AI-Powered Predictive Analytics For Injury Prevention In Sports: Integrating Wearable Sensors And Data-Driven Insights

Lidan Song*

Institute of Physical Education, Anyang Normal University, Anyang, Henan 455000, China

**Corresponding author. E-mail: lidansong01@outlook.com*

Received: May. 07, 2026; Accepted: Jun. 26, 2026

Sports injuries are increasingly common among athletes, impacting performance and recovery. Current models struggle to simulate the complex interactions of physiological, biomechanical, and workload factors, limiting their predictive accuracy. This study proposes a robust framework for accurately predicting sports injuries through multimodal monitoring. By integrating physiological, biomechanical, environmental, and workload parameters, the model processes data using algorithms for missing values, normalization, and noise reduction, followed by Principal Component Analysis. An Informer-based Deep Learning model with a ProbSparse attention mechanism effectively captures complex feature interactions. Experimental results show the model achieves accuracy, precision, recall, F1-score, and ROC-AUC of 99.63%, 99.63%, 99.63%, 99.62%, and 99.95%, respectively, outperforming traditional models like Logistic Regression and Support Vector Machine. This framework enhances early injury detection, supporting better athlete health management and training optimization.

Keywords: ports Injury Prediction, Informer Model, Wearable Sensors, Predictive Analytics, Athlete Monitoring

http://dx.doi.org/10.6180/jase.202611_34.062

1. Introduction

Sports injury prevention and athlete performance monitoring have become critical research areas due to the increasing physical demands placed on athletes in competitive sports environments [1]. Early sports monitoring relied on laboratory-based motion analysis and manual observations; however, wearable technologies now enable continuous physiological and biomechanical data collection during real sports activities [2]. Devices such as inertial measurement units, GPS trackers, pressure sensors, and smartphone-based sensing systems facilitate large-scale athlete movement and workload data acquisition in non-laboratory environments. The adoption of sensor-based monitoring systems by professional teams and sports organizations for injury prevention and performance optimization has grown substantially across research and academic settings [3].

Portable sensing systems capturing biomechanical movement data through inertial sensors and camera-based devices have been explored for identifying injury-associated motion patterns, while wearable sensor-based workload indices have been investigated to detect fatigue and performance dynamics [4]. Field-based motion analysis has enabled high-risk movement pattern detection during sport-specific activities, and machine learning predictive frameworks have been developed to estimate injury risk levels and provide early warning feedback during training and competition [5].

BioMAT, a transformer-based model for joint kinematic estimation from wearable IMU signals [6], demonstrated accurate kinematic motion representation but remained limited in injury risk prediction applicability. [7] explored sensor feature inputs for joint angle prediction across simple movements, but was constrained by dataset size and feature integration limitations. [8] achieved superior injury risk prediction by modeling complex multimodal relationships across training, health, and psychological monitoring data.

The objectives of this study are as follows:

- Develop an AI-based framework for injury risk prediction using multimodal athlete data, including physiological, biomechanical, environmental, and workload features.

- Implement data preprocessing techniques, including missing value handling, normalization, and noise reduction, to ensure data quality and consistency.

- Apply PCA for dimensionality reduction and improved model efficiency.

- Design an Informer-based deep learning model with ProbSparse attention for effective temporal dependency learning.

- Evaluate the model using some metrics like Accuracy, Precision, Recall, F1-Score, and ROC-AUC for injury and no injury classification.

The remainder of this paper is organized as follows: Section 2 reviews related work on sports injury prediction and wearable sensor-based monitoring; Section 3 presents the proposed framework; Section 4 discusses experimental outcomes and performance assessment; Section 5 summarizes findings and suggests future research directions.

2. Theory and formula

Wearable sensing and machine-learning techniques have been extensively investigated for sports injury prediction, rehabilitation, and biomechanical monitoring. Studies [9] employed inertial measurement units, RGB cameras, depth sensors, and biomechanical modeling to estimate movement patterns and injury-related parameters. Although these approaches demonstrated the potential of wearable technologies for injury monitoring, their applicability was limited by small datasets, restricted generalizability, and the lack of standardized clinical tools. Similarly, injury-risk estimation frameworks [10] utilized athlete monitoring data, including training load, sleep, mood, and injury records, but were constrained by study-design limitations and potential bias.

