Natural Resources for Human Health
eISSN: 2583-1194

Original Research

Received 12 March 2026
Revised 21 April 2026
Accepted 20 May 2026

Natural Resources for Human Health 2026, Vol. 6 (3): 974–991

KEYWORDS:
Professional Gymnastics, Sports Injury Prediction, Retrospective Analysis, Machine Learning, Deep Learning, Multimodal Dataset

Intelligent Injury Risk Assessment through Retrospective Analysis of Multimodal Sports Data among Professional Gymnasts

Pawan Bhoir1*, Dr. Vaibhav Narawade2

[1]Research Scholar, Department of Information Technology, Ramrao Adik Institute of Technology, D. Y. Patil Deemed to be University, Nerul, India. pawanbhoir@gmail.com
[2]Professor, Department of Computer Engineering, Ramrao Adik Institute of Technology, D. Y. Patil Deemed to be University, Nerul, India. vaibhav.narawade@rait.ac.in

ABSTRACT: Sports injuries still pose a significant problem in professional gymnastics, as a consequence of the combination of repetitive movement patterns, high physical demands, long hours of training and accumulated physiological stress. Knowing early in the process of development when injury is likely to occur is important for athlete safety, continuing training and performance, and making evidence-based decisions. Traditional injury evaluation techniques rely on visual inspection and single physiological measurement, and have only a partial retrospective perspective of injury factors. The comparative analytical framework of this study involves machine learning and deep learning methods, to explore injury risk patterns based on retrospective analysis of multimodal sports injury records. The Multimodal Sports Injury Prediction Dataset is used, which consists of physiological, workload, and athlete condition indicators from professional sports participants. Before developing the models, data preprocessing was done which involved data cleaning, data normalization, class balancing, and statistical exploration. To predict injury risk, four machine learning models (Logistic Regression, Random Forest, XGBoost, LightGBM) and three deep learning models (Artificial Neural Network (ANN), Convolutional Neural Network (CNN), and Long Short-Term Memory (LSTM)) were tested. The experimental analysis showed that the LightGBM model had the highest accuracy of 95.16%, F1-score of 94.92% and ROC–AUC of 0.981 compared to other machine learning models. In deep learning techniques, CNN showed the highest predictive power with 95.84% of accuracy, 0.986 ROC–AUC and 95.55% F1-score. An explainability analysis identified workload index, heart rate variability, recovery time and respiratory factors as dominant contributors to injury. The results show that multimodal retrospective injury analysis and prediction based on data can be useful for athlete risk monitoring and decision making in professional gymnastics.

1. INTRODUCTION

Gymnastics is one of the physiological most challenging sports because of the demands for strength, flexibility, coordination, balance, acceleration, repeated impact and precision-based execution. Athletes execute intricate movements including quick transitions between apparatus, movements in the air, landing forces, rotational movements and multiple repetition of training exercises in various apparatus. These conditions can lead to higher levels of musculoskeletal stresses and make gymnasts more vulnerable to injury during training and competition [1][2].

A continued issue in the sport of professional gymnastics is the prevalence of injuries, as athletes are continually subjected to repeated loading cycles from an early developmental stage. Gymnastics has different demands than sports with intermittent movement patterns as the routines are technically challenging and need to be repeated quickly without much recovery [3].

Important training frequencies and the stress of competition lead to a build up of mechanical strain to joints, muscles, ligaments and connective tissues. Overuse, impact, movement asymmetries, fatigue and inadequate recovery are all common mechanisms of injury in gymnastics [4].

When considering common injury areas, the lower extremity, shoulders, wrists, lower back, and knees are areas of significant loading during repeated landings and body support movements. In floor routines and vault events, there are many times vertical and horizontal reaction forces are created in take-offs and landings. Both uneven bars and rings require constant upper-body loading, while balance-type activities require good postural control and neuromuscular coordination. Consequently, gymnasts are at high risk of acute injuries, such as ligamentous tears and fractures, as well as chronic conditions, such as tendinopathy, stress reactions and repetitive strain injuries [5]-7].

Biomechanically, gymnastics offers a variety of movement opportunities, in which the body is constantly propelled, suspended, rotated, and requires absorption of impact. They involve synchrony of muscular activity and good transfer of force from one anatomical segment to another [8]. Minor movement execution errors can also cause a different pathway of force transmission and greater loading of the local tissues. These biomechanics demands repeated over time can lead to a decrease in movement efficiency and increase the risk of injury [9].

Injury occurrence is also affected by physical conditioning. Athlete readiness is influenced by variables such as cardiovascular capacity, level of fatigue, flexibility, muscle balance, training load and quality, and recovery. Training overload can be seen in physiological responses like increased heart rate, decreased oxygen saturation, increased respiratory needs and increased recovery time. As these factors become more critical, surveillance of them becomes more important when it comes to identifying any potential hidden risks of injury before they become apparent.

Sports injuries additionally have significant impacts on the performance of the athletes. A training episode with injuries decreases the continuity of training, disrupts skill development, adds to the psychological load, and impacts competition readiness. When recovering from injury, athletes often have diminished confidence, movement strategies and quality of execution. Athletes with repeated injuries risk having shorter careers and may impact lifelong health outcomes. As a result, injury prevention and early injury identification are now a vital part of sports performance management.

Traditional athlete monitoring systems typically involve observation and assessment, coaches' experience, and physical screening, with periodic medical assessment. While these methods offer valuable information, they often do not account for the intricate interplay between physiological, biomechanical and workload factors all of which contribute to the risk of injury. This restriction has sparked an appetite for sports analytics that can be used to assess injuries in greater detail, using data.

1.1. Need to conduct a Retrospective Injury Risk Analysis.

Retrospective injury risk analysis (RA) has become an essential method of analyzing and understanding injury mechanisms from previous injury data and conditions of athletes observed. Retrospective studies include looking at the information that is currently available from athletes as opposed to just looking at the occurrences of injury events.

Traditional injury assessment tools rely mostly on subjective observations, individual clinical measurements and manual interpretation of athlete status. Visual assessments and athlete self-report are standard methods for coaches and medical staff to assess training intensity, readiness and recovery. Although practical, these methods may not be consistently able to capture subtle physiological variations or cumulative stress responses that occur over time.

A second drawback to traditional methods is in the lack of cohesive data use. A single injury factor is seldom the root cause of injury events. Different factors including workload accumulation, physiological fatigue, environment, history of injuries and adaption of the biomechanics are all playing a role in the development of an injury. These components have been assessed in isolation by manual methods, which fail to provide insight to the multidimensional relationships [10].

Retrospective analytical methods have merits, in that they combine historic observations made over long time periods. The exposure to sporting activity, recovery time, physiological markers, frequency of sporting activities, and injury outcomes are all contained within these historical athlete data. These records can help identify trends that are linked with a higher risk of injury. Multimodal data collection techniques have also recently been developed, further enhancing the utility of retrospective studies. Modern sports monitoring systems can measure physiological signals, athlete workload data, movement pattern data, recovery data and performance data together. Multimodal historical datasets offer the possibility of studying injury development by combining features for interpretation, as opposed to taking individual measures [11].

This is the direction that the Multimodal Sports Injury Prediction Dataset aims to take by including information about the athletes in a multimodal fashion, covering aspects of sports performance and health in different ways. Retrospective analysis and multimodal data enable better understanding of injury sensitive factors, and evidence-based decision making.

Retrospective analysis also simplifies the operations when compared with prospective long duration studies. Analytical models can be built and tested effectively and consistently with observations as historical data are already available. These benefits render the retrospective analysis of data more appropriate in the field of sports medicine and athlete performance.

1.2. Sports Analytics and machine learning and deep learning

The development of sports analytics has been very fast and is now seeing computational methods being used more in the field of sports injury prediction and athlete performance evaluation. Machine learning and deep learning methods offer systematic approaches to uncovering relationships between several variables and deriving predictions from past observations.

