Preprint

Suggested Citation:

Amendolara, A. B., Pfister, D.*, Settelmayer, M.*, Shah, M., Wu, V., Donnelly, S., Johnston, B., Peterson, R., Sant, D., Kriak, J., Bills, K. (2022). *An overview of machine learning applications in sports injury prediction* [Preprint]. SportRxiv

Title: An overview of machine learning applications in sports injury prediction

Running Heading: Machine learning in sports injury prediction

Authors: Alfred Amendolara[1,2], Devin Pfister[2*], Marina Settelmayer[2*], Mujtaba Shah[2], Veronica Wu[2], Sean Donnelly[2], Brooke Johnston[2], Race Peterson[2], David Sant[2], John Kriak[2], Kyle Bills[2]

1 – Federated Department of Biological Sciences, New Jersey Institute of Technology, 323 Dr Martin Luther King Jr Blvd, Newark, NJ, 07102, USA

2 – Department of Biomedical Sciences, Noorda College of Osteopathic Medicine, 1712 E Bay Blvd Building 5, Suite 300, Provo, Utah 84606, USA

* Indicates equal contribution

Correspondence: Kyle Bills, DC, PhD, kbbills@noordacom.org

ORCID

Alfred Amendolara: 0000-0001-9696-8961

Mujtaba Shah: 0000-0002-2442-4878

David Sant: 0000-0001-7372-9896

Abstract

Use injuries represent a serious and intractable problem in athletics that has traditionally relied on historic datasets and human experience for prevention. Existing methodologies have been frustratingly slow at developing higher precision prevention practices. Technological advancements have permitted the emergence of artificial intelligence and machine learning (ML) as promising toolsets to enhance both injury mitigation and rehabilitation protocols. This article provides a comprehensive overview of ML techniques as they have been applied to sports injury prediction and prevention to date. Literature from the last five years has been compiled and the findings presented. Given the current lack of open source, uniform data sets, as well as a reliance on dated regression models, no strong conclusions about the real-world efficacy of ML as it applies to sports injury prediction can be made. However, it is suggested that addressing these two issues will allow powerful, novel ML architectures to be deployed, thus rapidly advancing the state of this field and providing validated clinical tools.

Key Points

- Significant progress has been made in predictive analysis of sports injury, but the quality of literature is varied and much of it focuses on traditional, less capable regression models.
- In order to produce clinically usable models, well structured, uniform data sets should be created and validated.

Declarations

Funding

Not applicable

Conflicts of Interest

The authors report there are no competing interests to declare.

Availability of Data and Material

Not applicable

Ethics Approval

Not applicable

Consent to Participate

Not applicable

Code Availability

Not applicable

Author Contributions

Author contributions have been structured following CRediT (Contributor Roles Taxonomy) suggestions.

Alfred Amendolara: Conceptualization, Methodology, Investigation, Writing – Original Draft, Writing – Review & Editing, Project Administration **Devin Pfister:** Investigation, Writing – Original Draft **Marina Settelmayer:** Investigation, Writing – Original Draft **Mujtaba Shah:** Investigation, Writing – Original Draft **Veronica Wu:** Investigation, Writing – Original Draft **Sean Donnelly:** Investigation, Writing – Original Draft **Brooke Johnston:** Investigation, Writing – Original Draft **Race Peterson:** Conceptualization **David Sant:** Conceptualization, Investigation, Writing – Review & Editing, Supervision **John Kriak:** Conceptualization, Writing – Review & Editing, Supervision **Kyle Bills:** Conceptualization, Writing – Review & Editing, Supervision

1. Introduction

Machine Learning (ML) is a complex discipline broadly defined as the creation of a computer system able to experientially learn and adapt without explicit instructions to generate predictive analytics [1, 2]. As computational resources have continued to increase, ML application and implementation in varied fields has grown, sports medicine included. The assessment, mitigation, and prevention of injury is of primary importance as injuries are ubiquitous and may result in severe physical, emotional, and financial consequences, especially at the professional level. In order to elucidate the complex factors contributing to athlete injuries and to enable greater predictive precision, a variety of ML models have been proposed in the literature [3-6].

As computational technologies advance, larger and more complex ML algorithms, including application of previously theoretical techniques, are possible. It is therefore useful to periodically compile and review literature that has been, or may be, applied to injury prediction and prevention as newer systems are capable of implementing new algorithms more efficiently. Additionally, though recent literature reviews exploring niche aspects of this field, limitations exist: most articles are written from the perspective of data mining and without interest in recency [5], are sports-specific [7-9] are limited in scope [3, 4, 10], or are focused on team sports only [6]. We seek to provide a comprehensive overview of the state of ML in sports injury across many sports using a broad selection of algorithms.

To provide a basis for exploration of novel ML models and methodologies, algorithms have been categorized based on function, limitations, and current or potential implementation to sports medicine. Each of the selected algorithms includes a brief background and an overview of relevant literature from the last 5 years. While these background sections provide context for individual algorithms, it is useful to provide a brief explanation of general ML concepts.

1.1. What is an algorithm?

In the context of this review, “algorithm” will be defined as the entire set of mathematical equations and rules for a given ML approach. Each algorithm uses a unique set of rules and equations to mathematically calculate an outcome [2]. The systematic application of the defined rules and equations to a dataset is referred to as “training a model”.

1.2. Training a model.

ML algorithms must be selected and trained prior to use. Within this topic exist several terms briefly defined below:

1. Data set – The complete set of data used to train and validate an algorithm. This data may be in a variety of forms, but often must be formatted appropriately for a given algorithm.
2. Batches – A set of data selected to be passed through an algorithm, often necessary due to memory constraints and often desirable due to optimization and training requirements.
3. Feature and feature extraction – Features are individual, measurable properties of data. Feature extraction is the process by which predictive and unique features are chosen from a data set. The collection of extracted features used to train a model is called the feature set.
4. Labels – Human inputs used to provide context to a ML algorithm prior to training e.g., a picture of a dog may be manually labeled “dog”.
5. Supervised learning – The process of guiding training of an algorithm by providing “labeled” data.
6. Unsupervised learning – The process of allowing an algorithm to group and cluster data without labels.

7. Gradients and gradient optimization – Gradients are the derivative vectors of the multivariate functions used in ML and may be used as metrics to guide and assess training. Algorithms exist to optimize gradient descent, known as gradient optimization.
8. Overfitting – The tendency of ML models to "memorize" training data. In other words, a model learns only the patterns of training data whether a mathematical relationship between parameters exists or not. This reduces the generalizability of a model. It is often a concern when using data sets that contain large numbers of features.
9. Hyperparameters/parameters - Parameters are internal values of a model that are derived from the data set. Hyperparameters are permanent parameters set prior to model training that often have a large impact on other model parameters.
10. Error measurements – These are quantifiable measurements of error calculated using equations such as root mean squared error. [2, 11]

Prior to selection of an appropriate algorithm, a data set must be constructed. Data format directly impact the algorithm being used and the intended application. Data sets are generally split into training data and testing data. Training data may be labeled or unlabeled, depending on whether supervised or unsupervised learning is desired. Some data is reserved as validation or test data in order to confirm the algorithm has been successfully trained [2]. Larger datasets are nearly universally desirable to enhance model usefulness. However, when only smaller data sets are available, statistical methods are available to increase the number of data points available to improve predictive power. This method is more useful for testing ML approaches than for training new models and is less preferable to using real world data.

Once data has been selected and subdivided, features must be extracted. These features may be manually identified, a time-consuming process, or automatically identified as a function of a given algorithm. This often represents a critical stage in model development [5, 12].

Finally, after the above steps have been completed, a model may be trained. Training is guided by rules or equations that seek to balance speed, performance, and generalizability. Training data is often passed through an algorithm in batches that allow massive data sets to be partitioned in smaller chunks and processed without overwhelming computer hardware. It can also aid in training optimization [2].

1.3. Proper validation and evaluation.

Following model training, validation and evaluation can occur. Proper validation and evaluation rely on several components: distinct training and testing data sets, an appropriate error metric, simulated data in the case of smaller data sets, and an understanding of common pitfalls in ML [11, 13]. The current standard for validation is *K*-fold cross validation. With *K* equal to 10, for example, the data is randomly split into 10 equal sections with 9 used for training and 1 reserved for validation. These sections are then shuffled to ensure generalizability [14]. Other techniques commonly used for validation are outside the scope of this discussion, but it is important to note that most approaches are based on shuffling or randomization of training data.