Subsequent research expanded to field-based injury-risk screening and biomechanical analysis. Studies [11] applied wearable sensing and performance-testing techniques to identify ACL injury risks and monitor athletic activities in real-world settings. While promising results were reported, limitations such as small sample sizes, limited validation, and single-athlete evaluations restricted broader applicability. Likewise, machine-learning approaches for biomechanical prediction, including foot-pronation prediction during running [7], achieved encouraging performance but relied on relatively small datasets and lacked validation across diverse athlete populations.

Studies using GNSS-IMU, inertial sensors, in-shoe systems, and deep learning face challenges like limited datasets, evaluation duration, sensor dependency, and misclassification, motivating robust multimodal frameworks for accurate sports injury prediction.

3. Experimental setup

Fig. 1 illustrates an AI-based sports injury framework that preprocesses multimodal data, applies PCA for feature reduction, trains an Informer model, and evaluates injury-risk prediction using Precision, Recall, F1-score, and ROC-AUC.

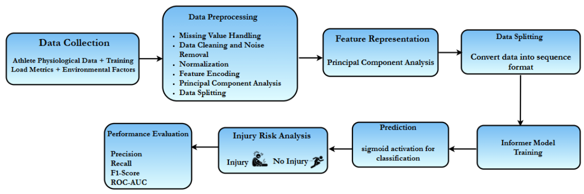

Fig. 1. Proposed workflow for sports injury risk prediction

3.1. Data Collection

The dataset includes 6000 athlete records with more than 30 features [12]. Athlete records with complete multimodal monitoring data were selected to represent diverse physiological characteristics, performance levels, injury histories, and sport disciplines, supporting representative population coverage and model generalizability across heterogeneous athletes. Biomechanical features include fatigue index, acceleration, angular velocity, and movement metrics. The dataset contains a binary target variable, injury_risk (1 = injury, 0 = no injury), with approximately 6000 records, representing diverse physiological, biomechanical, workload, and environmental conditions for reliable predictive analysis. The dataset was divided into 80% training and 20% testing subsets. Imputation, Z-score normalization, and PCA were applied only to training data to prevent leakage. The test set contained 1,086 samples with a moderately imbalanced distribution. The dataset includes team and individual sports with varying contact exposure, workload demands, and biomechanical loading patterns, captured through physiological, biomechanical, and workload features to support injury risk prediction across both settings. Wearable sensor placement affects biomechanical measurements, as different body locations capture varying acceleration, angular velocity, and ground reaction force patterns. This variability was addressed during preprocessing to ensure consistent multimodal feature representation.

3.2. Data Pre-Processing

3.2.1. Missing Value Handling

Missing values were mean-imputed to preserve data integrity, per Eq. (1).

$$x_i = \frac{1}{n}\sum_{j=1}^{n} x_j \tag{1}$$

Here, x_i is imputed value, x_j denotes the observed feature values, and n is the number of observations. Mean imputation was selected because physiological and biomechanical features exhibited approximately symmetric distributions, making the mean a representative replacement that preserves sample size, minimizes computational overhead, and supports stable Informer model training. Missing values occurred across physiological and biomechanical variables, including heart rate, galvanic skin response, and ground reaction force. These patterns were analysed before imputation to ensure reliability and consistent injury-risk prediction performance.

3.2.2. Data Normalization

Features were Z-score normalized to prevent magnitude dominance, per Eq. (2).

$$X_{\text{norm}} = \frac{X-\mu}{\sigma} \tag{2}$$

Here, X is original value, X_{norm} represents the normalized value, μ denotes the feature mean, and σ represents the standard deviation. Z-score normalization ensured equal feature contribution during gradient updates, preventing scale dominance. This standardization
accelerated convergence, reduced weight oscillations, maintained feature consistency across batches, and supported stable Informer model training.