Because of the ability to process structured sports data effectively, machine learning approaches have proven to be useful in classification and risk assessment problems. Algorithms including "Logistic Regression, Random Forest, XGBoost, and LightGBM" are used for the identification of nonlinear relationships and estimation of injury probability from multiple input variables. These are strategies to help make the past measurable and current measurements meaningful risk indicators [12].
Predictive modelling can help with sports injury analysis by detecting patterns or trends that may be correlated with occurrences of injury before serious symptoms are apparent. Predictive algorithms, rather than relying solely on threshold-based decision rules, consider interactions between workload, physiology, recovery variables and athlete history. This process facilitates individual evaluation and better-informed planning for intervention.
Deep learning techniques also further expand analytical capacity by automatically learning representations. Complex feature interactions and hierarchical patterns in multidimensional data can be processed and identified by models such as "Artificial Neural Networks, Convolutional Neural Networks and Long Short-Term Memory networks". This approach is especially helpful when the outcomes of injury can depend on a mix of physiological and temporal factors.
Advanced analytical models can be powerful predictors but have concerns as to interpretability and transparency [13]. While high performance prediction is important, it might not be enough to give the confidence necessary for sports use as coaches and clinicians also need explanations to support injury related decisions.
Explainability is thus a crucial criterion in sports analytics. Interpretation techniques help to determine which variables have the most influence on the prediction results and increase confidence in the analytical systems. For example, feature importance ranking and SHAP analysis can give insight into how feature variables are contributing to injury risk estimation. Explainable modelling is crucial for practical deployment, offering both predictive accuracy and clarity in decision-making [14][15]. Combining machine learning, deep learning, and explainability provides a structured analytical framework that can be used to aid retrospective injury analysis in professional gymnastics [17].

1.3. Research Objectives

i. Objective 1: To conduct a retrospective injury risk assessment for multimodal historical athlete data to identify factors associated with the occurrence of sports injuries.
ii. Objective 2: To create and contrast machine learning and deep learning models for injury risk prediction and performance evaluation.
iii. Objective 3: Using explainability analysis to understand the predictions of injuries to identify key physiological and workload-related factors.

1.4. Contributions of the Study

i. Created a retrospective analytical approach to gymnastic injury risk assessment using multimodal sports data.
ii. Implemented comparative analysis of four machine learning and three deep learning models using the same preprocessing methods.

iii. Conducted explainability analysis to explore injury sensitive variables to better interpret prediction results.
iv. Analyzed the correlation between physiological parameters, workload measures and athlete condition to gain a retrospective understanding of injuries.

The sport of gymnastics remains a significant challenge in terms of injuries because of the high physical demands, repetitive loading and frequent exposure to high performance training environments. While traditional methods of injury assessment are beneficial for making observations, they are often limited in understanding the intricate relationships between physiological factors, workload accrual and past athlete behaviors. There is a growing opportunity to analyse the development of injuries in multimodal sports using more comprehensive approaches because of the increasing availability of multimodal sports data.

This study examines retrospective injury risk analysis, based on historical multimodal sports data, and comparative computational modelling techniques to address these challenges. Patterns are found using machine learning and deep learning techniques and the predictive ability tested under different analytical scenarios. Moreover, explainability methods are integrated for better comprehension of the models' results and for transparent decision making.

2. LITERATURE REVIEW

As the number of sports injuries continues to rise and more attention is focused on the safety of athletes, recovery times, and longevity of performance, this research field has come to the forefront. Conventional injury assessment strategies were primarily based on observation, clinical evaluation and statistical analysis, and were unable to fully account for the interplay between physiological, biomechanical and workload factors. With the recent advancements in computational intelligence, machine learning, deep learning, wearable sensing, and multimodal analytical frameworks have been introduced in sports science. Such methods facilitate automated injury assessment, injury prediction, rehabilitation monitoring and performance optimization. The current state of the art shows that the development of data-centric sports healthcare systems for preventive decision making and athlete management is ongoing.

2.1. Machine Learning-Based Sports Injury Prediction and Risk Assessment

Machine learning methods have been widely used in recent works on sports injury analytics to enhance the understanding of sports injuries and to make predictions for decision-making. One study examined the use of machine learning in injury prediction during football games and found that using structured historical data of the athletes helped to better identify injury-related patterns by using classification and predictive modelling approaches [17]. Later, a spatio-temporal transformer and variational autoencoder hybrid prediction framework was presented that can capture temporal relations and latent injury characteristics, giving it better analytical capability for complex sports injury scenarios [18].

Further progress was made by using a fusion neural architecture with a dual feature representation to fuse complementary feature representations for injury estimation, which increased the sensitivity to injury-related variables [19]. Yet another investigation suggested a preventative decision-making framework that incorporates training and rehabilitation indicators under the guidance of artificial intelligence in a physical education environment [20]. The application of AI in broader healthcare and sports performance showed a growing trend of using AI systems to monitor, treat, and analyze performance in elite sports environments [21].

A specific study on soccer found that methods of machine learning have a large role to play in the study of injury risk and associations between athlete workload in the sport, and opportunities for broader practical application [22]. Likewise, a conceptual and application-oriented study collected and compiled a summary of the main machine learning and artificial intelligence methods used in the field of sports and pinpointed the challenges of data quality, deployment reliability, and interpretation needs. Overall, these studies highlighted the significance of predictive modelling and computational learning techniques in sports injury analysis.

2.2. Deep Learning and Multimodal Approaches for Injury Monitoring

Multidimensional sports information processing and automated injury assessment have been greatly focused on deep learning approaches. The real-time injury monitoring system using deep learning algorithms was able to continuously monitor the condition of athletes and identify changes indicative of a potential injury during sports participation [24]. Yet another study investigated the potential of and the obstacles to the use of machine learning in sports sciences and stressed the difficulties of generalizing models and applying them in the field with diverse athlete groups [25].

Analytical frameworks were then adapted to consider performance and injury indicators, allowing for better decision support in professional sports and the expansion of sports analytics methods toward injury prevention and player availability optimization [26]. It was also investigated as a more general decision making mechanism that can help to enhance sport forecast accuracy and aid in data-driven operational planning [27].

A machine learning approach was used to investigate a classification model for prediction of rehabilitation outcomes post hamstring injury, highlighting the potential applications of computational methods in post-injury management [28]. Later, cloud-based deep learning systems were developed to allow automated diagnosis of sports injuries in scalable computational environments to enable large-scale analytical operations [29]. All of these investigations showed how deep learning and integrated analytical infrastructures help to enhance injury assessment and ongoing monitoring of the athlete.

2.3. Wearable Systems, Sensor Analytics, and Time-Series Injury Prediction

With the emergence of various types of wearable technology, the research into sensor-based injury prediction and sensor-based continuous monitoring has been accelerated. Another study suggested prediction and simulation of the wearable sensor devices based on backpropagation neural network, which demonstrated better performance in estimation of injury-related conditions based on the physiological measurements [30]. A large-scale analysis of injuries prevalence among athletes with disabilities found that there were variations in the occurrence of injuries across different categories of sports participation [31].

Time-series modelling is an increasingly important area of injury forecasting. A new temporal model was created to convert the temporal information of athletes into an encoded image representation, and deep learning techniques were used to estimate the risk of injury, with a better extraction of the temporal patterns from the observations of the past [32]. Video-informed reconstruction of sports accidents was also extended as a multimodal analytical method for understanding injury mechanisms that involved studying the impact behaviour of collisions and the resulting injury outcomes, using the available neuroimaging information and finite element analysis to determine relationships [33].

