

A review of sports data analytics field: A bibliometric and network analysis of the articles published from 1997 to 2020

Mojtaba Safari¹, AmirHossein Masoumi¹, Rouzbeh Ghousi^{1*}, Hassan Sadeghi Naeini²

¹ *School of Industrial engineering, Iran University of Science & Technology, Tehran, Iran*

² *School of Architecture and Environmental Design, Iran University of Science & Technology, Tehran, Iran*

safari_mojtaba@ind.iust.ac.ir, amir.hossein.masoumi@gmail.com, ghousi@iust.ac.ir, naeini@iust.ac.ir

Abstract

Data analysis in competitive sports has increased significantly in recent years, and a significant number of studies have been done during the last decades. In the sports Data analysis field, bibliometric analysis and maps have not yet been used to analyze the production and visualize evolution and trends. Therefore, the primary purpose of this article is to review data science analysis in sports activities with network embedding-based visualization on a large-scale dataset. 805 articles were published between 1997 and 2020 and written by 3141 different authors from 1181 institutions, and 60 different countries were extracted from WOS by using R, Cite Space, and VOS viewer. Articles, journals, authors, countries, and universities that have played a significant role in developing this field are identified. Following that, meaningful knowledge of the communication networks among articles, authors, and keywords are illustrated by scientific paper mining. Moreover, articles have been divided into six groups based on the subject and methodology, which provide a comprehensive sight for researchers in this emerging field of sports.

Keywords: Sport, network science, bibliometric analysis, Web of Science, document co-citation analysis

1-Introduction

Data mining is the process of searching and analyzing to discover unknown patterns from data sources so that it has been known as an interdisciplinary science. Data mining can play a crucial role in businesses, medicine, biology, DNA analysis, fraud detection, long-distance communications, sports and entertainment, library, and information science (Araghi Niknam et al., 2021; Jiwei & Kamber, 2012; Masoumi et al., 2021). Before the advent of data mining, sports organizations relied entirely on human expertise and believed that coaches and executives could transform the gathered data into knowledge through their own experience. Nevertheless, the volume of data has increased daily, and sports executives have faced a considerable mass of data. Therefore, adopting a scientific approach for analyzing the data has risen considerably (Schumaker et al., 2010a). The relationship between the data

*Corresponding author

ISSN: 1735-8272, Copyright c 2022 JISE. All rights reserved

and sport can be classified into five levels:

1. There is no relationship,
2. Coaches and sports executives make decisions based on experiences and instinct,
3. Coaches and sports executives make decisions based on historical data,
4. Using statistical results in decision making,
5. Using data mining results in decision-making.

Most sports executives utilize the third and fourth methods and have achieved good results. Only a few choose the last one (Schumaker et al., 2010b). Recently, the sports industry has faced a considerable amount of data for each player, team, and match. The analysis system in sports is mainly implemented to help coaches and sports executives to predict matches, evaluate the teams, players, and coaches' performance, predict players' injuries, discover top talents, and evaluate the applied strategies (Haghighat et al., 2013; Schumaker et al., 2010b). Many kinds of research show the direct relation between data mining and the sports industry; for instance, the most common problems mentioned in the literature present a model for predicting sports matches. It can be considered one of the most rudimentary applications of data mining in the sports industry. Although many unknown factors predict the outcomes, this field has not developed enough (Cao, 2012). In predicting the competition results, using data mining, Haghighat et al. examined the related articles between 2003 and 2012 (Haghighat et al., 2013). We added to their studies and examined the articles up to 2020 (table 1). In this table, we have reviewed the sports matches' outcomes prediction systems using data mining techniques by explaining these systems, introducing their advantages and disadvantages, and presenting the results achieved.