3.2.3. Outlier Detection

Outliers were detected using the standardized Z-score method, as defined in Eq. (3).

$$Z = \frac{X-\mu}{\sigma} \tag{3}$$

Here, Z represents the standardized score, X is the observed data value and μ is the mean. Outlier detection identified extreme physiological and biomechanical measurements using a Z-score threshold of $|Z| > 3$. Abnormal values in features such as heart rate, acceleration, fatigue index, and ground reaction force were replaced using median-based imputation to reduce noise while preserving data distribution before PCA and Informer training.

3.2.4. Feature Representation using Principal Component Analysis (PCA)

PCA reduces dimensionality of physiological, biomechanical, and workload variables using standardized matrix $X \in R^{n\times d}$, to compute the covariance matrix C, as shown in Eq. (4).

$$C = \frac{1}{n}X^T X \tag{4}$$

The top k = 15 principal components accounted for 94.58% of total variance, capturing key injury-relevant patterns across physiological, biomechanical, and workload features. Components beyond PC15 contributed less than 2.08% variance individually and were excluded from subsequent Informer-based classification. Thus, PCA reduces computational complexity while enhancing feature representation for Informer model learning.

3.3. Informer-Based Architecture for Athlete Injury Risk Prediction

The Informer architecture was adapted for binary injury-risk classification by replacing the forecasting output with a sigmoid layer, where $p \geq 0.5$ indicates injury-risk and $p < 0.5$ indicates no-injury, using $k = 15$ PCA-reduced features as encoder inputs. The Informer model was trained using Adam optimization with a 0.0001 learning rate, batch size of 32,100 epochs, ReduceLROnPlateau scheduling, 2 encoder layers, 8 attention heads, GELU activation, 0.1 dropout, and binary cross-entropy loss, as summarized in Table 1. The selected configuration ensured stable convergence and consistent validation performance. Learning-rate adaptation and regularization improved generalization and prevented overfitting, while final hyperparameters were determined through preliminary experiments balancing accuracy, stability, and computational efficiency.

Fig. 2 shows Informer architecture processing PCA features through encoder-decoder ProbSparse attention and fully connected layer for injury risk classification.

3.3.1. ProbSparse Self-Attention (Encoder)

ProbSparse attention captures long-range dependencies selectively, per Eq. (5).

$$\text{Attention}(Q,K,V) = \text{Softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V \tag{5}$$

In Eq. (5), d_k is the key-vector dimension for scaling. ProbSparse attention selects the top-u most informative

Table 1. Hyperparameter configuration of the Informer-based injury prediction model

Parameter	Value
Optimizer	Adam
Learning Rate	0.0001
Batch Size	32
Epochs	100
Learning Rate Scheduler	ReduceLROnPlateau
Encoder Layers	2
Decoder Layers	1
Attention Heads	8
Hidden Dimension	512
Dropout Rate	0.1
Activation Function	GELU
Loss Function	Binary Cross Entropy
Classification Threshold	0.5

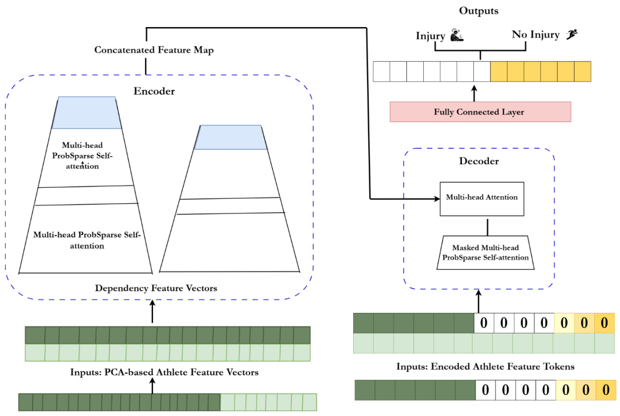

Fig. 2. Informer encoder–decoder model for athlete injury risk prediction

queries per attention head, reducing redundant computations while preserving discriminative feature information. This improves computational efficiency by lowering memory consumption and inference latency during training and deployment. The query Q, key K, and value V matrices are derived from 15-dimensional PCA feature vectors. ProbSparse attention selects only the top-u informative queries, reducing computational cost while emphasizing discriminative attributes for injury classification. ProbSparse attention prioritizes informative physiological, biomechanical, and workload features, assigning greater focus to injury-relevant indicators. This selective mechanism improves computational efficiency and enhances learning of important injury-related patterns from multimodal athlete data.