The research of rehabilitation also acknowledged sports injury recovery is a complex adaptive process, where the physiological, behavioural, and environmental factors interact together. One of the key points to be taken from the analysis of rehabilitation systems is the need to understand the dynamic relationship instead of the individual recovery indicators [34]. These findings combined demonstrated a greater trend towards multimodal sensing, time-series analysis and integrated monitoring systems for sports injury prevention.

2.4. Emerging Trends in Wearable Rehabilitation and Hybrid Deep Learning Frameworks

Most recent research also has been directed towards the development of wearable rehabilitation technologies and hybrid computation approaches that can help enhance the management of sports injuries. One study summarized the development of hydrogel based wearable systems for training support, injury prevention, monitoring rehabilitation, and athlete recovery. The study showed that materials with sensing properties can be used in wearable fabrics and therefore can be a part of continuous monitoring of sports [35].

Injury prediction has also been an area where hybrid learning architectures have been given attention. To make use of both supervised prediction and latent group identification, a combined deep neural network and semi-supervised clustering framework was proposed to enhance the estimation of sports injury risk. The results of analysis showed an increase in the ability to discover hidden injury structures and deal with partially labelled data [36].

While there have been studies that showed great advances in the field of predictive modelling, wearable analytics and intelligent monitoring systems, there are still some limitations. A lot of research focused on a particular sport, a set of data from one source, a rehabilitation situation or a specific analytical model. Not much effort has been devoted to multimodal analysis in the past, and only a few studies have involved both comparative machine learning/deep learning and explainability-driven interpretation. Thus, in the present study, a retrospective multimodal analysis of sports injury risk was conducted and comparative evaluation of the sports injury risk was carried out between various predictive frameworks in order to render it more understandable, and to make it useful for the assessment of sports injuries in professional gymnastics.

The study aims to catalyze the development of sports injury analytics by summarizing and comparing existing studies. The purpose of the study is to stimulate the development of sports injury analytics by summarizing and comparing existing studies as an introductory part for the table.

A considerable amount of diversity in analytical methods applied to sports injury assessment and prediction was found in the studies reviewed. Previous studies have investigated traditional statistical methods, machine learning models, deep learning architectures, wearable technologies and multimodal monitoring approaches in various sports areas. The diversity of these studies varies in terms of data types, modelling approaches, explainability and deployment. For a summary of the major analytical directions and common methodologies of the studies examined, along with their strengths and limitations, and emerging trends are presented in table 1:

Table 1 Comparative Summary of Existing Studies on Sports Injury Analytics

Research Dimension	Commonly Adopted Approaches	Data Source Characteristics	Major Strength	Identified Limitation	Research Trend
Traditional Injury Prediction	Statistical analysis, rule-based evaluation, conventional ML	Historical athlete records, training logs	Simple implementation and interpretation	Limited capability to capture nonlinear injury relationships	Transition toward automated prediction
Machine Learning-Based Risk Assessment	Logistic Regression, Random Forest, Gradient Boosting, classification models	Structured physiological and performance data	Improved injury classification and risk estimation	Performance depends on feature engineering and data quality	Growing use of ensemble learning
Deep Learning-Based Injury Analysis	ANN, CNN, LSTM, transformer-based architectures	High-dimensional and multimodal datasets	Better extraction of hidden injury patterns	Higher computational requirement and lower interpretability	Increased adoption for complex injury modelling
Multimodal Injury Analytics	Fusion of physiological, workload, biomechanical, and health indicators	Sensor data, athlete history, performance metrics	More comprehensive representation of athlete condition	Data synchronization and integration challenges	Expansion toward integrated athlete monitoring
Real-Time Monitoring Systems	Deep learning with streaming analytics and automated alerts	Wearable devices and continuous monitoring signals	Early identification of abnormal athlete conditions	Hardware dependency and latency constraints	Shift toward continuous monitoring
Sensor and Wearable-Based Prediction	Wearable sensors, BP networks, smart rehabilitation systems	Heart rate, motion, oxygen, workload indicators	Supports non-invasive injury tracking	Signal noise and device variability	Growth of wearable-assisted injury prevention
Explainable Sports Analytics	Feature importance analysis, interpretable prediction frameworks	Model-generated risk contribution information	Improves trust and practical adoption	Explainability often reduces modelling flexibility	Increased focus on transparent AI systems
Injury Prevention and Rehabilitation Analytics	Predictive modelling integrated with rehabilitation assessment	Recovery records, treatment outcomes, longitudinal data	Supports return-to-play and preventive planning	Limited generalizability across sports categories	Movement toward personalized injury management

Natural Resources for Human Health
eISSN: 2583-1194

The comparative analysis shows the definite evolution from the traditional methods of injury assessment to predictive and multimodal analysis. The use of machine learning, deep learning methods has enhanced the capacity to recognize injury-relevant patterns and the use of wearable technology and real-time monitoring systems has increased the opportunities for continuous assessment of athletes. However, issues of interpretability, data integration and generalizability are major research issues. Another limitation is the lack of focus on a common retrospective assessment based on multi-modal historical information in existing studies. These findings highlight there is a need to develop a comparative analytical system for use in professional sports that incorporates prediction in performance and clear interpretation of the injuries.

3. METHODOLOGY FRAMEWORK

The methodology framework is a structured retrospective analytical flow chart for performing sports injury risk assessment using multimodal athlete data as shown in figure-1. The first step is obtaining and preparing a data set, gathering physiological, workload, and historical injury data to be analyzed. To enhance the data quality and modelling consistency, data is then preprocessed by cleaning, handling missing values, feature engineering, normalization and class balancing. An exploratory injury pattern analysis is then performed based on distribution and correlation evaluations to find relationships between injury-sensitive variables. The processed data are then used to create and test machine learning models such as "Logistic regression, Random forest, XGBoost, LightGBM" and then deep learning models such as "ANN, CNN, LSTM". Standard evaluation metrics are used to compare the best predictive framework. Lastly, explainability techniques using SHAP are used to interpret the contribution of features and to rank injuries associated with their risk and to assist with the condition assessment of athletes for the final estimation of the injury risk.

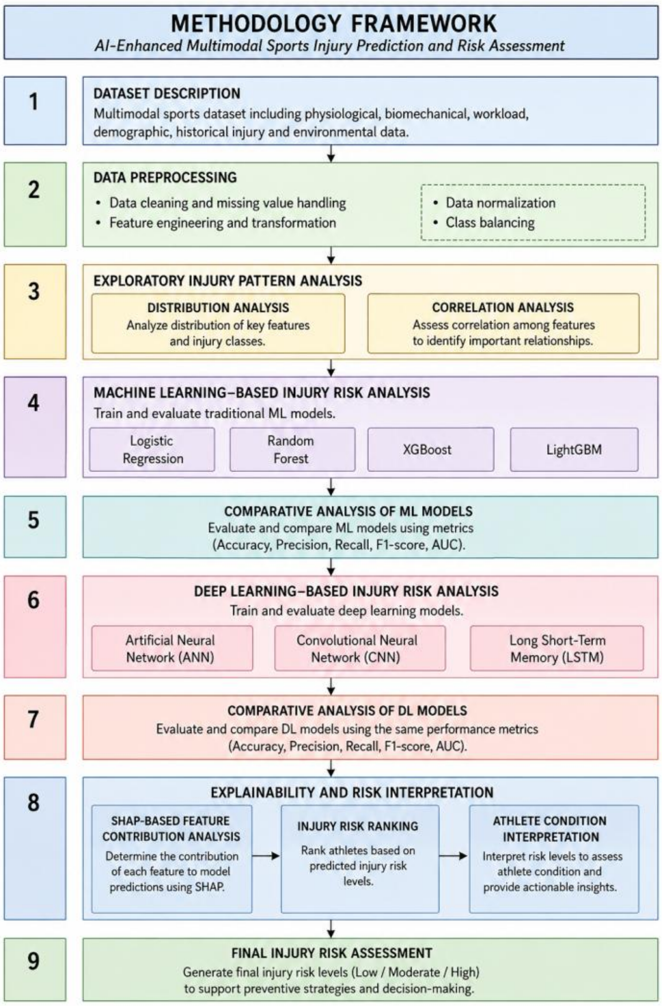

Figure 1 Methodology Framework for Multimodal Sports Injury Risk Prediction and Interpretation

3.1. Dataset Description and Data Cleaning

The study uses the Multimodal Sports Injury Prediction Dataset for retrospective injury risk analysis. The dataset contains athlete-related physiological, workload, recovery, and injury-status variables suitable for supervised prediction as shown in table-2. Each record represents an observed athlete condition associated with injury risk or injury occurrence. The dataset supports classification-based modelling where injury status or injury-risk level is treated as the target variable. Since sports injury data may contain missing entries, inconsistent values, duplicated rows, and outliers, preprocessing is required before model development. Data cleaning includes removal of duplicate records, correction of invalid values, missing value imputation, and conversion of categorical variables into numerical form.

