A basketball player injury risk warning system based on physical modeling and statistical analysis

Yutong Liu

Beijing No.2 Middle School, Beijing, China

kevinbfh3017@outlook.com

Keywords: Physical Modeling, Statistical Analysis, Injury Risk Warning, Graph Convolutional Neural Network, GCN-INJ

Abstract: This paper proposes an injury risk warning system for basketball players based on physical modeling and statistical analysis. This paper combines sensor technology and deep learning technology to propose a deep graph convolution model called GCN-INJ. It uses a variety of information such as players' physiological parameters, training intensity, and game data to capture the mutual influence between players through a graph convolutional network (GCN) to more accurately predict the risk of individual injuries. Experimental results show that the model is superior to traditional statistical models in both prediction accuracy and stability. The effectiveness of the model is verified through visual analysis of a large amount of data, and its limitations in practical applications are explored. Future research will focus on further optimizing model performance, introducing more dimensional data sources, improving prediction accuracy, and developing user-friendly applications so that coaches and athletes can monitor and manage injury risks in real time, thereby better supporting sports performance and health maintenance.

1. Introduction

Basketball players often experience injuries such as sprains, strains, fractures, dislocations, and ligament tears during intense games and rapid movements [1]. Common injuries include ankle sprains, which frequently occur when players step on opponents' feet, and anterior cruciate ligament (ACL) injuries, often caused by sudden directional changes or unprepared high-intensity activities [2]. These injuries significantly impact players' careers. Minor injuries may require weeks of recovery, but severe cases like fractures or ligament tears can take months. During this time, players lose game opportunities and the chance to enhance their skills and visibility. Prolonged injuries can also lead to psychological challenges like anxiety, depression, and reduced confidence, negatively affecting their careers and market value, making them less attractive to teams and sponsors.

To mitigate injury risks, basketball players and coaching teams implement preventive measures such as proper warm-ups and stretching to improve muscle flexibility and joint mobility; strength training, especially for the legs and core, to enhance stability; and using protective gear like knee and ankle braces. Players also focus on mastering correct techniques and landing skills to avoid technical errors. However, these traditional methods have limitations. Warm-ups may vary in effectiveness due to individual differences, and protective gear might hinder performance. Additionally, traditional methods struggle to monitor and predict risks from factors like overtraining and cumulative fatigue.

Recent advances in biomechanical modeling and data analysis have provided new solutions for injury risk prediction. By creating physical models of athletes in various motion states and analyzing historical data, researchers can identify risk factors for injuries. Early warning systems based on physical modeling and statistical analysis offer more accurate assessments and provide coaches and medical teams with scientific training adjustments, thereby reducing injury risks [3][4]. These systems have transformed injury prevention by combining large datasets and predictive tools to better understand athletes' physical conditions and potential hazards.

  DOI: 10.25236/icmmct.2025.026

Future early warning systems will become more intelligent and efficient, capable of real-time monitoring and delivering personalized injury warnings. This advancement will improve athlete safety, prolong careers, and benefit the broader sports industry by advancing sports science and protecting athletes' health rights. As data accumulates, these systems will become more sophisticated, offering tailored services for specific sports and athlete groups, further enhancing their practical and social value.

This paper aims to collect extensive data on NBA players, analyze injury risks through physical modeling and statistical methods, and predict future risks using machine learning algorithms and physical modeling techniques. The study includes five chapters: Chapter 1 introduces the background, Chapter 2 reviews related research on physical modeling and statistical analysis, Chapter 3 presents the construction of the injury risk warning system, Chapter 4 evaluates its results, and Chapter 5 summarizes the findings.

2. Related Research

2.1. Physical modeling

Physical modeling employs mathematical equations to simulate physical systems, aiding in injury prediction and prevention in basketball. Biomechanical tools like motion capture systems and force plates analyze movements such as jumping and landing, enabling precise measurement of biomechanical loads [5][6]. Modern sensing technologies, including wearable sensors, high-speed cameras, and IMUs, enrich data collection by monitoring real-time motion parameters and providing multidimensional data for personalized training and recovery [7].

Injury risk systems rely on data analysis using techniques like finite element analysis (FEA) and multibody dynamics software to simulate body mechanics, revealing joint and bone interactions during complex movements [8]. Machine learning algorithms, such as neural networks, complement these methods by identifying patterns in large datasets and predicting injury risks. Statistical methods like regression analysis and PCA further enhance insights into injury trends and critical risk factors.

Advancements in predictive analytics combine physical modeling and statistical analysis to improve injury forecasting. VR, AR, and smart wearables enable safe practice of high-risk actions and real-time feedback. As technology evolves, physical modeling will become pivotal in basketball training, safeguarding athletes' health and careers.