3.3.2. Dependency Pyramid Representation

The Informer encoder uses a layered dependency pyramid to progressively abstract feature relationships. Each encoder layer transforms 15-dimensional PCA representations into injury-relevant patterns by integrating physiological, biomechanical, and workload information for decoding. The dependency pyramid extracts hierarchical temporal features, where lower layers capture short-term fluctuations and deeper layers encode long-range dependencies, enabling discrimination between acute biomechanical stress and gradual physiological deterioration for injury-risk prediction.

3.3.3. Decoder Attention Mechanism

The decoder refines injury-relevant feature representations using masked multi-head ProbSparse self-attention and cross-attention, preventing future information leakage while focusing on discriminative athlete monitoring features before classification. Certain attention heads capture physiological, biomechanical, and workload relationships, including heart rate, fatigue index, acceleration, gait symmetry, and training intensity. Concatenated head outputs enable simultaneous recognition of multiple injury-relevant patterns, improving classification robustness.

3.3.4. Injury Risk Prediction Layer

A fully connected layer maps features to injury-risk probability via Eq. (6).

$$y = \sigma\left(W_o h + b_o\right) \tag{6}$$

Here Eq. (6), h is the feature vector, W_o is the weight matrix, and b_o is the bias term. A decision threshold of 0.5 is applied to the sigmoid output $\hat{y}$ for binary injury classification, while binary cross-entropy loss directly optimizes discrimination between injury-risk and no-injury classes.

4. Results and discussion

Informer framework effectively predicts sports injury risk. The confusion matrix records 1032 true negatives and 50 true positives, with 4 false negatives and zero false positives, yielding a total misclassification rate of 0.37% across

1086 test samples. Informer outperforms baselines: 99.63% accuracy, 99.95% ROC-AUC.

4.1. System Specification

Framework runs on Intel i5-14400 with 16 GB RAM, UHD 730, 477 GB storage, Windows 11 Pro, supporting PCA and Informer training.

4.2. Model Evaluation

The Informer achieves reliable injury classification using Accuracy, Precision, Recall, F1-score, and ROC-AUC, outperforming Logistic Regression, Naive Bayes, MLP, and SVM.

Fig. 3 shows Informer metrics: 0.9963 accuracy, precision, recall, 0.9962 F1-score, and 0.9995 ROC-AUC, indicating strong, balanced injury risk prediction.

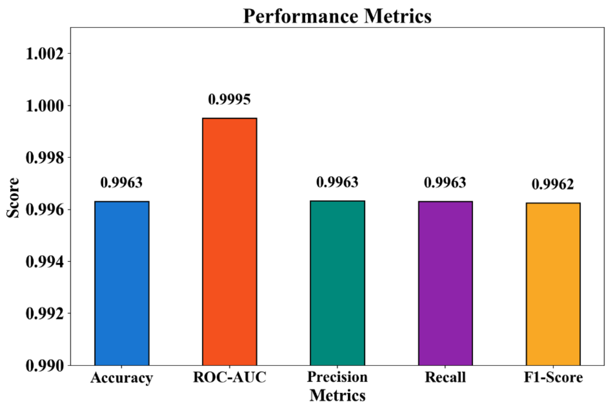

Fig. 3. Proposed Model Performance Metrics

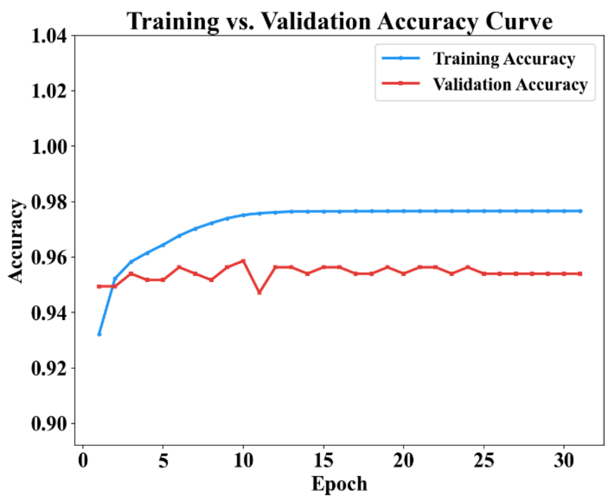

Fig. 4. Training–Validation Accuracy Trend

Fig. 7 shows stable convergence with training accuracy of 0.97 and validation accuracy of 0.95, with both curves