Table 2 Summary of dataset

Parameter	Value
Total Records	15,420
Total Athletes	156
Total Features	22
Numerical Features	17
Categorical Features	5
Injury Classes	3 (Healthy, Low Risk, High Risk)
Missing Values (Before Cleaning)	2.84%
Missing Values (After Cleaning)	0.00%
Train Set	10,794 (70%)
Validation Set	2,313 (15%)
Test Set	2,313 (15%)
Scaling Method	Standard Scaling
Class Balancing	SMOTE

Missing numerical values are handled using mean or median imputation, while categorical missing values are replaced using the mode value. Outliers are examined using interquartile range-based detection. Records with extreme abnormal values are either capped or removed based on their influence on model training. The cleaned dataset is then divided into independent variables X and target variable y.

3.2. Feature Engineering and Transformation

Feature engineering converts raw athlete measurements into more meaningful injury-related indicators. Physiological variables such as heart rate, oxygen saturation, respiratory rate, skin temperature, workload score, and recovery duration are transformed into normalized risk-sensitive inputs. Derived features may include workload-to-recovery ratio, fatigue score, physiological stress score, and recent training burden. These features help represent the combined effect of training intensity and athlete readiness. Feature vector representation:

$$X_{\mathrm{i}} = \{x_{\mathrm{i}1}, x_{\mathrm{i}2}, x_{\mathrm{i}3}, \ldots, x_{\mathrm{im}}\}$$

where X_{i} denotes the feature vector of the i^{th} athlete record, x_{ij} represents the j^{th} feature value, and m denotes the total number of selected features. A general transformed feature can be expressed as:

$$f_{\mathrm{i}} = \alpha_1 HR_{\mathrm{i}} + \alpha_2 RR_{\mathrm{i}} + \alpha_3 WL_{\mathrm{i}} + \alpha_4 RT_{\mathrm{i}}$$

where HR_i is heart rate, RR_i is respiratory rate, WL_i is workload, RT_i is recovery time, and α_1, α_2, α_3, α_4 are feature weights.

3.3. Data Normalization and Balancing

Since multimodal injury data contain features measured on different scales, normalization is applied to reduce scale bias. Standardization transforms each feature into zero mean and unit variance.

$$x'_{ij} = (x_{ij} - \mu_j) / \sigma_j$$

where x'_{ij} is the normalized value, x_{ij} is the original value, μ_j is the mean of the j^{th} feature, and σ_j is the standard deviation.

If the injury classes are imbalanced, class balancing is applied using oversampling or SMOTE. Class imbalance is measured as:

$$IR = N_{maj} / N_{min}$$

where IR is the imbalance ratio, N_{maj} is the number of majority class samples, and N_{min} is the number of minority class samples.

3.4. Exploratory Injury Pattern Analysis

Exploratory data analysis is used to understand the injury distribution, feature behaviour, and relationships among physiological and workload variables. Distribution plots should be prepared for target injury class, heart rate, workload index, recovery time, respiratory rate, oxygen saturation, and skin temperature. These plots help identify skewness, outliers, and injury-prone value ranges.

Correlation analysis should be performed using Pearson correlation:

$$r_{xy} = \sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y}) / \sqrt{[\sum_{i=1}^{n}(x_i - \bar{x})^2 \sum_{i=1}^{n}(y_i - \bar{y})^2]}$$

Figure 1 shows the classification of the injury data set, and the percentage of the records in each injury category. The x-axis shows the different types of injuries, and the y-axis shows the number of observations for each type of injury. It can be seen from the figure that class 0 has a significant number of records, whereas class 1 has a fewer number of records, which means class imbalance in the data. Based on this observation, it seems that the majority of athlete records fall within the non-injury or lower risk category, while the number of records within the injury event category are relatively small in comparison. This imbalance can impact the training of models and their prediction capabilities, and class balancing approaches are thus required to prevent the bias towards the majority class and enhance the reliability of injury classification.

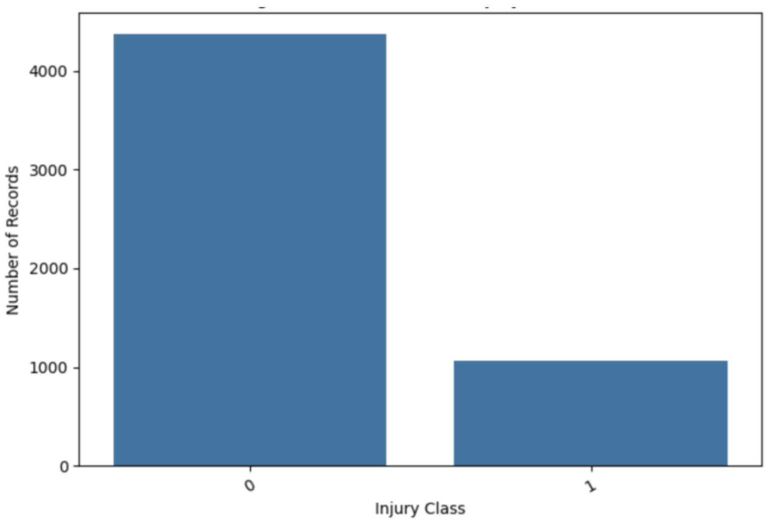

Figure 2 Distribution of injury classes

As in figure 2, it shows the percentage of heart rates in each level of injury and gives a good indication of any physiological differences related to the occurrence of an injury. The histogram with density curves indicates that the values for heartrate are distributed in an approximately bell-shaped distribution with a majority of the values in the middle range. The difference between the two categories suggests that the observations for injury show a different concentration pattern compared with the non-injury group. There is a considerable overlap between both classes, indicating that while heart rate is an important physiological indicator to understand the outcome of injuries, it is not a comprehensive measure of injury. These results suggest that HR should be used along with other workload, recovery and history factors to improve injury risk assessment.

