Information Systems 99 (2021) 101750

Contents lists available at ScienceDirect

Information Systems

journal homepage: www.elsevier.com/locate/is

A Data Science approach analysing the Impact of Injuries on Basketball Player and Team Performance

Vangelis Sarlis [a], Vasilis Chatziilias [b], Christos Tjortjis [a,*], Dimitris Mandalidis [b]

[a] The Data Mining & Analytics Research Group School of Science and Technology, International Hellenic University, 14th km Thessaloniki – N. Moudania, 57001 Thermi, Greece
[b] Sports Physical Therapy Laboratory, School of Physical Education and Sports Science, National and Kapodistrian University of Athens, Isminis 8, Dafni 17237, Athens, Greece

ARTICLE INFO

Article history:
Received 30 January 2021
Received in revised form 9 February 2021
Accepted 10 February 2021
Available online 22 February 2021
Recommended by Dennis Shasha

Keywords:
Sports injuries
Sports Analytics
Sports Data Mining
Data analysis
Statistics
Data Science

ABSTRACT

The sports industry utilizes science to improve short to long-term team and player management regarding budget, health, tactics, training, and most importantly performance. Data Science (DS) and Sports Analytics play key roles in supporting teams, players and experts to improve performance. This paper reviews the literature to identify important attributes correlated with injuries and attempts to quantify their impact on player and team performance, using analytics in the National Basketball Association (NBA) from 2010 up to 2020. It also provides an overview of Machine Learning (ML) and DS techniques and algorithms used to study injuries. Additionally, it provides information for coaches, sports and health scientists, managers and decision makers to recognize the most common injuries and investigate possible injury patterns during competitions. We identify teams and players who suffered the most, and the type of injuries requiring more attention. We found a high impact from injuries and pathologies on performance; musculoskeletal impairments are the most common ones that lead to decreased performance. Finally, we conclude that there is a weak positive relationship between performance and injuries based on a holistic multivariate model that describes player and team performance.

1. Introduction

Analytics are very well known in Sports [1]. NBA player performance is negatively affected by sport injuries. Despite the improvement of prevention and rehabilitation strategies, injury rates remain high. Therefore, team senior management are required to estimate and reduce injury risks, while optimizing team tactics and strategies regarding workload and rest days [2]. Especially, performance forecasting for betting companies or sports clubs is highly substantial either in terms of value or improvement. Therefore, Sports Analytics help and support domain experts to predict possible dire circumstances with the purpose to reduce costs and increase team or player performance [3].

This study examines in detail basketball health data over the latest decade (2010–20) through Data Mining (DM) in correlation with player and team performance. In addition, it attempts to find commonalities by performing classification according to injury and pathology criteria. Furthermore, it shows important insights about team injuries impact associated with advanced basketball analytics. Finally, it presents the performance impact of specific longevity injury types, through DM techniques.

Injuries constitute the biggest concern for teams, management and fans, especially for the best players, because they could dramatically affect the overall team performance [4]. The increased ambiguity in injury prediction depends on many parameters, which are difficult to recognize and quantify.

Sports participation is very popular all over the world with benefits on physical, psychological and social aspects. Regular engagement in sports has been found to enhance: (i) the musculoskeletal system by increasing muscular strength, endurance and power and contributing to the bone mineral content and density, (ii) the cardiorespiratory function, e.g. reducing the risk of coronal heart disease, (iii) mental health by promoting self-esteem and generally by improving the quality of life. A healthy and sports active lifestyle for an athlete can reduce also any association with injury in the preparation before or during the game [5]. Despite the health benefits, taking part in sports exposes the athlete to high injury risks. The occurrence of injuries on recreational and competitive athletes is affected by multiple parameters, such as the age, gender, sport type (contact or non-contact), training workload, moving patterns that each sport includes and other

* Corresponding author.
E-mail addresses: e.sarlis@ihu.edu.gr (V. Sarlis), vasilischatz@phed.uoa.gr (V. Chatziilias), c.tjortjis@ihu.edu.gr (C. Tjortjis), dmndldis@phed.uoa.gr (D. Mandalidis).

https://doi.org/10.1016/j.is.2021.101750
0306-4379/

important factors that are analysed further through this study [6–15].

Basketball is widespread all over the world, from recreational to professional gaming levels. Its popularity is mostly due to the NBA competition and promotion [2], as it is recognized as the top basketball competition in the world [4]. The performance of NBA players increases the popularity and the socio-economic condition of the player, the club, as well as the whole league [16]. Basketball is considered generally as non-contact game with increased physical demands, high intensity and speed moving patterns ([2] and [17]). However basketball has evolved over time to become an increasingly physical game in which contact is accepted and expected from players ([18] and [19]).

Big Data collection and information management are emerging sectors for trainers, physicians, doctors and health domain experts in Sports, but they could help in injury risk factor identification [20]. Hence, the use of biometric tracking technologies, such as accelerometer, Radio-Frequency Identification (RFID), Heart Rate (HR), and Global Positioning System (GPS) sensors and wearables can help to understand player or team disadvantages and optimize the performance, reduce potential injuries and enhance the recovery process [21].

The structure of the remainder of this manuscript is as follows:

2. ***Background***: reviews the literature on important sports analytics concepts in correlation with players' and teams' health data.
3. ***Methodology — Research Design:*** The definition of the problem, noteworthy characteristics and analytics, important injury and pathology analytics that influence team and player performance.
4. ***Findings:*** provides the appropriate data analysis of basketball injury analytics in correlation with team and player performance.
5. ***Discussion:*** Results illustration and critical analysis based on the inferred observations. In addition, this study provides a clear experimental analysis and evaluation over the existing and past basketball injury analytics.
6. ***Conclusion & Future Work:*** The evaluation purpose in the association of basketball performance analytics with injury analytics for teams and players and how their expansion can be applied to different sports domains. According to these directions future work is proposed.

Appendices: Segmentation and classification of important basketball injury analytics based on different features of advanced basketball analytics.

2. Literature Review — Background

Sports injuries are one of the biggest challenges in athletics for many years. The identification and risk assessment of this uncertain factor is crucial for the clubs not only for restructuring of the team selection but most important in economical ways because it costs big investments for the sport organization. Therefore, the avoidance or limitation or even the good accuracy in prediction for players injuries can lead into large cost savings and help them in their performance improvement. Through this literature review will be examined the benefits that provided through sports analytics as for important injury analytics that associated with performance, the effect in sports injuries and the illustration of most crucial injuries that influence the basketball performance in NBA the last decade 2010–2020.

2.1. Sports Analytics

Sports Analytics is an upcoming and promising segment of Analytics that combines Information Technology (IT) and Science. Its scope is to explore athlete performance based on past and current data and provide projections with purpose to improve business decisions [22]. Sports Analytics contains the processes of data collection and management, computational methods and forecasting modelling in order to utilize knowledge extracted from sports data and support decision making [3].

Domain experts, managers, technical staff, owners and players pay attention to advanced analytics with serious consideration for key decisions based on data, advanced metrics, Artificial Intelligence (AI) and technology. Sports Analytics has high attention not only from sports industry but in general the whole business by using data analysis, machine learning (ML) algorithms and technology methodologies that used as a good example to apply the application of AI approaches for its purpose based on their needs [23] and [24]. Sports industry stakeholders can take significant insights regarding team and players performance through these advanced analytics with the purpose to decide faster and more accurately during tough game/season moments. Consequently, a team can increase their chances to win. Team or player performance is the outcome of many parameters such as training and condition, psychology, body and mind preparation and other important factors. The appropriate statistical analysis and critical review of findings with the assistance of the state of the art technological tools and methods are the substantial ingredients for success [25] and [26].

Generally, in sports there are many features that determine charismatic talent, including training time, past performance, sports IQ, and basic skills that can be objectively measured such as reaction time, height, hand length, wingspan, body weight and many other anthropometric or game related characteristics. For instance researches showed that players with long arms and better agility drafted higher in the NBA draft lottery [27].

Sports Businesses invest large amount of money, so every single data column is important in decision making from ticket pricing, roster selection, opponent analysis and many match day decisions [28]. Therefore, it is a sport science field that manages data retrieval as well as the analysis of past and current advanced statistics [29]. The data collection and the combination of Data Science (DS) and ML methods can be a competitive advantage for a team and player in the targeted milestones. Sports Analytics can provide quantitative and qualitative analyses to team owners, technical team staff and players with the purpose to understand in more detail the past, make proper decisions in the current situations and predict future circumstances for maximizing the accuracy of the goal [30].

Data in sports can be retrieved from multiple sources (quantitative and/or qualitative). The variety of sports data can be for example boxscores, health, injury, videos, biometric and other important metrics. Data retrieval process can be integrated, standardized, and analysed through diverse sports advanced statistics with the purpose to support decision making in critical situations [31,32].

There is high complexity in the effort to correlate and analyse different type of sports data resources, including analysis in tactics, video-tracking, roster formation, injury prediction, physical performance measurement through Electronic Performance and Tracking Systems (EPTS) [33].

2.2. Player Injuries in Sports and Data Analysis Techniques

One of the most crucial segments directly related with player and team performance are injuries. Health and injury analytics are the vital ingredients to understand the past but also to use it for further analysis that can lead into team success. Hence, sports analytics and data can incorporate health engagement of players with analytics with purpose to avoid illness and injury in future similar circumstances [23].

Data Mining (DM) is the discovery of rules and patterns from large amounts of data, as well as the process of searching for valuable information within data [34]. Therefore, it is the engagement of one or more techniques into automated analysis and knowledge extraction from data. Furthermore, it is the process of analysing the examined datasets with the purpose to solve the described problems of performance and injuries association that will be referred in findings and discussion section. Sports teams use DM methodologies either for interpretation or segmentation purposes that will finally help them in decision making. Assembling DM techniques with important information can boost a team and give a competitive advantage against to an opponent who does not use these technological and scientific methodologies ([35] and [36]).

The classification of individual players or teams based on their performance can show different perceptions or ways of play. Once the preferences in each position have been decided, gameplay is set, the managers and coaching staff finally can understand, drill down and analyse insights and choose the best option for each situation. Through these advanced basketball analytics, this sophisticated approach can be customized and personalized for each team/player preference and/or performance [37]. Hence, the implementation of these concepts can automate such procedures for optimized classification, segmentation and forecasting.

Soccer is the innovator in prediction of injury possibility with the use of biomedical tracking and monitoring technologies with focus on parameters of workout performance, injury rate, history of injury and odds of injury. Nowadays, other sports (such as NFL, cricket, basketball etc.) understand that it is significant investment on that area because it is cost savvy factor in the long run of a team or a player career [38].

Sports business innovation can be applied in monitoring and tracking player health utilization which leads to injury deduction. Hence, athletes can prolong their careers in their maximum efficiency and performance. Best players of the team and especially all-stars invest large amounts of money for body recovery before and after the match and focus more when they suffer an injury. The career longevity of a player on that level is crucial and needs special care in terms of training, physiotherapy, diet, nutrition and everything relevant to body care. For example, one of greatest players in NBA league LeBron James spends 1.5 million dollars per year to take care of his body for training, massage, therapists, gym, appliance, chefs and other relevant costs [39].

