ISSN (Online): 2455-3662
EPRA International Journal of Multidisciplinary Research (IJMR) - Peer Reviewed Journal
Volume: 10| Issue: 4| April 2024|| Journal DOI: 10.36713/epra2013 || SJIF Impact Factor 2024: 8.402 || ISI Value: 1.188

THE PROGRESS IN THE RESEARCH OF MACHINE LEARNING IN SPORTS MEDICINE

Katherine Ning LI
Associate Professor, Institute of Sports science, Xi'an Physical Education University, China

Article DOI: https://doi.org/10.36713/epra16367
DOI No: 10.36713/epra16367

ABSTRACT

To explore the prospects and challenges of applying artificial intelligence and its machine learning subfield in sports medicine, to drive knowledge innovation in this domain. Research Content includes Applications of machine learning in sports medicine: Clustering and classifying athlete data, developing predictive models to optimize training and prevent injuries, and providing interpretable decision support for medical professionals. Challenges of machine learning in sports medicine: Issues with data availability and quality, model interpretability and transparency, as well as the integration with existing workflows. In summary, the potential of AI and machine learning in sports medicine is immense, but to fully harness their transformative value, interdisciplinary collaboration, data sharing, rigorous validation, and the establishment of ethical guidelines are essential. Only through these collective efforts can the field optimize athlete training, prevent injuries, and drive overall innovation in sports medicine.

KEYWORD: *Machine Learning,Sports Medicine , Artificial intelligence ,Knowledge Representation, Decision Support*

A. INTRODUCTION

The term "Artificial Intelligence (AI)" was first proposed by McCarthy et al. in 1955. They defined AI as "the science and engineering of making intelligent machines" that can perform tasks previously thought to be possible only for humans, such as abstract reasoning and high-level problem-solving. AI also refers to the scientific and technological efforts to develop intelligent computers capable of performing functions usually associated with human effort. A subset of AI is known as Machine Learning (ML), which is a process of automatically generating symbolic knowledge representations from data, and acquiring meaningful representations or symbols to capture the underlying characteristics or concepts of the data. In the fields of machine learning and AI, the goal is to extract higher-level knowledge and understanding from raw data, enabling computational systems to reason, induce, and make informed decisions.

B. ADVANCEMENTS IN THE APPLICATION OF MACHINE LEARNING

1. In data representation, machine learning can help use symbolic representations to represent various dimensions of sports medicine data, such as athlete profiles, injury records, performance metrics, and training regimes. This may involve the creation of knowledge graphs or ontologies to capture the relationships and dependencies between different variables.

1.1 Clustering and classification: Machine learning can be used to cluster athletes or patients based on characteristics such as age, physical build, assessment tests, etc. By leveraging the symbolic representations of electronic medical records, training regimes, and performance metrics, machine learning models can be developed to predict the future performance of athletes based on various training strategies. These models can consider factors such as training intensity, duration, frequency, recovery periods, and specific exercises, and recommend personalized training plans to maximize performance.

There have been studies that summarized the epidemiological characteristics of DL injury trends from 2000 to 2017[1], compiling MLB player data from 4 online baseball databases, including age, performance metrics, and injury history. A total of 84 ML algorithms were developed. The output of each algorithm reported whether a player would be injured in the next season and the anatomical location of the injury. The results showed that the machine learning models could predict the likelihood of injury in the next season with reasonable reliability, especially for defensive players.

Using machine learning techniques in the field of knee joint biomechanics data classification can help with the diagnosis of knee joint diseases. Studies have aimed to establish and validate machine learning models to diagnose patients with Generalized Joint Hypermobility (GJH) and normal individuals. Gait data and kinematic data were collected using a three-dimensional motion capture system. A deep neural network (GJHdnet) was proposed for GJH detection and evaluated in several aspects. The model achieved an accuracy of 95.77%, a specificity of 98.68%, and a recall rate of 76.84%, outperforming traditional machine learning methods. The trained model can run on cost-effective devices, assisting in the instant and accurate diagnosis of GJH[2].

1.2 Interpretability: Machine learning provides interpretable representations that allow sports medicine professionals to understand and explain the reasoning behind the decisions made by machine learning models. This transparency and interpretability are crucial for gaining trust and acceptance in this field, as they allow practitioners to validate and refine the knowledge encoded in the symbolic representations. Machine learning models can be used to generate interpretable rules or decision trees that capture the relationships between input features and outputs. These rules can provide clear guidance on training strategies based on the athlete's characteristics.