2.2. Statistical analysis of injury prediction

Statistical analysis is essential in injury prediction, revealing patterns and trends in historical data. In basketball, injuries range from acute to chronic types, with statistical methods identifying risk factors and guiding preventive measures. Key steps include data collection, preprocessing, modeling, and evaluation. Data is gathered from physiological indicators, training records, and wearable devices, while preprocessing ensures quality through cleaning, interpolation, and outlier detection.

Statistical models like linear regression, logistic regression, and survival analysis explore injury relationships and timing [9]. Machine learning algorithms, such as decision trees, random forests, and neural networks, enhance prediction accuracy by identifying complex patterns [10]. Feature selection methods like PCA and RFE improve model interpretability and performance [11]. Optimization through hyperparameter tuning and validation techniques, including cross-validation and LOOCV, ensure stability and reliability [12].

Statistical analysis effectively forecasts injuries, enabling data-driven adjustments to training and identifying high-risk actions. Advances in big data, AI, and IoT devices will further improve prediction accuracy and provide richer insights into injury prevention and management.

3. Injury risk early warning system for basketball players

Basketball is a global sport invented by Canadian James Naismith in 1891. It quickly became one of the most popular sports worldwide. Basketball games involve two teams of five players each, aiming to score by throwing the ball into the opponent's basket. Beyond basic rules, basketball encompasses skills and tactics like dribbling, passing, shooting, and defense. It promotes physical and mental health and fosters community cohesion. Whether professional or amateur, basketball attracts players globally with its unique appeal.

The NBA (National Basketball Association), founded in 1946, is one of the highest-level professional basketball leagues. It gathers the world's top players and captivates millions of fans yearly. The league consists of the Eastern and Western Conferences, each divided into three divisions. The season includes a regular season from October to April and playoffs from late April to mid-June. As the NBA transitions into the small-ball era, the pace of the game has accelerated, increasing players' physical burden and injury risks, especially to the lower limbs.

3.1. Graph Convolutional Neural Networks

Graph Convolutional Networks (GCNs) are deep learning models designed to process graph-structured data. Unlike traditional convolutional neural networks, which handle spatial or temporal data like images, GCNs manage complex forms like social networks or molecular structures. GCNs transmit information between nodes and learn local and global node feature representations. They use the graph Laplacian matrix to capture graph topology, enabling models to embed node representations.

GCNs have evolved with variants addressing diverse graph structures and tasks, including non-uniform graphs and large-scale computations. They combine with techniques like attention mechanisms to enhance performance. GCNs excel in fields like molecular property prediction, social network analysis, and recommendation systems. Using GCNs for basketball injury prediction is innovative. By modeling player interactions, each player becomes a node, and connections represent interactions like passing or defending. This structure captures individual performance data and teamwork dynamics.

3.1.1. Data preparation and preprocessing

Relevant data includes physiological (e.g., heart rate, blood pressure), performance (e.g., running distance, jumping height), video analysis (e.g., movement frequency), and injury records. Wearable devices collect real-time physiological data, GPS trackers record movement, and video analysis software evaluates technical actions. Rigorous preprocessing ensures data accuracy by cleaning, unifying formats, and handling outliers.

3.1.2. Model construction

We constructed a GCN model using players' physiological, performance, video, and injury data. Nodes represent athletes, and edges signify relationships, such as teamwork. Feature vectors for each node include physiological indicators and historical data. Graph convolution layers iteratively update node features by aggregating information from neighbors, capturing complex relationships.

Attention mechanisms enhance focus on critical features. During training, labeled data optimizes parameters to predict injury risks. Regularization prevents overfitting, while cross-validation fine-tunes hyperparameters. After training, the model’s performance is evaluated using unknown data. Successful models predict injury risks, guiding coaches in preventive strategies and training adjustments.

3.1.3. Graph Convolutional Neural Networks

GCNs specialize in processing graph-structured data by extracting and propagating features. In a graph G=(V,E), V and E represents nodes and edges, respectively. Nodes have feature vectors, and the adjacency matrix indicates connections. Spectral graph theory defines graph convolution using the Laplacian matrix , where is the degree matrix.

3.2. Injury Prediction Graph Convolutional Neural Network

As Figure 1 shows, the proposed GCN-INJ model predicts basketball injuries using graph-structured data. Nodes represent players, with features like speed, jump height, and position. Edges indicate interactions, capturing relationships crucial for injury prediction. The model's GCN blocks update node features via adjacency matrix-based aggregation. Graph pooling reduces complexity, while normalization stabilizes training.