2. Methods

A comprehensive literature review was conducted using Ovid Discovery Search and Google Scholar, which provided compiled results from many databases. PubMed/Medline, Institute of Electrical and Electronics Engineers (IEEE)/Institute of Engineering and Technology (IET), and ScienceDirect were accessed individually as well. A focus was placed on papers published from 2017-2022, although older papers were referenced for background. Algorithms were selected based on a preliminary literature review and included K-Nearest Neighbor (KNN), *K*-means, decision tree, random forest, gradient boosting and Adaboost, and neural networks. Search terms were "*algorithm name"* + "sport" + "injury" for each algorithm e.g., “neural network” + “sport” + “injury”. An attempt was made to include variations in algorithm name and abbreviation. Papers concerning prediction and analysis of sports injuries were included. Any papers that could not be accessed or where not available in English were excluded. Forty original research papers and eight review articles were selected based on the criteria described. A brief background on each algorithm was incorporated to provide context. Of note, we have excluded papers primarily relying on linear or logistic regression as we feel these algorithms do not represent the cutting edge of predictive analysis and have been addressed elsewhere in the literature.

3. Results

Results of the comprehensive literature review are summarized below. Each section includes a brief background on the relevant algorithm to provide context. Results of articles surveyed are then summarized in each *Applications* section. Papers were sorted into these sections based on algorithm tested. When more than one algorithm was explored, papers were included in the section with the most effective algorithm and in sections with algorithms that were nearly as successful where appropriate. Due to variable study design, and often disparate aims, no attempt has been made to directly compare or otherwise aggregate results quantitatively. Instead, we present overall trends in the discussion. Likewise, trends of shortcomings or pitfalls have been addressed in the discussion section. Note that due to the diversity of neural network implementations, papers pertaining to neural networks have been further subdivided following a brief introduction to general algorithm architecture.

3.1. KNN

3.1.1. Background

K-Nearest Neighbor is a supervised ML algorithm that uses similarity to group data points together to solve regression and classification problems. It is widely used in other fields of medicine. For example, in oncology, research using KNN has been able to classify different subtypes of acute myeloid leukemia cells which aid in identifying blood cell ratios [15]. K-Nearest Neighbor has also been used to evaluate and classify degenerative knee joint vibroarthrographic signals [16]. The algorithm assumes that similar data points will be found in close proximity to one another with respect to a given distance function. So, in a basic classification problem, KNN will assign a class to any given data point based on the class of its neighbors. In practice, KNN applies a weighted smoothing function to estimate data density. Weighting is based on *K* number of neighbors, in essence setting the bin size, resulting in small bins in high density areas and large bins in low density areas. Kernel functions may be applied to further smooth the density estimates. The advantages of KNN include its relative simplicity and ease of implementation, as well as its ability to make accurate predictions using a small data set [17]. However, when applied to very large data sets, the KNN algorithm becomes proportionally more complex and inefficient. While this problem is not insurmountable, it does necessitate mathematical condensing as well as dimensionality reduction [2, 18].

3.1.2. Application

In sports medicine, special sensors like accelerometers, gyroscopes, infrared sensors, and magnetometers can be attached to athletes to collect data. Using data collected from different body parts of athletes, KNN analyzes and determines certain behaviors for athletes in unique sporting events. With this recognition model, patterns predisposing to injury can be determined, allowing for potential injury prevention [19]. In addition to their general use as comparison algorithms, a 2018 paper applied KNN as part of a larger model, including both *K*-means and SVM, for injury prediction [20].

3.2. *K*-Means

3.2.1 Background

Due to its simplicity, K means is one of the most widely used clustering algorithms. *K*-means is an iterative algorithm designed to partition a data set into subgroups called clusters. These clusters are organized such that the sum of the squared distance between the data points and the clusters' centroids, the arithmetic mean of all the data points that belong to that cluster, is minimized. The less variation within a cluster, the more homogeneous the data points are within that cluster [21].

In practice, *K*-means relies on initial random selection of some number *K* centroids chosen from a dataset containing *n* cluster objects [22]. Once selected, Euclidean distance is calculated between all individual data points and each centroid. Points are then assigned to a cluster based on this distance (see Fig. 1). Using the calculated mean of each cluster, centroids are adjusted. This process occurs iteratively until clustering improvement plateaus, identified by the stabilization of centroids [23].

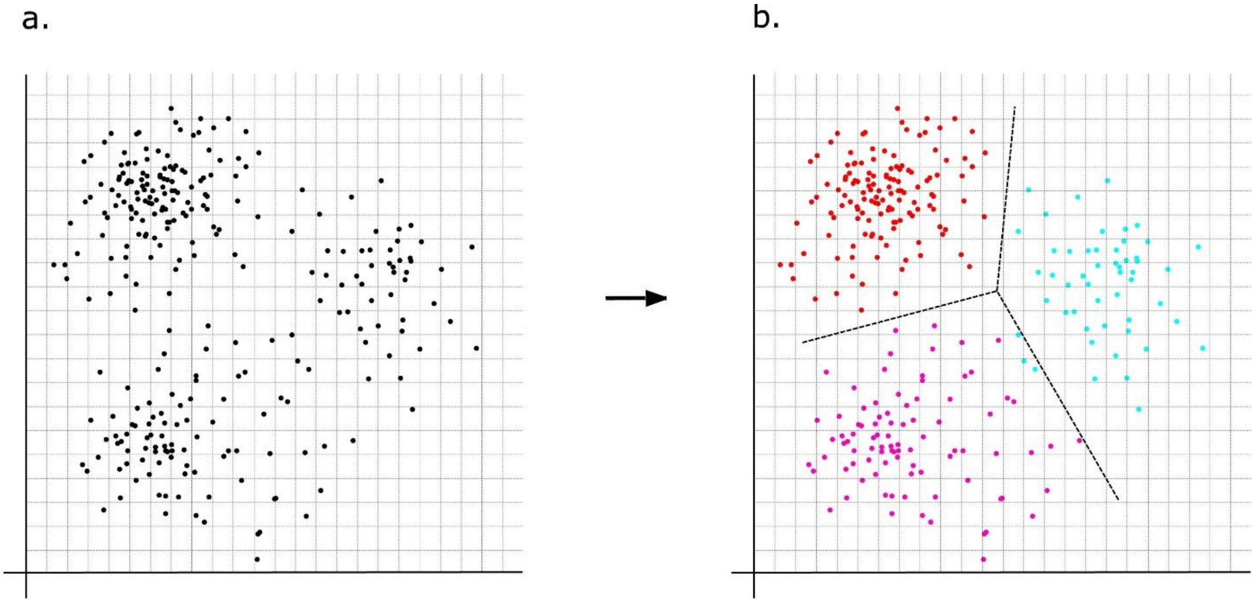

Fig. 1 Visualization of a 2-dimensional clustering. (a) shows un-clustered data. (b) shows data separated into 3 clusters represented by different colors and separated by dotted lines

3.2.2. Application

In 2020, a study by Dingenen et al. used *K*-means to establish that runners with the same injuries could be clustered into two different subgroups with a mean silhouette coefficient of 0.53 [24]. These subgroups were used to illustrate variable kinematic causes of running re*lated injury*. *K*-means was also used by Ibáñez et al in 2022 as a data separation technique for grouping women's basketball players into first and second divisions. This study effectively used *K*-means to analyze thresholds of deceleration, acceleration, speed, and impact on the players and determined a difference between the first and second division[25]. These so-called divisions were proposed to aid in personalization of training to prevent injuries and improve performance. As seen in these recent articles, and likely due to its simplicity and familiarity, *K*-means remains effective when applied to traditional clustering problems and may be suited to exploring injury risk factors or player characteristics.

3.3. Support Vector Machines (Devin)

3.3.1. Background

Support vector machines (SVM) are supervised learning algorithms that separate data points into distinct groups using hyperplanes. Hyperplanes' orientation and position are influenced by data points known as support vectors. Support vector machines map points in order to maximize the gap between the two categories (see Fig. 2A) known as the maximal margin [26, 27]. Once trained on a data set, SVM may be used to classify new data points and to discover informative patterns within data [28].

