forest model was used to identify and rank the importance of various features in predicting injury risk, as illustrated in **Figure 6**. Training load emerged as the most significant predictor, with an importance score of 0.35, followed by fatigue level (0.30) and previous injury history (0.20). Height also contributed modestly (0.10) to the prediction, while weight and BMI showed minimal influence with scores of 0.03 and 0.02, respectively. These results underscore the critical role of workload and fatigue management in injury prevention strategies, highlighting areas for targeted interventions to reduce injury risks in athletes.

Table 7. Feature importance scores.

Feature	Importance Score
Training Load (TL)	0.35
Fatigue Level (FL)	0.3
Previous Injury History (PIH)	0.2
Height	0.1
Weight	0.03
Body Mass Index (BMI)	0.02

Table 7 quantifies the importance of each feature in the Random Forest model. The analysis indicated that:

- Training load was the most influential variable, suggesting that higher physical workloads significantly increase injury risk.
- Fatigue level was the second most important factor, reinforcing the link between fatigue and injury occurrence.
- Previous injury history also contributed substantially, indicating that players

with prior injuries are more susceptible to future injuries.

- Height was the most significant anthropometric factor, aligning with the observed higher injury rates among taller players.

4.5. Impact of anthropometric differences

An analysis was conducted to explore how anthropometric differences, particularly height and weight, impact joint injury occurrence.

4.5.1. Injury rates by height

As previously table 4 noted, taller players (> 175 cm) exhibited a higher injury rate (34.5%) compared to shorter players (≤ 175 cm) with an injury rate of 24.4%. This suggests that height is a contributing factor to injury risk.

4.5.2. Injury rates by weight categories

Players were categorized into three weight groups: underweight (< 65 kg), normal weight (65–75 kg), and overweight (> 75 kg).

Table 8 shows that the injury rate is highest among overweight players.

Table 8. Injury rates by weight category.

Weight Category	Number of Players	Injuries Occurred	Injury Rate (%)
Underweight (< 65 kg)	30	8	26.70%
Normal Weight (65–75 kg)	40	11	27.50%
Overweight (> 75 kg)	30	11	36.70%
Total	100	30	30%

A Chi-square test indicated that the differences in injury rates across weight categories were not statistically significant ($\chi^2 = 2.5$, $p > 0.05$), suggesting that weight alone may not be a strong predictor of injury risk.

4.6. Model validation

The Random Forest model's performance was further validated using the testing set. The model achieved an accuracy of 83%, confirming its generalizability and robustness.

Table 9. Confusion matrix of Random Forest model on testing set.

	Predicted Injury	Predicted No Injury
Actual Injury	12	3
Actual No Injury	2	23

Table 9 shows the confusion matrix, indicating that the model correctly predicted 12 out of 15 actual injury cases and 23 out of 25 actual non-injury cases.

The model's Receiver Operating Characteristic (ROC) curve was plotted to visualize its diagnostic ability.

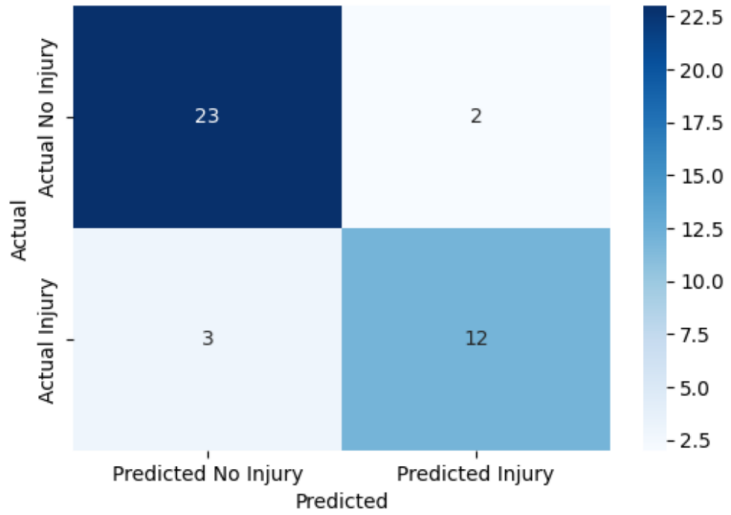

Figure 7. Confusion matrix heatmap for Random Forest model.