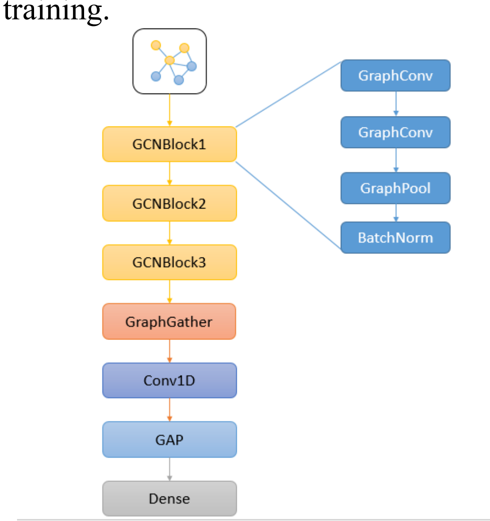

Figure 1 GCN-INJ Model Structure.

The GraphGather layer aggregates node features into a global vector, processed by Conv1D and pooling layers to identify patterns. A dense layer outputs the injury probability, helping coaches and medical teams take preventive measures and optimize training. By combining GCNs with traditional neural networks, the model accurately predicts injuries, supporting athletes' health and team performance.

4. Analysis of model experiment results

4.1. Data Visualization

We first visualized the dataset. We collected sports injury data of NBA players from 2010 to 2020. As Figure 2 shows, 2012 had the highest number of injuries.

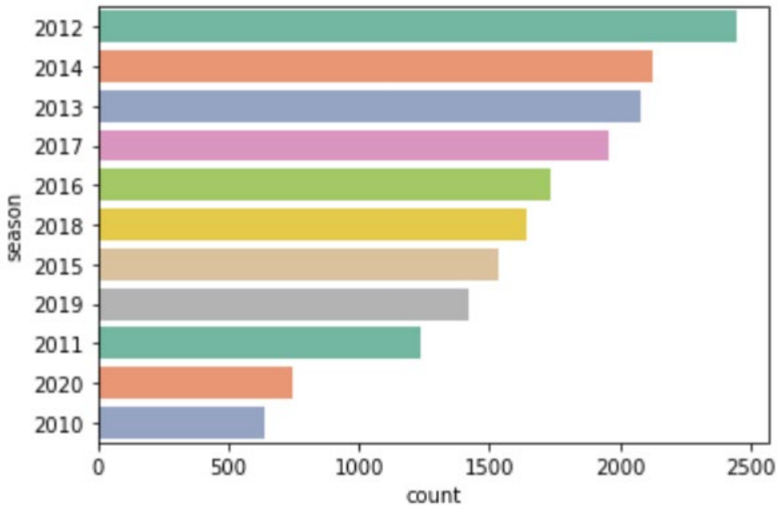

Figure 2 Visualization of total number of injuries.

Then in Figure 3 we visualized the top ten teams with the most injuries in 2012, and we can see that the Hornets had the most injuries in 2012 .

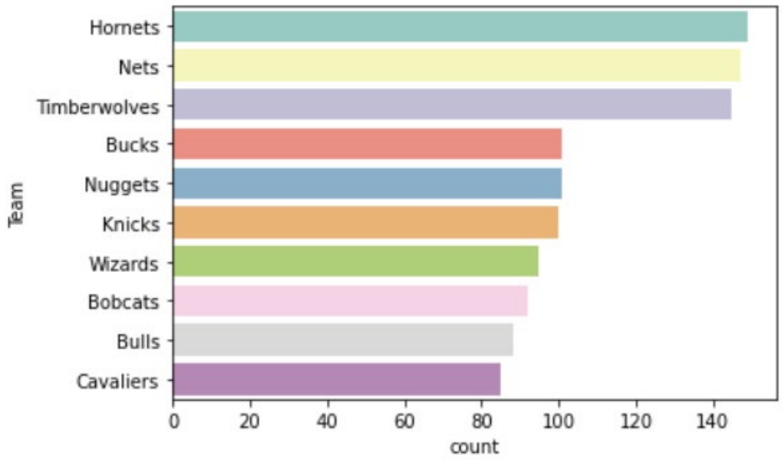

Figure 3 Visualization of the total number of team injuries.

4.2. Model training and testing results

The model was trained for 100 epochs with an initial learning rate of 0.001, following a gradual decay strategy where the rate decreased by 1% every 10 epochs. This approach accelerates early-stage loss reduction while refining weight adjustments later for better convergence.The training process used the Adam optimizer, an adaptive method combining momentum and RMSProp. Adam updates parameters by maintaining first-order and second-order moment estimates of the gradient, making it effective for sparse gradients and non-stationary objective functions.The primary evaluation metrics were Mean Squared Error (MSE) and Mean Absolute Error (MAE), both common loss functions in regression tasks. These metrics penalize large deviations, guiding the model toward more accurate predictions.