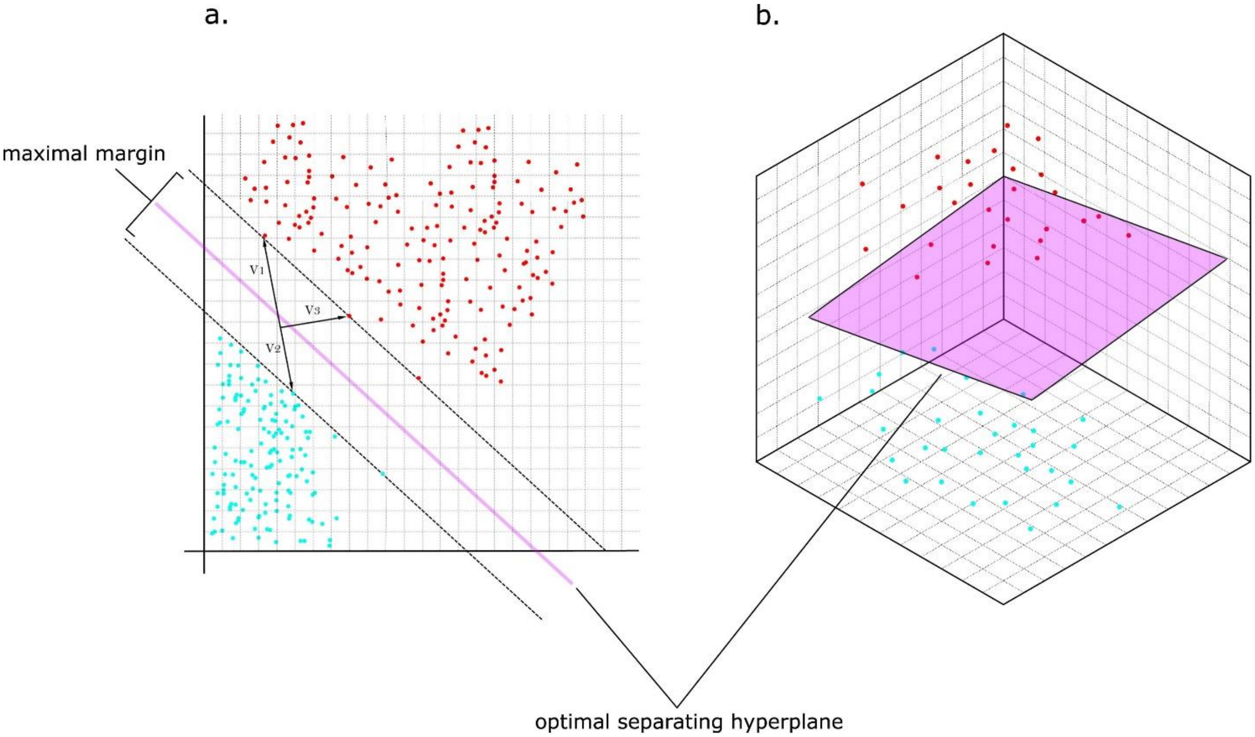

Fig. 2 Diagram of a theoretical support vector machine in 2 (a) and 3 (b) dimensions. Hyperplanes separate data. Note support vectors labeled *V1*, *V2*, and *V3*

3.3.2. Application

For sports specific applications, SVMs have been trained using modifiable metrics such as training load, performance techniques, psychological and neuromuscular assessments, and non-modifiable metrics such as anthropometric measurements, previous injury history, and genetic markers to accurately predict future injuries [29, 30]. Identification of injury risk factors such as these allows coaches and medical personnel to modify training loads, regiments, and techniques to potentially prevent future injuries [6]. For example, a 2018 paper by Ruddy et al. used a number of ML algorithms, including SVM, to assess risk factors identified in hamstring strain injuries [31]. In another 2018 paper by Carey et al., also exploring hamstring injury prediction and risk factors, SVM benefited substantially from data pre-processing, although it was ultimately outperformed by simple logistic regression [32]. Using non-physiological data, a 2017 paper predicting in-game injuries in Major League Soccer found that SVM were the most accurate of several tested algorithms, including logistic regression, multilayer perceptron, and random forest [33]. However, in recent literature, including two 2021 papers comparing efficacy of ML algorithms, SVMs have proven less effective than other algorithms [34, 35]. Despite this, SVM may still be valuable given their suitability for predicting high-dimensionality data sets, especially when combined with other techniques as in a 2022 paper by Wang et al. predicting triple jump injury [36].

3.4. Decision Tree

3.4.1 Background

A decision tree is a type of supervised ML that uses an iterative process of segregating datasets on specific features to predict an output category based on a set of input features. Beginning with the input node (the root node), data points are split into separate bins based upon their values for a specific feature. Each of these bins are then tested recursively to determine if the data points can be further split into separate smaller bins to achieve better accuracy until all nodes have reached a specified size or purity. Bins that can be further split are called decision nodes, while those that cannot denote an ultimate decision are known as leaf nodes [37].

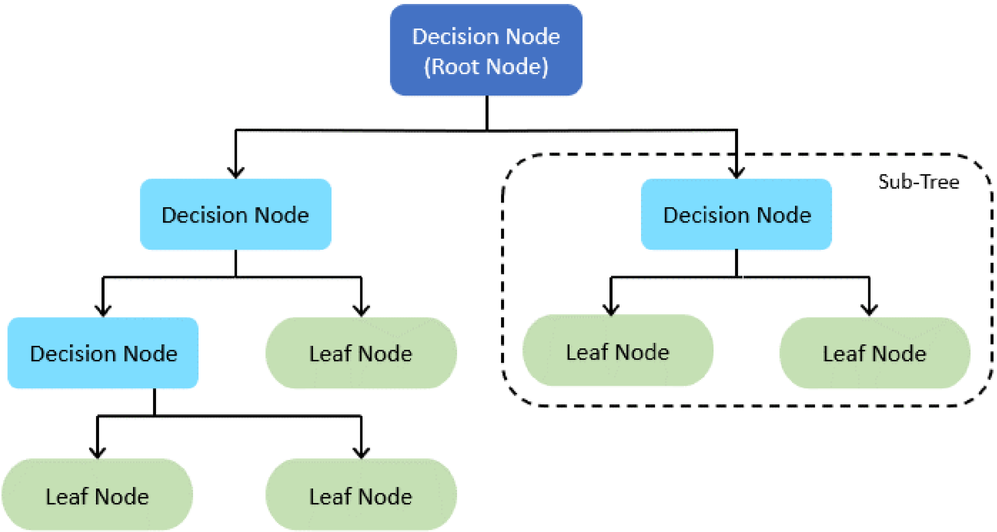

Fig. 3 Schematic diagram of a simple decision tree showing several decision nodes branching from a root node and terminating in leaf nodes

3.4.2. Application

Modern evolutions of the classic decision tree algorithm have been broadly applied in recent years. In 2018, Connaboy et al. used decision trees built with Chi-squared Automatic Interaction Detection (CHAID) to analyze factors contributing to lower extremity injury in military personnel. Using their model, the authors identified several factors leading to increased injury risk over a 365-day period [38]. Using a classification and regression decision tree (CART), Mendonca et al. investigated associations between various risk factors and patellar tendinopathy in volleyball and basketball players [39]. A 2021 paper by Kolodziej et al. applied a CART decision tree to predict youth soccer injuries, achieving a sensitivity of 0.73 and a specificity of 0.91 [40]. Another 2021 paper by Ruiz-Perez et al. attempted to reproduce a 2020 model by Rommers et al., which used field data collected via GPS. While they favorably compared C4.5 decision trees with several modeling approaches including KNN, SVM, and ADTree, they did not use the same algorithm as

Rommers et al. and did not achieve comparable performance (AUC 0.767 vs 0.850) [41, 42]. Contrary to these relatively promising results, Rossi et al. found that decision trees, although outperforming comparison algorithms, were not able to achieve a precision greater than 50% when forecasting soccer injuries [43]. Decision trees undoubtedly have a place in sports injury prediction, though their performance varies with data and model structure. Additionally, they can lack generalizability and overfit during training, thus limiting their accuracy [44].

3.5. Random Forest

3.5.1. Background

Because decision trees can lack generalizability and tend to overfit during training [44], random forests, which are a collection of random decision trees, offer a potential advantages. Random forest models rely on the creation of an ensemble of decision trees that vote on the final output (see Fig. 4).