2.3. Major Injuries that influence Performance in NBA league

The last two decades many studies have been conducted with the purpose of analysing injuries or general health pathologies that require physician referral, medication, emergency care or both, and cause game or practice being missed. The majority of the studies examined all types of injuries, as well as the general health pathologies ([17,40] and [18]), whereas some studies examined only one type of injury, like meniscal injuries ([41]), pelvis–hip–thigh injuries ([19] and [16]), facial injuries ([42]) and ankle injuries ([43]). Also, the most of the studies analysed the league injuries of all teams and players ([16–18,40,41] and [19]), but only one study included 53 NBA players [42].

In general, the data were collected through the National Basketball [Athletic] Trainers' Association (NBTA) online database [44] for studies [17–19,41] and [45], while some authors used publicly available records [40] and [16]. The number of seasons that were analysed and criticized in bibliography has a variety and there is a specific pattern; for example examined injuries in 1 season [40], in 3 seasons [45], in 10 seasons [17] and [16], in another study for 17 seasons [18], in [41] for 21 seasons, in study [19] for 24 seasons and in [42] for 34 seasons. In addition, the whole period of NBA season (Preseason, Regular season, Playoffs) was analysed by [45].

The injury rates were examined according to the age and anthropometric characteristics (height, weight and body mass), the player's position, the exposure against an opponent, the body's region, the mechanism of the injury (contact or not), the phase of the season, the days missed, the years of participation to the league, as well as the NBA experience according to the seasons [16–18,40,41] and [19]. Similarly, the injuries influence on player performance can take place based on game availability and career longevity (points, rebounds, assists, steals etc.) through playing periods of the season [16,42] and [45].

According to the results of studies, players in age greater than 30 years old demonstrated more injuries through a season [40] and [16]. Regarding the players' position and the injury prevalence, the literature review shows that "Forwards" presented higher number of injuries per 1000 player of exposures, while "Centers" missed the most playing time [40]. Although in the adductor injuries, "Guards" were injured more frequently than "Forwards" or "Centers" (49% vs 25% vs 25%, respectively [16].

The lower extremity was the most common injured body area and in particular the knee, ankle and foot joints [40] and [17]. Moreover, patellofemoral inflammation was the type of injury that caused more days missed, after that ankle injuries and knee sprains follow [17] and [18].

With regards to the studies that examined exclusively one region of the human body, they have showed that in seasons 1988–1989 till 2011–2012, the quadriceps group (hip region), was the most common injured structure and had a statistically significant game-related injury rate compared with other structures. Furthermore, strains were most frequent in the first month of the season and the greatest risk of strains occurred during the player's eighth year in the league as it shown in study [19]. One other study [16] refers that following adductor injuries NBA players returned to gameplay after missing an average of 16 to 17 days, or 7 to 8 games [16]. Correspondingly, it does not affect player performance, nor game availability or career longevity. Similarly, the facial fractures did not cause significant decline in performance regardless of operative or nonoperative management [42]. Finally, ankle sprains affected on 26% performance decline on average in NBA players with each season and this is the reason for many missed NBA games on aggregate. Most ankle sprains occurred during games (n = 565, 71.0%) and they involved based on contact mechanism of injury [43].

Some studies conducted research in the application of advanced ML techniques in order to make predictions for players injuries based on aggregation of box-score basketball analytics, SportsVU cameras data, player physio metrics and workload management [46].

Long missed game periods give a negative impact in players and teams performance. Therefore, a red alert in sport team's decision makers for an upcoming injury could be very supportive in medical and economical ways. Unfortunately, injuries are caused by various random events (exogenous or endogenous) and they have high level of ambiguity for predictions [47]. Therefore, the segment of medical and injury analytics can be caused by multivariate factors, which makes it difficult for prediction. Injury forecasting is uncertain according to physiologists and sports data analyst because it can be associated with factors of sleeping productive time, psychological, social and nutrition apart from training and physical activities [23].

3. Methodology — Research Design

This section introduces the key research questions, states the aims and objectives, and outlines the applied methodology. This study focuses on associating injury statistics with advanced basketball analytics. DM techniques were used in order to classify various non-homogeneous attributes resulted from data collection into distinct and well explained categories. In addition, the manuscript aimed to recognize insightful health and injury analytics in terms of performance during a period of 10 years.

3.1. Research Questions

1. Which were the most common types of general health pathologies and injuries in the NBA league, during the last decade? (RQ1)
2. Which teams and players suffered from the most injuries? (RQ2)
3. How did injuries affect NBA player and team performance from 2010 up to 2020? (RQ3)
4. Which are the key Data Science methods used for injury analysis? (RQ4)

The answers to the above research questions are crucial for club owners, team staff and domain experts in decision making, roster formation and budget control and/or investment, with the purpose to be more proactive through the association of advanced basketball and injury analytics [37].

3.2. Aim and Objectives

The main aim of this research study is to attain insights based on the correlation of basketball performance analytics with injury analytics. Due to the complexity of sports and the huge amounts of unstructured data retrieved, there is a lack of specificity and context, which can exploit valuable information in depth, through the help of analytics [48]. Team and player performance forecasting by gathering data from different perspectives (training, matches, injury, psychological) is a common practice used by the sports industry and betting companies, and can be used in short/mid/long term predictions [49].

It is crucial for sports teams to be able to understand team/player performance and make appropriate decisions [50]. An objective of this research, is to review basketball performance analytics used worldwide [51] associated with injury perspective or influence. The manuscript also strives to analyse NBA data in order to find useful intuitions in micro-level for players and how can this information be optimized to provide insights into the game and help decision making [52].

Based on retrieved advanced performance basketball analytics in combination with injury analytics this research targets to investigate into main and common injuries that impact on players and teams regarding their performance. In addition, the study presents which are the common injury types and which teams and players suffer most injuries (RQ1 and RQ2). Finally, the recovery time needed is significant for the team.

3.3. Methodology

The data were primarily retrieved and scrapped from various accredited basketball online sources using Python [44,53,54]. They were in a variety of unstructured, structured and hybrid data formats with qualitative and quantitative attributes. Hence, these data were combined, transformed, and analysed. Data mining techniques and state of the art machine learning algorithms were executed via the KNIME Analytics Platform and MS Excel. The code for data scraping, and the relevant KNIME workflows and Excel files for this data analysis can be accessed on GitHub at: https://github.com/vsarlis/nbainjuryanalytics

More specifically, for the period from 2010 up to March 2020 we collected advanced performance analytics for 1298 players and 11 225 records with daily injury reports for all players in the same period. Each player's data include position, team, characteristics (height, weight, age, origin etc.) apart from rating or performance indicators such as Net Rating (NetRtg), True Shooting percentage (TS%), Assists percentage (AST%), Steals percentage (STL%), Blocks percentage (BLK%), Turnover percentage (TOV%), Performance Estimate Rating (PER), WinShare (WS), Plus Minus (Plus/Minus or +/-) and Value Over Replacement (VORP) that are used in the literature as the most appropriate rating metrics [3, 55–57] and [58].

For the case of injury data, the situation was more difficult, because they were retrieved in raw unstructured text format with generic descriptions, with scientific medical jargon deprived of reasoning, and time series reports for basketball games. During data pre-processing we conducted cleansing, feature selection and engineering methods. This helped in depth data understanding and improvement with the purpose to make comprehensive classifications using a homogeneous architecture based on performance and injury criteria.

4. Results

Using the proposed methodology, we produced results addressing RQ1–3 in the first subsection dealing with injury data analysis side. The second subsection comparatively reviews ML and DM state-of-the-art techniques used in the bibliography.

4.1. Sport injuries in NBA league for 2010–2020 (RQ1)

During data collection and feature selection we identified 11 225 records indicating missed NBA league games by players during the last decade (2010 up to 2020). Further review of the retrieved data revealed records related (n = 8667) and records unrelated (n = 2558) to a specific injury or pathology. The latter group included records corresponding to decisions with regard to player injury status or their readiness for competition (e.g. "activated from or placed on injury list (IL) for rest", "returned to line up", "do not play (DNP)", "did not dress (DND)", "Day to Day" (DTD), "Rest" or "conditioning"). From all the records indicating an injury or pathology, 3753 referred to repeated recordings of the same injury suffered by a player and were therefore excluded from further data analysis. Eventually, only 5414 records corresponded to an actual injury or pathology with 596 being related to **general-health pathologies,** 185 to **head injuries** and 4112 to **musculoskeletal injuries**.

The pathology or injury was either **undisclosed or unspecified** in 521 records (Table 4.1). The general-health-related pathologies, such as appendicitis, flu, or gastroenteritis were classified according to the main organ system affected, i.e. the respiratory and digestive system. Other pathologies (n = 19) related to the circulatory (n = 7), the integumentary (n = 2) and the reproductive system (n = 1) as well as oral/teeth (n = 8) and ear problems (n = 1) were reported collectively in a separate group. Several pathologies remained unclassified, as there was no indication of the affected organ system (e.g. illness, general soreness, viruses, allergic reactions). Head-related injuries were also classified according to the system organ affected that is the integumentary (e.g. lacerations), the nervous (e.g. concussions) and the musculoskeletal system organ (e.g. facial bone fractures, bruises). Finally, musculoskeletal injuries were classified according to four major anatomical areas, i.e. the neck, the trunk, and the upper and lower extremity.

A separate group of records (n = 21) consisted musculoskeletal injuries that occurred in multiple distinct major anatomical areas (e.g. ankle and abdominal area, upper and lower limb). Further classification of the musculoskeletal injuries in anatomical sub-areas was performed for the trunk (chest, abdominal, thoracolumbar and pelvis area including the sacral, pubic and buttock sub-area), the upper extremity (shoulder, upper arm & forearm, elbow, and hand, thumb & fingers sub-area) and the lower extremity (hip, thigh, knee, calf, fibular, shin, ankle, heel, foot, toes sub-area) with the last major area presenting in several cases injuries in multiple distinct anatomical sub-areas (Table 4.1).

All injuries were classified according to the major area or sub-area that occurred based on the direct or indirect reference of a record to an injury. Direct references to a specific injury included (i) strains of a particular muscle, (ii) sprains of a specific ligament, (iii) cartilaginous damages (bulging or herniated discs, meniscal tears), (iv) subluxations, dislocations, or hyperextensions of a joint, (v) tendinopathies or (vi) bone fractures located in any of the major anatomical areas or sub-areas specified in this study. Injuries were also classified in a specific anatomical area when a record referred to an injury indirectly. Such records mainly concerned either symptoms such as pain, soreness, inflammation, muscle spasm, stiffness, tightness, bruise or swelling that occurred in a particular major anatomical area or sub-area or a method performed for specific treatment purposes (e.g. surgery to repair a herniated disc, arthroscopic surgery performed in a specific anatomical area, ultrasonic treatment to remove scar tissue).

Table 4.1 shows that musculoskeletal injuries are the most frequent ones (75.95%) followed by general health problems (11.01%), unspecified injuries (9.62%) and head injuries (3.42%).

Injury analytics classified distinct injury/pathology reasons of player absence into four different categories (health problems, organ systems, major anatomical areas, and anatomical sub-areas (Table A.2). Based on the injury classifications that applied in the collected data, they were transformed in a way to be associated with basketball performance analytics. Therefore, this correlation has showed that there is negative impact on player and team performance as injuries occurred more often.

