Molecular & Cellular Biomechanics 2025, 22(2), 1000.
https://doi.org/10.62617/mcb1000

Article

New paths to promote athletic injury prevention by integrating statistics and sports biomechanics

Yiming Zhao

School of Statistics, Jilin University of Finance and Economics, Changchun 130117, Jilin, China; Z2034345106@163.com

CITATION

Zhao Y. New paths to promote athletic injury prevention by integrating statistics and sports biomechanics. Molecular & Cellular Biomechanics. 2025; 22(2): 1000. https://doi.org/10.62617/mcb1000

ARTICLE INFO

Received: 4 December 2024
Accepted: 24 December 2024
Available online: 17 January 2025

Abstract: Athletic injuries are a common problem in sports. Due to the insufficient processing of multimodal biomechanical data by traditional prevention strategies, personalized risk prediction cannot be achieved. To this end, this paper adopts an athletic injury prevention method based on sparse principal component analysis (SPCA) and spatio-temporal graph convolutional network (ST-GCN). The Vicon Vantage V5 3D motion capture system and the Noraxon Ultium EMG electromyography acquisition device are used to obtain the athlete's joint angle change rate, ground reaction force (GRF) and electromyographic activity data, and the SPCA method is used to extract key biomechanical features, thereby reducing data redundancy and improving the representativeness of features. Subsequently, ST-GCN is used to construct a dynamic risk prediction model to capture the temporal changes and spatial dependencies in the motion sequence to achieve precise and efficient risk assessment. In the experimental verification, the prediction accuracy of the model reaches 95.3% when the number of features was 20, and the ability to provide risk feedback in real-time is realized to generate personalized injury prevention strategies. Studies have shown that the integration of statistics and sports biomechanics has effectively improved the efficiency of athletic injury prevention and provided new ideas for scientific and precise sports management.

Keywords: athletic injury prevention; sparse principal component analysis; spatio-temporal graph convolutional network; biomechanical data analysis; dynamic risk prediction

1. Introduction

Athletic injuries are an important issue that affects athletes' competitive ability and career life. How to scientifically and effectively prevent injuries has become one of the core issues in sports biomechanics research. With the increase in sports intensity [1,2] and complexity [3,4], traditional preventive measures are difficult to meet the personalized needs of athletes due to the lack of precise analysis of individual differences [5] and comprehensive understanding of multimodal data [6,7]. The current mainstream methods mainly rely on the basic theory and empirical judgment of sports biomechanics [8,9], and are highly dependent on biomechanical data [10,11]. However, it is difficult to fully explore the potential relationship between complex movements [12] and injury risks [13,14] in sports. Although data acquisition devices [15] are constantly being optimized, the multimodal data they collect has not been fully utilized, limiting the accuracy of risk prediction [16,17]. The redundancy of high-dimensional data and the complexity of potential features also further increase the difficulty of analysis, making it difficult to directly apply multimodal data to personalized injury risk modeling. For complex dynamic action sequences [18], existing analysis methods often ignore the correlation between time and space, making it difficult to precisely capture subtle changes in action patterns.

In the field of athletic injury prevention, existing research mainly relies on macroscopic biomechanical analysis such as joint angles and ground reaction forces. Cellular biomechanical properties are equally important for understanding the functions and injury mechanisms of tissues and organs. Integrating cellular and macroscopic biomechanical data provides a more comprehensive understanding of sports injuries. Cellular biomechanical changes reveal changes in the microstructure and function of tissues, which will predict or explain the occurrence and development of sports injuries. Therefore, exploring biomechanical changes at the cellular level will add depth to athletic injury prevention and reveal new prevention and treatment strategies. Athletic injury prevention strategies based on traditional methods are inefficient and cannot meet the needs of high-level competitive sports for precise risk assessment. Traditional athletic injury prevention strategies rely on standardized training plans and empirical evaluations. Static posture analysis only provides certain information, cannot reflect dynamic changes in exercise, and cannot accurately capture the movement deviations of athletes in competition, so it is difficult to effectively predict injury risks. Physical fitness tests and sports performance assessments are limited to single physiological data, ignoring changes in factors such as fatigue and environment during exercise, and cannot fully assess the actual risks of athletes. Although stretching and warm-up before and after exercise can help with short-term relief, they lack real-time monitoring of subtle changes in complex action sequences and cannot effectively intervene before injuries occur. In this context, finding a new method that integrates statistics and biomechanics to achieve precise and efficient athletic injury prevention has important theoretical and practical value.

In the field of athletic injury prevention, researchers have used a variety of statistical methods to try to find new solutions. Zhang [19] studied the athletic injury situation and risk prevention and control of college students by searching relevant literature, distributing questionnaires to college students and experts, and using SPSS (Statistical Product and Service Solutions) software for data statistics. He found that the longer students participated in sports, the less athletic injuries they suffered. This showed that strengthening physical education and improving students' awareness of sports risk prevention and control had positive significance for avoiding students' athletic injuries. Athletic injury prevention [20,21] depends on individual behavior and is closely related to the structural improvement of the education system. Keavanloo et al. [22] interviewed physics education teachers and constructed a theoretical model, and then conducted a questionnaire survey and structural equation modeling analysis on physics education students. They concluded that the incidence of athletic injuries among sports majors can be significantly reduced by improving enrollment, education planning, venue security, and teaching notification. With the increasing complexity of athletic injury prevention strategies [23,24], personalized intervention measures that take into account the biological characteristics of athletes [25] are particularly important. Su [26] used the fault tree analysis method to explore the statistical correlation between the biorhythm status of athletes and athletic injuries, and pointed out that formulating appropriate exercise patterns for athletes based on biorhythm theory can effectively improve their sports performance and reduce the risk of injury. These studies provide theoretical support and practical basis for athletic injury prevention [27,28], but there are still some deficiencies in personalized protection strategies for different types of sports, long-term effect evaluation, and interdisciplinary integration.

The application of sports biomechanics in athletic injury prevention has gradually become an important research direction in the field of sports science. Vancini et al. [29] reviewed and analyzed the latest progress in biomechanics and its impact on sports performance and injury prevention, summarized the key developments in this field, and emphasized its important value in optimizing sports performance and reducing the risk of injuries. To achieve more accurate injury prevention, more and more studies are focusing on how to combine the principles of biomechanics [30,31] with individualized training programs [32] to improve the intervention effect. Chang et al. [33] explored the effect of anterior cruciate ligament (ACL) injury prevention strategies on young female football players through biomechanical analysis and intervention studies, and proposed that ACL injury prevention training should be optimized based on individual kinematics and sports coordination. In biomechanical research, combining deep learning [34,35] and statistical analysis can further improve the accuracy and effect evaluation of injury prevention strategies. Wang et al. [36] used biomechanical analysis, deep learning technology, and statistical analysis to explore injury prevention and rehabilitation strategies in physical education, and verified that personalized teaching strategies based on biomechanical characteristics can effectively reduce the risk of athletic injuries and improve rehabilitation effects. These studies demonstrate the potential of biomechanics in athletic injury prevention, but current research still lacks in-depth interdisciplinary integration and personalized research on the characteristics of different sports, and more exploration and verification are needed in this regard.