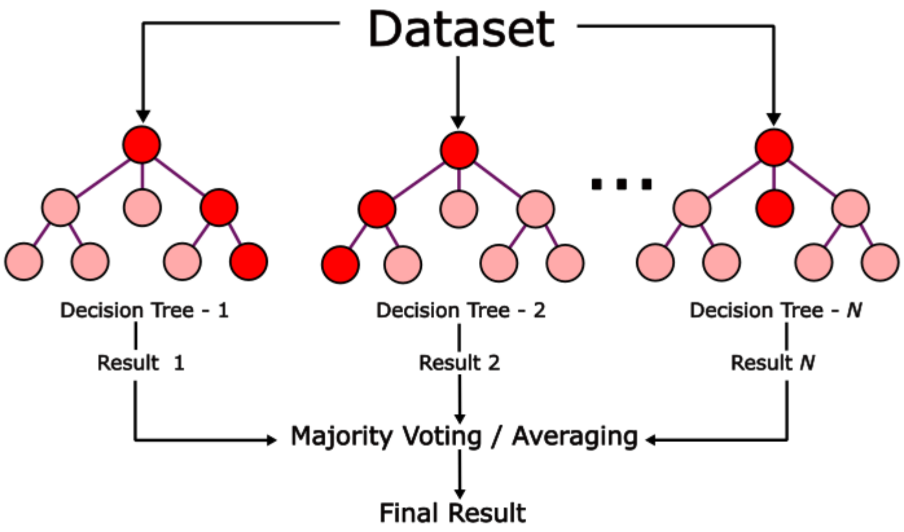

Fig. 4 A random forest model with *N* decision trees aggregating results to produce a final output

Implementation of a random forest model begins with modification of the original data using random sampling with replacement i.e., bootstrapping. This ensures that the same data is not used for every tree, increasing the model's sensitivity. Next, decision trees are independently trained using a random subset of features, reducing the correlation between trees. Finally, predictions are made by passing data through each tree and aggregating the results. [45]. Unfortunately, random forest models lack the transparency of decision trees, necessitating secondary methods of calculating feature importance. Random forests may also struggle when interpreting high-dimensionality data as uninformative features may be used when node-splitting [46].

3.5.2. Application

Random forest models have been applied to injury prediction with mixed success. In a study of sports-related dental injuries in children, random forest algorithms had slightly higher prediction accuracy when compared to the traditional regression methods [47]. A 2020 paper sought to address inconsistency in predictive performance by identifying key risk factors prior to training of the model. They were able to achieve an AUC of 0.79 [48]. A 2022 paper built a random forest model and achieved similar performance with an AUC of 0.72 [49]. In an investigation of paralympic swimmers classifying participants with and without brain injury to determine eligibility, random forests successfully classified 96% of the 51 participants [50]. Contrary to these studies, a 2021 paper found that random forest predicted ankle injuries in young athletes with similar performance to a logistic regression (ROC 0.63 versus 0.65, respectively) [51]. With proper application and unbiased feature selection, random forest models may be tuned to outperform existing classification methods, though they are sensitive to variations in data sets.

3.6. Gradient boosting and AdaBoost

3.6.1. Background

Gradient boosting is a generalization of the earlier AdaBoost algorithm, first described in a 1996 paper by Freund and Schapire [52]. AdaBoost is an ensemble technique that seeks to combine multiple weak learners, traditionally single decision trees known as stumps, into a more complex algorithm. This is desirable as it solves many of the problems present with decision trees [52]. Gradient boosting applies boosting as a gradient descent, improving the network with each subsequent iteration, and allowing for the use of a generic loss function. It solves several weaknesses of AdaBoost, including intolerance of outliers and inability to perform multiclass classification [53]. Both AdaBoost and gradient boosting are powerful algorithms that have been continuously refined since their conception allowing them to be applied broadly to regression and classification problems.

3.6.2. Application

Gradient boosting regularly outperforms baseline regression and various ML algorithms including decision tree and SVM for certain classification problems [54-59]. Nicholson et al. found Gradient boosting to be the most effective of several algorithms in assessing elbow valgus torque and shoulder distraction force in 168 high school and college pitchers [57]. Remarkably, a 2019 study predicting skier injuries found that gradient boosting produced a 0.25 increase in accuracy over logistic regression with an AUC of 0.76 vs 0.52 [54]. Hecksteden et al., in a 2022 prospective observation cohort study, also found that gradient boosting performed better than comparison algorithms when forecasting non-contact time-loss injuries in 88 soccer players [58].

Expanding beyond standard gradient boosting, a 2022 study used XGBoost (extreme gradient boost) to predict post-concussion injuries in 74 college football players with an accuracy of 91.9% [60]. Rommers et al. in a 2020 paper also used XGBoost, this time predicting injuries in 734 youth soccer players with a precision and recall of 84% and 83%, respectively. The authors also were able to classify injuries as either overuse or acute with a precision and recall of 82% [42]. Additionally, a recent retrospective review used an XGBoost model to explore the relationship between biomechanics and self-reported athlete injury [61]. Notably, only one recent paper was found to use AdaBoost, a 2022 study predicting injury in CrossFit practitioners. AdaBoost was found to perform better overall than comparison algorithms with an AUC of 77.93% [56].

A 2018 paper by Valenciano et al. found a modified boosting algorithm called SMOTEBoost (Synthetic Minority Oversampling Technique) was able to predict musculoskeletal injuries in 132 football and handball players with an AUC of 0.747, a true positive rate of 65.9%, and a true negative rate of 79.1% [55]. Another similar algorithm called SmooteBoostM1 was used to predict hamstring injuries in professional soccer players, producing a model with an AUC of 0.837 [62]. Overall, gradient boosting, including the earlier AdaBoost and other modified boosting algorithms, represents a pronounced upgrade over classic logistic regression as well as ML algorithms such as decision tree, KNN, SVM, and multilayer perceptron when applied to the limited-class classification problem presented by predicting sports injury.

3.7. Neural Networks

Neural networks provide some distinct advantages over other predictive techniques. They are structured as an interconnected network of nodes called neurons (see Fig. 4). These neurons represent self-contained sets of algorithms that output values based on their input. Neural networks allow models to learn vast amounts of data and detect patterns that would be otherwise impossible to extract. Two main types of neural networks exist, feed-forward and recurrent. In feed-forward networks, the output of the previous node is fed into the next node. In recurrent networks results are fed back to previous nodes [12, 63].

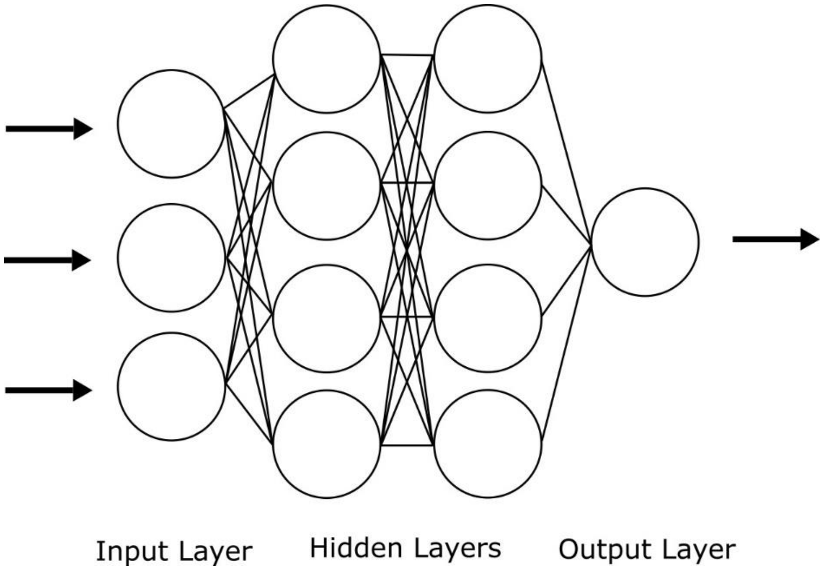

Fig. 5 General structure of a forward feeding, deep, fully connected neural network including an input layer, two hidden layers, and an output layer. Note that all nodes represent a discrete function and are connected to all nodes of both the previous and the next layer

Neural networks have a huge variety of available node algorithms and structures. An overview of these techniques is outside of the scope of this paper, but several processes are explored in more depth including application of convolutional neural networks (CNN), long-short term memory (LSTM), deep Gaussian covariance network (DGCN), and radial basis functions (RBF).

3.8. Convolutional Neural Networks

3.8.1. Background

Convolution is a mathematical process that applies a kernel matrix to transform an image pixel-by-pixel (see Eq. 1). This technique is useful for filtering images as well as image classification. In addition to image classification, convolution can be applied to any 2-dimensional array of numerical data. In the context of

ML, a convolutional neural network relies on alternating convolution and pooling layers to generate a feature map and eventually generate an output [64].

$$\begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1n} \\ x_{21} & x_{22} & \cdots & x_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ x_{m1} & x_{m2} & \cdots & x_{mn} \end{bmatrix} * \begin{bmatrix} y_{11} & y_{12} & \cdots & y_{1n} \\ y_{21} & y_{22} & \cdots & y_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ y_{m1} & y_{m2} & \cdots & y_{mn} \end{bmatrix} = \sum_{i=0}^{m-1} \sum_{j=0}^{n-1} x_{(m-i)(n-j)} \, y_{(1+i)(1+j)}$$

Eq. 1 Generalized equation for the convolution of a given 2-dimensional array of size (*n*,*m*)

Convolutional neural networks have been classically used in image analysis where the 2-dimensional structure and high feature density of pictures lend themselves to convolution. However, CNNs may be applied to any appropriately structured data to allow for a wider range of applications outside of traditional image analysis.