This study examined in more detail the major anatomical areas of musculoskeletal injuries as these take up the biggest part (75.95%) of health problems in the NBA. We found, as seen in Table A.4, that "Lower extremity" is the most frequent type, with 69.82% of the total musculoskeletal injuries. "Trunk" and "Upper extremity" injuries are next, with 14.45% and 14.25% respectively, while "Neck" (0.97%) and "Multiple anatomical areas" (0.51%) demonstrated the lowest injury frequency. Thus, we conclusively answer **RQ1** (Which were the most common types of general health pathologies and injuries in the NBA league, during the last decade?).

We also performed null hypothesis significance testing with the purpose to validate the hypothesis with **RQ3** that injuries affect NBA player and team performance, correlating injury and performance analytics. Injury analytics data were categorical while performance analytics were numerical. So, the former was transformed into numerical, in order to execute Null Hypothesis testing using the Pearson model. Table A.3 presents the correlation and p-values between injury and performance analytics. According to these results, all performance variables can be considered to have weak positive relationship with injury analytics. There is statistical significance (with p-value less than 0.05) which disproves the null hypothesis, so we can accept that certain injury analytics data are correlated with basketball performance analytics. However, we cannot conclude that they are the major factor influencing performance. For example, in the case of the "Value Over Replacement" (VORP) performance metric in relation with health problems, we noticed minor positive correlation (Correlation value = 0.2025 and p-value = 0), whereas the "Box Plus-Minus" (BPM) performance metric in relation with anatomical sub-areas shows non-relationship (Correlation value = −0.03358 and p-value = 0.0066). Fig. A.1 illustrates the correlation matrix based in coloured correlations of injury and performance analytics.

Table 4.1
Injury analytics categorization based on health problems, organ systems and anatomical areas/sub-areas.

Type of Injuries/Pathologies	Total Counts Health Problems	% of Health Problems of Grant Total
General health problems	**596**	**11.01%**
Digestive system	**103**	**1.90%**
Other system	**19**	**0.35%**
Respiratory system	**139**	**2.57%**
Unclassified	**335**	**6.19%**
Head injuries	**185**	**3.42%**
Integumentary system	**6**	**0.11%**
Musculoskeletal system	**42**	**0.78%**
Head	42	0.78%
Eye area	10	0.18%
Nose area	19	0.35%
Other facial areas	13	0.24%
Nervous system	**137**	**2.53%**
Head	137	2.53%
Cranial area	112	2.07%
Eye area	25	0.46%
Musculoskeletal Injuries	**4112**	**75.95%**
Musculoskeletal system	**4112**	**75.95%**
Lower extremity	2871	53.03%
Ankle area	801	14.79%
Calf area	133	2.46%
Fibular area	11	0.20%
Foot area	242	4.47%
Heel area	126	2.33%
Hip area	180	3.32%
Knee area	849	15.68%
Multiple anatomical areas	23	0.42%
Shin area	80	1.48%
Thigh area	350	6.46%
Toes area	76	1.40%
Multiple anatomical areas	21	0.39%
Neck	40	0.74%
Trunk	594	10.97%
Abdominal area	168	3.10%
Chest area	51	0.94%
Pelvic area	13	0.24%
Thoracolumbar area	362	6.69%
Upper extremity	586	10.82%
Elbow area	66	1.22%
Hand, Thumb & Fingers area	203	3.75%
Shoulder area	204	3.77%
Upper arm and Forearm area	14	0.26%
Wrist area	99	1.83%
Undisclosed or unspecified	**521**	**9.62%**
Total Unique reasons of Player Absence	**5414**	**100.00%**

According to the results, the hypothesis denotes strong evidence with statistical significance between performance and injury analytics, but there is minor positive or zero correlation due to the fact that the performance of players is a multivariate phenomenon. Hence, performance can be influenced by injuries, but it is just one factor out of several. Based on the aforementioned statements, we can hypothesize that team performance also has weak relationship with injuries due to the fact that players are parts of a team.

Through the analysis in team scope, the following Pareto chart (Fig. 4.1) plots the distribution of pathology and injury events of teams in the period 2010–20 in descending order of frequency. The cumulative line on a secondary axis is shown as a percentage

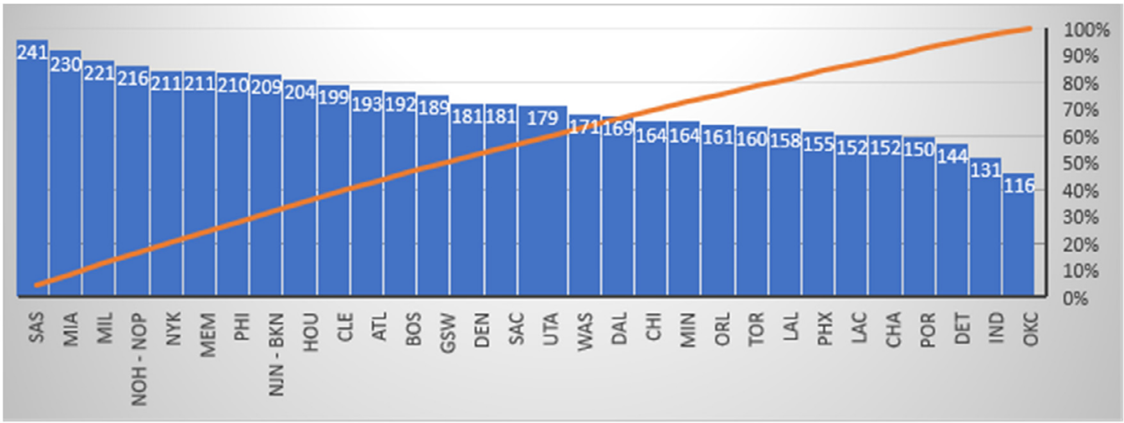

Fig. 4.1. Pareto chart of NBA teams of pathology/injury events in the periods 2010–2020.

Table 4.2
DS and ML techniques used in injury association with basketball analytics.

Data Mining\Machine Learning technique	Teams and Players injury relation
(1) Random Forest predictive model with autoregressive application with logistic regression [4]. (2) Random Forest algorithm for injuries forecasting [33]	(1) Based on that RF model teams can avoid players injuries with strategic decision for proper resting. (2) Injury analysis for prediction purposes based on time series data. It requires enough data, difficult to find due to high complexity.
(3) Linear and non-linear regression models [33,56] (4) Multiple regression models [59].	(3) Difficult in practise to apply feature selection methods with the use of r-squared to intercorrelate and interpret the results. (4) Quantify the correlation between various game parameters with the probability to win. Optimal players allocation to optimize planning and avoid injuries.
(5) Pattern recognition methods to identify optimal teams' substitutions over the season [38]	(5) The performance impact based on players loss of game(s) due to injury status
(6) Least Absolute Shrinkage and Selection Operator (LASSO) and Ridge regression methods [60]	(6) Optimization of prediction accuracy in risk assessment and performance forecasting by applying variable selection and regularization methods. In case of useless variables LASSO performs better, in other cases it is better to use Ridge regression
(7) Deep Learning and Artificial Neural Networks (ANN) for health and injury forecasting [56]	(7) ANN is a non-linear ML method based on the complex correlation between inputs and outputs, trying to identify patterns
(8) Support Vector Machine (SVM) with non-linear classifiers [4,56]	(8) Within per game statistics and injury information used for basketball outcome prediction
(9) Naïve Bayes networks probabilistic models classifiers [56,60]	(9) Diagnosis of sports injuries used for performance evaluation
(10) Decision trees used for classification purposes using models with class labels and branches [56]	(10) The output of that process is to help decision making with the criterion of which option is acting well based on the trained algorithm.
Unsupervised methods: (11) Fuzzy clustering [60] (12) K-means clustering [60] (13) K-Nearest Neighbour [60,61]	(11) Alternative non-linear method based on the possibility that one point belongs to more than one classes. (12) Data clustered according to feature similarity. Clustering is performed based on centroids. (13) The neighbours are selected from a set of attributes in k-NN class or the k-NN regression value
(14) Markov Blanket classifier [60]	(14) Used for injury risk identification and further investigation in sport performance

of the total events. Fig. 4.1 illustrates the cumulative injuries and health problems from 2010 up to 2020. Consequently, San Antonio Spurs (SAS) had the most injuries and pathologies (241), while the Oklahoma City Thunder (OKC) team had the least distinct impairments (116), partly answering **RQ2** (Which teams and players suffered from the most injuries?). Table A.2 shows the occurrences of player injuries based on the proposed classification derived from day-to-day injury reports. For instance, the class of "Anatomical sub-areas" showed that the most frequent musculoskeletal injuries in this time frame were "Knee area" and "Ankle area", with 849 and 801 occurrences respectively.

4.2. Data Mining and Machine Learning methods used for Sports Injury Analytics

Nowadays, there is an increasing demand in sports for new techniques with the purpose to provide significant insights and advanced analytics for teams, technical staff and players. Player performance prediction is correlated with many variables, such as psychology, injury risk [62], bad shots in the starting minutes of the match, good condition and opponents match-ups. These can influence performance, but are difficult to explain and quantify, due to the high level of uncertainty [63]. Nevertheless, technical staff and data analysts need to evaluate these metrics, monitor and track performance and make important decisions for team selection or future acquisitions through scouting [64].

This part of our research provides an overview of injury analytics and correlations with performance using basketball analytics. In addition, it reviews state of the art techniques and methodologies of DM and ML in bibliography used for performance evaluation and prediction using injury associations.

Injuries are difficult to predict because they present high level of ambiguity and uncertainty. CARMELO methodology is a blend of analytics forecasting using regression models that input Real

Table 4.3
Performance and injuries analytics based on absence reasons accuracy.

Data Mining/Machine Learning algorithms & Techniques	Accuracy
XGBoost Tree Ensemble	99.98%
XGBoost Linear Ensemble – Regression	99.76%
Linear Regression	99.75%
Naïve Bayes	96.30%
Tree Ensemble – Regression	90.64%
Random Forest – Regression	90.06%
AutoRegressive Integrated Moving Average (ARIMA)	83.30%
Fuzzy Rule	79.75%
k-nearest Neighbour (k = 2)	78.38%
Gradient Boosted Trees – Regression	77.16%
Polynomial Regression	55.58%
Support Vector Machine (SVM)	31.60%

Plus Minus (RPM) and Box Plus Minus (BPM) statistics to estimate performance. The method does not take into account attributes like psychology, injuries and ergometric statistics [3,65].

A comparative analysis of DS and ML techniques mentioned in the bibliography over the last 10 years as sophisticated state of the art methods for data analysis and prediction in injury analytics are listed in Table 4.2. We reviewed the literature on ML and DM techniques used for players and teams' performance evaluation in correlation with injury analytics **(RQ4)**.

In this paper we review the application of several data mining and machine learning methods and/or algorithms regarding performance and injury analytics, based on the Absence decision (Rest, DTD, DNP, Out indefinitely, and Out for season) and benchmarked their accuracy. For each technique, we assess if absence values need to be transformed from categorical into numerical ones, in preparation for execution. In addition, we partitioned the joined aggregated Performance and Injury dataset, according to the 80/20 Pareto rule [66], that is with 80% of the dataset for training and 20% for testing. The results of this comparative analysis depicted in Table 4.3, show that XGBoost Tree Ensemble, XGBoost Linear Ensemble (Regression) and Linear Regression were more accurate in Absence prediction, based on Performance and Injury analytics.

In order to find important insights into load and rest management of teams, we attempted to explain and quantify the impact of "Rest" in team performance. For that reason, and the purposes of further data analysis, we excluded from the initial dataset players with playing time less than 20 min per game for each season. This can be justified, as only these players are expected to significantly influence team performance through Rest/Load management. Therefore, performance analytics clustered players into three groups, based on Rest Management, that is days of absence due to Rest.