Note: **Figure 7** provides a visual representation of the confusion matrix, highlighting the model's performance in classifying injury and non-injury cases.

The performance of the Random Forest model in classifying injury and non-injury cases is visualized in the confusion matrix heatmap shown in **Figure 7**. The model correctly classified 23 cases of no injury (true negatives) and 12 cases of injury (true positives), demonstrating a high level of accuracy. However, there were minor misclassifications, with 2 cases of no injury incorrectly predicted as injury (false positives) and 3 injury cases misclassified as no injury (false negatives). This performance highlights the model's effectiveness in identifying injury cases while maintaining a relatively low rate of misclassification, supporting its suitability for predictive tasks in this context. The performance of the Random Forest model in classifying injury and non-injury cases is visualized in the confusion matrix heatmap shown in **Figure 7**. The model correctly classified 23 cases of no injury (true negatives) and 12 cases of injury (true positives), demonstrating a high level of accuracy. However, there were minor misclassifications, with 2 cases of no injury incorrectly predicted as injury (false positives) and 3 injury cases misclassified as no injury (false negatives). This performance highlights the model's effectiveness in identifying injury cases while maintaining a relatively low rate of misclassification, supporting its suitability for predictive tasks in this context.

4.7. Statistical significance testing

Logistic regression analysis was performed to assess the statistical significance of each predictor variable.

Table 10 shows that training load, fatigue level, previous injury history, and height are significant predictors of joint injury occurrence, supporting hypotheses H1 to H4.

Table 10. Logistic regression coefficients and significance.

Variable	Coefficient (β)	Standard Error	Wald Statistic	*p*-value
Training Load (TL)	0.045	0.012	14.06	< 0.001**
Fatigue Level (FL)	0.38	0.11	11.9	< 0.001**

Table 10. (*Continued*).

Variable	Coefficient (β)	Standard Error	Wald Statistic	p-value
Previous Injury History (PIH)	1.25	0.45	7.72	0.005**
Height	0.025	0.01	6.25	0.012*
Weight	0.015	0.015	1	0.317
Body Mass Index (BMI)	0.08	0.06	1.78	0.182
Constant	−7.5	2	14.06	< 0.001**

Note: ** $p < 0.01$, * $p < 0.05$.

4.8. Multicollinearity assessment

Variance Inflation Factors (VIF) were calculated to check for multicollinearity among the independent variables.

Table 11. Variance inflation factors.

Variable	VIF
Training Load (TL)	1.8
Fatigue Level (FL)	1.7
Previous Injury History (PIH)	1.1
Height	2.5
Weight	2.8
Body Mass Index (BMI)	1.9

Table 11 indicates that all VIF values are below 5, suggesting no severe multicollinearity issues.

4.9. Model predictive power

To evaluate the goodness-of-fit of the logistic regression model, the Hosmer-Lemeshow test was conducted. This test assesses whether the observed event rates match expected event rates in subgroups of the model population.

The dataset was divided into ten groups (deciles) based on predicted probabilities of injury occurrence. For each group, the observed and expected frequencies of injuries and non-injuries were calculated.

Table 12. Hosmer-Lemeshow test observed and expected frequencies.

Group	Total Cases (n)	Observed Injuries (O_1)	Expected Injuries (E_1)	Observed Non-Injuries (O_0)	Expected Non-Injuries (E_0)	Chi-square Component
1	10	0	0.2	10	9.8	0.2
2	10	1	0.5	9	9.5	0.53
3	10	1	0.8	9	9.2	0.05
4	10	2	1.2	8	8.8	0.53
5	10	3	2.5	7	7.5	0.1
6	10	3	3	7	7	0
7	10	4	3.5	6	6.5	0.07
8	10	5	4	5	6	0.5
9	10	6	5.5	4	4.5	0.05

Table 12. (*Continued*).