Table 1. Classification of articles that have predicted the results of sports competitions by using data mining technics between 2003 and 2020

Authors (year)	Data Set	Technique	Accuracy	Advantages	Disadvantages
(Kahn, 2003)	The first 15 weeks of National Football League, 2003	Artificial Neural Network	75%	* Compare the designed system with previous ones * Relatively high forecast system	* Small data set
(McCabe & Trevathan, 2008)	Australian Football League, National Rugby League	Artificial Neural Network	67.5%	* Participated in the Top Tipper (international betting competitions) in 2006-07 season	* Low prediction accuracy
(Zdravevski & Kulakov, 2009)	Asociacion de Clubes de Baloncesto (ACB) Basketball League statistics from 2008-2009 season	The decision tree, Bayesian network, Artificial Neural Network(ANN)	Best accuracy %72,80	* Collecting data from NBA league, designing the system as a module	* Low prediction accuracy * Manual feature selection * No comparison with previous studies results
(Davoodi et al., 2010)	A hundred AQUEDUCT Race Track games during January 1st-29th, Ney York, USA	Clas-Fuzzy-Chi RW	71. 5%	* Comparing different fuzzy algorithms * Calculate the standard deviation for algorithms * Considering the comprehensibility of the commands of fuzzy algorithms	* Use a small local data set * Low prediction accuracy

Table 1. Continued

Disadvantages	Advantages	Accuracy	Technique	Data Set	Authors (year)
(Ivanković et al., 2010)	Basketball League of Serbia B, 2005-2006 season until 2009-2010 season	Artificial Neural Network	80.96%	<ul style="list-style-type: none"> * Analyze the effect of shooting from different parts of the Court * Evaluate the future opponent by specifying the effect of the parameter in winning the game * high prediction accuracy 	* The data is not available to the public so the accuracy of this system is not comparable to other works
(Cao, 2012)	NBA League, 2005-2006 season until 2010-2011 season	classification	69.67%	<ul style="list-style-type: none"> * Use valid datasets * Compare your work with similar work 	* Low prediction accuracy
		Naïve Bayes	66.25%		
		SVM	67.70%		
		Artificial Neural Network	68.01%		
(Shi et al., 2013).	National College Athletics Association Basketball (NCAAB) matches	Naïve Bayes	74%	<ul style="list-style-type: none"> * Using NCAAB's data * Compare model accuracy with similar tasks 	* Low prediction accuracy
(Arabzad et al., 2014)	the Iran Pro League (IPL) 2013-2014 football matches	Artificial Neural Network	none	<ul style="list-style-type: none"> * Using IPL's big data * High accuracy in data gathering * Using previous studies results 	* No comparison with previous studies results
(Tax et al., 2015)	The Dutch Football Competition	classification algorithms	Best accuracy 85%	<ul style="list-style-type: none"> * Using Eredivisie's big data * High accuracy in input data * Using dimensionality reduction algorithms * High prediction accuracy 	* No comparison with previous studies results
(Wiseman, 2016)	2016 PGA Tour	linear regression	36%	<ul style="list-style-type: none"> * Using PGA Tour's data * Introducing a web Application * Using dimensionality reduction algorithms * Using different data mining algorithms and comparing their results 	* Low prediction accuracy
		Neural Network Regression	57.2%		
		Bayesian Linear Regression	58%		
		Decision Forest Regression	57.7%		
		Boosted Decision Tree Regression	42%		

Table 1. Continued

Disadvantages	Advantages	Accuracy		Technique	Data Set	Authors (year)
(Akarçeşme & Medicine, 2017)	Turkish Men's Volleyball League (TMVL) and the Turkish Women's Volleyball League (TWVL)	logistic regression	TMVL	83.45%	* Considering gender * High prediction accuracy	* Small data set
			TWVL	78.23		
(Tümer & Koçer, 2017)	male professional volleyball league in Turkey	Artificial Neural Network		98%	* High prediction accuracy *considering the sport of volleyball in a professional league	--
(Sujatha et al., 2018)	the Bundes Liga during the period 2005 to 2011	logistic regression	TMVL	83.45%	* Considering gender * High prediction accuracy	* Small data set
			TWVL	78.23		
(Thabtah et al., 2019)	National Basketball Association (NBA) of the United State	Naïve Bayes		76%	* High prediction accuracy * Using NBA's data *prediction NBA's games *Using feature selection for input data	* No comparison with previous studies results
		Artificial Neural Networks (ANNs)		83%		
		LMT (Logistical model OGISTIC MODEL TREE)		82%		
(Saavedra et al., 2020)	the 18th FINA World Championships (men)	classification-tree model		83.9%	* High prediction accuracy *to compare the match outcome after new law *Appropriate feature extraction	---