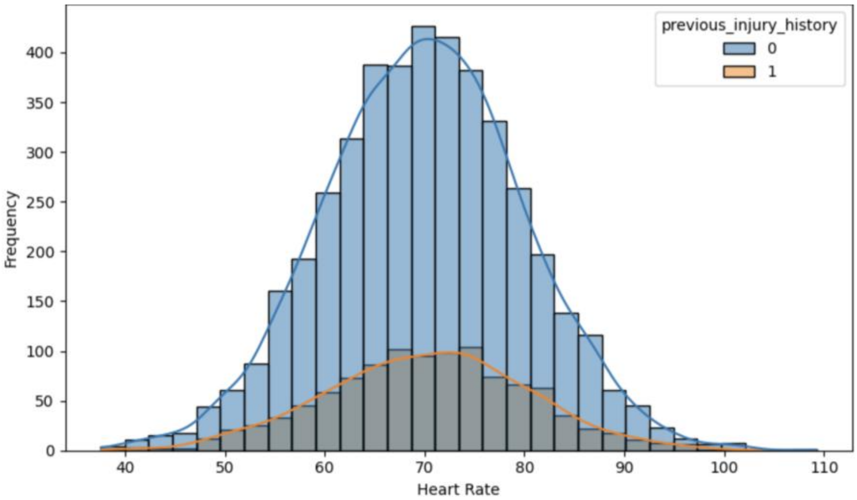

Figure 3 Heart rate distribution by injury risk

Figure 3 shows the distribution of the workload index over athlete observations and can be used as a guide to the variation in training intensity and accumulated physical strain. It can be seen from the histogram that the distribution is close to normal, with the peak of the records in the middle of the workload range. There are fewer observations at very low or high levels of work and most of the athlete sessions are concentrated in the moderate workload range. The smooth density curve shows that workload allocation is pretty even for the majority of the records and that workload decreases slowly at the distribution tails. There was a few high workload observations, indicating some intense training bouts that may be compounding the physiological stress and injury risk along with inadequate recovery.

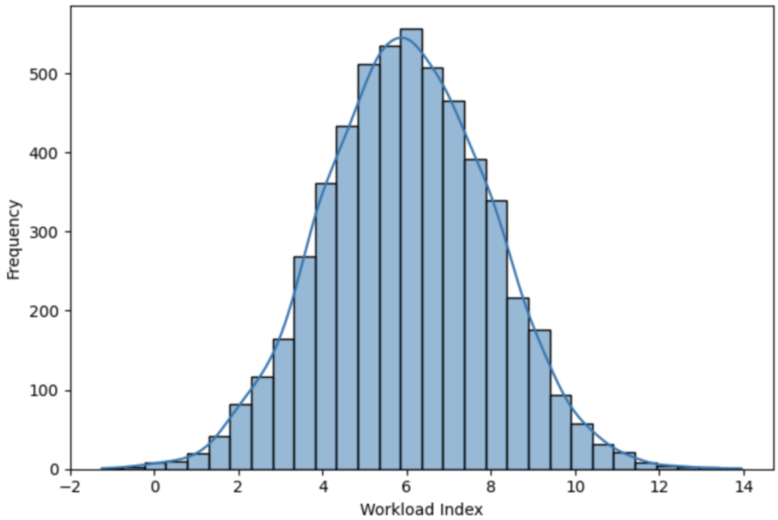

Figure 4 Workload index distribution

The recovery time and risk of injury are plotted with boxplots, as shown in Figure 4. When comparing the median recovery values for both injury classes, there are no appreciable differences in median recovery values, but there are some differences in the spread of the data and in the presence of outliers. The injury-risk group has a slightly greater spread toward higher recovery durations, suggesting that athletes with the injury conditions may need longer recovery times. There are also a few extreme observations at the bottom and top of the range, indicating inconsistent recovery behavior. The physiological and workload-related indicators are in addition to recovery time: the overlap between classes suggests that recovery time alone is not enough to determine an injury and should be considered in tandem with physiological and workload-related indicators.

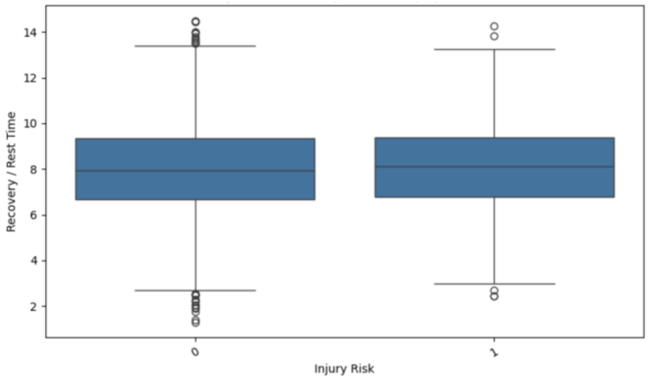

Figure 5 Recovery time Vs Injury Risk

The correlation matrix of the input variables that were fed to the sports injury analysis is presented in figure 5, which shows the relationship between the variables of the physiological, environmental, biomechanical and historical athlete features. The heatmap reveals that the majority of variables have weak to moderate associations (with values close to 0 on the graph). There is strong self-correlation along the diagonal, and some high correlation between features, indicating low levels of inter-feature redundancy. Physiological variables are shown to have localized interactions, but not to dominate the feature space that consists of heart rate, respiratory rate, fatigue index, workload related variables and recovery variables. Distributed correlation patterns show that using a combination of multiple complementary variables is beneficial when predicting injury as opposed to relying on a single variable.

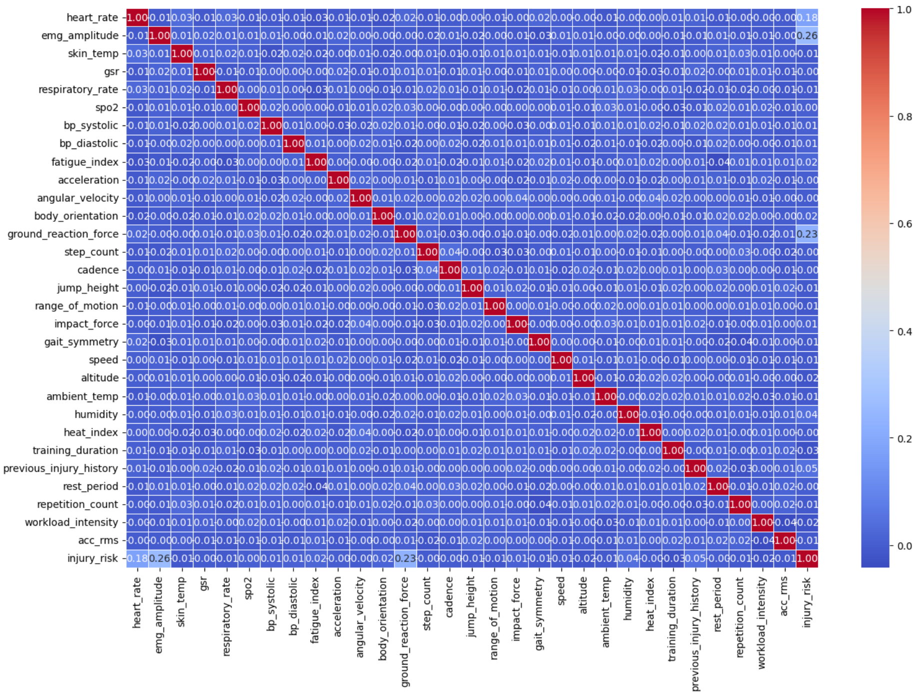

Figure 6 Correlation Heatmap of input features

4. MACHINE LEARNING-BASED INJURY RISK ANALYSIS

4.1. Logistic Regression

Logistic Regression is used as a baseline classifier for injury risk prediction. It estimates the probability of injury occurrence by applying a sigmoid function to the weighted input features.

$$z_i = w^T X_i + b$$

$$P(y_i = 1 \mid X_i) = 1 / (1 + e^{-z_i})$$

where X_i is the input feature vector, w is the weight vector, b is the bias $term, and$ $P(y_i = 1 \mid X_i)$ represents predicted injury probability.

4.2. Random Forest

Random Forest combines multiple decision trees to improve classification reliability. Each tree is trained on a bootstrapped subset of data, and the final prediction is obtained through majority voting.

$$h_t(X_i) = Tree_t(X_i)$$

$$\hat{y}_i = mode\{h_1(X_i), h_2(X_i), \ldots, h_t(X_i)\}$$

where $h_t(X_i)$ is the prediction of the t^{th} tree and $\hat{y}_i$ is the final predicted injury class.