SHAP (Shapley Additive explanations) is a method for explaining the predictions of machine learning models. It provides a framework for assigning importance values to different features or input dimensions, indicating their contribution to the model output[3]. SHAP values are based on the concepts of cooperative game theory, providing a unified approach to measure feature importance across different models. These methods can highlight the key features or input dimensions that contribute to a specific prediction or outcome. SHAP has been applied in sports medicine, where it can help identify the critical factors or input dimensions that influence an athlete's performance or injury risk. By understanding the importance of different features, coaches and sports medicine experts can make informed decisions on training strategies, injury prevention, or rehabilitation plans.

Locally Interpretable Model-Agnostic Explanations (LIME) is another technique used to explain the predictions of machine learning models[4]. It focuses on providing interpretable explanations for individual instances or data points, independent of the underlying model used. LIME approximates the behavior of the complex model with simpler, more interpretable models that are locally faithful to the original model's predictions.

Injury Risk Assessment: Using LIME, sports medicine experts can explain the factors that contribute to an athlete's risk of injury. LIME can identify which specific features (such as previous injuries, training load, or biomechanical data) have the most influence on the predicted likelihood of injury. This information helps to develop tailored injury prevention strategies for individual athletes. Studies have used pre-trained CaffeNet convolutional neural network (CNN) models to compare the accuracy of marker-based motion capture versus the average prediction of three key KJMs (knee joint moments) associated with anterior cruciate ligament (ACL) injury across three different sports-related motion types, demonstrating the feasibility of using deep learning for on-field knee injury assessment instead of laboratory-embedded force plates[5]. Other studies have validated LIME machine learning models to identify risk factors and quantify the overall risk of secondary meniscal injury in a longitudinal cohort following primary ACL reconstruction (ACLR). The results showed that machine learning models outperformed traditional prediction models and identified shorter time to return to sport, lower injury-time VAS, increased time from injury to surgery, proximal ACL tear location, and age >40 years at injury as risk factors for post-ACLR secondary meniscal tears. After detailed calculations, these models can be deployed in clinical settings to provide real-time, quantifiable risk for consultation and timely intervention.

Performance Prediction: LIME can explain the key factors driving an athlete's performance in a specific sport or event. By analyzing the local explanations provided by LIME, coaches and sports medicine experts can gain a deeper understanding of the training techniques, physiological attributes, or biomechanical factors that have the greatest impact on performance. This knowledge can guide the development of personalized training plans.

Treatment Effectiveness: LIME can be used to explain the predictors of the effectiveness of different treatment interventions or rehabilitation plans. By explaining the factors that contribute to successful outcomes, LIME can help sports medicine experts understand which therapies or exercises are most beneficial for specific athletes or injury types. This information can assist in optimizing treatment strategies and reducing rehabilitation time.

Performance Optimization: LIME can help identify the key features that contribute to poor athlete performance. By explaining the factors behind poor performance or stagnant results, coaches and sports medicine experts can make targeted adjustments to training plans, nutrition regimes, or recovery strategies. This can help athletes overcome performance barriers and fully realize their potential.

2. Decision Support Systems: Machine learning can aid in the development of sports medicine decision support systems. By encoding expert knowledge and guidelines into symbolic representations, machine learning models can assist coaches and sports medicine professionals in making informed decisions on training strategies. These systems can provide recommendations on athlete selection, load progression, recovery regimes, and injury prevention strategies based on analyses of symbolic representations and historical data.

Learning medical ontologies from unstructured data sources such as clinician notes, academic papers, and medical texts. Machine learning algorithms analyze the data to detect key concepts (e.g., injuries, treatments, risk factors) and the relationships between them. The result is an initial knowledge graph that can then be refined by subject matter experts.

Ontologies provide a formal representation of the concepts in a domain and the relationships between those concepts. In medicine, ontologies enable the standardized conceptualization of knowledge to support applications like clinical decision support systems. Traditionally, subject matter experts manually construct ontologies by defining concepts, taxonomies, and relationships based on existing knowledge. However, with the increasing availability of digital data (e.g., clinician notes, academic literature, web resources), machine learning techniques have the opportunity to automatically generate ontology drafts, which can then be refined by experts. Learning production rules that simulate clinical decision-making and standard procedures from sports medicine data. Algorithms detect patterns between attributes like symptoms, test results, diagnoses, patient profiles, and recommended treatments. These patterns are formulated as IF-THEN rules that can drive AI-based diagnostic and treatment recommendation systems[7]. Experts are still needed to validate and refine these production rules, representing knowledge in the IF-THEN form to link clinical conditions and attributes (IF) with conclusions and recommended actions (THEN). In sports medicine, production rules can establish decision pathway models for diagnosis and management based on symptoms, test results, medical history, and other factors. For example: If knee pain + swelling + Lachman's test positive, then likely anterior cruciate ligament injury → perform MRI; If ACL tear + desire to return to cutting sports, then recommend surgical reconstruction.