The neural network algorithm has been used more than other algorithms and usually has higher accuracy. Meanwhile, studies that used the feature selection process in the pre-processing step were more accurate than other studies. Some studies predicted injury time and time to return (Kerr et al., 2016; Lefevre et al., 2017). In another study about sports injuries, researchers collected data through YouTube videos to analyze the rate of body movement difficulties to prevent possible injuries in gymnastics athletes. They hypothesized that young rhythmic gymnasts with a high preference for hyper-extendable body movement difficulties have a high Potential for column vertebrae injuries (Agopyan, 2020). Some studies have also investigated the number of injuries and the type of them. In the sports psychology subject, Holt et al. discovered the pattern in developing sports talent and becoming a professional player (Holt & Dunn, 2004). In contrast, Lemyer et al. has developed coaching talent with a descriptive statistics approach (Lemyre et al., 2007).

Sports data analysis is a relatively recent type of research, which has grown significantly in recent times. However, bibliometric analysis has not yet been used to analyze the productions and visualize evolution and trends in this field. The presented article is a review study and analyzed the documents from different aspects to determine the literature's undiscovered parts. R, Cite space, and VOSviewer software, also known as the bibliometrix package, were used to analyze the extracted data from Web of Science. Then, the articles were categorized based on their topics. Thus, this study will allow researchers

to effectively understand the sports data analysis area and identify relevant topics, the most popular authors, articles, keywords, and their relationships. It will also allow researchers to identify current trends and future research areas in this field. After explaining the research goals, a description of data gathering is presented in the 2-1 section. The approach for reviewing the article is described in the methodology . The results section is consists of 3-1 and 3-6, respectively. The primary outcomes of the study are provided in the conclusion section.

2-Material and methods

2-1-Data collection

Searching for articles in databases, such as Web of Science and Scopus, is performed through keywords; therefore, choosing proper keywords to access articles related to a specific field and its results are essential. In order to find important keywords in this field, a general view of previous research is achieved by investigating the review articles, so the keywords used in the literature were found (Bonidia et al., 2018; Kos et al., 2018). A list of keywords, including "data mining," "sport," "anthropometric," "talent," "body fat," and "athletes," was obtained as the essential keywords through a comprehensive review of previous literature. The Web of Science database, the most credible database in scientometrics in organizations (Dzikowski, 2018; Fabregat-Aibar et al., 2019), was chosen to extract articles.

With the help of conjunctions "AND" and "OR" and the extracted list of keywords, we expanded the search range until we obtained 805 articles from 1997 to 2020, with the following combination. More precise information regarding the search results is given in (Fig).