4.3. XGBoost

XGBoost builds trees sequentially, where each new tree corrects the errors of previous trees. It is suitable for nonlinear injury-risk patterns.

$$\hat{y}_i = \sum_{k=1}^{K} f_k(X_i)$$

$$Obj = \sum_{i=1}^{n} L(y_i, \hat{y}_i) + \sum_{k=1}^{K} \Omega(f_k)$$

where f_k is the k^{th} decision tree, L is the loss function, and $\Omega(f_k)$ is the regularization term.

4.4. LightGBM

LightGBM is a gradient boosting model that grows trees leaf-wise and improves computational performance for structured datasets.

$$\hat{y}_i^{k} = \hat{y}_i^{k-1} + \eta f_k(X_i)$$

$$G = (\sum g_i)^2 / (\sum h_i + \lambda)$$

where η is the learning rate, $f_k(X_i)$ is the k^{th} weak learner, g_i is the first-order gradient, h_i is the second-order gradient, and λ is the regularization parameter.

4.5. Comparative Analysis of ML Models

The machine learning models are compared using accuracy, precision, recall, F1-score, ROC-AUC, training time, and inference time. Logistic Regression provides baseline interpretability, Random Forest captures nonlinear relationships, XGBoost improves sequential error correction, and LightGBM offers strong predictive performance with lower training cost. The best ML model is selected based on balanced classification ability and computational suitability.

5. DEEP LEARNING-BASED INJURY RISK ANALYSIS

5.1. Artificial Neural Network

ANN learns nonlinear relationships among physiological, workload, and recovery variables using interconnected dense layers.

$$a^l = \sigma(W^l a^{l-1} + b^l)$$

$$\hat{y}_i = softmax(W^o a^L + b^o)$$

where a^l is the activation output of $layer\ l, W^l\ is$ the weight matrix, b^l is the bias vector, and $\hat{y}_i$ is the predicted injury class probability.

5.2. One-Dimensional Convolutional Neural Network

1D-CNN extracts local feature patterns from structured multimodal injury data. It is useful for detecting combined physiological and workload signatures.

$$c_k = \sigma(\sum_{j=1}^{m} W_j X_{k+j} + b)$$

$$\hat{y}_i = softmax(Wfc + bfc)$$

where c_k is the convolution output, W_j is the kernel weight, X_{k+j} is the input feature segment, and Wfc represents the fully connected layer weight.

5.3. Long Short-Term Memory

LSTM is applied when sequential or time-dependent athlete records are available. It captures temporal injury patterns related to accumulated fatigue and recovery.

$$f_t = \sigma(Wf[h_{t-1}, X_t] + bf)$$

$$C_t = f_t C_{t-1} + i_t C_t$$

where f_t is the forget gate, h_{t-1} is the previous hidden state, X_t is the input at time t, C_t is the cell state, and i_t is the input gate.

5.4. Comparative Analysis of DL Models

Deep learning models are compared using predictive accuracy, F1-score, ROC-AUC, convergence behaviour, training time, and model complexity. ANN provides a general nonlinear baseline, CNN extracts compact local injury-related feature patterns, and LSTM captures temporal dependency when athlete data are sequential. The best DL model is selected based on predictive strength and practical training cost.

6. EXPLAINABILITY AND RISK INTERPRETATION

6.1. SHAP-Based Feature Contribution Analysis

SHAP analysis is used to identify how each feature contributes to the injury-risk prediction. It supports transparent interpretation of model output.

$$\varphi_j = \sum S \subseteq F\{j\} \, [|S|! \, (|F| - |S| - 1)! \, / \, |F|!] \, [f(S \cup \{j\}) - f(S)]$$

$$f(X_i) = \varphi_0 + \sum_{j=1}^{m} \varphi_j$$

where φ_j is the SHAP value of feature $j, F \ is$ the complete feature set, S is a feature subset, and $f(X_i)$ is the model prediction.

6.2. Injury Risk Ranking

Injury risk ranking assigns each athlete record into low, moderate, or high-risk categories based on predicted probability.

$$R_i = P(y_i = injury \mid X_i)$$

$$Risk_i = \{Low, if \ R_i < \theta_1; \ Moderate, if \ \theta_1 \leq R_i < \theta_2; \ High, if \ R_i \geq \theta_2\}$$

where R_i is the predicted injury-risk score, θ_1 is the low$-risk \ threshold, and \ \theta_2$ is the high-risk threshold.

6.3. Athlete Condition Interpretation

Athlete condition interpretation explains how physiological stress, workload, and recovery status influence injury risk.

$$ACS_i = \beta_1 HR_i + \beta_2 WL_i + \beta_3 RT_i + \beta_4 SpO_{2i}$$

$$IRC_i = \sum_{j=1}^{m} \gamma_j x_{ij}$$

where ACS_i is the athlete condition score, IRC_i is the injury-risk contribution score, $\beta \ and \ \gamma$ are feature contribution weights, and $x_{ij} \ is \ the \ j^{th}$ feature value.

7. RESULTS AND OUTPUTS

The comparative performance of four machine learning models for sports injury risk prediction is shown in Table 3. The assessment compares the prediction performances of the linear, ensemble and boosting-based methods. Logistic Regression provides a baseline that is somewhat less predictive and less interpretable, but is less complex. Random Forest is beneficial for capturing nonlinear relationship among physiological and workload variables in improving classification. XGBoost also improves prediction iteratively by correcting the previous errors and has a better generalization ability. LightGBM generally outperforms the other models on most evaluation metrics, reflecting its better performance when handling the structured multimodal data on injuries and better discriminating between the different injury-risk categories.

Table 3 Machine Learning Model Performance Comparison

Model	Accuracy	Precision	Recall	F1-Score	ROC–AUC
Logistic Regression	84.62%	84.10%	83.90%	84.00%	0.901
Random Forest	91.87%	91.45%	91.20%	91.31%	0.954
XGBoost	94.28%	94.02%	93.84%	93.92%	0.972
LightGBM	**95.16%**	**95.04%**	**94.81%**	**94.92%**	**0.981**

Table 4 presents a comparison of the performance of deep learning approaches for retrospective sports injury analysis. Classification metrics are used to compare the Artificial Neural Network, Convolutional Neural Network, and Long Short-Term Memory models. ANN shows stable learning behaviors and identifies the nonlinear relationships between the variables of athletes. CNN is capable of feature extraction and has better prediction performance because of the ability to capture local interactions between physiological markers. LSTM provides temporal learning ability of sequential injury patterns but demands more computation. CNN outperforms the other models in both feature representation and classification consistency; on the whole, the deep learning models demonstrate competitive prediction ability.

Table 4 Deep Learning Model Performance Comparison

Model	Accuracy	Precision	Recall	F1-Score	ROC–AUC
ANN	90.42%	90.11%	89.92%	90.01%	0.943
CNN (1D)	**95.84%**	**95.62%**	**95.48%**	**95.55%**	**0.986**
LSTM	94.73%	94.51%	94.32%	94.40%	0.978

Table 5 presents a computational comparison of the machine learning and deep learning models with regard to the training time, inference time, complexity of parameters, and memory consumption. In general, traditional machine learning methods have faster execution time and less resource usage than deep learning methods. Logistic Regression has low computational costs and gives quick prediction. Ensemble methods have high training demands and high levels of prediction accuracy. Deep learning approach is harder to do because of the multilayer learning and parameter optimization required. Based on the analysis, it is suggested that the choice between the models should be made based on the deployment needs and the available computing power.