To solve the above problems, this paper proposes an athletic injury risk prediction method based on the combination of statistical analysis and machine learning. The Vicon Vantage V5 3D motion capture system is used to collect the athlete's motion trajectory and joint angle change data; the Noraxon Ultium EMG electromyography device is used to record the electromyographic (EMG) activity signal; the Bertec force plate system is used to measure the ground reaction force. Combining these devices, a high-precision, multimodal dataset is constructed to cover key biomechanical features. In the data processing stage, the SPCA method is used to extract key features and screen biomechanical indicators that are highly correlated with injury risk while reducing the data dimension to alleviate the interference of redundant information. In the modeling stage, the ST-GCN is used to design a dynamic risk prediction framework, model the motion data as a graph structure, fully explore the dynamic relationship and spatial dependency characteristics in the time series, and predict the potential injury risk under different motion modes. This paper combines SPCA with ST-GCN for athletic injury prevention, innovatively integrates statistics and biomechanical methods, and realizes dynamic risk assessment through high-precision data collection and analysis, providing athletes with personalized injury prevention strategies and providing a new perspective for multimodal biomechanical data analysis. The method in this paper achieves an effective transformation from theoretical analysis to practical application from data collection, feature extraction to dynamic modeling, and provides theoretical support and technical reserves for future intelligent and precise athletic injury prevention and control research.

2. Athletic injury prevention methods

2.1. Data collection and processing

In data collection, this paper uses the Vicon Vantage V5 3D motion capture [37,38] system to record the athlete's joint angle change rate, obtains electromyographic activity signals through the Noraxon Ultium EMG electromyography device, and uses the Bertec force plate system to measure the ground reaction force to construct a multimodal biomechanical dataset. During the data collection process, all participants signed an informed consent form on the premise of fully understanding the purpose of the study and the use of the data, ensuring the legality and transparency of data use. In order to protect privacy, all personal identification information has been removed, and the data is stored and processed in an anonymous form, strictly following relevant ethical guidelines and privacy protection regulations. The time series data collected by each device are synchronously calibrated to ensure time alignment and spatial consistency. The athlete completes standardized movements, reduces random errors and captures biomechanical features, providing high-quality raw data for subsequent processing. **Figure 1** intuitively shows the effect of motion trajectory and joint point marking in multimodal data collection.

Figure 1. Motion collection and joint point marking.

In the data processing stage, the collected signal is subjected to wavelet threshold denoising to remove high-frequency noise caused by electromagnetic interference or device jitter. Assuming that the wavelet coefficient of the collected original signal after decomposition is $C_{j,k}$, the threshold processing is performed using the following equation:

$$\hat{C}_{j,k} = \begin{cases} C_{j,k} - \lambda \text{ if } C_{j,k} > \lambda, \\ C_{j,k} + \lambda \text{ if } C_{j,k} \leq -\lambda, \\ 0 \text{ otherwise.} \end{cases} \quad (1)$$

Among them, λ is the threshold parameter for noise removal. For missing points in the signal, linear interpolation is used to fill in the missing values to ensure data integrity. The normalization operation maps the multimodal data to the interval [0,1] to eliminate the influence of dimensional differences. The processed standardized data provides a consistent input for subsequent feature extraction and modeling.

2.2. SPCA sparse feature extraction method

SPCA is a feature extraction method that applies sparsity constraints based on traditional principal component analysis (PCA), aiming to achieve effective selection of key features through sparse regularization. Compared with traditional PCA, SPCA introduces sparsity constraints in feature extraction, which makes it more focused on retaining features closely related to injury risk while reducing redundant information when processing high-dimensional data. This method improves the efficiency of feature selection and enhances the interpretability of data in low-dimensional space. For athletic injury prevention tasks, this means that key biomechanical indicators can be identified more accurately. In athletic injury risk assessment, features extracted by SPCA are more representative, which helps to reduce the computational burden in the model training process and prevent overfitting caused by high-dimensional data. The goal of SPCA [39] is to project high-dimensional data into a low-dimensional space while ensuring that the projection matrix is sparse, thereby retaining important information related to damage risk and reducing data redundancy. Traditional PCA extracts principal components by maximizing the variance of data projection, and the objective function is:

$$\max_{w} w^{\top}\Sigma w\,, s.t. \|w\|_2 = 1, \tag{2}$$

Among them, w is the projection vector, and Σ is the data covariance matrix. However, the projection vector generated by PCA is usually dense, which makes the extracted features difficult to interpret and contains a lot of redundant information.

To apply sparsity, SPCA adds an L1 norm constraint in the optimization process and changes the optimization objective to:

$$\max_{w} w^{\top}\Sigma w - \alpha\|w\|_1, s.t. \|w\|_2 \leq 1, \tag{3}$$

Among them, α is the sparsity adjustment parameter, which is used to balance the maximization of the projection variance and the sparsity requirement. Through this optimization form, SPCA can select fewer but important features, significantly improve the efficiency of feature selection, and retain key information related to damage risk. The optimization problem is solved by an iterative algorithm, and finally a sparse projection vector w is obtained, which maps the original high-dimensional data to a low-dimensional subspace.

The application of sparse regularization effectively reduces feature redundancy, optimizes computational efficiency, and improves the interpretability of data in low-dimensional space. In the damage risk assessment task, the features extracted by SPCA are more representative, which reduce the computational burden during model training and avoid the risk of overfitting caused by high-dimensional data. This process lays a solid foundation for subsequent dynamic risk prediction modeling.

2.3. ST-GCN model construction and risk assessment

Compared with time series models such as LSTM, ST-GCN can simultaneously capture spatiotemporal dependencies and dynamic changes, which is more effective in analyzing complex motion patterns and predicting potential injury risks. For multimodal biomechanical data, ST-GCN [40] uses graph convolution operations to model the spatial relationship between joints and extracts the temporal dependencies in motion sequences through temporal convolution layers, thereby providing more accurate risk prediction than single temporal or spatial dimension analysis. The model comprehensively considers the spatial structure and temporal evolution of motion data, providing more accurate predictions for dynamic risk modeling, showing great potential in athletic injury prevention and rehabilitation applications.

2.3.1. ST-GCN model construction

The construction process of the graph structure involves the close combination of spatial dependency and time series. In the risk assessment process, ST-GCN first preprocesses the collected raw data, which includes denoising, data normalization, and division of motion stages to ensure the quality of the input data. Next, the model constructs the athlete's spatio-temporal graph model through spatio-temporal graph convolution operations. Each node represents a joint of the athlete's body, and the edges between nodes represent the spatial relationship or mutual dependence between joints. By performing convolution operations on the graph structure, ST-GCN extracts the athlete's posture features and their dynamic changes over time. These dynamic features include the changes in angles, speeds, and accelerations of each joint during different movements of the athlete, which are directly related to the risk of injury to the athlete in different training and competition environments. The graph convolution operation of ST-GCN fuses these spatio-temporal features to capture the athlete's posture changes at a specific time step and the dynamic changes across time steps. The model framework is shown in **Figure 2**.