2.1 Injury Risk Assessment: Machine learning can help identify the factors that increase the risk of injury for athletes. By analyzing historical injury data, training loads, and other relevant variables, machine learning models can be trained to predict the likelihood of future injuries. This information can be used to optimize training strategies by adjusting training loads, incorporating recovery periods, and modifying exercise programs to reduce the risk of injury.

2.2 Personalized Training Plans: Through machine learning, customized training plans can be developed for individual athletes. By considering the athlete's circumstances, including age, fitness level, injury history, and performance goals, machine learning models can generate training recommendations that maximize performance improvement while minimizing injury risk. These personalized plans can be adjusted over time based on the athlete's condition and objectives. Studies have used artificial intelligence (AI) methods with machine learning (ML) techniques to provide more in-depth feedback to athletes by personalizing their health status, especially considering the individual multi-factorial data used in predictive models, and conducted an athletics individual athlete ICPR risk assessment (i.e., I-REF)[8].

3.Knowledge Discovery: Machine learning can help uncover new insights and patterns from sports medicine data. By applying logical reasoning and inference techniques to the symbolic representations, hidden relationships and correlations can be discovered. This can help identify factors that contribute to improved performance, injury prevention, or rehabilitation strategies.

3.1 Semantic networks are a knowledge representation method that uses nodes to represent concepts and links to represent the relationships between these concepts. In sports medicine, semantic networks can capture the complex network of interacting factors that influence health risks and outcomes. For concussions, a semantic network could link factors such as impact forces, genetic variations, hormone levels, neck strength, previous injuries, helmet use, age, and medications.

Traditionally, subject matter experts have constructed semantic networks based on various research findings and their own experiences. However, the abundance of biomedical data now available, especially "holistic" data that links biological components to health attributes, allows machine learning algorithms to automatically generate draft semantic networks. These data-driven networks can then be validated and refined by experts to become a powerful risk landscape model. Data sources that can be used to extract a concussion risk network include research datasets containing genetic data, blood-based biomarkers, neurocognitive test results, injury histories, and other athlete attributes. While relationships between factors can be determined, the datasets may be limited or produce spurious associations without proper analysis.

Discuss the biomedical literature on how genetics, biomarkers, impacts, and other factors relate to concussion risk, severity, and recovery. Machines can scan large text corpora to detect statistical relationships and patterns, but the networks they generate require expert review to reliably represent knowledge and elucidate contradictions or knowledge gaps in the literature.

Wearable sensor data measures head accelerations and impacts during competitions. This data can directly illustrate the relationship between impact forces and concussion probability but may be limited by the monitored sports and athlete types. It also depends on medical evaluations to confirm actual concussion outcomes.

The main benefit of using machine learning to extract semantic networks is the ability to scale and detect subtle or complex relationships that may be difficult to determine through human knowledge engineering alone. However, machine-generated networks are subject to biases and limitations in the source data and require curation to become clinically useful knowledge models. Experts must determine how to integrate machine learning networks with existing expert-authored concussion risk models while avoiding inaccuracies or logical inconsistencies.

3.2 Symbolic Representation Learning

Symbolic representation learning is a type of machine learning that refers to the process of acquiring meaningful representations or symbols to capture the underlying characteristics or concepts of data. In the fields of machine learning and artificial intelligence, symbolic representation learning aims to extract higher-level knowledge and understanding from raw data, enabling computational systems to reason, induce, and make informed decisions. Symbolic representation learning focuses on capturing the inherent structure and semantics of data through symbolic representations. These representations can take the form of symbols, rules, logical expressions, or graphs, providing a more interpretable and human-understandable way to represent and operate on knowledge. Symbolic representation learning often involves techniques such as symbolic logic, knowledge graphs, ontologies, and rule-based systems. It leverages the power of logical reasoning and inference to derive new knowledge from existing representations and perform logical deduction. By learning symbolic representations, AI can perform tasks like knowledge discovery, knowledge integration, semantic understanding, and decision-making in a more interpretable and explainable manner.

Machine learning models are trained on data to detect patterns and relationships that can be used for prediction or to gain insights. However, the features and knowledge encoded in the algorithms and parameters of the models themselves are not inherently human-understandable. Interpretability is crucial for applications like medicine, where transparency and explainability are necessary conditions for trust and validation of the results.

One approach to improving model interpretability is to map the model's features to symbolic knowledge representations like ontologies. A machine learning model may predict the probability of an athlete's ACL injury based on features like knee swelling, instability, limited range of motion, and age[9]. The features selected by the model can be mapped to concepts and relationships in an ontology, with associated probabilities. The result is an explanation of how the model linked key factors to the predicted ACL injury risk based on trends in its training data.