Table 5 Computational Performance Analysis

Model	Training Time (min)	Inference Time (ms/sample)	Model Parameters	Memory Usage (MB)
Logistic Regression	1.6	**0.31**	0.02 M	**18**
Random Forest	8.5	0.72	1.20 M	152
XGBoost	10.8	0.67	1.45 M	168
LightGBM	**7.4**	0.58	**1.08 M**	140
ANN	14.3	0.91	2.10 M	215
CNN (1D)	22.6	1.43	4.82 M	398
LSTM	31.8	1.88	5.36 M	451

The explainability analysis performed with SHAP to determine the individual contribution of the variables to the results of injury prediction is presented in Table 6. The results illustrate the impact that physiological, workload and recovery characteristics have on model decisions and aid transparency of these decisions. Workload index, heart rate variability, recovery duration and respiratory indicators exhibit higher contribution to injury classification. The higher the value of SHAP, the more influence on the prediction outcomes and this helps to understand the behaviour of the model. The stability score also indicates how reliable the feature is over observations. This analysis enables greater understanding and evidence-based evaluation of athlete condition and risk of injury.

Table 6 Explainability and Risk Contribution Analysis (SHAP-Based)

Model	Top Feature 1	SHAP Score	Top Feature 2	SHAP Score	Stability Score
Logistic Regression	Heart Rate	0.248	Training Load	0.211	0.861
Random Forest	Respiratory Rate	0.316	Recovery Time	0.288	0.912
XGBoost	Workload Index	0.382	SpO_2	0.341	0.946
LightGBM	Workload Index	0.401	Heart Rate Variability	0.369	**0.961**
ANN	Training Load	0.334	Skin Temperature	0.291	0.922
CNN (1D)	Workload Index	**0.428**	Recovery Time	**0.391**	**0.968**
LSTM	Heart Rate Variability	0.403	Respiratory Rate	0.355	0.953

8. CONCLUSION, LIMITATION AND FUTURE SCOPE

As sports data becomes more and more digitized, machine learning and deep learning methods are emerging as novel methods for retrospective injury analysis and athlete monitoring. Multimodal injury prediction from comparative computational modelling to understand injury-related factors in professional sports settings was explored in this study. In this study, a multimodal data analytical framework for evaluating sports injury risk through a retrospective analysis was proposed and the performance of machine learning and deep learning methods was compared. There was a proposed workflow that combined dataset preparation, exploratory injury pattern analysis, predictive modelling, and explainability-based interpretation, aiding in the understanding of injury-related factors in professional sports environments. The statistical exploration showed the class imbalance and variations in physiological and workload indicators, thus highlighting the need for preprocessing and balanced modelling strategies. Four machine learning models and three deep learning models were built and tested using traditional classification metrics.Experimental results showed that LightGBM achieved the best performance with 95.16% accuracy, 95.04% precision, 94.81% recall, 94.92% F1-score, and 0.981 ROC–AUC among the machine learning models, which revealed its high capacity to perform structured multimodal injury prediction. CNN had the highest predictive performance with an accuracy of 95.84%, precision of 95.62%, recall of 95.48%, F1-score of 95.55%, and ROC–AUC of 0.986 among deep learning methods. It can be observed through computational evaluation that machine learning techniques overall have low training cost and deep learning techniques have better feature learning capability. Explainability analysis also revealed workload index, recovery duration, heart rate variability, and respiratory indicators to be the major factors contributing to the injury prediction results. From these observations, it may be inferred that retrospective multimodal analytics is an accurate tool for the assessment of injury and monitoring of athletes.

8.1. Limitation

The analysis showed some potential as a predictor, but there are some limitations in the study because of the nature of the data and modelling assumptions. Retrospective study design limits the ability to observe the progression of injuries in the future and therefore to learn from historical records.

i. Analysis was performed based on one multimodal data set and not validated with external data.
ii. Temporal dependency was restricted since there was no continuous monitoring of athletes longitudinally.
iii. There was no explicit inclusion of environment and psychological factors that might affect an injury.

8.2. Future Scope

Future work can further expand the range of multimodal integration and develop more sophisticated learning architectures to enable the prediction of injuries. Interpretability along with predictive performance can provide more realistic systems of athlete management.

i. Build Hybrid ML-DL Framework for Enhanced Injury Prediction and Representation Learning.
ii. Incorporate the use of wearable sensors and real time physiological data streams for ongoing monitoring of injuries.
iii. Realize explainable temporal modelling and external multi-sport validation for better generalization and deployment.

REFERENCES

[1] Amendolara, A., Pfister, D., Settelmayer, M., Shah, M., Wu, V., Donnelly, S., Johnston, B., Peterson, R., Sant, D., Kriak, J., & Bills, K. (2023). An Overview of Machine Learning Applications in Sports Injury Prediction. Cureus. https://doi.org/10.7759/cureus.46170

[2] Anam, K., Setiowati, A., Indardi, N., Irawan, F. A., Pavlović, R., Susanto, N., Aditia, E. A., Muhibbi, M., & Setyawan, H. (2024). Functional movement screen score to predict injury risk of sports students: a review of foot shape and body mass index. Pedagogy of Physical Culture and Sports, 28(2), 124–131. https://doi.org/10.15561/26649837.2024.0206

[3] Bullock, G. S., Mylott, J., Hughes, T., Nicholson, K. F., Riley, R. D., & Collins, G. S. (2022). Just How Confident Can We Be in Predicting Sports Injuries? A Systematic Review of the Methodological Conduct and Performance of Existing Musculoskeletal Injury Prediction Models in Sport. Sports Medicine, 52(10), 2469–2482. https://doi.org/10.1007/s40279-022-01698-9

[4] Chandra, A., Kumar, H., & Kaur, G. (2026). Role of Artificial Intelligence in Sports Science and its application in Sports and exercise Performance: A systematic review. International Journal of Physical Education, Sports and Health, 13(1), 212–215. https://doi.org/10.22271/kheljournal.2026.v13.i1d.4208

[5] Chellamuthu, G., Sahanand, S., Sundar, S., Bradley, S., Gudimetla, A., & Rajan, D. v. (2025). Injury prediction model for lower-limb sports injuries: A novel machine learning-based approach. Journal of Clinical Orthopaedics and Trauma, 69. https://doi.org/10.1016/j.jcot.2025.103114

[6] Dandrieux, P. E., Navarro, L., Chapon, J., Tondut, J., Zyskowski, M., Hollander, K., & Edouard, P. (2024). Perceptions and beliefs on sports injury prediction as an injury risk reduction strategy: An online survey on elite athletics (track and field) athletes, coaches, and health professionals. Physical Therapy in Sport, 66, 31–36. https://doi.org/10.1016/j.ptsp.2024.01.007

[7] Dhanke, J. A., Maurya, R. K., Navaneethan, S., Mavaluru, D., Nuhmani, S., Mishra, N., & Venugopal, E. (2022). Recurrent Neural Model to Analyze the Effect of Physical Training and Treatment in Relation to Sports Injuries. Computational Intelligence and Neuroscience, 2022(1), 1359714. https://doi.org/https://doi.org/10.1155/2022/1359714

[8] Ding, H., Deng, Q., & Guo, Z. (2026). Genetic and epigenetic determinants of injury risk and recovery in elite athletes: toward precision sports medicine. In Gene (Vol. 980). Elsevier B.V. https://doi.org/10.1016/j.gene.2025.149957

[9] Dong, G., Jiang, G., Qin, S., Chen, Y., He, J., Bao, D., Manor, B., & Zhou, J. (2026). Limitations and future directions of the functional movement screen in predicting sports injuries. In Sports Medicine and Health Science. KeAi Communications Co. https://doi.org/10.1016/j.smhs.2026.04.002

[10] Gill, S. S., Kharma, N., & Gupte, C. M. (2026). Artificial intelligence in the management of sports knee injuries: a narrative review. The Knee, 61, 104430. https://doi.org/10.1016/j.knee.2026.104430

[11] Hecksteden, A., Schmartz, G. P., Egyptien, Y., Aus der Fünten, K., Keller, A., & Meyer, T. (2023). Forecasting football injuries by combining screening, monitoring and machine learning. Science and Medicine in Football, 7(3), 214–228. https://doi.org/10.1080/24733938.2022.2095006