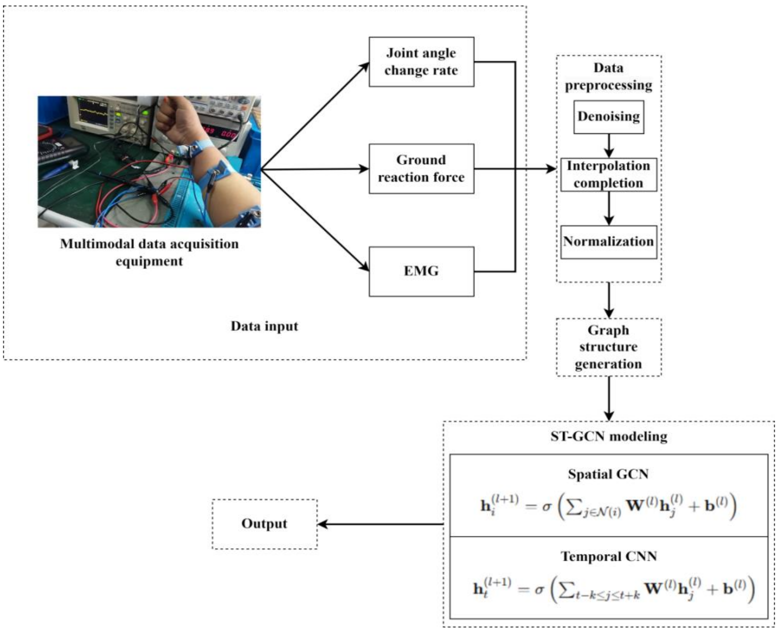

Figure 2. Model framework diagram.

The core modules of the ST-GCN model include graph convolution layer and temporal convolution layer. The graph convolution layer can capture the spatial dependencies between joints by performing convolution operations on the graph structure. The graph convolution operation of each layer helps the model learn the complex interactions and biomechanical laws between joints by weighted averaging the features of adjacent nodes. In graph convolution, the update equation of the representation vector h_i of a node is:

$$h_i^{(l+1)} = \sigma(\sum_{j \in \mathcal{N}(i)} W^{(l)} h_j^{(l)} + b^{(l)}) \tag{4}$$

Among them, $h_i^{(l)}$ is the feature vector of node i at the l-th layer; $\mathcal{N}(i)$ is the set of neighbor nodes of node i; $W^{(l)}$ is the weight matrix of the l-th layer; $b^{(l)}$ is the bias term; σ is the activation function.

In the time dimension, the temporal convolution layer is responsible for extracting the dynamic changes between consecutive frames. By sliding the convolution window on the time axis, the temporal convolution layer can capture the temporal dependencies in the action sequence, thereby accurately modeling the evolution of the action. The node feature representation of the temporal convolution layer can be updated by the following equation:

$$h_t^{(l+1)} = \sigma(\sum_{t-k \leq j \leq t+k} W^{(l)} h_j^{(l)} + b^{(l)}) \tag{5}$$

Among them, $h_t^{(l)}$ is the node feature at time step t; k is the size of the convolution window, which represents the span of adjacent frames in time; $W^{(l)}$ and $b^{(l)}$ are the weight matrix and bias term of the temporal convolution layer. The interaction of spatial and temporal information provides the model with stronger feature representation capabilities.

To optimize the performance of the model, the loss function design considers multiple objectives in spatial and temporal dimensions. The loss function not only includes the prediction error, but also includes the regularization constraint on the spatial structure and the smoothness constraint on the temporal dependency. Assuming that the output of the model is $\hat{y}$, then the loss function can be written as:

$$\mathcal{L} = \mathcal{L}_r + \mathcal{L}_s + \mathcal{L}_t \tag{6}$$

Among them, $\mathcal{L}_r$ is the conventional prediction loss; $\mathcal{L}_s$ is the regularization loss for spatial dependency; $\mathcal{L}_t$ is the smoothness loss on the time series. These constraints help avoid overfitting and enhance the model's learning ability for long time series. The gradient descent method and adaptive learning rate adjustment mechanism are used in the optimization process to ensure that the model can effectively extract useful features from a large amount of action data. Through these designs, the ST-GCN model can comprehensively consider the spatial structure and temporal evolution of the action, provide more accurate predictions for dynamic risk modeling, and show great potential in the application of athletic injury prevention and rehabilitation.

2.3.2. Risk assessment

The ST-GCN model captures the dynamic characteristics of athletes' posture changes, movement trajectories, load intensity, etc., during training or competition, and then evaluates the injury risk of athletes under different sports conditions. The injury risk score R is calculated based on the extracted spatial-temporal feature Z, and the equation is as follows:

$$R = \text{Sigmoid}(W_r Z + b_r) \quad (7)$$

Among them, W_r is the weight of the regression layer; b_r is the bias term; Sigmoid is the activation function, which maps the score to the range of [0, 1]. The higher the score value, the greater the injury risk in the current sports state. According to the different score intervals, the system conducts a comprehensive analysis of the athlete's state, and combines the previously collected biomechanical characteristics to build a targeted strategy generation module.

2.4. Personalized injury risk strategy generation mechanism

The personalized injury risk strategy generation mechanism aims to transform the injury risk prediction results based on the ST-GCN model into effective prevention strategies, thereby providing athletes with tailored training and rehabilitation guidance. In the personalized injury risk strategy generation mechanism, the assessment and management of psychological factors are an indispensable part. Psychological states such as stress and anxiety have a significant impact on athletes' sports performance and injury risk. By integrating psychological assessment tools into biomechanical data analysis, a more comprehensive risk prediction model is constructed to optimize the effectiveness of preventive measures. The design of psychological intervention programs should be closely integrated with physical training programs to ensure that the two complement each other and jointly promote the overall health of athletes. The impact of individual differences such as age, gender, and weight on sports performance and injury risk cannot be ignored. Full consideration of these parameters in the model can improve the accuracy of personalized risk prediction and ensure that each athlete receives training and rehabilitation guidance that fits their own characteristics. Athletes from different backgrounds will show different risk characteristics when facing the same sports, so the model design needs to be flexible enough to adapt to various situations. Through a comprehensive analysis of the above variables, the most appropriate prevention plan can be tailored for each athlete.

For the purpose of meeting the needs of different athletes, it is crucial to develop diverse injury prevention strategies. Each strategy should be customized according to the specific situation of the athlete to ensure its relevance and effectiveness. The system will conduct in-depth analysis based on the high-risk score, identify abnormal indicators associated with high risk, and make corresponding adjustment suggestions accordingly. The feedback information not only covers the current risk score and risk factor distribution, but also provides specific guidance on how to improve movement patterns or adjust training loads. The real-time feedback mechanism ensures that athletes are always within a safe training range, while promoting closed-loop management from problem detection to feedback optimization. This process provides a scientific basis for action adjustment, training optimization, and recovery management, and also lays the foundation for the construction of intelligent and personalized injury prevention strategies.