The results of this analysis revealed that teams which do not rest their most utilized, in terms of time, players, demonstrate the lowest performance. The bottom of Table 4.4 comprises 10 teams with the worst performance: 15.21 for PER and 3.21 WS, having BPM −0.19 and VORP 0.88. The best outcomes, with respect to performance analytics, were achieved by teams with balanced rest management, as indicated by the middle group of 10 teams in Table 4.4.

5. Discussion

Digital platforms quantify muscular soreness, nutrition quality, sleeping quality and other Key Performance Indicators (KPIs) for biometrics or physical conditions, which can be used with limited effort and cost. Hence these platforms are data repositories which can be used to identify important information for technical staff and players through ML.

Table 4.4
Team performance in relation with Rest management in the period of 2010–20.

Teams	per_avg	per_class	ws_avg	ws_class	bpm_avg	bpm_class	vorp_avg	vorp_class	REST
SAS	16.87		4.81		1.61		1.64		67
CLE	15.43		3.27		0.38		1.12		31
DAL	15.63		3.33		0.41		1.00		27
GSW	15.26		4.13		0.92		1.42		25
PHI	15.24	15.51	2.42	3.53	-0.19	0.35	0.67	1.12	21
SAC	15.82		3.30		0.08		1.03		21
ATL	15.29		4.40		0.34		1.29		19
LAC	16.10		4.73		1.21		1.67		17
BKN - NJN	14.73		2.52		-0.55		0.64		16
LAL	14.72		2.38		-0.75		0.70		13
TOR	14.99		3.37		0.08		0.97		13
HOU	15.69		4.83		0.76		1.47		12
MIA	17.74		5.43		1.67		1.91		12
MEM	16.08		4.20		1.06		1.46		12
BOS	15.41	15.92	3.84	4.09	0.46	0.67	1.17	1.3	11
MIN	16.50		2.90		0.15		0.90		10
OKC	15.72		4.04		0.83		1.35		10
CHI	15.31		4.03		0.65		1.27		9
DEN	15.45		3.92		0.42		1.16		9
POR	16.32		4.33		0.67		1.38		9
NYK	16.60		4.06		0.37		1.15		9
IND	13.97		3.73		-0.15		0.99		8
NOH - NOP	17.29		3.32		0.70		1.19		7
WAS	15.31		2.65		-0.11		0.73		7
DET	14.93	15.21	2.90	3.21	-0.28	-0.19	0.76	0.88	7
MIL	14.58		2.81		-0.60		0.67		6
PHX	15.30		3.21		-0.04		0.93		6
ORL	15.01		3.29		-0.51		0.74		0
UTA	15.08		3.96		0.23		1.16		0
CHA	14.01		2.20		-1.46		0.45		0
Grand Total	15.57		3.53		0.25		1.08		414

Table 5.1
Top30 most volatile players in NBA championship for the period 2010–2020.

Top30 - Injury volatile	DNP	DTD	Out for season	Out indefinitely	Rest	Total
Derrick Rose	4	19	4	2	3	32
Anthony Davis	7	21	3	3	1	35
Chandler Parsons	8	8	3	1	10	30
Andrew Bogut	9	10	3	1	2	25
Wesley Matthews	2	7	3	1	4	17
Darren Collison	2	4	3	1	2	12
Blake Griffin	2	3	2	5	4	16
Tyreke Evans	15	12	2	4	0	33
Kevin Love	9	19	2	3	4	37
Chris Paul	5	11	2	3	5	26
Nene Hilario	15	15	2	2	0	34
Gordon Hayward	4	5	2	2	0	13
Kobe Bryant	1	3	2	2	4	12
Devin Booker	1	6	2	2	0	11
Rudy Gay	6	12	2	1	2	23
Joel Embiid	1	9	2	1	10	23
Jerryd Bayless	6	10	2	1	0	19
DeMarcus Cousins	4	7	2	1	5	19
Michael Carter-Williams	3	11	2	1	0	17
Dirk Nowitzki	2	5	2	1	7	17
Nikola Pekovic	8	3	2	1	0	14
Mario Chalmers	5	5	2	1	0	13
Andrea Bargnani	7	2	2	1	0	12
Lance Stephenson	2	6	2	1	1	12
Kyrie Irving	4	11	1	6	4	26
Rajon Rondo	10	6	1	4	11	32
Danilo Gallinari	3	13	1	4	1	22
Joakim Noah	9	5	1	3	0	18
Patrick Beverley	2	11	1	3	0	17
John Wall	4	4	1	3	3	15

Table 5.2
Top30 players with the least important absences in NBA championship for the period 2010–2020.

Top30 with least absences	DNP	DTD	Out for season	Out indefinitely	Rest	Total
Domantas Sabonis	0	1	0	0	0	1
Bam Adebayo	0	1	0	0	0	1
Jayson Tatum	0	2	0	0	0	2
Nikola Jokic	1	3	0	0	0	4
Donovan Mitchell	0	4	0	0	0	4
Damian Lillard	0	4	0	0	1	5
DeAndre Jordan	0	3	0	0	3	6
Giannis Antetokounmpo	1	6	0	0	1	8
Jeff Teague	3	7	0	0	2	12
James Harden	8	4	0	0	1	13
Jimmy Butler	3	10	0	0	0	13
Zach Randolph	6	7	0	0	3	16
Draymond Green	1	14	0	0	2	17
Steven Adams	0	4	0	1	0	5
Otto Porter	1	4	0	1	0	6
Rudy Gobert	0	6	0	1	0	7
Khris Middleton	1	5	0	1	0	7
C.J. McCollum	0	5	0	1	1	7
Serge Ibaka	2	5	0	1	1	9
Bradley Beal	3	6	0	1	1	11
Russell Westbrook	2	5	0	1	4	12
DeMar DeRozan	3	5	0	1	3	12
Paul Millsap	6	5	0	1	2	14
Myles Turner	0	4	0	2	0	6
Paul George	1	6	0	2	2	11
Kawhi Leonard	3	12	0	3	4	22
Karl-Anthony Towns	0	1	1	0	0	2
Ben Simmons	0	2	1	0	0	3
Dwight Howard	5	6	1	0	6	18
LeBron James	3	7	1	0	15	26

Table 5.3
NBA performance analytics for Derrick Rose.

Seasons	Avg_mpg	Sum_gp	avg_pts	avg_reb	avg_ast	avg_ts_pct	avg_net_rating	avg_per	avg_ws	avg_bpm	avg_vorp
Derrick Rose	**28.11**	**411**	**16.67**	**2.87**	**4.66**	**0.52**	**-0.84**	**16.66**	**3.07**	**-0.08**	**1.32**
2010-11	37.36	81	25.00	4.10	7.70	0.55	8.30	23.50	13.10	6.80	6.70
2011-12	35.26	39	21.80	3.40	7.90	0.53	10.60	23.00	6.00	6.40	2.90
2013-14	31.10	10	15.90	3.20	4.30	0.45	-3.30	9.70	-0.20	-3.10	-0.10
2015-16	31.77	66	16.40	3.40	4.70	0.48	-4.20	13.40	0.40	-2.50	-0.30
2016-17	32.53	64	18.00	3.80	4.40	0.53	-3.90	17.00	3.00	-1.00	0.50
2017-18	15.85	50	8.40	1.40	1.50	0.51	-6.20	11.40	-0.10	-5.45	-0.20
2018-19	27.29	51	18.00	2.70	4.30	0.56	0.70	19.50	3.00	1.40	1.20
2019-20	25.96	50	18.10	2.40	5.60	0.56	-3.40	21.00	2.50	2.20	1.40

Table 5.4
"Out of season" and "Out indefinitely" injury analytics for Derrick Rose.

Date	Player	Season	TEAM _Abbr	Health problems	Organ systems	Major anatomical areas	Anatomical sub-areas	Decision	Notes
28/04/2012	Derrick Rose	2011–12	CHI	Musculoskeletal Injuries	Musculoskeletal system	Lower extremity	Knee area	Out for season	torn ACL in left knee (out for season)
23/11/2013	Derrick Rose	2013–14	CHI	Musculoskeletal Injuries	Musculoskeletal system	Lower extremity	Knee area	Out for season	torn meniscus in right knee (out for season)
24/02/2015	Derrick Rose	2014–15	CHI	Musculoskeletal Injuries	Musculoskeletal system	Lower extremity	Knee area	Out indefinitely	torn medial meniscus in right knee (out indefinitely)
30/09/2015	Derrick Rose	2015–16	CHI	Head injuries	Musculoskeletal system	Head	Eye area	Out indefinitely	surgery to repair fractured left orbital bone (out indefinitely)
05/04/2017	Derrick Rose	2016–17	NYK	Musculoskeletal Injuries	Musculoskeletal system	Lower extremity	Knee area	Out for season	surgery on left knee to repair torn meniscus (out for season)
12/03/2019	Derrick Rose	2018–19	MIN	Musculoskeletal Injuries	Musculoskeletal system	Upper extremity	Elbow area	Out for season	sore right elbow (out for season)

Table 5.5
Giannis Antetokounmpo NBA performance analytics.

Seasons	Avg_mpg	Sum_gp	avg_pts	avg_reb	avg_ast	avg_ts_pct	avg_net_rating	avg_per	avg_ws	avg_bpm	avg_vorp
Giannis Antetokounmpo	**32.40**	**522**	**20.50**	**9.11**	**4.39**	**0.58**	**3.86**	**22.94**	**9.19**	**5.00**	**4.33**
2013-14	24.64	77	6.80	4.40	1.90	0.52	-4.40	10.80	1.20	-2.50	-0.20
2014-15	31.37	81	12.70	6.70	2.60	0.55	1.10	14.80	6.20	0.00	1.20
2015-16	35.29	80	16.90	7.70	4.30	0.57	-2.60	18.80	7.10	2.10	2.90
2016-17	35.56	80	22.90	8.80	5.40	0.60	1.50	26.10	12.40	7.30	6.70
2017-18	36.75	75	26.90	10.00	4.80	0.60	2.80	27.30	11.90	6.20	5.70
2018-19	32.75	72	27.70	12.50	5.90	0.64	12.50	30.90	14.40	10.40	7.40
2019-20	30.43	57	29.60	13.70	5.80	0.61	16.10	31.90	11.10	11.50	6.60

This study applied pre-processing, cleansing, and aggregation methods to consolidate sports data into a single dataset containing advanced basketball analytics and injury statistics ready for analysis. Based on our results there is a strong association between performance analytics and injury attributes.

This research was conducted using a combination of data from different sources with different meanings and values; so, feature selection and data transformation were deemed essential in the process. Consequently, we identify important insights through the analysed data for teams and players analytics ***(RQ2 and RQ3)***.

Tables 5.1 and 5.2 indicate the top 30 most volatile players and top 30 "strongest" players in the period of 2010–20, focusing on the "out for season" and "out indefinitely" criteria. These criteria were set with suitable criticality weight. Primary weight was assigned to the "Out for season" attribute, as it takes players the longer to recover and get ready for action. Secondary weight was assigned to the "Out indefinitely" attribute due to the return to line-up being ambiguous. After that comes the "Did Not Play" (DNP) attribute as an indicator of a game not played. The "Date to Date" (DTD) attribute refers to questionable player participation to a match. The least weighted attribute is "Rest", as it is not clearly associated with injury or health pathology of the athlete ***(RQ2 and RQ3)***.