As Figure 4 shows, to further demonstrate the effectiveness of the model, we visualize the intermediate features of the model.

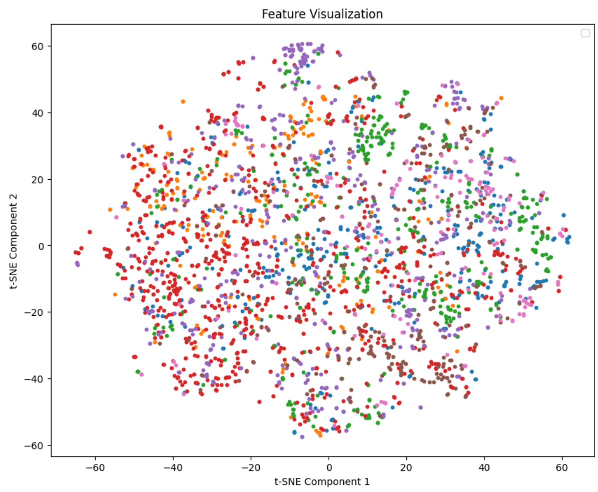

Figure 4 Input feature distribution.

Figure 4 shows the distribution of input features. It can be seen that the features that have not been processed by the model cannot present the results well. Then we feed the data into our model, and we can see that the features after being fed into the GCN model have obvious feature distribution, which shows the effectiveness of our model.

Finally, we tested the model. When we input the specific information of a player in each game, the model can output the injury probability of the corresponding player.

5. Summary and Outlook

This paper presents an injury risk warning system for basketball players using physical modeling and statistical analysis. Chapter 1 introduces the background, emphasizing the importance of injury prevention in professional sports and outlining current research challenges. Chapter 2 discusses the technical foundation, including sensor technology, data acquisition, and machine learning, particularly the advantages of graph convolutional networks (GCNs) for structured data. Chapter 3 details the GCN-based injury prediction model, which integrates players' physiological parameters, training intensity, and game data to predict individual injury risks. The model outperforms traditional statistical models in accuracy and stability. Chapter 4 visualizes the collected data, evaluates the model's effectiveness, and explores limitations such as data quality and sample size. Future research will enhance model performance by incorporating additional data sources like biomechanical parameters and social network analysis. The system's applicability to other sports will also be explored, along with the development of user-friendly applications for real-time injury risk management.

References

[1] Zhu Chao, and Dai Bin. College Students' Physical Education and Health. Chongqing University Electronic Audiovisual Publishing House Co., Ltd., 2021.

[2] Zhen Jie, and Xiao Tao. Concise Sports Biomechanics. Chongqing University Electronic Audiovisual Publishing House Co., Ltd., 2020.

[3] Wang L, Huang L. Analysis of the Causes and Prevention of Sports Injuries in School Physical Education and Training Based on Big Data Analysis[C]//2020 International Conference on Computers, Information Processing and Advanced Education (CIPAE). IEEE, 2020: 111-113.

[4] Bittencourt, Natalia FN, et al. "Complex systems approach for sports injuries: moving from risk factor identification to injury pattern recognition—narrative review and new concept." British journal of sports medicine 50.21 (2016): 1309-1314.

[5] Johnson, William R., et al. "Predicting athlete ground reaction forces and moments from motion capture." Medical & biological engineering & computing 56 (2018): 1781-1792.

[6] Chaaban, Courtney R., et al. "Combining inertial sensors and machine learning to predict vGRF and knee biomechanics during a double limb jump landing task." Sensors 21.13 (2021): 4383.

[7] Rana, Manju, and Vikas Mittal. "Wearable sensors for real-time kinematics analysis in sports: A review." IEEE Sensors Journal 21.2 (2020): 1187-1207.

[8] Adamski, Dirk. Simulation in Chassis Technology . Springer Fachmedien Wiesbaden: Wiesbaden, Germany, 2020.

[9] Bullock, Garrett S., et al. "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 (2022): 2469-2482.

[10] Van Eetvelde, Hans, et al. "Machine learning methods in sport injury prediction and prevention: a systematic review." Journal of experimental orthopaedics 8 (2021): 1-15.

[11] Li, Jundong, et al. "Feature selection: A data perspective." ACM computing surveys (CSUR) 50.6 (2017): 1-45.

[12] Bergstra, James, and Yoshua Bengio. "Random search for hyper-parameter optimization." Journal of machine learning research 13.2 (2012).

[13] Yan, Sijie, Yuanjun Xiong, and Dahua Lin. "Spatial temporal graph convolutional networks for skeleton-based action recognition." Proceedings of the AAAI conference on artificial intelligence . Vol. 32. No. 1. 2018.