First, the main factors leading to high risk scores need to be identified. Abnormalities in the spatial feature dimension are usually related to joint angles beyond the normal range and uneven distribution of mechanical loads, while abnormalities in the temporal feature dimension may be manifested in incoherent movements, excessive acceleration fluctuations, or unbalanced rhythms. By analyzing these abnormal indicators, the system determines the specific problems in the athlete's movements. Combined with the athlete's historical data, the system further distinguishes the essential causes of these problems, whether they are due to technical movement defects, excessive training load, potential muscle fatigue, or insufficient strength. After clarifying the cause of the problem, the system matches the corresponding intervention strategy. For joint angle or trajectory deviations, the system recommends optimizing the movement mode by prompting the athlete to adjust the range of motion of a specific joint to restore to the standard state. For situations where the load exceeds the standard, it is recommended to adjust the training plan, reduce the current intensity or reduce the number of repetitions to ensure that the load fluctuates within a reasonable range. If fatigue signals are found, the system gives priority to recovery suggestions.

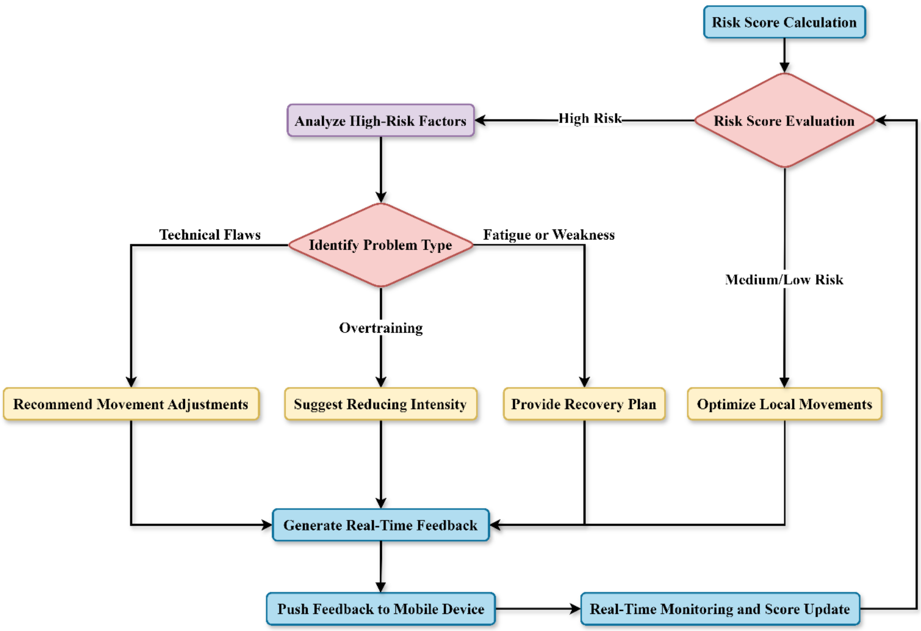

Figure 3. Prevention strategy generation process.

After the strategy is generated, the real-time feedback mechanism is activated, and the athlete receives feedback information through the mobile device. The feedback content is displayed in a graphical form, showing the current score, detailed distribution of risk factors and corresponding optimization plans, so as to understand the status of key joints, dangerous nodes of movements, and recommended adjustment directions. The feedback intuitively presents the comparison between the joint's out-of-range and ideal range, and is accompanied by specific text prompts to guide how to adjust the posture or change the training load. To improve the usability of feedback, the system supports real-time monitoring and iterative optimization, and dynamically updates the score and strategy by continuously collecting new data to ensure that athletes are always within the safe training range. The process of the prevention strategy generation mechanism is shown in **Figure 3**.

Through the prevention strategy generation process driven by injury risk score, athletes can obtain precise and efficient guidance in each training, and realize closed-loop management from problem detection to feedback optimization. This mechanism not only provides a scientific basis for action adjustment, training optimization, and recovery management, but also lays the foundation for the intelligent construction of personalized injury prevention strategies.

3. Experimental design

3.1. Dataset construction

To fully demonstrate the biomechanical characteristics of athletes when performing different sports movements, this study constructs a dataset containing core parameters such as joint angle change rate, electromyographic activity, and ground reaction force. To evaluate the generalization ability of the model, the constructed dataset is randomly divided into training set, validation set, and test set, with a ratio of 70%, 15%, and 15%, respectively. To further ensure the stability and reliability of the model, this study adopted the k-fold cross-validation method. A 5-fold cross-validation ($k = 5$) was selected, that is, the data set was evenly divided into 5 subsets. In each round of training, 4/5 of the data was used for training and 1/5 of the data was used for validation. This setting ensures that each sample has the opportunity to participate in the validation process and also balances the relationship between computational cost and validation effect to a certain extent. All experiments were run independently multiple times, and the mean and standard deviation were reported to fully reflect the performance fluctuations of the model. **Table 1** shows the biomechanical characteristics of each sample and the corresponding injury risk score, in order to provide data support for subsequent risk prediction and personalized prevention strategy generation.

Table 1. Athlete biomechanical characteristics and injury risk score data.

Sample ID	Movement Type	Joint Angle Change Rate (°/s)	EMG Activity (mV)	GRF (N)	Risk Score
1	Squat	45.2	0.76	1300.5	0.62
2	Lunge	50.8	0.85	1405.3	0.75
3	Vertical Jump	78.1	1.1	1608.2	0.88
4	Forward Step Down	32.4	0.67	1254.7	0.49
5	Single-leg Squat	58.9	0.92	1452.1	0.72

In **Table 1**, the sample number is used to uniquely identify individual data; the movement type represents the action classification; the joint angle change rate and electromyographic activity are the core features; the ground reaction force reflects the load distribution; the risk score is the result of the model prediction, ranging from 0 to 1, which is used to quantify the injury risk of the action. Through the above dataset structure, multimodal biomechanical information can be fully integrated to provide data support for personalized risk assessment and intervention strategy formulation.

3.2. Model parameter setting

To achieve precise and efficient dynamic risk prediction, this paper combines the SPCA and ST-GCN models to make full use of their advantages in feature extraction and spatio-temporal modeling. SPCA selects key features through sparsity constraints to improve the representativeness of the data, while ST-GCN comprehensively considers the time and space dimensions to capture the potential risks in complex actions. The model parameter settings are shown in **Table 2**.

Table 2. Parameter settings.