Tables 5.1 and 5.2 exhibit that serious impact on top class players performance in correlated with injuries. For example, Derrick Rose (2010–11 NBA MVP) demonstrated decreasing performance (Table 5.3) having suffered from serious injuries after the end of regular season 2011–12 (Table 5.4).

Tables 5.3 and 5.5 show selected important performance rating basketball analytics ([3,67] and [23]) with the purpose to benchmark the association between injury/pathology and performance. These attributes are "avg_mpg" (average minutes per game), "sum_gp" (sum of games played), "avg_pts" (average of scoring points), "avg_reb" (average rebounds), "avg_Ast" (average assist), "avg_ts_pct" (average true shooting percentage), "avg_net_rating" (average Net Rating), "avg_per" (average Performance Estimate Rating), "avg_ws" (average Win Share), "avg_bpm" (average box plus minus) and "avg_vorp" (average Value Over Replacement).

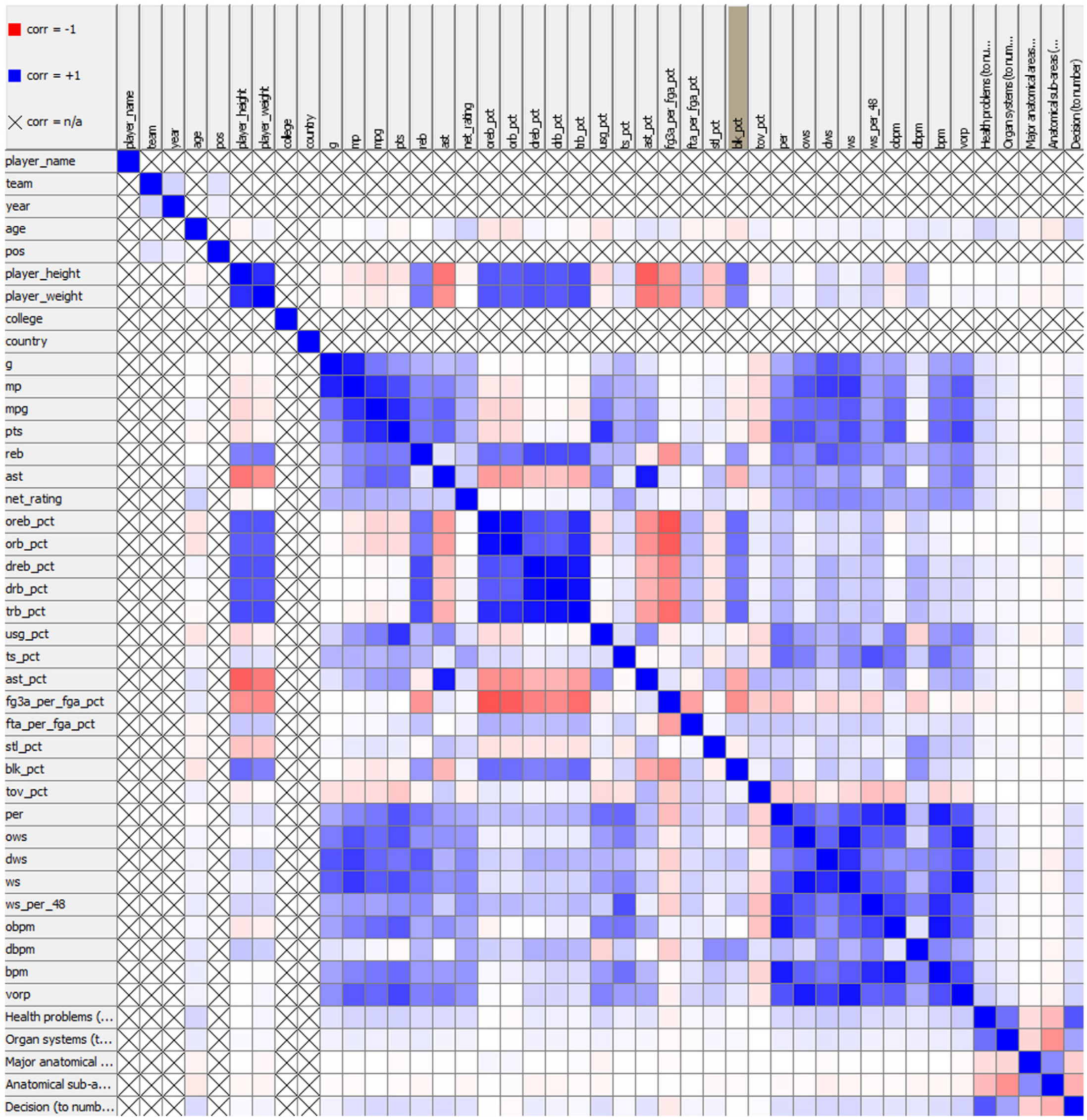

Fig. A.1. Correlation matrix between NBA performance and injury analytics.

On the other hand, players with few critical incidents of injuries (Table 5.2) such as Giannis Antetokounmpo, are demonstrating an increasing performance, which is not negatively correlated with the injury factor, as presented by advanced performance analytics in Table 5.5 ***(RQ2 and RQ3)***.

6. Conclusion & Future Work

6.1. Conclusion

The sports industry and Basketball in particular pays extra attention to small details with the purpose to avoid extra costs and maximize efficiency of teams and players. The objective of this work was an in-depth analysis on the relationship between performance and injury analytics. Data statistical analysis indicated a weak positive correlation between injuries and performance. This implies that injury is just one of several important variables affecting team and player performance.

This study showed that musculoskeletal system injuries are the most common injuries in the NBA. Specifically, for the 2010–2020 period, in terms of anatomical areas, lower extremity issues, trunk and upper extremity injuries were the most frequent and critical. This revealed insights, such as regular or critical injuries in the ankle or knee area which deserve further investigation. ***(RQ1)***.

It also examined how much teams and players suffered, as demonstrated by Fig. 4.1 and Table A.1 as well as Tables 5.1–5.5 and A.2, respectively. Analysing these data leads to the conclusion that even though performance is evidently impacted negatively from injuries, it is hard to quantify this impact as performance is affected by several other intertwined factors. Further in-depth analysis examining specific injuries or pathologies on a player per player case is proposed to achieve better estimations. For example, in the case of Derrick Rose, a 57.8% decline in performance, as assessed by the PER (avg_per) analytic, was identified from season 2011–12 to 2013–2014 after his injury. All in all, injuries and pathologies are one important parameter involved in the performance multivariate model and needs a holistic approach per player and\or team **(RQ2)**.

Hence, there are ways to avoid health issues by understanding bodily and training thresholds, consuming resting days appropriately, optimizing the load and managing as well as understanding in more depth all the parameters that affect the efficiency of player and team performance ***(RQ3)***. The results

Table A.1
Team injury issues in the period 2010–20 categorized per season.

Teams injury issues	Team Abbrev.	2010–11	2011–12	2012–13	2013–14	2014–15	2015–16	2016–17	2017–18	2018–19	2019–20	Total
Atlanta Hawks	**ATL**	8	21	29	21	21	11	26	25	18	13	**193**
Boston Celtics	**BOS**	33	21	23	21	9	14	18	20	16	17	**192**
Cleveland Cavaliers	**CLE**	17	13	25	19	18	27	28	21	20	11	**199**
New Orleans Hornet–Pelicans[a]	**NOH–NOP**	10	27	22	22	17	37	20	20	19	22	**216**
Chicago Bulls	**CHI**	7	20	15	16	9	28	21	17	17	14	**164**
Dallas Mavericks	**DAL**	9	14	14	10	20	29	32	12	23	6	**169**
Denver Nuggets	**DEN**	13	20	13	18	35	22	35	10	8	7	**181**
Golden State Warriors	**GSW**	11	13	17	19	14	27	23	18	20	27	**189**
Houston Rockets	**HOU**	10	11	16	33	20	22	19	43	18	12	**204**
Los Angeles Clippers	**LAC**	8	14	15	16	9	31	17	17	11	14	**152**
Los Angeles Lakers	**LAL**	6	11	24	26	27	10	19	11	18	6	**158**
Miami Heat	**MIA**	14	18	12	19	31	29	37	20	27	23	**230**
Milwaukee Bucks	**MIL**	30	28	6	37	24	25	19	14	25	13	**221**
Minnesota Timberwolves	**MIN**	14	20	26	22	29	5	13	10	12	13	**164**
New Jersey Nets–Brooklyn Nets[b]	**NJN–BKN**	15	29	12	30	20	21	25	23	23	11	**209**
New York Knicks	**NYK**	13	20	25	28	27	24	22	14	24	14	**211**
Orlando Magic	**ORL**	23	12	28	17	16	14	11	21	9	10	**161**
Indiana Pacers	**IND**		14	12	13	20	26	15	10	10	11	**131**
Philadelphia 76ers	**PHI**	4	15	13	15	34	33	42	26	17	11	**210**
Phoenix Suns	**PHX**	9	11	15	11	9	23	23	19	15	20	**155**
Portland Blazers	**POR**	23	20	9	6	21	7	23	12	11	18	**150**
Sacramento Kings	**SAC**	14	13	12	11	18	47	34	14	10	8	**181**
San Antonio Spurs	**SAS**	15	27	30	28	25	30	34	30	16	6	**241**
Oklahoma City Thunder	**OKC**	5	9	15	9	18	15	10	9	15	11	**116**
Toronto Raptors	**TOR**	20	15	16	16	8	22	18	8	21	16	**160**
Utah Jazz	**UTA**	30	20	17	7	18	14	24	26	13	10	**179**
Memphis Grizzles	**MEM**	12	13	18	14	23	22	29	43	25	12	**211**
Washington Wizards	**WAS**	21	12	25	17	15	24	15	15	10	17	**171**
Detroit Pistons	**DET**	20	10	10	12	7	23	13	12	15	22	**144**
Charlotte Hornets	**CHA**	14	19	13	14	20	19	20	17	9	7	**152**
Total		**428**	**510**	**527**	**547**	**582**	**681**	**685**	**557**	**495**	**402**	**5414**

[a]New Orleans Hornets have changed name in the end of season 2012–13. From 2013 to 14 season they play with the name of New Orleans Pelicans.
[b]New Jersey Nets have changed name in the end of season 2011–12. From 2012 to 13 season they play with the name Brooklyn Nets.

of this research indicated that teams with balanced rest\load management achieved better performance based on basketball analytics (Table 4.4).

The comparative review regarding ML and DM techniques and algorithms have shown that LASSO and Ridge regression methods are used to optimize prediction. Random Forest needs big data to achieve accurate predictions with the purpose to gain insights into avoiding injuries. Neural Networks and Pattern recognition methods identify complex correlations and patterns through injury analytics. SVM are used for game outcome prediction, based on injury information. These methods could be combined or aggregated producing for better insights and optimized accuracy **(RQ4)**. In this research we also examined Performance analytics based on player absence from games. Result indicated that there is high accuracy with XGBoost (for Tree Ensemble and Linear Ensemble), Linear Regression and Naïve Bayes (Table 4.3).

