Proceedings of CONF-MLA 2025 Symposium: Applied Artificial Intelligence Research
DOI: 10.54254/2755-2721/2025.BJ27632

The Advancements of Artificial Intelligence Application in Sports Injury Prediction

Qifan Huo

Guangzhou Zhixin High School, Guangzhou, China
ekxehx@gmail.com

Abstract. This article intends to explore the development prospect of Artificial Intelligence (AI) application in sports injury prediction. This field is not only important for sports industry but also for medical fields since the results of injury prediction could be an instruction for medical care. Specifically, the article states some AI techniques to make predictions. The article will elaborate on these techniques. The article states steps for data cleaning which are a vital step for AI models. In addition, the article lists machine learning models including random forests, Support Vector Machines, K-Nearest Neighbors and deep learning models including Convolutional Neural Networks, Artificial Neural Networks, Recurrent Neural Networks. The article would focus on the principles of these models. According to principles, models are really useful and promising for personalized injury prediction since AI models can analyze complex patterns of variables. Some models like Recurrent Neural Networks that require sequential data can get accurate relationships between data and time and make better prediction.

Keywords: Deep learning, machine learning, sports injury prediction

1. Introduction

Sports always play an important role in human societies. Nowadays, with the improving standard of living, sports are becoming more meaningful. Sports can not only help promote individual development, but can also promote technological advancement, social cohesion and cultural communication. However, while loving sports, many people suffer from injuries caused by playing sports. Since sports are becoming more important, there are many studies and methods about how to handle injuries. For example, there are many methods like acute injury management that focus on principles of Protection, Optimal Loading, Ice, Compression, Elevation (PRICE). In addition, Nordic Hamstring Exercise that can reduce the risk of recurrence. As Artificial Intelligence (AI) becomes increasingly popular, researches on sports' injuries have increased and achieved significant results. In order to decrease risks or negative effect of injuries and help people enjoy sports better, it is important not only to learn basic methods for managing injuries, but also to develop a deeper understanding based on relevant research. To that end, injury-related factors can be studied and categorized systematically.

According to the study carried out by Bahr [1], risk factors can be divided into intrinsic and extrinsic or modifiable and non-modifiable factors. Factors can be understood by a model called

© 2026 The Authors. This is an open access article distributed under the terms of the Creative Commons Attribution License 4.0 (https://creativecommons.org/licenses/by/4.0/).

DOI: 10.54254/2755-2721/2025.BJ27632

Meeuwisse. It's a model that explains factors for injuries and indicates factors' weights. According to the Meeuwisse model, internal factors are prerequisites of injuries, and internal factors are not sufficient to cause injuries. Instead, external factors are enabling factors. So injuries are caused by the interaction of internal factors and external factors. In addition, the article uses a new way to describe injuries. Instead of only focusing on biomechanical, the article considers the real circumstances like two men block in volleyball. And ignoring contact or non-contact could oversimplify injuries. Furthermore, AI is applied to researches into injuries.

AI applications in sports are promising, and it can be reflected in many researches. For sports' injuries, AI can help with injuries' prediction, prevention etc. According to the study carried out by Claudino et al. [2], artificial neural network, decision tree classifier, and support vector machine have been used in soccer, basketball and other sports events. The article mentions that AI performs better than traditional methods in predicting injury risk in studies. It indicates that AI applications have great potential to predict performance. In addition, Dindorf et al. also points out that traditional studies which are based on sequence prevention are too simple to recognize some potential risks of injuries [3]. In real situations, AI can be advantageous since AI can handle a large volume of data, find hidden patterns and so on. Therefore, AI will probably advance the field of sports injuries.

2. Method

2.1. The steps for injury prediction based on data analytics

Data analytics are the basis of injury prediction. High-quality data are directly related to the performance of prediction. The steps of data analytics are preprocessing, feature selection, data integration, visualization [4]. Specifically, preprocessing requires standardizing data and handling missing data. Feature selection requires identifying weights of different injury factors or finding important injury factors. Data integration is to combine data from different systems into one dataset. For example, combining heart rate from tracking devices and injury history record. Integrating data is very important because it helps models handle all relevant injury factors. Visualization is using charts, graphs, tables to indicate the patterns pf data. By visualization, the prediction could be clearer and people can understand the prediction based on models better [4].