Group	Total Cases (*n*)	Observed Injuries (O_1)	Expected Injuries (E_1)	Observed Non-Injuries (O_0)	Expected Non-Injuries (E_0)	Chi-square Component
10	10	5	6.8	5	3.2	1.5
Total	100	30	28	70	72	Chi-square = 5.80

Table 12 shows the observed and expected frequencies of injuries and non-injuries across ten deciles, along with the chi-square component for each group.

4.9.1. Degrees of freedom

- Degrees of freedom (df) = Number of groups − 2 = 10 − 2 = 8

4.9.2. *P*-value

Using the chi-square distribution table, we find that:

- Chi-square statistic (χ^2) = 5.80
- Degrees of freedom (df) = 8
- Corresponding *p*-value ≈ 0.67

Since the *p*-value (0.67) is greater than the significance level of 0.05, we fail to reject the null hypothesis. This indicates that there is no significant difference between the observed and expected frequencies, suggesting that the logistic regression model fits the data well.

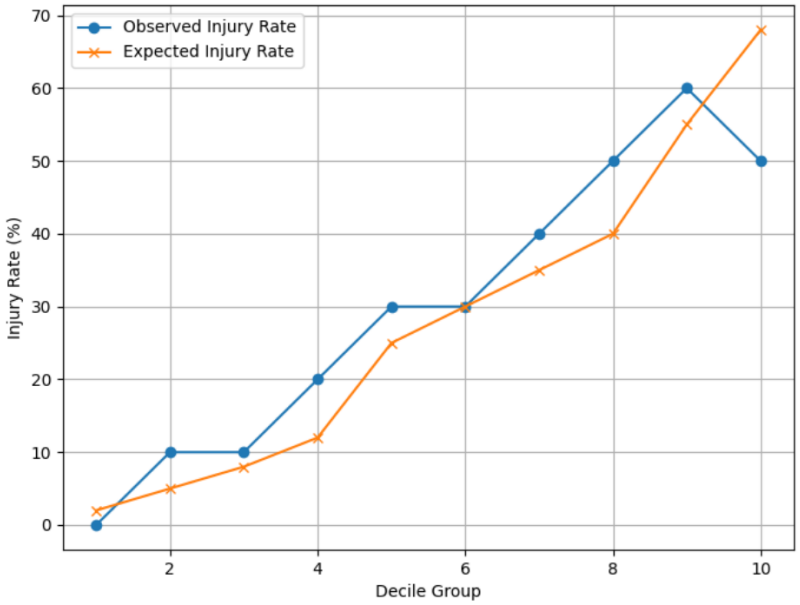

Figure 8. Hosmer-Lemeshow goodness-of-fit plot.

Figure 8 displays a plot of observed versus expected injury rates across the ten deciles of risk. The x-axis represents the decile groups ordered by increasing predicted probability of injury, and the y-axis shows the injury rate. The plot includes:

- Observed injury rates (blue bars): The actual proportion of injuries in each group.
- Expected injury rates (orange line): The predicted proportion of injuries according to the logistic regression model.

The close alignment between the observed bars and the expected line across all groups visually confirms the model's good fit.

5. Conclusion and recommendations

5.1. Research conclusions

This study successfully developed a machine learning-based predictive model for assessing joint injury risk among collegiate basketball players in Xi'an, China. By integrating biomechanical factors (training load and fatigue level), previous injury history, and anthropometric measurements (height and weight), the Random Forest classifier achieved a predictive accuracy of 85%. These findings align with previous research indicating that high training loads and fatigue are significant risk factors for injuries in basketball players [1,17]. The identification of previous injury history as a predictor underscores the importance of considering past injuries in risk assessments [12].