Basketball analytics are used to optimize performance on court by identifying the most efficient players and teams, but also optimal combination on pairs of teams [50]. Team rotation also affects player selection. Hence, a team has a roster of 12 players that are ready to be productive and efficient for every minute they play. The new coaching trend shows that the technical staff desires 12 eligible players to be ready for each match instead of old-fashioned coaching style that used 5 to 7 players for most of the game. Over the last years there is an award for the 6th player in basketball associations, and this means that bench players can make a huge impact. Therefore, the proper balance of team rotation, injury prevention and role distribution is a key factor for team success and there is a huge difference between this approach in comparison with the previous decade perception [68].

Injuries negatively impact a team's finance. Proper player usage management, workload tracking and monitoring, rest management and recovery based on training history/status, age, history of injuries, psychology, stress acceptance can affect performance. The business understanding can be achieved with the proper analysis of data requirements and after that the suitable data transformation with purpose to model valuable information. This data modelling can give a competitive advantage to players, technical staff, scouters and team owners and can lead them

Table A.2
Classification of distinct injury/pathology reasons of player absence during the NBA periods 2010–20.

Health problems	Organ systems	Major anatomical areas	Anatomical sub-areas
General-health-related pathologies (596)	Digestive (103)		
	Respiratory (139)		
	Others (19)		
	Unclassified (335)		
Head injuries (185)	Musculoskeletal (42)		Eye (10)
			Nose (19)
			Other facial sub-areas (13)
	Nervous (137)		Eye (25)
			Cranial (112)
	Integumentary (6)		Eye (3)
			Mouth (2)
			Forehead (1)
Musculoskeletal injuries (4112)	Musculoskeletal (4112)	Neck (40)	
		Trunk (594)	Abdominal (168)
			Thoracolumbar (362)
			Chest (51)
			Pelvis (13)
		Upper extremity (586)	Shoulder (204)
			Upper arm & Forearm (14)
			Elbow (66)
			Wrist (99)
			Hand, Thumb & Fingers (203)
		Lower extremity (2871)	Hip (180)
			Thigh (350)
			Knee (849)
			Calf (133)
			Fibular (11)
			Shin (80)
			Heel (126)
			Ankle (801)
			Foot (242)
			Toes (76)
			Multiple sub-areas (23)
		Multiple major areas (21)	
Undisclosed or unspecified (521)			
5414 unique reasons of absence			

Table A.3
Pearson hypothesis correlation with performance and injury analytics.

Performance analytics	Injury and pathology analytics	Correlation value	p-value
dws	Health problems (to number)	0.203733647	0
vorp	Health problems (to number)	0.20250302	0
ws	Health problems (to number)	0.198798205	0
pts	Health problems (to number)	0.17777791	0
vorp	Decision (to number)	0.175652937	0
mpg	Health problems (to number)	0.174924883	0
mp	Health problems (to number)	0.168113505	0
ows	Health problems (to number)	0.167714983	0
age	Health problems (to number)	0.163223241	0
ws	Decision (to number)	0.158759013	0
dws	Decision (to number)	0.157230572	0
per	Health problems (to number)	0.155746406	0
bpm	Health problems (to number)	0.154361048	0
usg_pct	Health problems (to number)	0.147925275	0
mpg	Decision (to number)	0.147874715	0
reb	Health problems (to number)	0.144807622	0
ast	Health problems (to number)	0.143563138	0
pts	Decision (to number)	0.141766282	0
age	Decision (to number)	0.139955705	0
ows	Decision (to number)	0.136814476	0
net_rating	Health problems (to number)	0.136644511	0
ws_per_48	Health problems (to number)	0.136627497	0
ast	Decision (to number)	0.136620921	0
obpm	Health problems (to number)	0.1311837	0
usg_pct	Decision (to number)	0.13117092	0
bpm	Decision (to number)	0.130693744	0
per	Decision (to number)	0.1285853	0
reb	Decision (to number)	0.117934248	0
mp	Decision (to number)	0.117664308	0
net_rating	Decision (to number)	0.113934376	0
g	Health problems (to number)	0.110995604	0
ast_pct	Decision (to number)	0.110866792	0

(*continued on next page*)

Table A.3 (*continued*).

Performance analytics	Injury and pathology analytics	Correlation value	p-value
ws_per_48	Decision (to number)	0.110397258	0
obpm	Decision (to number)	0.106864139	0
dws	Organ systems (to number)	0.106758716	0
ast_pct	Health problems (to number)	0.105522419	0
dbpm	Health problems (to number)	0.102594894	0
vorp	Organ systems (to number)	0.101908744	2.22045E−16
ws	Organ systems (to number)	0.097689685	2.44249E−15
dbpm	Decision (to number)	0.097167679	3.55271E−15
age	Organ systems (to number)	0.09015643	2.84217E−13
age	Anatomical sub-areas (to number)	−0.089672837	3.80448E−13
dreb_pct	Health problems (to number)	0.083457486	1.40734E−11
reb	Organ systems (to number)	0.082683879	2.16711E−11
fg3a_per_fga_pct	Decision (to number)	−0.080625669	6.7052E−11
dreb_pct	Decision (to number)	0.080128018	8.77431E−11
ows	Organ systems (to number)	0.078831405	1.75485E−10
mp	Organ systems (to number)	0.074845304	1.38008E−09
ts_pct	Health problems (to number)	0.070277343	1.292E−08
per	Organ systems (to number)	0.069363988	1.98816E−08
bpm	Organ systems (to number)	0.069342822	2.00799E−08
pts	Organ systems (to number)	0.06750502	4.7014E−08
dbpm	Organ systems (to number)	0.065649338	1.08569E−07
drb_pct	Health problems (to number)	0.062810931	3.74188E−07
dws	Anatomical sub-areas (to number)	−0.062264943	4.71941E−07
mpg	Organ systems (to number)	0.061443927	6.66656E−07
dreb_pct	Organ systems (to number)	0.058226514	2.47608E−06
ws_per_48	Organ systems (to number)	0.057772346	2.96415E−06
usg_pct	Organ systems (to number)	0.056301228	5.26113E−06
fg3a_per_fga_pct	Health problems (to number)	−0.056259761	5.34584E−06
g	Organ systems (to number)	0.053937845	1.28501E−05
stl_pct	Health problems (to number)	0.05287222	1.9E−05
g	Decision (to number)	0.052080337	2.52903E−05
net_rating	Organ systems (to number)	0.052062684	2.54509E−05
dreb_pct	Anatomical sub-areas (to number)	−0.051530406	3.07747E−05
obpm	Organ systems (to number)	0.050901092	3.84348E−05
age	Major anatomical areas (to number)	−0.050143444	5.00617E−05
vorp	Anatomical sub-areas (to number)	−0.049117925	7.11823E−05
drb_pct	Organ systems (to number)	0.04825817	9.51309E−05
player_weight	Decision (to number)	0.047511161	0.000121933
drb_pct	Decision (to number)	0.047027887	0.000142908
player_weight	Anatomical sub-areas (to number)	−0.046443881	0.000172788
stl_pct	Decision (to number)	0.046167381	0.000188899
usg_pct	Major anatomical areas (to number)	−0.044375222	0.000332783
ast	Organ systems (to number)	0.043646037	0.000416623
drb_pct	Anatomical sub-areas (to number)	−0.040134267	0.001174033
fta_per_fga_pct	Decision (to number)	0.039983352	0.001225394
dbpm	Anatomical sub-areas (to number)	−0.039786753	0.001295415
ws	Anatomical sub-areas (to number)	−0.038080969	0.002077187
trb_pct	Health problems (to number)	0.037807769	0.002236661
player_weight	Health problems (to number)	0.037737029	0.002279743
orb_pct	Major anatomical areas (to number)	0.037007513	0.002770512
usg_pct	Anatomical sub-areas (to number)	−0.036567991	0.003110959
ast_pct	Organ systems (to number)	0.036027463	0.003581753
net_rating	Anatomical sub-areas (to number)	−0.035587916	0.004011387
player_height	Anatomical sub-areas (to number)	−0.035275804	0.004344304
oreb_pct	Major anatomical areas (to number)	0.034547813	0.005220198
reb	Anatomical sub-areas (to number)	−0.033888492	0.006147913
stl_pct	Organ systems (to number)	0.033854165	0.00620005
bpm	Anatomical sub-areas (to number)	−0.033585879	0.006621433
trb_pct	Organ systems (to number)	0.032572231	0.008455642
player_weight	Major anatomical areas (to number)	−0.031821649	0.01009376
ts_pct	Decision (to number)	0.031363578	0.011227093
player_weight	Organ systems (to number)	0.030653644	0.013207344
ts_pct	Organ systems (to number)	0.030340712	0.014174171
per	Anatomical sub-areas (to number)	−0.028859666	0.019646245
blk_pct	Organ systems (to number)	0.028507416	0.02119188
stl_pct	Anatomical sub-areas (to number)	−0.028247593	0.022398827
fta_per_fga_pct	Health problems (to number)	0.027434308	0.026570882
trb_pct	Decision (to number)	0.02587134	0.036494712
vorp	Major anatomical areas (to number)	−0.025155282	0.042005134
ws_per_48	Anatomical sub-areas (to number)	−0.025124245	0.042259092
player_height	Organ systems (to number)	0.023887064	0.053492849
trb_pct	Anatomical sub-areas (to number)	−0.023593594	0.056495811
fg3a_per_fga_pct	Organ systems (to number)	−0.023592211	0.056510297
blk_pct	Anatomical sub-areas (to number)	−0.023253739	0.060147393
orb_pct	Decision (to number)	−0.022456837	0.069481249
player_height	Decision (to number)	0.022280403	0.071700952

(*continued on next page*)

Table A.3 (*continued*).