2.2. The Machine Learning-based injury prediction

Machine Learning (ML) plays an important role in Artificial Intelligence (AI)-based injury prediction. ML can reveal the relationships between different injury factors like physical performance metrics, previous injuries, training intensity. Musat et al. describe techniques of ML and each technique's property: Firstly, The Random Forests (RF) model is the model that can create decision trees and average different decision trees' results. The Random Forests model can reduce over-fitting and provides robust predictions. To that end, athletes can use smart wearables to record data and save data in USB flash drive or cloud storage. After analysis of activity and probability of injury, a database can be found. In addition, discrete wavelet transform (DWT) is very important. DWT can seperate wavelet's scale so that three-dimensional data can be analyzed by computers in a binary manner. After these steps, random forests can use bootstrap sampling to create k training sets and form decisions trees. Assuming there are total M features, extract m features. Calculate weight for each leaf node [5].

Secondly, the Support Vector Machines (SVM) can find an optimal boundary and classify data categories. By doing this, SVM can find out the degree of different injury risks. The injury

DOI: 10.54254/2755-2721/2025.BJ27632

prediction can depend on the degree of injury risks. For instance, this model can discern high-risk groups and make predictions. The last technique of ML is K-Nearest Neighbors (KNN). KNN will sample similar athletes. For instance, this method would analyze another athlete with similar risk profile when studying one athlete's circumstance. These ML techniques can make patterns of injury risks and make prediction.

2.3. The Deep Learning-based injury prediction

Deep Learning (DL) models can handle large datasets and figure out patterns without manual feature extraction [6]. There are some techniques of [4]: Firstly, Convolutional Neural Networks (CNNs) are used to analyze lattice-like data structures. Specifically, the model uses filters to extract regularities that are invariant with respect to their position. CNNs can effectively decrease computational complexity because of local connectivity and training. The model can analyze complex data like images of training, video about gait. Athletes can wear sensors so that it will help analysis. By analyzing those patterns, injuries can be predicted from abnormal gaits or any other postures [4, 7]. Secondly, Artificial Neural Networks ANNs) are like biological neural networks. They consist of different layers of interconnected neurons that can extract features and analyze patterns of injury factors. The ANNs can make predictions for a variety of injuries by analyzing complex and new variables. This flexibility enables the ANNs to make personalized injury predictions because it can analyze variables like individual injury history. Lastly, the Recurrent Neural Networks (RNNs) are designed to analyze sequential data so that these models can analyze the relationship between variables and time. Therefore, RNNs can make predictions by tracking the changes of variables over time. Basically, input signals with different lengths. The RNNs will process signal sequence, that is not only present input but also previous input. The hidden layer would calculate the sequential with respective weight to get results. In addition, RNNs are divided into four classes. The many-to-one architecture would be effective for injury prediction. Input variables like heartrate, training volume and muscle fatigue will be used to get results of injury risks [4,7].

3. Discussion

Artificial intelligence (AI) models can analyze large datasets to identify patterns of risks and make personalized training plans or predictions. AI can also achieve real-time monitoring and risk alerts that can increase reliability. Furthermore, AI has been applied into a variety of sports programs. AI injury prediction is applied in both team sports and individual sports. For team sports, AI can monitor collision data like concussions, predict muscle injuries and optimize training loads. AI can analyze real-time biomechanics and detect signs of injuries. For instance, AI can detect some vulnerable areas like knees in alpine skiing. AI can monitor collision data, analyze muscle fatigue or circumstances of joints to predict probability of injuries like strains.

However, some challenges still exist. Firstly, AI models' reliance on past data could be a restriction because previous data would be more likely to be low-quality or even missing. This could affect models' accuracy and effect.