3.8.2. Application

Kautz et al., in their 2017 paper, use CNN to analyze wearable sensor data and allow for automated player monitoring in beach volleyball players. Compared to algorithms including SVM, KNN, Gaussian, and Decision Tree, the CNN provided significantly increased classification accuracy [65]. Pappalardo et al. developed a CNN to analyze multivariate time series extracted from Electronic Performance and Tracking Systems worn by professional soccer players. Their approach allowed for automated feature extraction, an advantage over more traditional time series analysis. Additionally, they were able to develop an injury forecaster that was explainable, which is a necessity for a deployable, real-world model [66]. Similarly, Chen et al. describe a process of converting time series data acquired from player-worn sensors to 2-dimensional images for analysis using a CNN. Notably, they validate using only acceleration data from a single sensor and were able to achieve acceptable levels of accuracy in classification [19]. Song et al. in their 2020 paper developed an optimized-CNN to predict and assess injuries in volleyball players. Using multidimensional sports data, they found that their algorithm was more accurate than comparison algorithms. Additionally, they described a framework for cloud-based deployment and integration with Internet of Things [67]. Ma et al. in a 2019 paper also proposed a CNN for analysis of sports data using a real time cloud-based system and Internet of Things [68]. Ghazi et al. in a 2021 paper describe the use of CNN to estimate peak maximal principal strain in traumatic head injuries. Using data from the National Football League, they were able to achieve >90% accuracy in prediction of concussion vs non-concussion [69].

3.9. Long-Short Term Memory Based Neural Networks (LSTM)

3.9.1. Background

A common feature of feed-forward and recurrent neural networks is the use of gradients in training. Gradients affect the "on/off" signals of the individual nodes of a neural network. Depending on the data set and hyper-parameters of the model, gradients can produce NA values. Several solutions to this problem, known as exploding and disappearing gradients, have been developed, including the use of LSTM nodes which introduce a constant error carousel (CEC) [70]. The CEC allows for gradients to remain unchanged from one node to the next. The more recent addition of a "forget gate" allows the LSTM node to reset, further reducing gradient runaway [71]. Neural networks integrating these types of nodes allow powerful time series analysis.

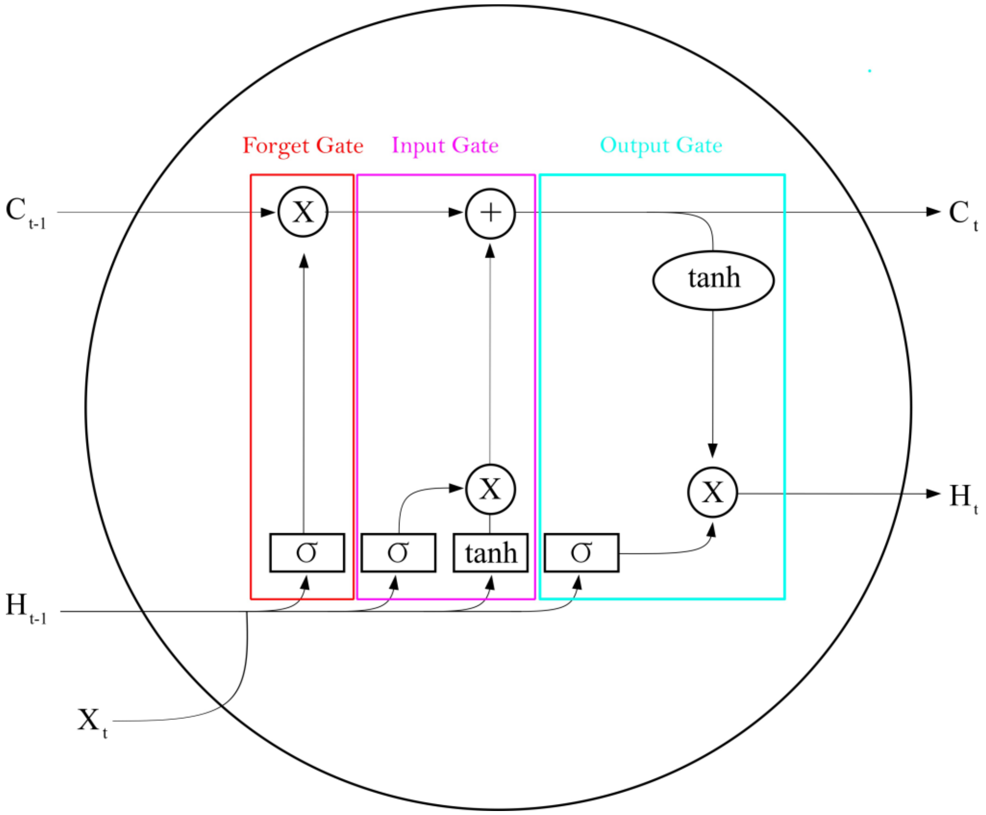

Fig. 6 Diagram of a single LSTM node including input, output, and forget gate [72]

3.9.2. Applications

While LSTM nodes are primarily used for time series analysis, they may be combined with other algorithms to provide an advantage in prediction and classification problems because of their unique nature. In 2021, Meng et al. combined CNN with LSTM to allow for reliable analysis of 2-dimensional data by the LSTM nodes. Using images of professional athletes, they were able to achieve 97.0% classification accuracy for risk stratification broken into No Risk, Low Risk, Medium Risk, and High Risk of injury. The model achieved a sensitivity of 95.70% and a specificity of 97.54% [34]. A combined architecture model such as this may ultimately yield more accurate algorithms.

3.10. Deep Gaussian Covariance Neural Networks

3.10.1. Background

A Gaussian process is a non-parametric, stochastic process defined such that a finite collection of random variables has a multivariate normal distribution. Critically, Gaussian processes can be described by their second order statistics. Defining a covariance function will completely describe the behavior of the original process. By adding a final layer of nodes containing covariance functions to a neural network, the Gaussian process hyperparameters can be treated as outputs of the neural net. This has the advantage of allowing the neural net to solve an easier problem, the tuning of Gaussian hyperparameters, rather than the actual regression which is left to the final layer of covariance functions [73].

3.10.2. Application

A 2022 paper by Rahlf et al. outlined a prospective study protocol using a deep Gaussian covariance network to analyze the relationship between internal and external factors contributing to runner injury. Recruitment for this study was ongoing at the time of publication [74]. This promises to provide real world data on predictive performance of a neural network.

3.11.4. Radial Basis Function Neural Networks

3.11.1. Background

Radial basis functions allow interpolation of multi-dimensional data by calculating the Euclidean distance between data points and a known center point. These functions may be used as activation functions in a neural network. Networks using radial basis functions may be applied to a variety of tasks including regression and classification [75, 76].

3.11.2. Application

In a 2021 paper, Xiang applied an RBF-based neural network to injury predictions. They stratified injury risk and validated using questionnaires sent to expert coaches [77]. Another 2021 paper proposes a similar RBF-based neural network to predict sports injuries. Injury risk is stratified into low risk, at risk, and high risk of injury [78]. Notably, the author looked to determine which factors may contribute most to injury risk. Despite their novel premise, both papers lack robust validation or large data sets and are largely methodological.

3.12. Fuzzy and Grey Neural Network

3.12.1. Background

Fuzzy set theory applies degrees of membership to elements contained within so-called fuzzy set. This contrasts with the “crisp”, or dichotomous, membership assumed in traditional mathematics [79]. Grey theory proposes that systems without information are black while systems with complete information are white. Most real systems, then, are grey, implying incomplete information. Various grey models have been proposed to address this [80]. Fundamentally, both grey and fuzzy theory deal with uncertainty in statistics. Although they are different mathematically, they deal with similar datasets and have been included in the same section for brevity.

3.12.2. Application

A 2021 paper by Wang et al. describes use of a Fuzzy neural network to evaluate degree of injury in sports. They found that the Fuzzy neural network outperformed Bayesian and Lagrange models. However, this was a theoretical proposal using simulated data [81]. Another 2021 paper by Zhang et al. proposed a grey neural network which inputs the results of n-grey models into a neural network for final prediction. This too was a theoretical algorithm tested and validated with simulation data [82]. Despite their lack of real-world application, both papers present intriguing possibilities for integrating Fuzzy and Grey theory as a method of dealing with the inherent variability in sports injury data.

4. Discussion

4.1. Limitations

Many of the articles examining neural networks were theoretical in that they proposed a novel algorithm but validated on a small, artificial data set. These papers are useful to determine new avenues of research and were included. However, without transparent, real-world data or clear explanations of the proposed data collection and preparation, they do not provide concrete information on algorithm efficacy. Additionally, while most articles detail the equations used, many do not explicitly present the model structure, nor do they provide code.