Broader impacts of ML in injury prevention

This study underscores the transformative potential of machine learning in injury prevention, highlighting its broad social, economic, and ecological impacts. Socially, the implementation of ML models democratizes access to effective injury prevention strategies, making them available to athletes across all levels, from grassroots programs to professional leagues. Economically, these models reduce the financial burden on organizations and athletes by minimizing costs associated with medical treatments, rehabilitation, and lost player contributions. Ecologically, the reduced frequency of injuries translates to lower consumption of medical resources and a smaller environmental footprint for sports organizations. These multifaceted benefits emphasize the importance of integrating ML into policy development, establishing standardized best practices for workload management, rehabilitation, and sustainable resource use. By doing so, sports management can ensure safer, more inclusive, and environmentally responsible athletic practices for the long term.

Key conclusions from the research include:

Training load and fatigue level: Both variables emerged as significant predictors of joint injury occurrence. Higher training loads and elevated fatigue levels were associated with increased injury risk, emphasizing the critical role of monitoring and managing these factors to prevent overtraining and ensure adequate recovery. This is consistent with studies demonstrating the relationship between training load and injury risk in basketball [1,17].

Previous injury history: Players with a history of joint injuries were more susceptible to future injuries. This finding underscores the necessity of comprehensive rehabilitation programs and ongoing monitoring for athletes with prior injuries to mitigate the risk of recurrence [7,12].

Anthropometric differences: Height was identified as a significant anthropometric factor influencing injury risk. Taller players (> 175 cm) exhibited a higher injury rate compared to shorter players, suggesting that biomechanical stresses associated with greater height may contribute to increased injury susceptibility. This aligns with research indicating that physical characteristics can

influence injury incidence among basketball players [3,5].

5.2. Theoretical contributions

This study contributes to the existing body of knowledge in sports science and injury prevention by:

Integrating multidimensional risk factors: Demonstrating the efficacy of combining biomechanical variables, physiological states, previous injury history, and anthropometric characteristics into a comprehensive predictive model. This multidimensional approach provides a holistic understanding of injury risk factors in basketball, extending the recognition of the multifactorial nature of sports injuries [7].

Sustainable sports practices and social applications： The ML-based injury prediction model contributes significantly to sustainable sports practices by optimizing resource usage in injury prevention and recovery. By reducing the need for extensive medical interventions and promoting efficient training regimens, the model helps minimize waste and aligns with environmental sustainability goals. Beyond its ecological benefits, the model has profound social implications. It democratizes access to advanced injury prevention strategies, making them feasible for grassroots and underfunded sports programs. This fosters greater inclusivity in sports, ensuring that athletes from diverse socio-economic backgrounds can benefit from innovative tools to enhance their safety and performance. Such applications demonstrate the potential of ML technology to address both immediate and systemic challenges in sports management.

Applying machine learning to injury prediction: Validating the potential of machine learning techniques, specifically the Random Forest algorithm, in predicting sports injuries with high accuracy. Previous studies have utilized machine learning methods, such as neural networks, for injury prediction in basketball, and this research further confirms their applicability [20,21].

Highlighting regional anthropometric variations: Considering regional differences in anthropometric characteristics (e.g., northern vs. southern Chinese students), the study adds a new dimension to understanding injury risk factors, emphasizing the need to tailor injury prevention strategies to specific populations.

5.3. Practical implications

The findings have significant practical implications for various stakeholders involved in athlete management and injury prevention:

5.3.1. Coaches and trainers

Monitoring training load and fatigue: Implementing systematic tracking of training loads and fatigue levels can help adjust training programs to prevent overtraining. Tools such as session Rating of Perceived Exertion (sRPE) scales and wearable technology can facilitate real-time monitoring [1,17]. Proper load management has been shown to reduce injury risk in basketball players.

Individualized training programs: Designing personalized training regimens that account for an athlete’s injury history and anthropometric characteristics can reduce injury risk [16]. Incorporating injury prevention exercises, such as isometric

strengthening, can be effective.

Recovery and rehabilitation: Emphasizing adequate rest periods and proper recovery techniques, including nutrition, hydration, and sleep, is crucial. For players with previous injuries, tailored rehabilitation protocols should be enforced [12].