Data collection could threaten athletes' privacy. Specifically, models may memorize some sensitive information like private information of athletes. According to some studies, models would sometimes memorize data instead of learning patterns when there is too much data. Sometimes data can be accessed by figuring out regularities or rules of models. For instance, decision tree leakage can be caused by analyzing routes of decision trees. So, data minimization, anonymization, and

DOI: 10.54254/2755-2721/2025.BJ27632

controlled access are important for the protection of sensitive data. For example, try to collect as few sensitive data as possible and only collect essential information or store data with encryption. Some organizations like General Data Protection Regulation (GDPR) [8], Health Insurance Portability and Accountability Act (HIPAA) have done something like that. Specifically, GDPR requires personal sensitive data from wearables to be anonymized and pseudonymized. Both GDPR and HIPAA have strict rules for the access of data so that data won't be used on appropriate purpose. However extreme data preservation may create data sharing dilemmas and be harmful for AI models' training, so it's necessary to find a balance between protection of data and enough training.

Lack of Transparency. The lack of model transparency is another problem. For models, the interpretation needs to be improved. Some models are too complicated for athletes to understand. For instance, artificial neural networks (ANNs) have hidden layers and hundreds of neurons. Nonlinear activation functions will also increase the complexity. And convolutional neural networks (CNNs) are also complicated because CNNs are used to analyze data like images. CNNs need to detect millions of edges of images. Furthermore, the weights are not readable for athletes. Weights are numerical matrices without explanations. For example, if there is a personalized injury prediction, AI models may combine heart rate, historical records and other factors to make indicators for injuries. However, the prediction made by the model can't be understood by others. In addition, random forests include hundreds of decision trees. However, random trees models require combing data of decision trees. Those calculations are invisible for athletes. These examples prove the complexity and incomprehensibility of AI models. To improve the transparency of models, visualization can help people understand models better. Furthermore, there are some AI explainable tools that can help people find out routes or pathways of models. With clear models, adequate interpretation, users or athletes would have greater trust in the decision made by models.

4. Conclusion

This article concludes some applications of AI models and elaborates on them. Then the article analyzes some development prospects. Basically, data quality is important for AI models. It is necessary to clean data for AI models. In terms of model selection, deep learning models and machine learning models are suitable options to make injury prediction. AI models are effective for injury prediction and have significant implications for sports healthcare. This would be helpful for both professional athletes and regular sports lovers. AI models can analyze large and complex data, make real-time monitoring to make personalized injury prediction which promotes personal healthcare. In addition, AI models can be applied into a variety of sports like individual and teamwork sports. On the other hand, there are some problems to be solved like data privacy and transparency. However, some organizations like General Data Protection Regulation (GDPR) have made some rules to improve data privacy. And some AI explainable tools can help people better understand models. Therefore, the AI application in injury prediction is promising.

References

[1] Bahr, R., & Holme, I. (2003). Risk factors for sports injuries—a methodological approach. British journal of sports medicine, 37(5), 384-392.

[2] Claudino, J. G., Capanema, D. D. O., de Souza, T. V., Serrão, J. C., Machado Pereira, A. C., & Nassis, G. P. (2019). Current approaches to the use of artificial intelligence for injury risk assessment and performance prediction in team sports: a systematic review. Sports medicine-open, 5(1), 28.

[3] Dindorf, C., Bartaguiz, E., Gassmann, F., & Fröhlich, M. (Eds.). (2024). Artificial Intelligence in Sports, Movement, and Health. Springer.

DOI: 10.54254/2755-2721/2025.BJ27632

[4] Van Eetvelde, H., Mendonça, L. D., Ley, C., Seil, R., & Tischer, T. (2021). Machine learning methods in sport injury prediction and prevention: a systematic review. Journal of experimental orthopaedics, 8(1), 27.
[5] Halder, R. K., Uddin, M. N., Uddin, M. A., Aryal, S., & Khraisat, A. (2024). Enhancing K-nearest neighbor algorithm: a comprehensive review and performance analysis of modifications. Journal of Big Data, 11(1), 113.
[6] Dindorf, C., Bartaguiz, E., Gassmann, F., & Fröhlich, M. (Eds.). (2024). Artificial Intelligence in Sports, Movement, and Health. Springer.
[7] Musat, C. L., Mereuta, C., Nechita, A., Tutunaru, D., Voipan, A. E., Voipan, D., ... & Nechita, L. C. (2024). Diagnostic applications of AI in sports: a comprehensive review of injury risk prediction methods. Diagnostics, 14(22), 2516.
[8] Regulation, P. (2018). General data protection regulation. Intouch, 25, 1-5.