Problems with data transparency are not limited to neural network focused papers. Many of the other papers discussed in this review rely on small or artificial data sets. Additionally, there is a lack of consistent validation techniques and a large potential for mishandling of data. It is also worth mentioning that there exists a persistent problem with multicollinearity in physiological data sets.

Inter-article variability in algorithm efficacy may also prevent strong conclusions from being drawn. Models must be carefully built and algorithms specially selected. Additionally, variations in data quality and structure can impact model performance. Thus, it is difficult to compare any two papers unless they use functionally identical model architectures, parameters and data. Most papers do not fit this criterion. It should be noted that this does not make such papers useless, only difficult to compare directly. Instead, algorithms must be judged based on technical characteristics and capabilities and selected based on individual circumstances.

Because of increased interest in applying ML models to critical decisions in health care and society generally, an ethical imperative has emerged for transparent algorithm. Transparency provides a necessary check and balance to mitigate the risks associated with artificial intelligence-informed decisions. Having addressed these general limitations, each algorithm will be discussed individually.

4.2. Algorithms

K-Nearest Neighbor has some practical limitations to the sample sizes it can efficiently analyze. However, its simplicity and versatility are clear. Integration of special sensors allowing for more precise data collection has improved KNN injury recognition models and increase their ability to identify factors that contribute to injury. Enhanced identification of predictive injury features at the resolution of an individual athlete allows coaches and medical personnel to alter training methods to avoid the identified injury risk. However, KNN has been relegated to the role of comparison algorithm in many of the papers discussed in this article. This should not dissuade future researchers from considering it for use, though.

Another simple algorithm, *K*-means lends itself well to feature extraction. Based on recent work in the literature, *K*-means can be used to classify biokinetic data. Alternatively, *K*-means can effectively be used to predict future high performing players. However, a more interesting application may be found in the preprocessing of data. *K*-means clustering may be applied to data sets early in the exploration phase, rather than as a final predictive algorithm. In any case, *K*-means should be considered when possible.

Support vector machines can be used to both predict the occurrence of an injury as well as elucidate the risk factors that contribute to injury. However, in recent literature, SVM based models have met with mixed success. Even so, SVM should be considered when predicting sports injury events, especially when dealing with high dimensionality data. Notably, the best performing SVM models are built as ensemble models, combining the advantages of several algorithms.

Decision trees may also be suitable in medical decision making as they provide reasonable classification accuracy combined with simple representation of gathered knowledge. More importantly, they provide a remarkably transparent decision-making process, allowing deep exploration of features. And, due to this transparency, the decision-making process can be easily validated by an expert which greatly enhances its utility in situations containing high uncertainty. Random forest models increase predictive accuracy compared to decision trees at the expense of reduced transparency. Additionally, they may struggle when data contains high dimensionality, though condensing may provide adequate abatement. Even with the stated limitations, both decision tree and random forest have performed reasonably well in specific situations and their application should be considered.

Gradient boosting and Adaboost represent significant improvements in predictive capabilities over classic regression as well as the decision trees on which they are based. They are easier to implement and more transparent than neural networks while possessing a capacity for large feature sets. Additionally, they are particularly useful when applied in the context of injury prediction where classification can be limited to a binary choice. In cases where transparency is less critical than predictive accuracy, gradient boosting provides a balance between complexity and performance.

While gradient boosting provides various advantages over simpler models, neural networks tend to be the most accurate and powerful ML algorithms currently available. This performance comes at the price of increased complexity, training time, data requirements, and computational resources. Despite these drawbacks, papers rank CNN, RNN, and other NN architectures favorably against comparison algorithms. However, there is a lack of robust real-world validation largely due to lack of readily available large data sets. Researchers are also using player mounted sensors to collect raw time series data. While this is a valid approach to data collection, it fails to make use of the powerful image recognition and pose-estimation potential of CNN and limits player enthusiasm for data collection in real-world scenarios. There is a clear route to explore more novel approaches to data collection and structuring, as well as to develop robust studies using real-world data. Any given model architecture or combination of architectures could be applied to any given properly tuned data set. This knowledge alone is of little practical value; however, it demonstrates the need for larger sets of real-world data to further triage algorithm utility between situations. Even with the stated limitations, if the data and computational resources are available, neural networks should be heavily considered.

To illustrate one final observation, it is worth examining a recent systematic review by Bullock et al. The review in question presented 30 studies applying ML to sports injury prediction. Notable in their selection criteria was the inclusion of logistic and Poisson regression, both valid but dated approaches to predictive analysis, as well as the exclusion of novel methodologies for modeling. In fact, 22 of the 30 papers included logistic regression, and 2 of the remaining 8 used Poisson regression [3]. We believe this succinctly illustrates a major bottleneck in the application of ML to sports medicine. A significant number of quality studies are failing to make full use of modern, powerful ML algorithms. Instead, they rely on well-studied but potentially inadequate regression techniques in addition to falling prey to some other pitfalls discussed earlier. Recent research that does attempt to move past these relatively simple models often fails to produce reliable, generalizable results. Additionally, these papers are often of limited value to those looking for practical applications of ML. Despite these drawbacks, we feel that it is unreasonable to dismiss the usefulness or real-world applicability of ML based on decidedly outdated methodologies.

5. Conclusion

There appear to be several issues relating to the application of ML as a form of predictive analytics in sports medicine. For example, there is a lack of uniform data sets related to sports injury, resulting in an inability to easily test and validate novel approaches to modeling. Data is being collected inefficiently, particularly with respect to the use of cumbersome player-worn sensors. Studies are difficult to compare due to the individualized nature of ML model architectures and a lack of transparent reporting regarding algorithm construction. In some cases, outdated or inappropriate models are being applied for the sake of ease of implementation. For example, logistic regression is often considered a ML algorithm due to its ability to produce a categorical output, but it is not adaptive like other ML techniques and is consistently outperformed by modern ML algorithms. Surprisingly, even logistic regression models, which are outdated and not considered ML, continue to be used as a prediction tool, often with poor performance. Many injury prediction studies rely entirely on these older techniques, resulting in the appearance that ML is of little clinical use. Importantly, this emphasizes the early stage of the research into ML applications in sports injury and the potential for positive future exploration into its use.

Potential solutions to the aforementioned issues include the creation of open-source, uniform data sets that can be tailored to the strengths of targeted algorithms. The vast amounts of data available to sports teams and sports casting agencies, notably, high quality video footage, could be used to generate large databases for the training of CNN to a variety of ends. This solution would eliminate two of the above problems simultaneously. It would provide researchers with a large, reliable, uniform data set with which to train and validate. It would also eliminate the need to collect data using unreliable athlete-worn sensors. An additional benefit of pose estimation-based prediction is the generalizability that will likely result, allowing pre-trained networks to be tuned to multiple sports with relative ease.

Another potential solution is a reduced reliance on older regression analysis models. While logistic regression models can be powerful tools, they often break down when applied to the complex, multivariate problems presented by sports injury prediction. We have shown this to be the case in the literature generally, as logistic regression is a common baseline comparison model, as emphasized in our discussion of the recent review article by Bullock et al. Though these older models still hold a great deal of utility, they shouldn't be conflated with ML models. Further, modern ML models likely hold greater potential to provide solutions to especially complex problems in injury prediction.

Despite the outlined challenges, significant potential exists within this space. By thoughtfully selecting algorithms and by building adequate data sets, researchers will be able to explore more novel approaches and continue to push the boundaries of ML capability in improving sports medicine outcomes.