Performance analytics	Injury and pathology analytics	Correlation value	p-value
mp	Anatomical sub-areas (to number)	−0.022103126	0.073989431
dws	Major anatomical areas (to number)	−0.021561294	0.081356304
obpm	Anatomical sub-areas (to number)	−0.021244279	0.085935534
orb_pct	Health problems (to number)	−0.01991998	0.107360779
ows	Anatomical sub-areas (to number)	−0.019283029	0.119073516
pts	Major anatomical areas (to number)	−0.018035752	0.144885187
fg3a_per_fga_pct	Major anatomical areas (to number)	−0.017405252	0.159466936
tov_pct	Decision (to number)	0.01722846	0.163748543
pts	Anatomical sub-areas (to number)	−0.016157157	0.191562974
g	Anatomical sub-areas (to number)	−0.015912807	0.198367601
trb_pct	Major anatomical areas (to number)	0.01583842	0.200473785
ast_pct	Major anatomical areas (to number)	−0.012802184	0.300778859
orb_pct	Anatomical sub-areas (to number)	0.012120232	0.327263754
fta_per_fga_pct	Anatomical sub-areas (to number)	−0.011610022	0.348039546
tov_pct	Health problems (to number)	−0.010742624	0.385237965
mpg	Anatomical sub-areas (to number)	−0.010702631	0.387009723
ws	Major anatomical areas (to number)	−0.01052102	0.395117611
net_rating	Major anatomical areas (to number)	−0.010057219	0.416284564
fta_per_fga_pct	Organ systems (to number)	0.010054429	0.416413865
ws_per_48	Major anatomical areas (to number)	0.009682129	0.433882255
reb	Major anatomical areas (to number)	0.008950359	0.469423575
player_height	Health problems (to number)	0.008913841	0.471238395
tov_pct	Major anatomical areas (to number)	−0.0071492	0.563376572
blk_pct	Major anatomical areas (to number)	−0.006774381	0.58400911
ts_pct	Major anatomical areas (to number)	0.006182223	0.617303226
mp	Major anatomical areas (to number)	−0.005917625	0.632444104
oreb_pct	Anatomical sub-areas (to number)	0.005870525	0.635155718
player_height	Major anatomical areas (to number)	−0.005809271	0.638689565
fta_per_fga_pct	Major anatomical areas (to number)	−0.005295322	0.668657325
dreb_pct	Major anatomical areas (to number)	−0.005114549	0.67932794
blk_pct	Health problems (to number)	0.004989319	0.686757938
tov_pct	Anatomical sub-areas (to number)	0.004912462	0.691333032
fg3a_per_fga_pct	Anatomical sub-areas (to number)	0.004499526	0.716103672
ast	Anatomical sub-areas (to number)	0.004483195	0.717089723
orb_pct	Organ systems (to number)	−0.004218233	0.733152137
ts_pct	Anatomical sub-areas (to number)	0.004096462	0.740573945
oreb_pct	Decision (to number)	0.00391622	0.751603804
ast	Major anatomical areas (to number)	−0.00345627	0.779976227
g	Major anatomical areas (to number)	0.003394718	0.783796398
ows	Major anatomical areas (to number)	−0.003162445	0.798258395
blk_pct	Decision (to number)	0.00314294	0.799476069
oreb_pct	Health problems (to number)	−0.002845255	0.818118933
drb_pct	Major anatomical areas (to number)	0.002448377	0.843132958
stl_pct	Major anatomical areas (to number)	0.002328131	0.850744305
obpm	Major anatomical areas (to number)	0.002219088	0.857658523
mpg	Major anatomical areas (to number)	0.001650596	0.893870996
bpm	Major anatomical areas (to number)	0.001384868	0.910878567
dbpm	Major anatomical areas (to number)	−0.001236421	0.920397998
oreb_pct	Organ systems (to number)	−0.000330038	0.978718989
per	Major anatomical areas (to number)	0.00018176	0.988279064
ast_pct	Anatomical sub-areas (to number)	−0.000128081	0.991740451
tov_pct	Organ systems (to number)	2.62675E−05	0.99830606

one step ahead in terms of strategy and tactics management in comparison with opponents choices [38].

Game results and championship titles are influenced by player and team management and use. Small details impact results considerably. Therefore, basketball and sports in general blend various uncertainties. Injuries is an important factor, because it is associated with playing and substitutions but also impact club finances ***(RQ3)***. Reducing this ambiguity can lead to process optimization and improved player and team efficiency [56].

6.2. Future Work

Accurate injury prediction is hard to be estimated. This uncertainty is useful to be examined in detail for future projections. It contains many parameters for research in terms of cost saving due to injuries, but more importantly in order to maintain a healthy and stable team roster [3]. Further work could seek an "injury factor" based on a structured form, with an ontology for injuries involving related hierarchies and taxonomy, that will help data entry, as well as injury identification and classification. Such and ontology will help data scientists and domain experts to identify injury patterns, injury reproducibility and reasoning or important insights into new injuries.

Physiological characteristics in terms of physical and training improvement can be assisted through video data analysis to assess the movements of athletes with pattern recognition. In addition, GPS sensors and wearables can be used in more specialized and focused way for monitoring and tracking activities, in comparison with other teams and players. Metrics can be improved with the association of advanced basketball analytics, sleep patterns, health analytics and other important biometric analytics. Therefore, injury limitation could optimize athlete performance and maximize player and team efficiency [23].

Future work can focus on the psychological, physical or additional injury available metrics for prediction. The technical team and coaches could be assisted to foresee situations that could affect team performance. There is a study on NBA player psychology and behaviour, which measured the athlete's social networks activity and correlated it with performance in future games [62]. Sentiment analysis can also be conducted on social network posts

Table A.4
Musculoskeletal injury analytics for NBA periods 2010–20.

	Total Counts Health Problems										% of Health Problems of Grant Total										Total Health Problems	Total % Health Problems of Grant Total
Type of Injury/Pathology	**10-11**	**11-12**	**12-13**	**13-14**	**14-15**	**15-16**	**16-17**	**17-18**	**18-19**	**19-20**	**10-11**	**11-12**	**12-13**	**13-14**	**14-15**	**15-16**	**16-17**	**17-18**	**18-19**	**19-20**		
Musculoskeletal Injuries	341	354	422	421	418	469	478	444	426	339	8.29	8.61	10.26	10.24	10.17	11.41	11.62	10.80	10.36	8.24	4112	100
Musculoskeletal system	341	354	422	421	418	469	478	444	426	339	8.29	8.61	10.26	10.24	10.17	11.41	11.62	10.80	10.36	8.24	4112	100
Lower extremity	237	241	290	290	273	324	350	317	315	234	5.76	5.86	7.05	7.05	6.64	7.88	8.51	7.71	7.66	5.69	2871	69.82
Ankle area	61	65	93	82	86	79	89	97	82	67	1.48	1.58	2.26	1.99	2.09	1.92	2.16	2.36	1.99	1.63	801	19.48
Calf area	11	12	14	18	15	14	19	10	9	11	0.27	0.29	0.34	0.44	0.36	0.34	0.46	0.24	0.22	0.27	133	3.23
Fibular area	1		1	1	1	1	3	2		1	0.02	0.00	0.02	0.02	0.02	0.02	0.07	0.05	0.00	0.02	11	0.27
Foot area	28	22	27	29	13	27	35	24	23	14	0.68	0.54	0.66	0.71	0.32	0.66	0.85	0.58	0.56	0.34	242	5.89
Heel area	7	13	8	9	16	14	14	14	18	13	0.17	0.32	0.19	0.22	0.39	0.34	0.34	0.34	0.44	0.32	126	3.06
Hip area	14	9	17	12	20	24	24	18	19	23	0.34	0.22	0.41	0.29	0.49	0.58	0.58	0.44	0.46	0.56	180	4.38
Knee area	77	69	89	97	71	105	99	95	92	55	1.87	1.68	2.16	2.36	1.73	2.55	2.41	2.31	2.24	1.34	849	20.65
Multiple anatomical areas	5	4	3	2	2		2	5			0.12	0.10	0.07	0.05	0.05	0.00	0.05	0.12	0.00	0.00	23	0.56
Shin area	6	5	8	6	12	10	8	10	9	6	0.15	0.12	0.19	0.15	0.29	0.24	0.19	0.24	0.22	0.15	80	1.95
Thigh area	22	33	28	28	29	41	46	34	49	40	0.54	0.80	0.68	0.68	0.71	1.00	1.12	0.83	1.19	0.97	350	8.51
Toes area	5	9	2	6	8	9	11	8	14	4	0.12	0.22	0.05	0.15	0.19	0.22	0.27	0.19	0.34	0.10	76	1.85
Multiple anatomical areas			6	4	2	2	6		1		0.00	0.00	0.15	0.10	0.05	0.05	0.15	0.00	0.02	0.00	21	0.51
Neck	2		8	6	3	2	2	4	7	6	0.05	0.00	0.19	0.15	0.07	0.05	0.05	0.10	0.17	0.15	40	0.97
Trunk	57	64	57	68	70	73	47	50	62	46	1.39	1.56	1.39	1.65	1.70	1.78	1.14	1.22	1.51	1.12	594	14.45
Abdominal area	15	24	21	14	17	17	17	12	15	16	0.36	0.58	0.51	0.34	0.41	0.41	0.41	0.29	0.36	0.39	168	4.09
Chest area	3	7	3	8	8	8	3	5	4	2	0.07	0.17	0.07	0.19	0.19	0.19	0.07	0.12	0.10	0.05	51	1.24
Pelvic area	2			2	1			2	3	3	0.05	0.00	0.00	0.05	0.02	0.00	0.00	0.05	0.07	0.07	13	0.32
Thoracolumbar area	37	33	33	44	44	48	27	31	40	25	0.90	0.80	0.80	1.07	1.07	1.17	0.66	0.75	0.97	0.61	362	8.80
Upper extremity	45	49	61	53	70	68	73	73	41	53	1.09	1.19	1.48	1.29	1.70	1.65	1.78	1.78	1.00	1.29	586	14.25
Elbow area	6	5	8	5	9	5	7	10	6	5	0.15	0.12	0.19	0.12	0.22	0.12	0.17	0.24	0.15	0.12	66	1.61
Hand, Thumb & Fingers area	15	14	25	19	25	19	26	19	19	22	0.36	0.34	0.61	0.46	0.61	0.46	0.63	0.46	0.46	0.54	203	4.94
Shoulder area	14	21	19	18	23	32	21	27	11	18	0.34	0.51	0.46	0.44	0.56	0.78	0.51	0.66	0.27	0.44	204	4.96
Upper arm and Forearm area	1		4	1		1	3	2		2	0.02	0.00	0.10	0.02	0.00	0.02	0.07	0.05	0.00	0.05	14	0.34
Wrist area	9	9	5	10	13	11	16	15	5	6	0.22	0.22	0.12	0.24	0.32	0.27	0.39	0.36	0.12	0.15	99	2.41
Total Unique reasons of Absence	**341**	**354**	**422**	**421**	**418**	**469**	**478**	**444**	**426**	**339**	**8.29**	**8.61**	**10.26**	**10.24**	**10.17**	**11.41**	**11.62**	**10.80**	**10.36**	**8.24**	**4112**	**100**

to understand the psychological status and behaviour of players. This work showed that injuries are events with high uncertainty and unlucky moments, but at the same time can be used in the future for pattern recognition [69].

In summary, an aggregation in the following segments could provide perspective and areas to focus on in order to maximize performance:

- Advanced predictive and prescriptive Sports Analytics based on statistics of box-score to analyse current and past performance.
- GPS and wearable sensor data analysis through sophisticated algorithms.
- Tactic analysis using state of the art algorithms for decision making.
- Biometric analysis for training and health consistency optimization.
- Health and Injury data analysis to avoid future mishaps and quick recovery.
- Video (SportVU) data analysis for tactics, scouting, player and team optimization.
- Sleep Analysis to recognize sleep patterns and optimize sleep quality.
- Trip duration, games and training workload and fatigue correlation.
- Social Network analysis to understand social behaviour and psychological status.
- Nutrition specialized approach based on the needs, potential and position of each player.
- Risk management and mitigation actions.
- Budget analysis, investments, and forecasting analysis.
- Leadership and clutch skills identification for teams and players.

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: The only known to us conflict of interest is with staff working at the International Hellenic University or the National and Kapodistrian University of Athens, Greece.

Appendix

The following tables detail injury categorization according to basketball analytics during 2010–20.

See Fig. A.1 and Tables A.1–A.3.