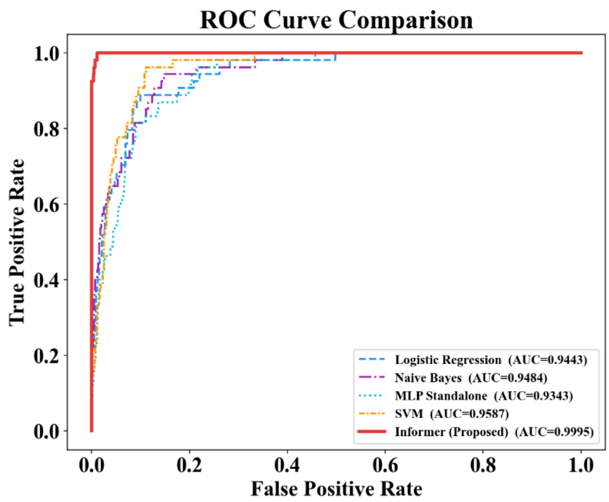

Fig. 5. ROC Model Comparison

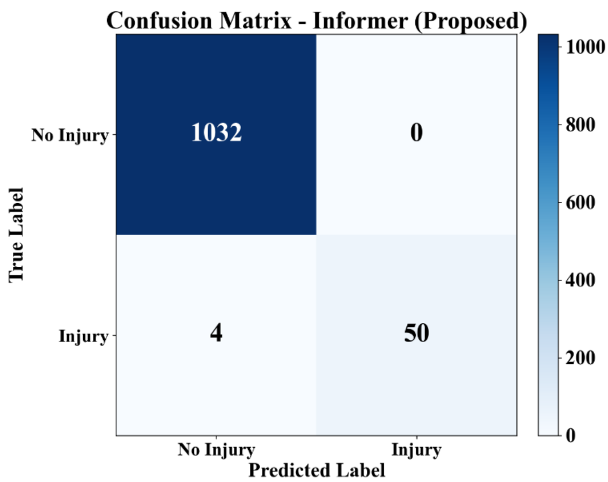

Fig. 6. Injury Prediction Classification Results

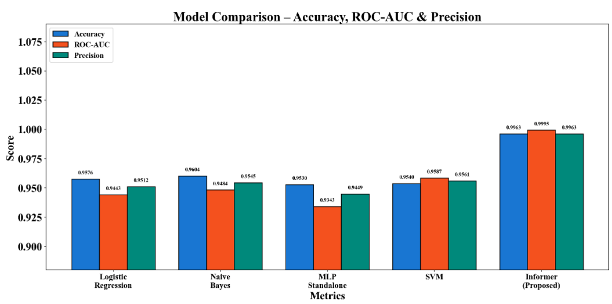

Fig. 7. Comparative Model Performance Metrics

exhibiting minimal fluctuations throughout training, indicating reliable learning behaviour and strong generalization, further confirmed by the final test accuracy of 99.63% on the held-out dataset.

Fig. 5 shows Informer outperforms Logistic Regression, Naïve Bayes, MLP, and SVM with near-perfect ROC-AUC 0.9995 for injury prediction.

Fig. 6 shows only 4 misclassifications (0.37%) among 1086 samples, with zero false positives, confirming effective injury classification.

4.3. Comparative Performance Analysis of Machine Learning Models for Injury Prediction

Informer outperforms traditional models through attention-based multimodal feature learning. The Informer requires higher training time of 148 seconds with $O(n \log n)$ complexity; however, its inference time of 0.91 ms/sample is lower than MLP (1.95 ms/sample) and SVM (1.42 ms/sample), and its ProbSparse attention ensures scalable real-time wearable sensor-based athlete monitoring without significant latency, as summarized in Table 2.