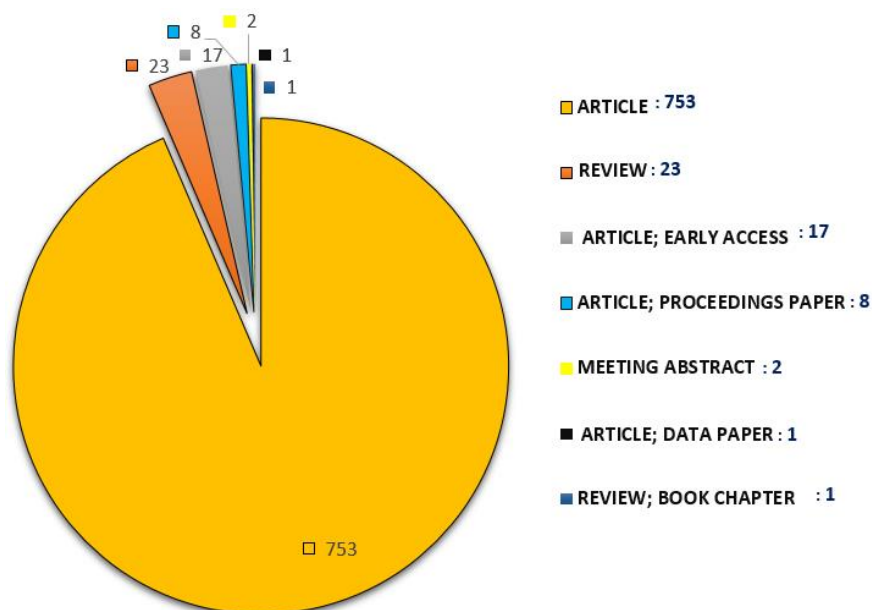


Fig 1. Details of Search Results

The mentioned combination certainly covers the articles. In the next stage, essential parts of the extracted articles, including title and abstract, were examined. Unrelated articles were removed from the articles list to avoid bias in results. Fifteen articles were removed and finally, 790 articles were reviewed with the help of scientometrics. Some steps must be introduced to choose the articles for analysis Fig shows the process used in these articles.

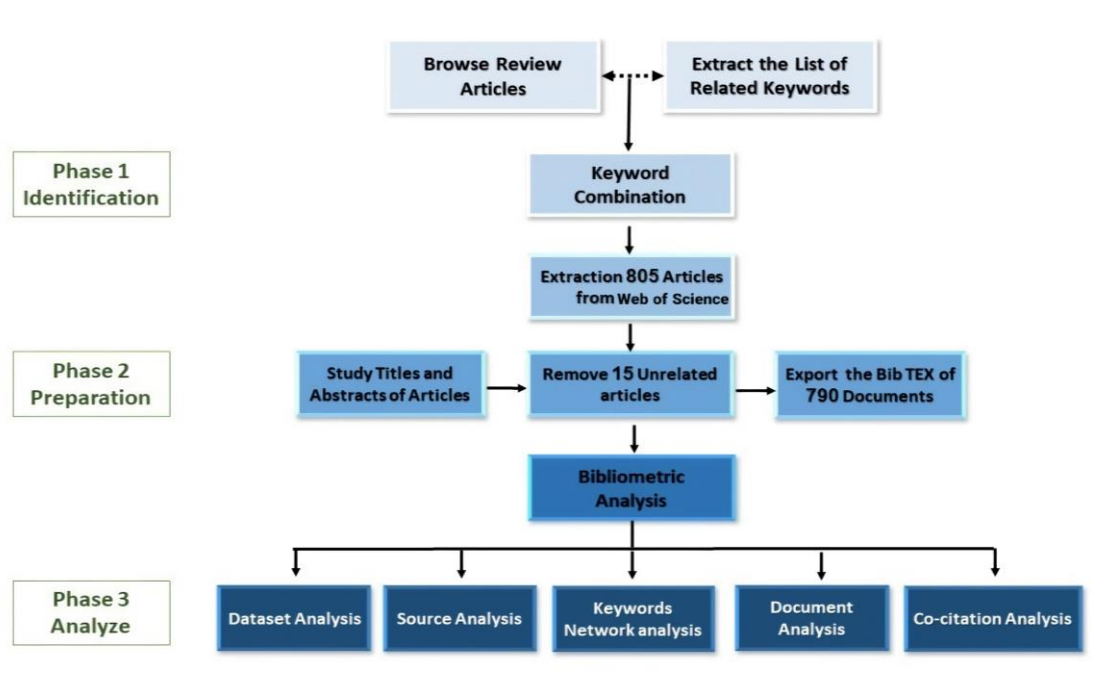


Fig 2. Research steps

2-2-Methodology

Bibliometrics is introduced as a tool for measurement and analysis. It is a statistical method that has a wide range of applications (Norton, 2000). Nowadays, the bibliometric approach is used to understand previous studies better and even predict the events related to the topic that may happen in the future (Morris et al., 2002). The primary purpose of the quantitative and scientometric study is to find the relational patterns in scientific societies, determine the level of cooperation among authors from different fields, determine outstanding publications, introduce top authors and frequently used articles, show the scientific structure of specific fields, and scientifically evaluate countries, universities, and institutes (McCulloh et al., 2013; Niu et al., 2014). In order to recognize the dependent patterns among authors, articles, and keywords, bibliometric techniques are divided into four categories; citation analysis, co-citation analysis, keyword analysis, and co-authorship analysis (Zupic & Čater, 2015).