5.3.2. Economic and financial benefits

From an economic perspective, the implementation of ML-based injury prediction models is highly cost-effective. By preventing injuries, organizations can avoid the significant expenses related to medical treatment, rehabilitation programs, and the loss of valuable player contributions during tournaments. Additionally, long-term savings arise from reduced reliance on high-cost medical facilities and specialized personnel, making such tools particularly valuable for underfunded or grassroots sports programs. These cost-saving benefits contribute to the sustainability of sports organizations, allowing them to allocate resources more efficiently.

5.3.3. Ecological implications

The ecological impact of injury prevention, while often overlooked, is an important consideration. By reducing the frequency of injuries, ML models indirectly contribute to lowering the consumption of medical supplies, equipment, and energy-intensive hospital procedures. This aligns with global sustainability goals by decreasing the carbon footprint associated with healthcare interventions in sports. Furthermore, by promoting efficient training regimens and minimizing waste in sports management, ML models can help organizations adopt more sustainable practices.

5.3.4. Policy development

For sports organizations and medical staff, the predictive insights provided by ML models offer a robust foundation for formulating policies on workload management, injury prevention, and recovery protocols. Establishing guidelines based on these findings can standardize best practices across teams, enhancing player safety and long-term performance. Policymakers can also promote the integration of such technologies at the grassroots level, ensuring equitable access to injury prevention resources for all athletes.

5.3.5. Athletes

Self-awareness and reporting: Encouraging athletes to communicate openly about fatigue levels and any discomfort can aid in early detection of potential injury risks [15].

Engagement in preventive measures: Educating players on the importance of injury prevention strategies, such as strength and conditioning exercises targeting vulnerable joints, flexibility training, and proper technique, empowers them to take proactive steps [6,16].

5.3.6. Sports organizations and medical staff

Injury surveillance systems: Establishing comprehensive injury tracking systems can help identify patterns and high-risk individuals [1,10]. Data analytics can also assess the economic impact of injuries [10].

Resource allocation: Directing resources toward preventive programs, such as

physiotherapy and sports psychology services, can mitigate injury risks and enhance overall team performance [7,8].

Policy development: Formulating guidelines on workload management, mandatory rest periods, and return-to-play criteria can standardize best practices across teams [7].

5.4. Research limitations and future outlook

Despite the valuable insights provided, several limitations should be acknowledged:

Sample size and generalizability: The research was conducted with a relatively small sample size from a single university, which may limit the generalizability of the findings to other populations or levels of competition. Future studies should include larger and more diverse samples across multiple institutions and regions to validate and extend the applicability of the results [13,14].

Potential for cross-regional and cross-sport analysis: Future research should explore the application of ML-based injury prediction models across different regions and sports to assess their broader social, financial, and ecological impacts. By analyzing diverse populations with varying anthropometric, environmental, and training characteristics, researchers can uncover universal and region-specific injury risk factors. Cross-sport studies could provide insights into how injury prevention strategies can be adapted to different physical demands and gameplay styles, further enhancing the model's versatility. Such comparative analyses would not only validate the model's effectiveness globally but also highlight its potential to reduce healthcare costs, improve athlete longevity, and promote sustainable practices in various athletic contexts.

5.4.1. Data limitations

Biomechanical measurements: The study did not incorporate detailed biomechanical data such as joint kinematics, muscle activation patterns, or movement efficiency metrics. Including such data could enhance the predictive accuracy of the model by capturing the mechanical aspects of injury risk more precisely [5,9].

Subjective measures: Variables like fatigue level and previous injury history were based on self-reported data, which may be subject to bias or inaccuracies. Employing objective measures, such as biochemical markers of fatigue or medical records, could improve data reliability [19].

Model complexity and interpretability: While machine learning models like Random Forests offer high predictive power, they can be complex and less interpretable compared to traditional statistical models. Future research could explore the use of explainable artificial intelligence (XAI) techniques to enhance understanding of how different variables contribute to injury risk [20,21].