Fig. 7 shows Informer outperforms traditional ML models with 0.9963 accuracy, 0.9995 ROC-AUC, and 0.9963 precision in injury prediction.

Table 3 presents five-fold cross-validation results, with accuracies ranging from 99.45% to 99.72%. A mean accuracy of 99.58% and 0.09 standard deviation confirm strong stability, robustness, and reliable generalization.

4.3.1. Calibration Analysis

Table 4 shows reliable probability estimates, with low Brier Score, ECE, and MCE indicating minimal calibration error, supporting the practical use of predicted injury-risk probabilities for decision-making in athlete monitoring and injury prevention.

4.4. Performance Comparison

Informer outperforms existing methods on standard metrics, thanks to multimodal integration and advanced techniques, as shown in Table 5.

The Informer model is selected over LSTM and GRU because they suffer from vanishing gradients over long monitoring sequences, making them unable to capture cumulative fatigue and workload spikes that precede injuries. The Temporal Fusion Transformer incurs higher computational cost without proportional performance gain. CNN-LSTM models combine convolutional feature extraction with LSTM-based sequence learning but are limited by fixed receptive fields and vanishing gradients. Informer's ProbSparse attention captures long-range temporal dependencies across entire sequences, enabling superior modeling of cumulative fatigue and workload escalation patterns critical for sports injury detection. Table 6 compares the proposed Informer with [15], where ProbSparse self-attention captures global multimodal dependencies and long-range feature interactions, enabling more effective attention-weighted injury classification than LSTM-FCN.

Table 7 compares dimensionality reduction methods, showing PCA achieved the highest accuracy and ROC-AUC. Despite capturing complex relationships, nonlinear approaches were less effective, supporting PCA as the optimal feature representation technique.

4.5. Discussion

Sports injuries are classified as overuse injuries from repetitive strain, fatigue, and workload escalation, or acute injuries from sudden trauma, acceleration spikes, and abnormal ground reaction forces. The proposed framework addresses both categories by integrating physiological indicators such as heart rate and fatigue index for overuse pattern detection, and biomechanical parameters such as acceleration and ground reaction force for acute injury risk identification. Multimodal Informer achieves 99.63% accuracy, 99.95% ROC-AUC, 99.63% precision. The high accuracy and ROC-AUC values may raise concerns regarding potential overfitting. The stable training-validation accuracy trend, gradual convergence behaviour, and strong performance on the held-out test set indicate effective generalization of the proposed model. The near-perfect performance values are further supported by the balanced class distribution across diverse physiological conditions, biomechanical movement patterns, and workload intensities, reducing the likelihood of class imbalance bias influencing predictive outcomes. Broader validation using independent athlete datasets collected under field conditions remains necessary to strengthen the interpretation and generalizability of these highly accurate predictive outcomes beyond the current controlled dataset evaluation. Informer-based models outperform traditional methods for injury prediction. The dataset is sourced from a public repository and may not fully represent physiological and biomechanical variability across diverse athlete populations, sport disciplines, and performance levels. Prospective evaluation across different sport disciplines and athlete populations is required before practical deployment in real-world sports monitoring environments. Performance across varying sensor configurations and sport-specific movement patterns requires further evaluation using larger and more diverse athlete datasets to improve model generalization and practical applicability.

5. Conclusion

An Informer-based multimodal approach utilizing athlete monitoring data was used to develop the framework to predict athletes who are at risk of sustaining an injury. The

Table 2. Computational complexity and inference efficiency comparison across models

Model	Training Time (s)	Inference Time (ms/sample)	Complexity
Logistic Regression	12	0.18	O(n)
Naïve Bayes	8	0.12	O(n)
SVM	84	1.42	$O\left(n^2\right)$
MLP	105	1.95	$O(n \cdot L)$
Informer	148	0.91	O(n log n)

Table 3. Five-Fold Cross-Validation Results

Fold	Accuracy (%)
Fold 1	99.45
Fold 2	99.54
Fold 3	99.72
Fold 4	99.63
Fold 5	99.58
Mean ± Std	99.58 ± 0.09