Co-citation is defined as the frequency with which two documents are cited together by other documents (Fabregat-Aibar et al., 2019). The purpose of co-citation analysis is to find a mental structure of the research topic and states that another article consecutively cites two articles. It shows the two articles' mutuality in the topic. The amount of mutuality depends on how many times the co-citation has been repeated. Therefore, the articles in one field can be clustered. In addition to article analysis, co-citation analysis can be used in other units, such as the author's journals (Small, 1973). Some researchers consider co-citation analysis the most critical tool in the bibliometric study. They have used this method to review the previous literature. (Bartolini et al., 2019; Fabregat-Aibar et al., 2019; Hjørland & Management, 2013; Morris et al., 2002; Small, 1973; Zupic & Čater, 2015). The keyword analysis is used to find the key topics in the intended field; also, it helps to find new and emerging fields since it depends on the article's concept. The citation cannot be a credible index on its own because old articles are usually presented as outputs. To find the most compelling articles, "Eigenvector centrality" and "citation per year" are used (McCulloh et al., 2013).

Our methodology consists of four main sections. The first part collects the metadata, including sentence, year of publication, keywords, title, authors' names, and institute/university. The second part analyzes the keywords in which keywords represent the main concepts mentioned in the article's main body. The third part is citation analysis. The article's main purpose is to present a comprehensive evaluation; therefore, researchers can present useful information by choosing one of these approaches. In our research, all three of these instances are evaluated. The last part is co-citation analysis. In this

part, we tried to cluster articles based on topic or applied methods and generally review each classification.

3-Results

3-1-Basic bibliometric indicators

In this section, metadata, including authors' names, titles, keywords, author's institution, and publication year, is selected as input data. This section is divided into four parts: Dataset analysis, authors and country of productions analysis, document analysis, and keywordsnetwork analysis.

3-2-Dataset analysis

Figure 3 depicts the number of articles published by year. As a result, data mining in sport has absorbed researchers' attention in the last two decades, with an annual growth rate of 13.34%. It is worth mentioning that the years 2019, 2020, and 2018 have the most published articles by 102, 89, and 83, respectively.