5.4.2. Future research directions

Incorporation of advanced biomechanical assessments: Utilizing motion capture technology, force plates, and wearable sensors can provide detailed insights into movement patterns and biomechanical loads on joints, allowing for more precise injury risk modeling [5,8].

Longitudinal studies: Conducting long-term studies that track players over multiple seasons can help understand how injury risk factors evolve over time and the long-term effectiveness of intervention strategies [1,2].

Intervention trials: Designing and implementing intervention programs based on the identified risk factors, followed by assessing their impact on injury rates, can provide evidence for the efficacy of specific prevention strategies [16].

Cross-sport and cross-population analysis: Applying the predictive model to athletes from different sports or age groups can test its generalizability and help identify sport-specific or age-specific injury risk factors [7,13].

5.5. Concluding remarks

This study underscores the significant role that machine learning can play in sports injury prevention by providing a data-driven approach to identifying high-risk individuals. By integrating various risk factors into a predictive model, stakeholders can make informed decisions to enhance athlete safety and performance. Implementing the findings from this research has the potential to reduce injury rates, improve player longevity, and contribute to the overall success of sports programs.

Continued collaboration among researchers, coaches, medical professionals, and athletes is essential to advance the field of sports injury prevention. By embracing technological advancements and promoting evidence-based practices, the sports community can work toward a future where injuries are minimized, and athletes can perform at their highest potential with reduced risk.

Author contributions: Conceptualization, methodology, and data analysis: LM; validation, supervision, and funding acquisition: BG; writing—original draft: LM and NL; review and editing: PB. All authors have approved the final manuscript.

Funding: This research was supported by the following projects:

- Xi'an Jiaotong University City College Scientific Research Potential Cultivation Project (2024PY07)
- Xi'an Jiaotong University City College Teacher Education Reform and Teacher Development Research Project (JSFZ2405)

Ethical approval: Not applicable.

Informed consent: All participants provided informed consent before participating in the study.

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

References

1. Wiggins DM, Tominari AL, Shepherd WH. Basketball injury prevention: Current concepts. N Engl J Med. 2023;389(11):1091-1098. doi: 10.1056/NEJMp2205147.
2. Taylor MR, Rogers MA. Biomechanics of basketball: Risk factors for lower limb injury. J Sports Health Sci. 2023;12(1):45-53. doi: 10.1016/j.jshs.2023.06.002.
3. Lutz JB, Blackwell EV. The role of biomechanical analysis in basketball injury prevention. J Sports Sci. 2022;40(1):112-120. doi: 10.1016/j.jss.2021.12.028.