Model Name	Key Parameter	Value	Purpose
SPCA	Sparsity regularization parameter	0.1	Balancing feature sparsity and projection variance
ST-GCN	Number of graph convolution layers	3	Extracting spatial dependency features
	Number of temporal convolution layers	2	Capturing temporal dynamic features
	Temporal convolution window size	5	Defining the temporal range
Model Optimization	Optimizer	Adam	Improving model training efficiency
	Initial learning rate	0.001	Controlling optimization step size
	Batch size	32	Handling data scale
	Maximum iterations	500	Limiting training cycles

3.3. Evaluation indicators

To fully verify the effectiveness and practical application value of the proposed method, the evaluation indicators used in the experiment include prediction accuracy, feature contribution rate (FCR), mean risk score error (MRSE), temporal-spatial modeling efficiency (TSME) and real-time feedback delay (RTFD). By quantifying the performance of the model at different levels, it provides systematic support for the research results.

Accuracy is used to evaluate the overall prediction ability of the model for athletic injury risk, which is defined as the ratio of the number of correctly predicted samples to the total number of samples:

$$\text{Accuracy} = {\text{TP} + \text{TN}}/{\text{TP} + \text{TN} + \text{FP} + \text{FN}} \tag{8}$$

Among them, TP and TN are the correct prediction numbers of positive and negative classes, respectively, and FP and FN are the number of incorrectly predicted positive and negative samples, respectively.

FCR is used to quantify the impact of biomechanical features on model performance and calculate the decrease in model accuracy after removing a feature:

$$\mathrm{FCR} = {\mathrm{ACC_{all}} - \mathrm{ACC_{w/oi}}}\Big/{\mathrm{ACC_{all}}} \tag{9}$$

Among them, $\mathrm{ACC_{all}}$ is the prediction accuracy when all features are included, and $\mathrm{ACC_{w/oi}}$ is the prediction accuracy after removing feature i.

MRSE is used to evaluate the precision of the model's personalized risk score, which is calculated by the mean square error between the predicted and true values:

$$\mathrm{MRSE} = \frac{1}{N}\sum_{i=1}^{N}(R_i^p - R_i^t)^2 \tag{10}$$

Among them, R_i^p and R_i^t represent the predicted risk score and the true risk score of the i-th sample, respectively, and N is the total number of samples.

TSME is used to evaluate ST-GCN's ability to model temporal dynamics and spatial dependencies, and is calculated by combining the significance of temporal and spatial features in the model:

$$\mathrm{TSME} = \frac{1}{N}\sum_{i=1}^{N}(W_i^t + W_i^s) \tag{11}$$

Among them, W_i^t and W_i^s are the temporal and spatial feature weights of sample i, respectively.

RTFD measures the speed at which the system generates personalized injury risk feedback, which is defined as the average delay time from input data to output results:

$$\mathrm{RTFD} = \frac{1}{M}\sum_{j=1}^{M}T_j \tag{12}$$

Among them, T_j is the delay time of the j-th prediction, and M is the total number of predictions. These evaluation indicators are used to comprehensively analyze the performance of the proposed method in multimodal feature extraction, dynamic modeling and real-time feedback capabilities, providing data support and theoretical basis for the experimental results.

4. Results

4.1. Verification of the effectiveness of multimodal feature extraction

To verify the optimization ability of SPCA for multimodal biomechanical data, this paper experimentally evaluates the impact of the number of features on the model prediction accuracy. The number of features increases from 1 to 20, and the corresponding prediction accuracy is recorded. The focus of the experiment is to analyze the specific effect of the change in the number of features on the model performance, and to provide a basis for feature selection for the dynamic risk prediction model. **Figure 4** is a curve showing the relationship between the number of features and the prediction accuracy.

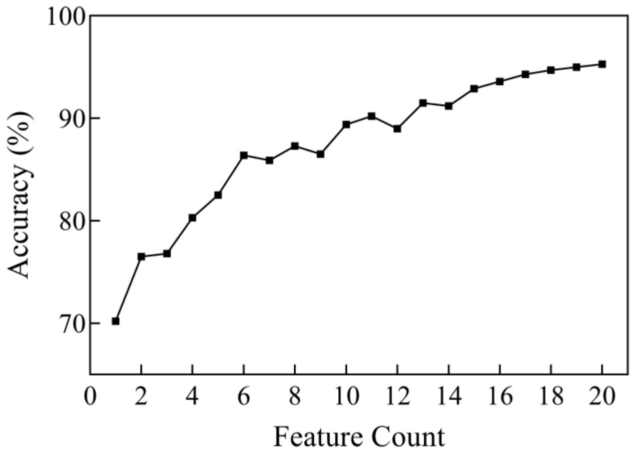

Figure 4. Accuracy trend of sparse feature extraction method.

The data analysis shows that as the number of features increases, the prediction accuracy of the model shows a steady improvement trend overall. Within the first 6 feature numbers, the accuracy increases rapidly from 70.2% to 86.4%, which shows that increasing the number of features effectively improves the prediction ability of the model. When the number of features is between 6 and 14, the accuracy fluctuates slightly, and the accuracy drops to 86.5% at the 9th feature. Later, with the application of more features, the model gradually stabilizes. When the number of features reaches 15, the accuracy begins to show a more stable growth, and finally reaches 95.3% at 20 features. This trend shows that the increase in features helps improve the accuracy of the model, but when there are too many features, the improvement effect slows down and brings certain redundant information. Therefore, a reasonable selection of the number of features is crucial to improving the performance of the prediction model.

Although the experimental results demonstrate the effectiveness of the SPCA method in feature extraction, potential sources of error are still worth analyzing. During the data collection phase, noise or bias introduced by sensor calibration accuracy and equipment synchronization issues can affect feature accuracy. Individual differences in athletes such as body shape, muscle strength, and flexibility increase the variability of the data. During the feature selection process, there is a subjective bias in the selection of sparsity parameters, which can easily lead to key biomechanical indicators not being fully captured. These factors interact with each other and affect model performance.

4.2. Comparison of different time series models

The experiment compares the performance of four time series models, Long Short-Term Memory (LSTM), Transformer, Graph Attention Network (GAT) and ST-GCN, in processing various types of action complexity, in order to reveal the differences in the ability of different models in modeling spatio-temporal dependencies. LSTM processes traditional time series data by capturing long-term temporal dependencies through a recursive structure, but it faces a bottleneck in computational efficiency in tasks with higher complexity. Transformer captures long-term dependencies through a self-attention mechanism to efficiently process sequence data in parallel, and is widely used in time series tasks. GAT effectively models spatial dependencies in data through a graph attention mechanism, and is suitable for graph structure data analysis. ST-GCN combines graph convolution with temporal convolution, and can simultaneously mine the temporal and spatial dependency characteristics in data, making it suitable for analyzing complex action patterns. In this experiment, the action complexity is divided into 10 levels from easy to difficult: Level 1 is standing; Level 2 is standing on one leg; Level 3 is squatting; Level 4 is walking in a straight line; Level 5 is running in a small range; Level 6 is jogging; Level 7 is fast running; Level 8 is jumping; Level 9 is fast change of direction running; Level 10 is full sprinting. The comparison results are shown in **Figure 5**.