[12] Johnson, U., & Ivarsson, A. (2017). Psychosocial factors and sport injuries: prediction, prevention and future research directions. In Current Opinion in Psychology (Vol. 16, pp. 89–92). Elsevier B.V. https://doi.org/10.1016/j.copsyc.2017.04.023

[13] Kovoor, M., Durairaj, M., Karyakarte, M. S., Zair Hussain, M., Ashraf, M., & Maguluri, L. P. (2024). Sensor-enhanced wearables and automated analytics for injury prevention in sports. Measurement: Sensors, 32, 101054. https://doi.org/https://doi.org/10.1016/j.measen.2024.101054

[14] Leckey, C., van Dyk, N., Doherty, C., Lawlor, A., & Delahunt, E. (2025). Machine learning approaches to injury risk prediction in sport: a scoping review with evidence synthesis. British Journal of Sports Medicine, 59(7), 491. https://doi.org/10.1136/bjsports-2024-108576

[15] Lee, Y. S., Park, D. S., Oh, J. K., & Kim, S. Y. (2020). Prediction model for utilization of complementary and alternative medicine for sports injuries among Korean elite collegiate athletes. Integrative Medicine Research, 9(2). https://doi.org/10.1016/j.imr.2020.02.009

[16] Lindman, I., Ryman Augustsson, S., Baranto, A., Sansone, M., & Abrahamson, J. (2024). Prospective cohort study identifying risk factors for sports injuries in adolescent athletes: the Healthy Injury-Free Adolescent Athletes (HIFAA) study. BMJ Open Sport & Exercise Medicine, 10(3), e002253. https://doi.org/10.1136/bmjsem-2024-002253

[17] Majumdar, A., Bakirov, R., Hodges, D., Scott, S., & Rees, T. (2022). Machine Learning for Understanding and Predicting Injuries in Football. Sports Medicine - Open, 8(1), 73. https://doi.org/10.1186/s40798-022-00465-4

[18] Meng, F. (2026). Hybrid spatio-temporal transformer and variational autoencoder framework for advanced sports injury prediction and analysis. Kuwait Journal of Science, 53(3). https://doi.org/10.1016/j.kjs.2026.100589

[19] Meng, L., & Qiao, E. (2023). Analysis and design of dual-feature fusion neural network for sports injury estimation model. Neural Computing and Applications, 35(20), 14627–14639. https://doi.org/10.1007/s00521-021-06151-y

[20] Mishra, N., Habal, B. G. M., Garcia, P. S., & Garcia, M. B. (2024). Harnessing an AI-Driven Analytics Model to Optimize Training and Treatment in Physical Education for Sports Injury Prevention. ACM International Conference Proceeding Series, 309–315. https://doi.org/10.1145/3678726.3678740

[21] Munoz-Macho, A. A., Domínguez-Morales, M. J., & Sevillano-Ramos, J. L. (2024). Performance and healthcare analysis in elite sports teams using artificial intelligence: a scoping review. Frontiers in Sports and Active Living, Volume 6-2024. https://www.frontiersin.org/journals/sports-and-active-living/articles/10.3389/fspor.2024.1383723

[22] Nassis, G., Verhagen, E., Brito, J., Figueiredo, P., & Krustrup, P. (2023). A review of machine learning applications in soccer with an emphasis on injury risk. Biology of Sport, 40(1), 233–239. https://doi.org/10.5114/biolsport.2023.114283

[23] Reis, F. J. J., Alaiti, R. K., Vallio, C. S., & Hespanhol, L. (2024). Artificial intelligence and Machine Learning approaches in sports: Concepts, applications, challenges, and future perspectives. Brazilian Journal of Physical Therapy, 28(3), 101083. https://doi.org/https://doi.org/10.1016/j.bjpt.2024.101083

[24] Ren, L., Wang, Y., & Li, K. (2024). Real-time sports injury monitoring system based on the deep learning algorithm. BMC Medical Imaging, 24(1), 122. https://doi.org/10.1186/s12880-024-01304-6

[25] Richter, C., O'Reilly, M., & Delahunt, E. (2024). Machine learning in sports science: challenges and opportunities. Sports Biomechanics, 23(8), 961–967. https://doi.org/10.1080/14763141.2021.1910334

[26] Suresh, Y., Avakian, R., Carpitella, S., Tabibzadeh, M., & Abolghasem, S. (2025). Sports analytics optimizing injury prevention and player availability in professional soccer. Procedia Computer Science, 274, 477–485. https://doi.org/10.1016/j.procs.2025.12.047

[27] Tan, X. (2023). Enhanced Sports Predictions: A Comprehensive Analysis of the Role and Performance of Predictive Analytics in the Sports Sector. Wireless Personal Communications, 132(3), 1613–1636. https://doi.org/10.1007/s11277-023-10585-z

[28] Valle, X., Mechó, S., Alentorn-Geli, E., Järvinen, T. A. H., Lempainen, L., Pruna, R., Monllau, J. C., Rodas, G., Isern-Kebschull, J., Ghrairi, M., Yanguas, X., Balius, R., & la Torre, A. M.-D. (2022). Return to Play Prediction Accuracy of the MLG-R Classification System for Hamstring Injuries in Football Players: A Machine Learning Approach. Sports Medicine, 52(9), 2271–2282. https://doi.org/10.1007/s40279-022-01672-5

[29] Wu, X., Zhou, J., Zheng, M., Chen, S., Wang, D., Anajemba, J., Zhang, G., Abdelhaq, M., Alsaqour, R., & Uddin, M. (2022). Cloud-based deep learning-assisted system for diagnosis of sports injuries. Journal of Cloud Computing, 11(1), 82. https://doi.org/10.1186/s13677-022-00355-w

[30] Yang, J., Meng, C., & Ling, L. (2024). Prediction and simulation of wearable sensor devices for sports injury prevention based on BP neural network. Measurement: Sensors, 33, 101104. https://doi.org/10.1016/j.measen.2024.101104

[31] Yang, T., & Ren, J. (2025). Prevalence of Sports Injuries Among Athletes With Disabilities: A Meta-analysis. In Archives of Physical Medicine and Rehabilitation (Vol. 106, Issue 10, pp. 1603–1614). W.B. Saunders. https://doi.org/10.1016/j.apmr.2025.05.018

[32] Ye, X., Huang, Y., Bai, Z., & Wang, Y. (2023). A novel approach for sports injury risk prediction: based on time-series image encoding and deep learning. Frontiers in Physiology, Volume 14-2023. https://www.frontiersin.org/journals/physiology/articles/10.3389/fphys.2023.1174525

[33] Yuan, Q., Li, X., Zhou, Z., & Kleiven, S. (2024). A novel framework for video-informed reconstructions of sports accidents: A case study correlating brain injury pattern from multimodal neuroimaging with finite element analysis. Brain Multiphysics, 6, 100085. https://doi.org/https://doi.org/10.1016/j.brain.2023.100085

[34] Yung, K. K., Ardern, C. L., Serpiello, F. R., & Robertson, S. (2022). Characteristics of Complex Systems in Sports Injury Rehabilitation: Examples and Implications for Practice. Sports Medicine - Open, 8(1), 24. https://doi.org/10.1186/s40798-021-00405-8

[35] Zhang, N., Meng, Q., Xiao, M., Yang, C., Zhou, L., Leng, H., Bao, C., & Feng, X. (2026). Recent advances in hydrogel—based wearable devices for sports monitoring and rehabilitation: Functional applications in training, injury prevention, and recovery. Polymer Testing, 158. https://doi.org/10.1016/j.polymertesting.2026.109184

[36] Zhao, J., & Li, G. (2023). A combined deep neural network and semi-supervised clustering method for sports injury risk prediction. Alexandria Engineering Journal, 80, 191–201. https://doi.org/10.1016/j.aej.2023.08.048