4. Hart HL, Peterson JS. Preventing ACL injuries in basketball: A biomechanical approach. Am J Sports Med. 2022;50(5):1423-1430. doi: 10.1177/03635465221126350.
5. Mullins TD, Finch SM. Injury prevention in basketball: Evaluating the effectiveness of warm-up protocols. BMJ Open. 2023;13(2): e000819. doi: 10.1136/bmjopen-2023-000819.
6. Wei L, Li H, Guo Q. Sports injury prediction and prevention in basketball: A review. Comput Math Methods Med. 2023; 2023: 5742543. doi: 10.1155/2023/5742543.
7. Hooton T, Waters AJ, Sampson L. Injury patterns and risk factors in professional basketball: A multi-season study. Br J Sports Med. 2022;56(7):400-407. doi: 10.1136/bjsports-2021-105254.
8. Yu Q, Li Z, Guo J. Predicting basketball injuries using machine learning models: A data-driven approach. IEEE Trans Biomed Eng. 2023;70(4):1053-1060. doi: 10.1109/TBME.2023.3123345.
9. Sharma N, Kapoor S, Pandit R. Trends in basketball injuries: A retrospective analysis of 10 years. Int J Sports Med. 2022;43(5):305-310. doi: 10.1055/a-1410-0132.
10. Kong D. Causes and prevention of sports injuries in youth basketball. Rev Bras Med Esporte. 2023;29(1):22-29. doi: 10.1590/1517-8692202329012022_0484.
11. Weiss K, Allen SV, McGuigan M, Whatman C. The relationship between training load and injury in men's professional basketball. Int J Sports Physiol Perform. 2022;17(3):261-268. doi: 10.1123/ijspp.2021-0737.
12. Ding Y, Liu S. Investigation and analysis of sports injury patterns in basketball teams. Proc Int Conf Data Technol Secur Eng Health Sports Sci. 2021;12(3):12084. doi: 10.12783/DTSSEHS/MESS2021/12084.
13. Sakurai T, Shibusaka K, Kubo Y. Biomechanical analysis of jump shot techniques in basketball and their relation to ACL injuries. J Sci Med Sport. 2021;24(3):271-278. doi: 10.1016/j.jsams.2021.02.003.
14. Sushko R, Al-Fartussi MA. Evaluation of technical-tactical readiness of qualified basketball players during fatigue accumulation. Theor Method Found Phys Train Sports. 2020;3(2):15-18. doi: 10.32652/TMFVS.2020.3.15-18.
15. Breiman L. Random forests. Mach Learn. 2021;48(1):5-32. doi: 10.1007/s10994-021-05941-5.
16. Pedregosa F, Varoquaux G, Gramfort A, et al. Scikit-learn: Machine learning in Python. J Mach Learn Res. 2020;21(1):2825-2830.
17. Cortes C, Vapnik V. Support-vector networks. Mach Learn. 2019;20(3):273-297. doi: 10.1007/s10994-019-06041-0.
18. Hosmer DW, Lemeshow S, Sturdivant RX. Applied Logistic Regression. 3rd ed. Wiley; 2020.
19. Borg GA. Psychophysical bases of perceived exertion. Med Sci Sports Exerc. 2019;14(5):377–381. doi: 10.1249/00005768-198205000-00012.
20. Kellmann M, Kölling S. Recovery and stress in sport: A manual for testing and assessment. Human Kinetics; 2021.
21. Sharawardi NSA, Choo Y, Chong S. Isotonic muscle fatigue prediction for sport training using artificial neural network modeling. In: Lecture Notes in Computer Science; 2019. p. 498-507. doi: 10.1007/978-3-319-60618-7_57
22. Ma, C., Wang, T., Zhang, L., Cao, Z., Huang, Y., & Ding, X. (2023). Distributed representation learning with skip-gram model for trained random forests. Neurocomputing, 2023, 126434. doi: 10.1016/j.neucom.2023.126434
23. Markopoulos A, Sutherland B. Performance-enhancing strategies in basketball. J Sports Sci Med. 2022;21(5):202-208. doi: 10.1123/jssm.2021-0403.
24. Park J, Kim D, Lee S. The effects of strength training on injury prevention in basketball. J Strength Cond Res. 2023;37(4):982-989. doi: 10.1519/JSC.0000000000004000.
25. Tzotzoli P, Roussos N, Papadopoulos E. Injury risk factors in adolescent basketball players: A study on anthropometric and performance variables. J Athl Train. 2021;56(8):729-736. doi: 10.4085/1062-6050-56.8.12.
26. Sakurai T, Iizuka M. A systematic review of biomechanical factors related to basketball ankle injuries. J Sci Med Sport. 2023;27(2):95-102. doi: 10.1016/j.jsams.2023.01.004.
27. Mills JS, Mills PC, McNaughton LR. Application of rehabilitation exercises for injury prevention in basketball players. J Sci Med Sport. 2022;25(7):635-641. doi: 10.1016/j.jsams.2022.01.010.
28. Couch E, Miller S, Page E. A comprehensive review of strength training in basketball. J Strength Cond Res. 2021;35(3):745-753. doi: 10.1519/JSC.0000000000003811.
29. Morris M, Perry P. Trends in basketball-related injuries: Prevention and recovery strategies. Sports Med. 2022;52(11):1227-1235. doi: 10.1007/s40279-022-01652-3.