References

1. Samuel, A.L., *Some Studies in Machine Learning Using the Game of Checkers.* IBM Journal of Research and Development, 1959. **3**(3): p. 210-229.DOI: 10.1147/rd.33.0210.
2. Alpaydin, E., *Introduction to machine learning*. 2020: MIT press.
3. Bullock, G.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 Med, 2022.DOI: 10.1007/s40279-022-01698-9.
4. Van Eetvelde, H., et al., *Machine learning methods in sport injury prediction and prevention: a systematic review.* J Exp Orthop, 2021. **8**(1): p. 27.DOI: 10.1186/s40634-021-00346-x.
5. Horvat, T. and J. Job, *The use of machine learning in sport outcome prediction: A review.* WIREs Data Mining and Knowledge Discovery, 2020. **10**(5): p. e1380.DOI: 10.1002/widm.1380.
6. Claudino, J.G., et al., *Current Approaches to the Use of Artificial Intelligence for Injury Risk Assessment and Performance Prediction in Team Sports: a Systematic Review.* Sports Med Open, 2019. **5**(1): p. 28.DOI: 10.1186/s40798-019-0202-3.
7. Rico-González, M., et al., *Machine learning application in soccer: A systematic review.* Biology of Sport, 2023: p. 249-263.DOI: 10.5114/biolsport.2023.112970.
8. Nassis, G., et al., *A review of machine learning applications in soccer with an emphasis on injury risk.* Biology of Sport, 2023: p. 233-239.DOI: 10.5114/biolsport.2023.114283.
9. Koseler, K. and M. Stephan, *Machine learning applications in baseball: A systematic literature review.* Applied Artificial Intelligence, 2017. **31**(9-10): p. 745-763.
10. Seow, D., I. Graham, and A. Massey, *Prediction models for musculoskeletal injuries in professional sporting activities: A systematic review.* Translational Sports Medicine, 2020. **3**(6): p. 505-517.DOI: 10.1002/tsm2.181.
11. Liu, Y., et al., *How to Read Articles That Use Machine Learning: Users' Guides to the Medical Literature.* JAMA, 2019. **322**(18): p. 1806-1816.DOI: 10.1001/jama.2019.16489.
12. Grossberg, S., *Nonlinear neural networks: Principles, mechanisms, and architectures.* Neural Networks, 1988. **1**(1): p. 17-61.DOI: 10.1016/0893-6080(88)90021-4.

13. Redyuk, S., et al., *Learning to Validate the Predictions of Black Box Machine Learning Models on Unseen Data*, in *Proceedings of the Workshop on Human-In-the-Loop Data Analytics - HILDA'19*. 2019, Association for Computing Machinery: Amsterdam, Netherlands. p. 1-4.DOI: 10.1145/3328519.3329126.

14. Fushiki, T., *Estimation of prediction error by using K-fold cross-validation.* Statistics and Computing, 2009. **21**(2): p. 137-146.DOI: 10.1007/s11222-009-9153-8.

15. Prakisya, N.P.T., et al., *Utilization of <i>K</i>-nearest neighbor algorithm for classification of white blood cells in AML M4, M5, and M7.* Open Engineering, 2021. **11**(1): p. 662-668.DOI: 10.1515/eng-2021-0065.

16. Liu, K., et al. *Classification of knee joint vibroarthrographic signals using k-nearest neighbor algorithm*. in *2014 IEEE 27th Canadian Conference on Electrical and Computer Engineering (CCECE)*. 2014.DOI: 10.1109/CCECE.2014.6900933.

17. Bressan, M. and J. Vitrià, *Nonparametric discriminant analysis and nearest neighbor classification.* Pattern Recognition Letters, 2003. **24**(15): p. 2743-2749.DOI: 10.1016/s0167-8655(03)00117-x.

18. Zhang, Z., *Introduction to machine learning: k-nearest neighbors.* Annals of Translational Medicine, 2016. **4**(11): p. 218-218.DOI: 10.21037/atm.2016.03.37.

19. Chen, X., G. Yuan, and F. Khan, *Sports Injury Rehabilitation Intervention Algorithm Based on Visual Analysis Technology.* Mobile Information Systems, 2021. **2021**: p. 1-8.DOI: 10.1155/2021/9993677.

20. Naglah, A., et al. *Athlete-Customized Injury Prediction using Training Load Statistical Records and Machine Learning*. in *2018 IEEE International Symposium on Signal Processing and Information Technology (ISSPIT)*. 2018.DOI: 10.1109/ISSPIT.2018.8642739.

21. MacQueen, J. *Classification and analysis of multivariate observations*. in *5th Berkeley Symp. Math. Statist. Probability*. 1967.

22. Hong, X., *Basketball Data Analysis Using Spark Framework and K-Means Algorithm.* J Healthc Eng, 2021. **2021**: p. 6393560.DOI: 10.1155/2021/6393560.

23. Likas, A., N. Vlassis, and J. J. Verbeek, *The global k-means clustering algorithm.* Pattern Recognition, 2003. **36**(2): p. 451-461.DOI: 10.1016/s0031-3203(02)00060-2.

24. Dingenen, B., et al., *Subclassification of recreational runners with a running-related injury based on running kinematics evaluated with marker-based two-dimensional video analysis.* Phys Ther Sport, 2020. **44**: p. 99-106.DOI: 10.1016/j.ptsp.2020.04.032.

25. Ibanez, S.J., C.D. Gomez-Carmona, and D. Mancha-Triguero, *Individualization of Intensity Thresholds on External Workload Demands in Women's Basketball by K-Means Clustering: Differences Based on the Competitive Level.* Sensors (Basel), 2022. **22**(1): p. 324.DOI: 10.3390/s22010324.

26. Noble, W., *What is a support vector machine?* 2006: Nature Biotechnology.

27. Cortes, C. and V. Vapnik, *Support-vector networks.* Machine Learning, 1995. **20**(3): p. 273-297.DOI: 10.1007/bf00994018.

28. Guyon, I., et al., *Gene Selection for Cancer Classification using Support Vector Machines.* Machine Learning, 2002. **46**(1/3): p. 389-422.DOI: 10.1023/a:1012487302797.

29. Van Eetvelde, H., et al., *Machine learning methods in sport injury prediction and prevention: a systematic review*. Journal of Experimental Orthopaedics, 2021. **8**(1): p. 27.DOI: 10.1186/s40634-021-00346-x.

30. Rodas, G., et al., *Genomic Prediction of Tendinopathy Risk in Elite Team Sports.* Int J Sports Physiol Perform, 2019. **15**(4): p. 1-7.DOI: 10.1123/ijspp.2019-0431.

31. Ruddy, J.D., et al., *Predictive Modeling of Hamstring Strain Injuries in Elite Australian Footballers.* Med Sci Sports Exerc, 2018. **50**(5): p. 906-914.DOI: 10.1249/MSS.0000000000001527.

32. Carey, D.L., et al., *Predictive Modelling of Training Loads and Injury in Australian Football.* International Journal of Computer Science in Sport, 2018. **17**(1): p. 49-66.DOI: 10.2478/ijcss-2018-0002.

33. Landset, S., M.F. Bergeron, and T.M. Khoshgoftaar. *Using Weather and Playing Surface to Predict the Occurrence of Injury in Major League Soccer Games: A Case Study*. in *2017 IEEE International Conference on Information Reuse and Integration (IRI)*. 2017.DOI: 10.1109/IRI.2017.86.

34. Meng, L. and E. Qiao, *Analysis and design of dual-feature fusion neural network for sports injury estimation model.* Neural Computing and Applications, 2021.DOI: 10.1007/s00521-021-06151-y.

35. Shen, H., *Prediction simulation of sports injury based on embedded system and neural network.* Microprocessors and Microsystems, 2021. **82**: p. 103900.DOI: 10.1016/j.micpro.2021.103900.

36. Wang, S. and B. Lyu, *Evidence-based sports medicine to prevent knee joint injury in triple jump.* Revista Brasileira de Medicina do Esporte, 2022. **28**: p. 195-198.

37. Kingsford, C. and S.L. Salzberg, *What are decision trees?* Nat Biotechnol, 2008. **26**(9): p. 1011-3.DOI: 10.1038/nbt0908-1011.

38. Connaboy, C., et al., *Employing machine learning to predict lower extremity injury in US Special Forces.* Medicine and science in sports and exercise, 2018.

39. Mendonça, L.D., et al., *Association of hip and Foot Factors with Patellar Tendinopathy (Jumper's knee) in athletes.* journal of orthopaedic & sports physical therapy, 2018. **48**(9): p. 676-684.

40. Kolodziej, M., et al., *Identification of Neuromuscular Performance Parameters as Risk Factors of Non-contact Injuries in Male Elite Youth Soccer Players: A Preliminary Study on 62 Players With 25 Non-contact Injuries.* Front Sports Act Living, 2021. **3**: p. 615330.DOI: 10.3389/fspor.2021.615330.

41. Ruiz-Perez, I., et al., *A Field-Based Approach to Determine Soft Tissue Injury Risk in Elite Futsal Using Novel Machine Learning Techniques.* Front Psychol, 2021. **12**: p. 610210.DOI: 10.3389/fpsyg.2021.610210.

42. Rommers, N., et al., *A Machine Learning Approach to Assess Injury Risk in Elite Youth Football Players.* Med Sci Sports Exerc, 2020. **52**(8): p. 1745-1751.DOI: 10.1249/MSS.0000000000002305.

43. Rossi, A., et al., *Effective injury forecasting in soccer with GPS training data and machine learning.* PLoS One, 2018. **13**(7): p. e0201264.DOI: 10.1371/journal.pone.0201264.