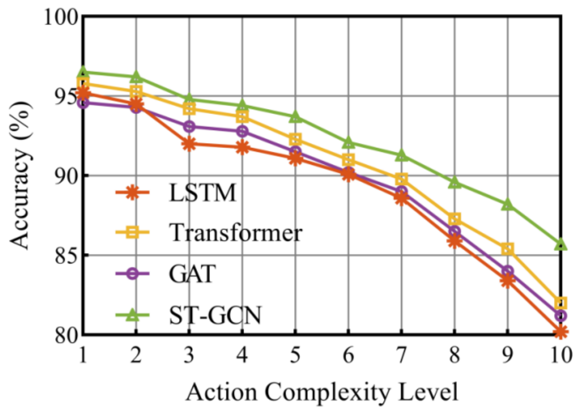

Figure 5. Comparison of the accuracy of different models in the action complexity.

As can be seen from **Figure 5**, with the increase of the action complexity, the prediction accuracy of ST-GCN is always higher than that of other models. In the high-complexity action of full sprinting, the accuracy of ST-GCN reaches 85.7%, far exceeding other models. The accuracy of LSTM performs better than GAT in low-complexity actions such as standing and standing on one leg, but it decreases in complex actions such as fast change of direction running and full sprinting, and finally only 80.2% in the action with a complexity of 10. Transformer and GAT perform similarly in most cases, but Transformer has a slight advantage in accuracy in high-complexity actions. ST-GCN shows obvious advantages in modeling spatio-temporal dependencies, indicating that it is suitable for prediction tasks of multimodal and complex dynamic data.

4.3. Correlation between biomechanical characteristics and injury risk

In athletic injury risk assessment, it is crucial to identify different risk types and effectively distinguish them. This experiment uses the Pearson correlation coefficient analysis method to evaluate the correlation between the five biomechanical characteristics of athletes, namely, joint angle change rate, ground reaction force, electromyographic activity, movement speed and movement duration, and the injury risk level (low, medium-low, medium, high, and very high) to quantify athletic injury risk. This analysis reveals the correlation between each mechanical characteristic and the injury risk level, reflecting the degree of influence of different characteristics in each risk level. The results are shown in **Figure 6**.

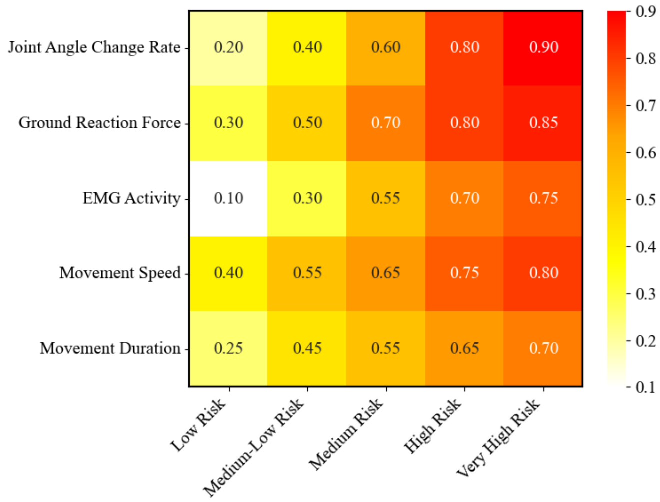

Figure 6. Correlation between biomechanical characteristics and injury risk.

Figure 6 shows that the joint angle change rate, ground reaction force and movement speed are prone to high and very high risks. Among them, the joint angle change rate shows the strongest correlation of 0.9 in the very high risk level, indicating that there is a strong correlation between drastic joint angle changes and very high risks. The correlation between medium risk and ground reaction force reaches 0.7, indicating that improper use of force is prone to medium risk injuries. Improper control of movement speed is prone to low risk injuries, with a correlation of 0.4. These results show that there is a close relationship between different biomechanical characteristics and athletic injury risks, which provides data support for the formulation of athletic injury prevention strategies.

4.4. Comparison of feature extraction methods

The experiment compares the performance of five feature extraction methods in athletic injury prediction: SPCA, Local Linear Embedding (LLE), *t*-Distributed Stochastic Neighbor Embedding (*t*-SNE), Independent Component Analysis (ICA) and Autoencoder. SPCA applies L1 norm constraints for feature selection, effectively reducing redundant data and improving feature interpretability; LLE is good at processing nonlinear data structures, but has low computational efficiency; *t*-SNE is mainly used for dimensionality reduction and data visualization and is suitable for visualization of high-dimensional data; ICA can effectively separate independent components in the signal and performs well in processing independent features; Autoencoder automatically extracts features through unsupervised learning. The experiment compares their performance in five indicators: prediction accuracy, feature contribution rate, computational efficiency, model complexity, and interpretability, in order to evaluate their applicability and advantages in athletic injury risk prediction. The relevant experimental results are shown in **Table 3**.

Table 3. Comparison of different feature extraction methods.

Method	Accuracy (%)	FCR (%)	Computational Efficiency	Model Complexity	Interpretability
SPCA	95.3	85	High	Low	High
LLE	92.1	75	Medium	Medium	Medium
t-SNE	88.6	70	Low	High	Low
ICA	91.4	78	Medium	Medium	Medium
Autoencoder	93.2	80	High	High	Low

Table 3 shows that the SPCA method has the highest risk prediction accuracy of 95.3%, and its FCR also reaches the highest 85%, indicating that it has advantages in reducing data redundancy and improving feature representativeness. The accuracy and FCR of t-SNE are the lowest, only 88.6% and 70%, indicating that this method is not suitable for risk prediction. The performance of Autoencoder is second only to SPCA. Like SPCA, it has high computational efficiency, but its model complexity is high and its interpretability is low, which limits its application in this regard. By comprehensively considering the accuracy, feature contribution rate, computational efficiency and interpretability, SPCA is the best method choice in the application of athletic injury prevention.

4.5. Impact of optimization strategy on model performance

Five optimization strategies are compared in the experiment, which are optimized for feature extraction, network structure and data processing, respectively, to improve the performance of the athletic injury prediction model. Strategy 1: Without optimization in the baseline model: the original SPCA feature extraction and ST-GCN modeling are used as the basis for performance comparison. Strategy 2: Feature extraction enhancement: by applying L1 regularization in SPCA, the sparsity of features is enhanced; redundant information is reduced; the representativeness of features is improved. Strategy 3: ST-GCN structure optimization: the ST-GCN network structure is adjusted; the number of graph convolution layers is increased; the convolution method is improved to optimize the spatio-temporal relationship modeling capability. Strategy 4: Comprehensive optimization: combining strategies 2 and 3, feature extraction and network structure are optimized simultaneously to maximize model performance. Strategy 5: Data enhancement optimization: data enhancement technology is used to expand the diversity of the training set and improve the robustness of the model. The experiment evaluates the effect of each strategy by comparing three indicators: damage prediction accuracy, real-time feedback delay, and spatio-temporal modeling efficiency. **Figure 7** shows the comparison results of different optimization strategies on these indicators.