Table 4. Calibration Analysis Results

Metric	Value
Brier Score	0.0064
Expected Calibration Error (ECE)	0.0081
Maximum Calibration Error (MCE)	0.0143

Table 5. Performance Comparison with Recent Sports Injury Prediction Studies

Method	Accuracy (%)	ROC-AUC (%)	Precision (%)
BioMAT Transformer [6]	94.20	95.10	93.60
Foot Pronation Prediction [13]	74.70	88.20	82.10
ACL Injury Risk Detection [14]	92.30	93.70	91.50
ML Injury Risk Estimation [8]	93.10	94.40	92.60
Proposed Informer Model	99.63	99.95	99.63

Table 6. Comparison of LSTM-FCN and Informer-Based Framework

Aspect	LSTM-FCN [15]	Proposed Informer Model
Feature Extraction	CNN local features; LSTM temporal dependencies	ProbSparse self-attention global dependencies
Classification Behavior	Recurrent feature representations	Attention-weighted feature interactions
Long-Range Dependency	Limited by recurrent memory	Direct full-sequence modeling

Table 7. Comparison of Dimensionality Reduction Techniques

Method	Accuracy (%)	ROC-AUC (%)	Features
PCA	99.63	99.95	15
Kernel PCA	99.51	99.82	15
t-SNE	98.88	98.91	15
UMAP	99.26	99.41	15

multimodal framework is a combination of physiological, biomechanical, environmental, and workload characteristics obtained from athlete monitoring systems to develop an accurate model for predicting injury. The process includes collecting data, preprocessing data, creating a multivariate representation of features using Principal Component Analysis, and applying the Informer model to classify samples as either having sustained an injury or not sustaining an injury. Results from the experimental study showed that the proposed framework produced the best predictive performance with 99.63% accuracy, 99.95% ROC-AUC, and 99.63% precision. Results showed that the Informer framework provides a method for capturing complex relationships among multimodal variables associated with athletes and that the prediction of risk for injury using the Informer framework significantly exceeds that achieved with traditional machine learning prediction models. Several advantages of the multimodal framework are that it provides improved accuracy in predicting risk for injury among athletes, allows for the efficient modeling of high-dimensional multimodal variables, and provides coaches and sports scientists with a method to support early detection of injury risk, which will allow coaches and sports scientists the opportunity for making preventative decisions and optimizing training programs.

Future work may expand on this study by including data from real-time wearable sensors for ongoing injury monitoring and prediction. Incorporating SHAP and LIME in future work would enhance interpretability by identifying key physiological and biomechanical features influencing injury risk predictions, thereby improving trust, usability, and practical adoption in sports settings. Additionally, advanced multimodal deep learning architectures and larger athlete populations from a variety of sports may enable model generalization and robustness. Providing explanation techniques to assist with the interpretation of injury prediction results may provide additional opportunities for practical use in sports analytics and athlete health management systems.

6. Acknowledgements

The author would like to acknowledge the institutional support provided by the Institute of Physical Education, Anyang Normal University, for facilitating the completion of this research. No external financial support was received for this study.

7. Declarations:

7.1. Data Availability

https://www.kaggle.com/datasets/anjalibhegam/multimodal-sports-injury-prediction-dataset

7.2. Conflicts of Interest

The authors declare that they have no known competing financial interests or personal relationships that could have influenced the work reported in this paper.

7.3. Funding Statement

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

7.4. Author Contribution

L.S. conceived and designed the study; conducted the literature review; developed the AI-powered predictive analytics framework; analyzed wearable sensor data and interpreted the results; prepared all figures and tables; drafted the manuscript; revised the manuscript critically for important intellectual content; and approved the final version of the manuscript for publication.

7.5. Ethical Approval

This article does not contain any studies involving human participants or animals performed by any of the authors.

7.6. Consent to Participate

Not applicable.

7.7. Consent to Publication

All authors have provided consent for publication of this manuscript.

7.8. Competing Interests

The authors declare no competing interests

References

[1] L. Yang, O. Amin, and B. Shihada, (2024) *"Intelligent wearable systems: Opportunities and challenges in health and sports"* **ACM Computing Surveys** ***56***(7): DOI: 10.1145/3648469.