44. Breiman, L., *Random forests.* Machine learning, 2001. **45**(1): p. 5-32.

45. Cutler, A., D.R. Cutler, and J.R. Stevens, *Random forests*, in *Ensemble machine learning*. 2012, Springer. p. 157-175.

46. Nguyen, T.T., J.Z. Huang, and T.T. Nguyen, *Unbiased feature selection in learning random forests for high-dimensional data.* ScientificWorldJournal, 2015. **2015**: p. 471371.DOI: 10.1155/2015/471371.

47. Farhadian, M., S. Torkaman, and F. Mojarad, *Random forest algorithm to identify factors associated with sports-related dental injuries in 6 to 13-year-old athlete children in Hamadan, Iran-2018 -a cross-sectional study.* BMC Sports Sci Med Rehabil, 2020. **12**(1): p. 69.DOI: 10.1186/s13102-020-00217-5.

48. Henriquez, M., et al., *Machine Learning to Predict Lower Extremity Musculoskeletal Injury Risk in Student Athletes.* Front Sports Act Living, 2020. **2**: p. 576655.DOI: 10.3389/fspor.2020.576655.

49. Goggins, L., et al., *Detecting Injury Risk Factors with Algorithmic Models in Elite Women's Pathway Cricket.* Int J Sports Med, 2022. **43**(4): p. 344-349.DOI: 10.1055/a-1502-6824.

50. Hogarth, L., et al., *Classifying motor coordination impairment in Para swimmers with brain injury.* J Sci Med Sport, 2019. **22**(5): p. 526-531.DOI: 10.1016/j.jsams.2018.11.015.

51. Jauhiainen, S., et al., *New Machine Learning Approach for Detection of Injury Risk Factors in Young Team Sport Athletes.* Int J Sports Med, 2021. **42**(2): p. 175-182.DOI: 10.1055/a-1231-5304.

52. Freund, Y. and R.E. Schapire, *A decision-theoretic generalization of on-line learning and an application to boosting.* Journal of computer and system sciences, 1997. **55**(1): p. 119-139.

53. Friedman, J.H., *Greedy function approximation: a gradient boosting machine.* Annals of statistics, 2001: p. 1189-1232.

54. Radovanović, S., et al. *Ski Injury Predictions with Explanations*. in *ICT Innovations 2019. Big Data Processing and Mining*. 2019. Cham: Springer International Publishing.

55. Lopez-Valenciano, A., et al., *A Preventive Model for Muscle Injuries: A Novel Approach based on Learning Algorithms.* Med Sci Sports Exerc, 2018. **50**(5): p. 915-927.DOI: 10.1249/MSS.0000000000001535.

56. Moustakidis, S., et al., *Prediction of Injuries in CrossFit Training: A Machine Learning Perspective.* Algorithms, 2022. **15**(3): p. 77.

57. Nicholson, K.F., et al., *Machine Learning and Statistical Prediction of Pitching Arm Kinetics.* Am J Sports Med, 2022. **50**(1): p. 238-247.DOI: 10.1177/03635465211054506.

58. Hecksteden, A., et al., *Forecasting football injuries by combining screening, monitoring and machine learning.* Sci Med Footb, 2022: p. 1-15.DOI: 10.1080/24733938.2022.2095006.

59. Luu, B.C., et al., *Machine Learning Outperforms Logistic Regression Analysis to Predict Next-Season NHL Player Injury: An Analysis of 2322 Players From 2007 to 2017.* Orthop J Sports Med, 2020. **8**(9): p. 2325967120953404.DOI: 10.1177/2325967120953404.

60. Mansouri, M., et al., *A predictive paradigm for identifying elevated musculoskeletal injury risks after sport-related concussion.* Sports Orthopaedics and Traumatology, 2022. **38**(1): p. 66-74.DOI: 10.1016/j.orthtr.2021.11.006.

61. Windsor, J., et al., *A Retrospective Study of Foot Biomechanics and Injury History in Varsity Football Athletes at the U.S. Naval Academy.* Mil Med, 2022. **187**(5-6): p. 684-689.DOI: 10.1093/milmed/usab370.

62. Ayala, F., et al., *A Preventive Model for Hamstring Injuries in Professional Soccer: Learning Algorithms.* Int J Sports Med, 2019. **40**(5): p. 344-353.DOI: 10.1055/a-0826-1955.

63. Kotsiantis, S.B., I. Zaharakis, and P. Pintelas, *Supervised machine learning: A review of classification techniques.* Emerging artificial intelligence applications in computer engineering, 2007. **160**(1): p. 3-24.

64. O'Shea, K. and R. Nash, *An introduction to convolutional neural networks.* arXiv preprint arXiv:1511.08458, 2015.

65. Kautz, T., et al., *Activity recognition in beach volleyball using a Deep Convolutional Neural Network.* Data Mining and Knowledge Discovery, 2017. **31**(6): p. 1678-1705.DOI: 10.1007/s10618-017-0495-0.

66. Pappalardo, L., et al., *Explainable Injury Forecasting in Soccer via Multivariate Time Series and Convolutional Neural Networks.* Barça Sports Anal. Summit, 2019.

67. Song, H., et al., *Secure prediction and assessment of sports injuries using deep learning based convolutional neural network.* Journal of Ambient Intelligence and Humanized Computing, 2021. **12**(3): p. 3399-3410.DOI: 10.1007/s12652-020-02560-4.

68. Ma, H. and X. Pang, *Research and Analysis of Sport Medical Data Processing Algorithms Based on Deep Learning and Internet of Things.* IEEE Access, 2019. **7**: p. 118839-118849.DOI: 10.1109/access.2019.2936945.

69. Ghazi, K., et al., *Instantaneous Whole-Brain Strain Estimation in Dynamic Head Impact.* J Neurotrauma, 2021. **38**(8): p. 1023-1035.DOI: 10.1089/neu.2020.7281.

70. Hochreiter, S. and J. Schmidhuber, *Long short-term memory*. Neural Comput, 1997. **9**(8): p. 1735-80.DOI: 10.1162/neco.1997.9.8.1735.

71. Gers, F.A., J. Schmidhuber, and F. Cummins, *Learning to forget: continual prediction with LSTM.* Neural Comput, 2000. **12**(10): p. 2451-71.DOI: 10.1162/089976600300015015.

72. Amendolara, A., *Predictive modeling of influenza in New England using a recurrent deep neural network*. 2019. *Theses*. 1739. https://digitalcommons.njit.edu/theses/1739.

73. Cremanns, K. and D. Roos, *Deep Gaussian covariance network.* arXiv preprint arXiv:1710.06202, 2017.

74. Rahlf, A.L., et al., *A machine learning approach to identify risk factors for running-related injuries: study protocol for a prospective longitudinal cohort trial.* BMC Sports Sci Med Rehabil, 2022. **14**(1): p. 75.DOI: 10.1186/s13102-022-00426-0.

75. Broomhead, D.S. and D. Lowe, *Radial basis functions, multi-variable functional interpolation and adaptive networks*. 1988, Royal Signals and Radar Establishment Malvern (United Kingdom).

76. Orr, M.J., *Introduction to radial basis function networks*. 1996, Technical Report, center for cognitive science, University of Edinburgh ….

77. Xiang, C., *Early Warning Model of Track and Field Sports Injury Based on RBF Neural Network Algorithm.* Journal of Physics: Conference Series, 2021. **2037**(1): p. 012084.DOI: 10.1088/1742-6596/2037/1/012084.

78. He, F. and W. Wang, *Early Warning Model of Sports Injury Based on RBF Neural Network Algorithm.* Complexity, 2021. **2021**: p. 1-10.DOI: 10.1155/2021/6622367.

79. Zimmermann, H.J., *Fuzzy set theory.* Wiley Interdisciplinary Reviews: Computational Statistics, 2010. **2**(3): p. 317-332.DOI: 10.1002/wics.82.

80. Ngo, H.A., T.N. Hoang, and M. Dik, *Introduction to the Grey Systems Theory and Its Application in Mathematical Modeling and Pandemic Prediction of Covid-19*, in *Analysis of Infectious Disease Problems (Covid-19) and Their Global Impact*, P. Agarwal, et al., Editors. 2021, Springer Singapore: Singapore. p. 191-218.DOI: 10.1007/978-981-16-2450-6_10.

81. Wang, D. and J.S. Yang, *Analysis of Sports Injury Estimation Model Based on Mutation Fuzzy Neural Network.* Comput Intell Neurosci, 2021. **2021**: p. 3056428.DOI: 10.1155/2021/3056428.

82. Zhang, F., Y. Huang, and W. Ren, *Basketball Sports Injury Prediction Model Based on the Grey Theory Neural Network.* J Healthc Eng, 2021. **2021**: p. 1653093.DOI: 10.1155/2021/1653093.