Figure 7. Comparison of the impact of optimization strategies on athletic injury prediction performance.

Figure 7 shows that with the strengthening of the optimization strategy, the accuracy of athletic injury prediction shows a significant upward trend, and Strategy 4 reaches the highest accuracy of 95.3%, which is much higher than other strategies. RTFD performs best in Strategy 4, with a delay of only 130 milliseconds, while the delay of Strategy 1 is 150 milliseconds, showing the positive impact of the optimization strategy on the timeliness of feedback. For TSME, Strategy 4 also performs well, with a value of 1.45, which is much higher than the 0.72 of the basic model. In contrast, the TSME and feedback delay of Strategy 3 are more balanced and robust. Overall, Strategy 4 shows the best performance in terms of accuracy, real-time feedback delay, and spatio-temporal modeling efficiency, indicating that it has significant advantages in optimizing athletic injury prevention models.

4.6. Case study

In a study on athletic injury prevention for professional football players, the research team applied the athletic injury risk prediction model proposed in this paper that combines statistical analysis and machine learning. In the study, the Vicon Vantage V5 3D motion capture system recorded the rate of change of joint angles of players when performing specific training movements, the Noraxon Ultium EMG electromyography device obtained muscle activity signals, and the Bertec force plate system measured ground reaction forces. After wavelet threshold denoising, linear interpolation to fill missing values, and normalization, these data formed a high-quality multimodal biomechanical dataset. Through the SPCA method, the research team extracted features that are highly correlated with injury risk, reduced data dimensions, and alleviated the interference of redundant information. Subsequently, the ST-GCN model modeled these features in space and time, fully explored the dynamic relationship and spatial dependence characteristics, and realized the prediction of potential injury risks under different movement modes.

accurately captures subtle changes in movement patterns when faced with complex dynamic action sequences, providing each player with a personalized injury risk assessment. For a player who was performing high-intensity change-of-direction running training, the model identified an abnormal rate of change in the angle of his knee joint, and combined the electromyographic signal and ground reaction force data to calculate a high injury risk score. Based on this score, the personalized strategy generation mechanism determined that the player had technical movement defects, mainly manifested in the uneven distribution of mechanical loads on the knee joint. Based on this, the system recommended adjusting the player's training plan, reducing the current intensity, and recommending exercises to strengthen the muscles around the knee joint. After receiving the feedback information, the athlete adjusted the training method according to the guidance. After a period of time, the re-evaluation showed that the risk of injury was significantly reduced, proving the effectiveness of the model in practical applications. Throughout the process, from data collection to feature extraction to dynamic modeling, the research team demonstrated the ability to transform theoretical analysis into practical applications, providing theoretical support and technical reserves for intelligent and precise research on athletic injury prevention and control.

5. Conclusions

The athletic injury prevention method based on SPCA and ST-GCN proposed in this paper optimizes feature extraction and dynamic risk prediction through multimodal data acquisition. Experiments show that when the number of features reaches 20, the model prediction accuracy is stable at 95.3%. Compared with other models such as LSTM and Transformer, which have an accuracy of 80.2% and 82% respectively in high-complexity action prediction, ST-GCN shows better spatio-temporal dependency modeling capabilities. The correlation analysis shows that the correlation between the joint angle change rate and the risk of very high injury is as high as 0.9, which verifies the important role of biomechanical characteristics in risk assessment. However, there is still room for improvement in the robustness of long-term training series and the optimization of targeted strategies for different types of sports. Future research will focus on integrating more advanced sensing technologies and computing methods to improve data acquisition accuracy and feature extraction depth. At the same time, by expanding data sets and adjusting model parameters to enhance generalization capabilities, personalized prevention and rehabilitation plans can be developed for different sports characteristics, further promoting the intelligent and precise development of athletic injury prevention.

Ethical approval: Not applicable.

Conflict of interest: The author declares no conflict of interest.

References

1. Jamnick, Nicholas A., Robert W. Pettitt, Cesare Granata, David B. Pyne, and David J. Bishop. "An examination and critique of current methods to determine exercise intensity." Sports Medicine 50.10 (2020): 1729-1756.
2. Rose, Grace L., Tina L. Skinner, Gregore I. Mielke, and Mia A. Schaumberg. "The effect of exercise intensity on chronic inflammation: a systematic review and meta-analysis." Journal of Science and Medicine in Sport 24.4 (2021): 345-351.
3. Reiser, Julian E., Edmund Wascher, Gerhard Rinkenauer, and Stefan Arnau. "Cognitive-motor interference in the wild: Assessing the effects of movement complexity on task switching using mobile EEG." European Journal of Neuroscience 54.12 (2021): 8175-8195.
4. Khan, Taha, and Peter G. Jacobs. "Prediction of mild cognitive impairment using movement complexity." IEEE journal of biomedical and health informatics 25.1 (2020): 227-236.
5. Matsuura, Yuki, and Genta Ochi. "The potential of heart rate variability monitoring for mental health assessment in top wheel gymnastics athletes: a single case design." Applied Psychophysiology and Biofeedback 48.3 (2023): 335-343.
6. Cheng, Jingming, Xie Wenjun, Shen Ziqi, Li Lin, and Liu Xiaoping. "Multimodal human motion synchronization dataset." Journal of Computer-Aided Design & Computer Graphics 34.11 (2022): 1713-1722.