C. LIMITATIONS AND CHALLENGES OF MACHINE LEARNING IN SPORTS MEDICINE

Data availability and quality are major challenges in applying machine learning to sports medicine. The sports medicine domain often lacks comprehensive and reliable datasets. Data collection methods such as wearable devices and medical imaging may have limitations, and data across different sports may be biased and inconsistent. Ensuring data quality and addressing these limitations is crucial for developing effective machine-learning models.

Another challenge is the interpretability and transparency of machine learning models. Complex models may be difficult to understand in terms of the underlying rationale behind their predictions and recommendations. This is particularly important in sports medicine, as decisions directly impact the health of athletes. Improving interpretability and transparency is key to gaining the trust of coaches, athletes, and professionals.

To drive the effective application of machine learning in sports medicine, sustained interdisciplinary collaboration, data sharing, and ethical framework development are needed in areas such as data quality, model interpretability, and privacy/security. Only then can the immense potential of machine learning be maximized in optimizing training and preventing injuries.

REFERENCES

1. *Karnuta, J. M., Luu, B. C., Haeberle, H. S., Saluan, P. M., Frangiamore, S. J., Stearns, K. L., Farrow, L. D., Nwachukwu, B. U., Verma, N. N., Makhni, E. C., Schickendantz, M. S., & Ramkumar, P. N. (2020). Machine Learning Outperforms Regression Analysis to Predict Next-Season Major League Baseball Player Injuries: Epidemiology and Validation of 13,982 Player-Years From Performance and Injury Profile Trends, 2000-2017. Orthopaedic journal of sports medicine, 8(11), 2325967120963046. https://doi.org/10.1177/2325967120963046*
2. *Zhong, G., Huang, S., Zhang, Z., Xie, Z., Liu, H., Huang, W., Zeng, X., Hu, L., Liang, H., & Zhang, Y. (2023). Diagnosis of generalized joint hypermobility with gait patterns using a deep neural network. Computers in biology and medicine, 164, 107360. https://doi.org/10.1016/j.compbiomed.2023.107360*
3. *Zhang, L., Zhao, S., Yang, Z., Zheng, H., & Lei, M. (2024). An Artificial Intelligence Platform to Stratify the Risk of Experiencing Sleep Disturbance in University Students After Analyzing Psychological Health, Lifestyle, and Sports: A Multicenter Externally Validated Study. Psychology research and behavior management, 17, 1057–1071. https://doi.org/10.2147/PRBM.S448698*
4. *Raptis, S., Ilioudis, C., & Theodorou, K. (2024). From pixels to prognosis: unveiling radiomics models with SHAP and LIME for enhanced interpretability. Biomedical physics & engineering express, 10(3), 10.1088/2057-1976/ad34db. https://doi.org/ 10.1088/2057-1976/ ad34db*
5. *Johnson, W. R., Mian, A., Lloyd, D. G., & Alderson, J. A. (2019). On-field player workload exposure and knee injury risk monitoring via deep learning. Journal of biomechanics, 93, 185–193. https://doi.org/10.1016/j.jbiomech.2019.07.002*
6. *Cristiani, R., Mikkelsen, C., Wange, P., Olsson, D., Stålman, A., & Engström, B. (2021). Autograft type affects muscle strength and hop performance after ACL reconstruction. A randomised controlled trial comparing patellar tendon and hamstring tendon autografts with standard or accelerated rehabilitation. Knee surgery, sports traumatology, arthroscopy : official journal of the ESSKA, 29(9), 3025–3036. https://doi.org/10.1007/s00167-020-06334-5*
7. *Greenhalgh, J., Dalkin, S., Gibbons, E., Wright, J., Valderas, J. M., Meads, D., & Black, N. (2018). How do aggregated patient-reported outcome measures data stimulate health care improvement? A realist synthesis. Journal of health services research & policy, 23(1), 57–65. https://doi.org/10.1177/1355819617740925*
8. *Edouard, P., Steffen, K., Peuriere, M., Gardet, P., Navarro, L., & Blanco, D. (2021). Effect of an Unsupervised Exercises-Based Athletics Injury Prevention Programme on Injury Complaints Leading to Participation Restriction in Athletics: A Cluster-Randomised Controlled Trial. International journal of environmental research and public health, 18(21), 11334. https://doi.org/10.3390/ijerph182111334*
9. *Mangone, M., Diko, A., Giuliani, L., Agostini, F., Paoloni, M., Bernetti, A., Santilli, G., Conti, M., Savina, A., Iudicelli, G., Ottonello, C., & Santilli, V. (2023). A Machine Learning Approach for Knee Injury Detection from Magnetic Resonance Imaging. International journal of environmental research and public health, 20(12), 6059. https://doi.org/10.3390/ijerph20126059*