[2] A. Alzahrani and A. Ullah, (2024) *"Advanced biomechanical analytics: Wearable technologies for precision health monitoring in sports performance"* **Digital Health** ***10***: 1–18.

[3] M. Kovoor, M. Durairaj, M. S. Karyakarte, M. Z. Hussain, M. Ashraf, and L. P. Maguluri, (2024) *"Sensor-enhanced wearables and automated analytics for injury prevention in sports"* **Measurement: Sensors** ***32***: 101054. DOI: 10.1016/j.measen.2024.101054.

[4] A. Arzehgar et al., (2025) *"Sensor-based technologies for motion analysis in sports injuries: A scoping review"* **BMC Sports Science, Medicine and Rehabilitation** ***17***(1): 15. DOI: 10.1186/s13102-025-01063-z.

[5] C. Tang, W. Yi, Z. Zhang, E. Occhipinti, and L. G. Occhipinti, (2026) *"AI-Driven Smart Sportswear for Real-Time Fitness Monitoring Using Textile Strain Sensors"* **IEEE Transactions on Biomedical Engineering** ***73***(2): 555–565. DOI: 10.1109/TBME.2025.3588051.

[6] M. Sharifi-Renani, M. H. Mahoor, and C. W. Clary, (2023) *"BioMAT: An open-source biomechanics multi-activity transformer for joint kinematic predictions using wearable sensors"* **Sensors** **23**(13): 5778. DOI: 10.3390/s23135778.

[7] D. Hollinger, M. C. J. S, H. Chen, and M. Zabala, (2024) *"The Effect of Sensor Feature Inputs on Joint Angle Prediction across Simple Movements"* **Sensors *24***(11): 3657. DOI: 10.3390/s24113657.

[8] P.-E. Dandrieux et al., (2023) *"Relationship between a daily injury risk estimation feedback based on machine learning techniques and actual injury risk in athletics: Protocol for a prospective cohort study over an athletics season"* **BMJ Open *13***(5): e069423. DOI: 10.1136/bmjopen-2022-069423.

[9] T. Tan et al., (2023) *"A scoping review of portable sensing for out-of-lab anterior cruciate ligament injury prevention and rehabilitation"* **NPJ Digital Medicine *6***(1): 46. DOI: 10.1038/s41746-023-00782-2.

[10] C. C. Claros-Olivares, M. N. Anderson, W. Qian, A. J. Brockmeier, and T. A. Buckley, (2025) *"A Machine Learning Model for Post-Concussion Musculoskeletal Injury Risk in Collegiate Athletes"* **Sports Medicine *55***(8): 1971–1982. DOI: 10.1007/s40279-025-02196-4.

[11] A. Shaw, P. Newman, J. Witchalls, and T. Hedger, (2023) *"Externally validated machine learning algorithm accurately predicts medial tibial stress syndrome in military trainees: A multicohort study"* **BMJ Open Sport and Exercise Medicine *9***(2-e001566): DOI: 10.1136/bmjsem-2023-001566.

[12] Anjalibhegam. *Multimodal sports injury prediction dataset*. Accessed March 19, 2026. 2026.

[13] L. Xiang, Y. Gu, A. Wang, V. Shim, Z. Gao, and J. Fernandez, (2023) *"Foot pronation prediction with inertial sensors during running: A preliminary application of data-driven approaches"* **Journal of Human Kinetics *87***: 29–40. DOI: 10.5114/jhk/163059.

[14] A. Benjaminse, M. Mendicino, E. M. Nijmeijer, P. Margheriti, A. Gokeler, and S. Di Paolo, (2026) *"On-Field Assessment of Joint Load in Football Using Machine Learning (Part II)"* **Sensors *26***(8): 2562. DOI: 10.3390/s26082562.

[15] X. Liu, Z. Chen, B. Zheng, Q. Guo, Y. Yu, D. Zhang, and A. Wulamu, (2022) *"Spatiotemporal and kinematic characteristics augmentation using Dual-GAN for ankle instability detection"* **Mathematical Biosciences and Engineering *19***(10): 10037–10059. DOI: 10.3934/mbe.2022469.