References

[1] R.B. Minton, Sports Math: An Introductory Course in the Mathematics of Sports Science and Sports Analytics, Chapman and Hall/CRC, 2017.
[2] S. Kaplan, et al., The economic impact of NBA superstars: evidence from missed games using ticket microdata from a secondary marketplace, in: MIT Sloan Sport. Anal. Conf. Conf., 2019, pp. 1–29.
[3] V. Sarlis, C. Tjortjis, Sports analytics — Evaluation of basketball players and team performance, Inf. Syst. 93 (2020).
[4] H. Talukder, et al., Preventing in-game injuries for nba players, Sport. Anal. Conf. 2015 (2016) 1–13.
[5] K. Lankhorst, et al., Sports participation related to injuries and illnesses among ambulatory youth with chronic diseases: Results of the health in adapted youth sports study, BMC Sports Sci. Med. Rehabil. 11 (1) (2019) 1–12.
[6] N. Maffulli, U.G. Longo, N. Gougoulias, D. Caine, V. Denaro, Sport injuries: A review of outcomes, Br. Med. Bull. 97 (1) (2011) 47–80.
[7] E. Cumps, E. Verhagen, R. Meeusen, Prospective epidemiological study of basketball injuries during one competitive season: Ankle sprains and overuse knee injuries, J. Sport. Sci. Med. 6 (2) (2007) 204–211.
[8] K.D. Barber Foss, G.D. Myer, T.E. Hewett, Epidemiology of basketball, soccer, and volleyball injuries in middle-school female athletes, Phys. Sportsmed. 42 (2) (2014) 146–153.
[9] M.K. Drew, J. Cook, C.F. Finch, Sports-related workload and injury risk: Simply knowing the risks will not prevent injuries: Narrative review, Br. J. Sports Med. 50 (21) (2016) 1306–1308.
[10] S. Habelt, C.C. Hasler, K. Steinbrück, M. Majewski, Sport injuries in adolescents, Orthop. Rev. (Pavia) 3 (2) (2011) 18.
[11] L. Abernethy, C. Bleakley, Strategies to prevent injury in adolescent sport: A systematic review, Br. J. Sports Med. 41 (10) (2007) 627–638.
[12] C.A. Emery, Injury prevention and future research, Med. Sport Sci. 48 (2005) 179–200.
[13] K. Herman, C. Barton, P. Malliaras, D. Morrissey, The effectiveness of neuromuscular warm-up strategies, that require no additional equipment, for preventing lower limb injuries during sports participation: a systematic review, BMC Med. 10 (2012) 1–12.

[14] M. Klügl, et al., The prevention of sport injury: An analysis of 12 000 published manuscripts, Clin. J. Sport Med. 20 (6) (2010) 407–412.
[15] N. Maffulli, U.G. Longo, F. Spiezia, V. Denaro, Sports injuries in young athletes: Long-term outcome and prevention strategies, Phys. Sportsmed. 38 (2) (2010) 29–34.
[16] B.H. Patel, et al., Adductor injuries in the National Basketball Association: an analysis of return to play and player performance from 2010 to 2019, Phys. Sportsmed. 00 (00) (2020) 1–8.
[17] C. Starkey, Injuries and illnesses in the National Basketball Association: A 10-year perspective, J. Athl. Train. 35 (2) (2000) 161–167.
[18] M.C. Drakos, B. Domb, C. Starkey, L. Callahan, A.A. Allen, Injury in the National Basketball Association: A 17-year overview, Sports Health 2 (4) (2010) 284–290.
[19] T.J. Jackson, C. Starkey, D. McElhiney, B.G. Domb, Epidemiology of hip injuries in the national basketball association: A 24-year overview, Orthop. J. Sport. Med. 1 (3) (2013).
[20] E. Malinowski, Betaball: How Silicon Valley and Science Built One of the Greatest Basketball Teams in History, Atria Books, 2017.
[21] R.S. Sikka, M. Baer, A. Raja, M. Stuart, M. Tompkins, Analytics in sports medicine: Implications and responsibilities that accompany the era of big data, J. Bone Jt. Surg. Am. 101 (3) (2019) 276–283.
[22] M. Cohen, M. Sloane, Predictive modeling and statistical analysis in sports, 2015.
[23] T.H. Davenport, Analytics in sports: The new science of winning, Int. Inst. Anal. 2 (February) (2014) 1–28.
[24] M. Kontaki, Y. Manolopoulos, A. Gounaris, A.N. Papadopoulos, K. Tsichlas, Efficient and flexible algorithms for monitoring distance-based outliers over data streams, Inf. Syst. (2015) 1–17.
[25] L. Martin, Sports Performance Measurement and Analytics, Pearson Education LTD, 2016.
[26] M. van Bommel, L. Bornn, Adjusting for scorekeeper bias in NBA box scores, Data Min. Knowl. Discov. 31 (6) (2017) 1622–1642.
[27] B. Cole, A.J.H. Arundale, J. Bytomski, A. Amendola, Basketball Sports Medicine and Science, Springer, 2020.
[28] W. Tichy, Changing the Game: Dr. Dave Schrader on Sports Analytics, No. May, Ubiquity, 2016, pp. 1–10.
[29] B.C. Alamar, Sports Analytics: A Guide for Coaches, Managers, and Other Decision Makers, Columbia University Press, 2013.
[30] G. Vinue, I. Epifanio, Archetypoid analysis for sports analytics, 31 (6), 2017.
[31] E. Morgulev, O.H. Azar, R. Lidor, Sports analytics and the big-data era, Int. J. Data Sci. Anal. 5 (4) (2018) 213–222.
[32] I. Bhandari, E. Colet, J. Parket, Z. Pines, R. Pratap, K. Ramanujam, Advanced scout: Data mining and knowledge discovery in NBA data, Data Min. Knowl. Discov. 125 (1997) 121–125.
[33] J. Fernández, From Training to Match Performance: An Exploratory and Predictive Analysis on F.C. Barcelona, Facultat d'Informàtica de Barcelona, 2016.
[34] M. Dunham, Data Mining - Introductory and Advanced Topics, Prentice Hall, 2003.
[35] A. Senderovich, A. Shleyfman, M. Weidlich, A. Gal, To aggregate or to eliminate, ? Optimal model simplification for improved process performance prediction, Inf. Syst. (2018) 1–16.
[36] I. Katakis, et al., Mining urban data (Part B), Inf. Syst. 57 (2016) 75–76.
[37] F. Bianchi, T. Facchinetti, P. Zuccolotto, Role revolution, : towards a new meaning of positions in basketball, Electron. J. Appl. Stat. Anal. 10 (03) (2017) 712–734.
[38] R.P. Schumaker, O.K. Solieman, H. Chen, Sports knowledge management and data mining, in: Annual Review of Information Science and Technology Develop, Vol. 44, Springer, 2010, pp. 115–158.
[39] LeBron James Diet, workouts, treatment cost him $1.5 million in body care - Business Insider, 2020, [Online]. Available: https://www.businessinsider.com/how-lebron-james-spends-money-body-care-2018-7. (Accessed 12 December 2020).
[40] J. Orchard, J. Hayes, Using the world wide web to conduct epidemiological research, : an example using the national basketball association, Int. Sport. J. 2 (2) (2001) 1–15.
[41] P.C. Yeh, C. Starkey, S. Lombardo, G. Vitti, F.D. Kharrazi, Epidemiology of isolated meniscal injury and its effect on performance in athletes from the National Basketball Association, Am. J. Sports Med. 40 (3) (2012) 589–594.
[42] P.P. Salehi, A. Heiser, S.J. Torabi, B. Azizzadeh, J. Lee, Y.H. Lee, Facial fractures and the national basketball association: Epidemiology and outcomes, Laryngoscope 130 (12) (2020) E824–E832.
[43] M.M. Herzog, et al., Ankle sprains in the national basketball association, 2013–2014 through 2016–2017, Am. J. Sports Med. 47 (11) (2019) 2651–2658.
[44] NBA.com. [Online]. Available: https://stats.nba.com.
[45] C.D. Mack, et al., The establishment and refinement of the national Basketball Association player injury and illness database, J. Athl. Train. 54 (5) (2019) 466–471.
[46] H. Talukder, et al., Preventing in-game injuries for NBA players Paper ID : 1590, in: MIT Sloan Sport. Anal. Conf., Vol. 2015, 2016, pp. 1–13.
[47] B.R. Humphreys, C. Johnson, The effect of superstars on game attendance: Evidence from the NBA, J. Sports Econom. 21 (2) (2020) 152–175.
[48] B. Gerrard, Moneyball and the role of sports analytics: A decision-theoretic perspective, in: North American Society for Sport Management Conference, No. Nassm, NASSM 2016, 2016, pp. 2010–2012.
[49] P.O.S. Vaz de Melo, V.A.F. Almeida, A.A.F. Loureiro, C. Faloutsos, Forecasting in the NBA and other team sports: Network effects in action, ACM Trans. Knowl. Discov. Data 6 (3) (2012) 1–27.
[50] L. Pham, S. Anand, J. Piette, Evaluating basketball player performance via statistical network modeling, in: MIT Sloan Sport. Anal. Conf. No. June, 2011.
[51] G. Vinue, I. Epifanio, Forecasting basketball players' performance using sparse functional data, Stat. Anal. Data Min. (1) (2018) 1–26.
[52] A.M. Franks, A. D'Amour, D. Cervone, L. Bornn, Meta-analytics: Tools for understanding the statistical properties of sports metrics, J. Quant. Anal. Sport. 12 (4) (2016) 151–165.
[53] NBA Basketball reference, 2019, [Online]. Available: https://www.basketball-reference.com/. (Accessed 20 August 2019).
[54] ESPN NBA Stats, 2019, [Online]. Available: https://www.espn.com/nba/stats. (Accessed 20 August 2019).
[55] V. Madhavan, Predicting NBA game outcomes with hidden Markov models, 2017.
[56] C. Cao, Sports data mining technology used in basketball outcome prediction, 2012.
[57] A.M. Franks, A. D'Amour, D. Cervone, L. Bornn, Meta-analytics: Tools for understanding the statistical properties of sports metrics, J. Quant. Anal. Sport. 12 (4) (2016) 151–165.
[58] Y. Yang, Predicting regular season results of NBA teams based on regression analysis of common basketball statistics, 2015.
[59] Y. Li, L. Wang, F. Li, A Data-Driven Prediction Approach for Sports Team Performance and Its Application to National Basketball Association, No. September, Omega - Elsevier, 2019, 102123.
[60] J.G. Claudino, D. de O. Capanema, T.V. de Souza, J.C. Serrão, A.C. Machado Pereira, G.P. Nassis, Current approaches to the use of artificial intelligence for injury risk assessment and performance prediction in team sports: a systematic review, Sport. Med. - Open 5 (1) (2019).
[61] O. Levchenko, et al., BestNeighbor: efficient evaluation of kNN queries on large time series databases, Knowl. Inf. Syst. (2020).
[62] P.O. Donoghue, Research methods for sports performance analysis, 2009.
[63] K. Pelechrinis, E. Papalexakis, Athlytics, in: Proc. Elev. ACM Int. Conf. Web Search Data Min., WSDM '18, 2018, pp. 787–788.
[64] T. Chartier, Coachable business results, J. Corp. Account. Financ. 27 (1) (2015) 83–85.
[65] N. Silver, CARMELO NBA Player projections, 2015.
[66] A.B. Downey, Think Stats: Probability and Statistics for Programmers, Vol. 70, No. 2, 2011.
[67] G. Vinué, I. Epifanio, Forecasting basketball players' performance using sparse functional data*, Stat. Anal. Data Min. 12 (6) (2019) 534–547.
[68] Statathlon.com, 2019, [Online]. Available: https://statathlon.com/.(Accessed 20 August 2019).
[69] N.F.N. Bittencourt, W.H. Meeuwisse, L.D. Mendonça, J.M. Ocarino, S.T. Fonseca, Complex systems approach for sports injuries: moving from risk factor identification to injury pattern recognition—narrative review and new concept, Br. J. Sports Med. (2016) 1309–1314.