7. Liu, Song. “Tennis players’ hitting action recognition method based on multimodal data.” International Journal of Biometrics 16.3-4 (2024): 317-336.
8. Huifeng, Wang, Achyut Shankar, and G. N. Vivekananda. “Modelling and simulation of sprinters’ health promotion strategy based on sports biomechanics.” Connection Science 33.4 (2021): 1028-1046.
9. Liu, Ying, and Tengqi Zhu. “Individualized new teaching mode for sports biomechanics based on big data.” International Journal of Emerging Technologies in Learning (iJET) 15.20 (2020): 130-144.
10. Clouette, Julien, Timothy Leroux, Ajaykumar Shanmugaraj, Moin Khan, Chetan Gohal, Christian Veillette, Patrick Henry, and Ryan A Paul, et al. “The lower trapezius transfer: a systematic review of biomechanical data, techniques, and clinical outcomes.” Journal of shoulder and elbow surgery 29.7 (2020): 1505-1512.
11. Nagar, Priyanka, Andriette Bekker, Mohammad Arashi, Cor-Jacques Kat, and Annette-Christi Barnard. “A dependent circular-linear model for multivariate biomechanical data: Ilizarov ring fixator study.” Statistical Methods in Medical Research 33.9 (2024): 1660-1672.
12. Souissi, Mohamed Abdelkader, Omar Trabelsi, Olfa Tounsi, Aymen Hawani, Sofien Fekih, Hichem Souissi, Adnen Gharbi, Ahlem Amor, Swantje Scharenberg, and Nizar Souissi, et al. “Optimizing video feedback for snatch technical error correction in young weightlifters: Comparing the effectiveness of different video playback speeds.” International Journal of Sports Science & Coaching 19.4 (2024): 1612-1621.
13. Hou, Zhihong, and Yuan Xue. “Sports training injury risk assessment combined with dynamic analysis algorithm.” Molecular & Cellular Biomechanics 21.3 (2024): 484-484.
14. Cao, Haixin, and Kun You. “Application of Functional Movement Screening in the Evaluation of Sports Ability and the Prediction of Sports Injury Risk.” Investigación Clínica 61.1 (2020): 203-211.
15. Shi, Bo, and Bin Lv. “Accuracy of badminton swing action recognition based on fractional time network.” Journal of Intelligent & Fuzzy Systems 45.2 (2023): 2409-2418.
16. Dong, Xuelian, and Yuanhang Sun. “Injury Risk Prediction and Prevention Algorithm for Athletes Based on Data Mining.” Journal of Electrical Systems 20.6s (2024): 1717-1728.
17. Chen, Qiang. “Identification of Potential Injury Risk Factors and Prediction Model Construction of Athletes using Data Mining Algorithm.” Journal of Electrical Systems 20.6s (2024): 2048-2058.
18. Hope, Ed R., Keval Patel, James Feist, Oliver R. Runswick, and Jamie S. North. “Examining the importance of local and global patterns for familiarity detection in soccer action sequences.” Perception 53.3 (2024): 149-162.
19. Zhang, Zhen. “Scientific sports medicine in preventing sports injuries of college students.” Revista Brasileira de Medicina do Esporte 27.8 (2021): 796-799.
20. Limos, Nasstassja, and Maria Minerva Calimag. “Lower Extremity Injury Prevention in Taekwondo Poomsae Athletes: An Educational Intervention.” Journal of Sports and Physical Education Studies 2.1 (2022): 01-09.
21. Bisht, Neha, and Swati Srivastava. “Impact of Physical and Mental Training on Overall Performance and Sports Injury Prevention in Female Volleyball Athletes.” Indian Journal of Physiotherapy & Occupational Therapy Print-(ISSN 0973-5666) and Electronic–(ISSN 0973-5674) 15.3 (2021): 64-70.
22. Keavanloo, Fahimeh, Esmaeil’Sharifian, Kourosh Ghahraman Tabrizi, Mohammad Seyedahmadi. “Presenting and Fitting a Sports Injury Prevention Model in Physical Education Students.” Journal of Sabzevar University of Medical Sciences 28.3 (2021): 413-424.
23. Ross, Andrew G., Alex Donaldson, and Roslyn G. Poulos. “Nationwide sports injury prevention strategies: A scoping review.” Scandinavian journal of medicine & science in sports 31.2 (2021): 246-264.
24. Rohilla, Rajesh. “Unlocking the Games: A Proactive Approach to Safeguarding Athletic Prowess through Cutting-Edge Sports Injury Prevention Strategies.” Sports Science & Health Advances 1.2 (2023): 140-142.
25. Celik, Onur, Gizem Ece Celik, Bekir Eray Kilinc, and Gulcan Harput. “Knowledge and awareness of anterior cruciate ligament injury among Turkish professional athletes: an online survey.” Research in Sports Medicine 32.5 (2024): 787-803.
26. Su, Shaohui. “Statistical calculation method and analysis of athletes’ biorhythm state and sports injury.” Mobile Information Systems 2021.1 (2021): 8313185.
27. Lutter, Christoph, Christophe Jacquet, Evert Verhagen, Romain Seil, and Thomas Tischer. “Does prevention pay off? Economic aspects of sports injury prevention: a systematic review.” British journal of sports medicine 56.8 (2022): 470-476.

28. Salemi, Samira, Mehrzad Hamidi, Yousef Moghadas, Mohammad Hossein Alizadeh, and Mohammad Karimizadeh. "Strategy of Preventing Sports Injuries of Students in Schools with Grounded Theory Approach." Physical Treatments-Specific Physical Therapy Journal 13.4 (2023): 287-294.
29. Vancini, Rodrigo Luiz, Marilia Santos Andrade, Claudio Andre Barbosa De Lira, Thais Russomano. "Recent Advances in Biomechanics Research: Implications for Sports Performance and Injury Prevention." Health Nexus 1.3 (2023): 7-20.
30. Lu, Hua. "A study on Modeling and Simulation of Sports Training Injury Optimization from a Biomechanical Perspective." Molecular & Cellular Biomechanics 21.2 (2024): 233-233.
31. Huang, Yufei, and Qingxiu Jia. "Impact biomechanics in sports and exercise: Understanding the mechanical forces behind sports-related injuries." Molecular & Cellular Biomechanics 21.1 (2024): 325-325.
32. Hao, Zhenzhu. "Biomechanical analysis and rehabilitation strategies of common lower limb injuries in sprinters." Molecular & Cellular Biomechanics 21.3 (2024): 489-489.
33. Chang, Wenqing, and Zhaoxia Wang. "Biomechanics of athlete movement: kinematic analysis and injury prevention." Journal of Electrical Systems 20.3 (2024): 1075-1084.
34. Guan, Lu. "Intelligent rehabilitation assistant: Application of deep learning methods in sports injury recovery." Molecular & Cellular Biomechanics 21.2 (2024): 384-384.
35. Iduh, Blessing Nwamaka, Maryrose Ngozi Umeh, Overcomer Ifeanyi Anusiuba, and Fraser Anwaitu Egba. "Development of a Predictive Modeling Framework for Athlete Injury Risk Assessment and Prevention: A Machine Learning Approach." European Journal of Theoretical and Applied Sciences 2.4 (2024): 894-906.
36. Wang, Mingtao, and Zhen Pei. "Injury prevention and rehabilitation strategies in physical education: A machine learning-based approach using biomechanical characteristics." Molecular & Cellular Biomechanics 21.2 (2024): 412-412.
37. Ishida, Tomoya, Takumi Ino, Yoshiki Yamakawa, Naofumi Wada, Yuta Koshino, Mina Samukawa, Satoshi Kasahara, and Harukazu Tohyama, et al. "Estimation of Vertical Ground Reaction Force during Single-leg Landing Using Two-dimensional Video Images and Pose Estimation Artificial Intelligence." Physical Therapy Research 27.1 (2024): 35-41.
38. Zwolfer, Michael, Martin Mössner, Helge Rhodin, Werner Nachbauer, and Dieter Heinrich. "Integration of a skier-specific keypoint detection model in a hybrid 3D motion capture pipeline." Current Issues in Sport Science (CISS) 9.4 (2024): 013-013.
39. Chen, Fan, and Karl Rohe. "A New Basis for Sparse Principal Component Analysis." Journal of Computational and Graphical Statistics 33.2 (2024): 421-434.
40. Lovanshi, Mayank, and Vivek Tiwari. "Human skeleton pose and spatio-temporal feature-based activity recognition using ST-GCN." Multimedia Tools and Applications 83.5 (2024): 12705-12730.