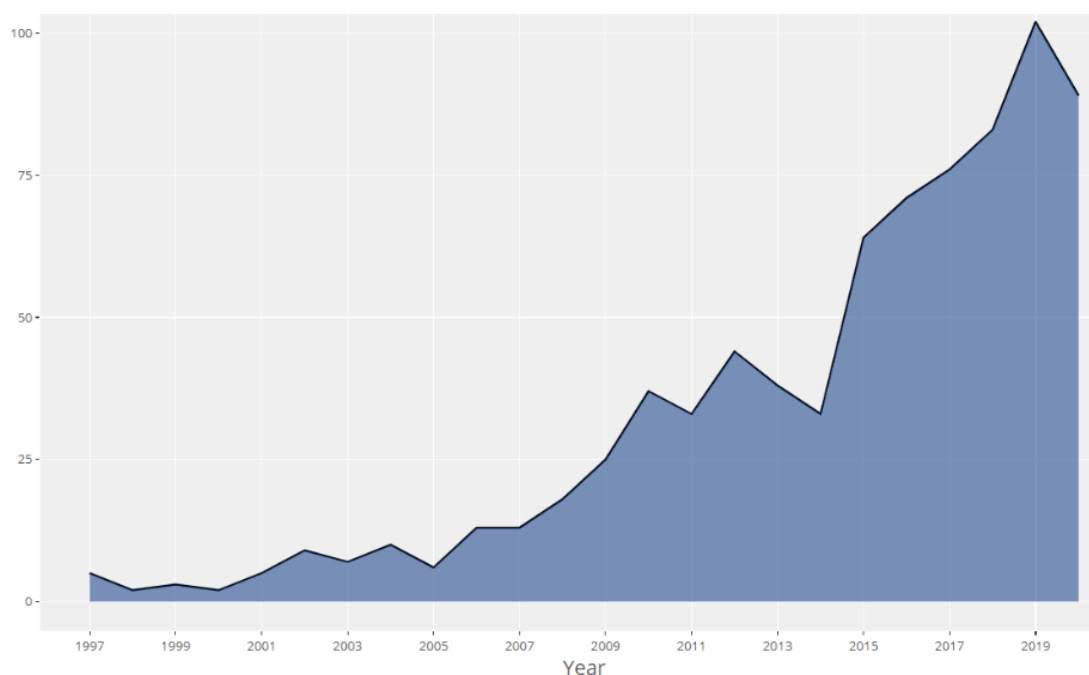


Fig 3. Published papers

As a result of the Bibliometric analysis, the tree-field plot extracted by R software is presented in figure 4. The graph illustrates the most used subjects by authors and the journals by which most articles have been published. For example, Mr. Bargary n, Draper c, Colby s, Warmenhoven j have focused on analyzing biomechanical data in the sports industry in recent years. These articles have often been published in the “Scandinavian Journal of Medicine & Science in Sports” and “Journal of Science and Medicine in Sport.” As another result, most of the articles about coaching and sports data mining published in the “Journal of Physical Education and Sport Pedagogy”, so that Camir.m, Woods.ct, and Robert.s have published the most articles in this category. In addition, Fig illustrates that there are many sources for articles in football data analysis Because there is a robust connection between the keyword soccer and journals.

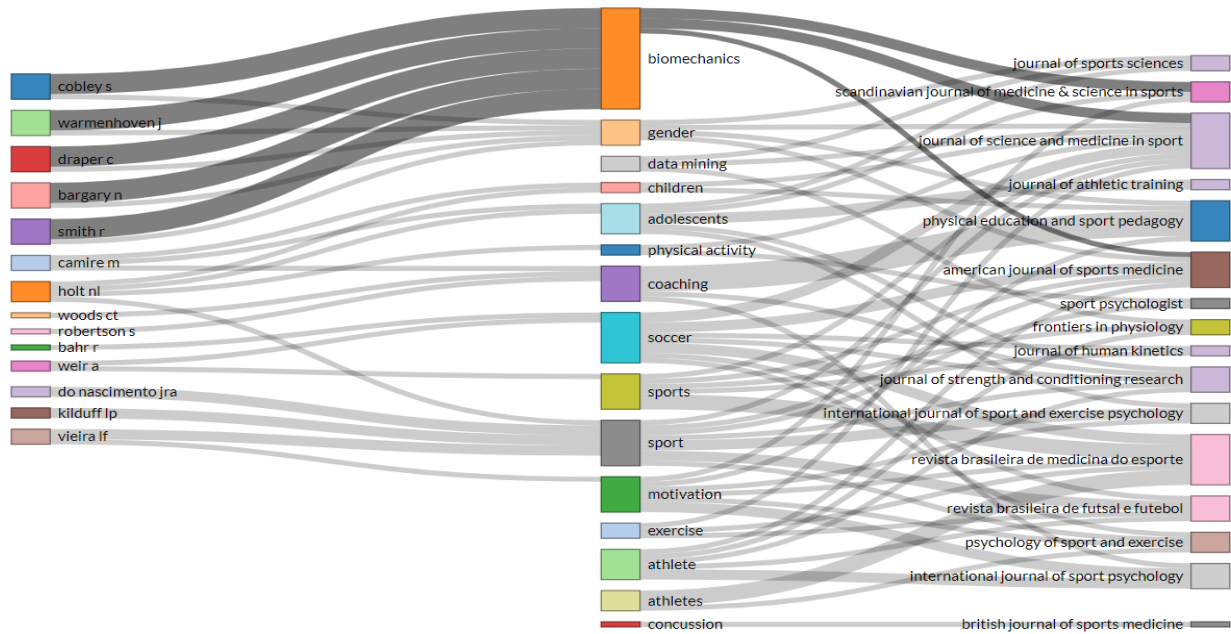


Fig 4. Tree -field plot

3-3-Source analysis and H-INDEX

The top 10 journals are listed in Table 2. The H-index has been adopted as a valuable tool for reviewing publication statistics in journals. It has been introduced as a numerical method for evaluating the overall effective performance of researchers, journals, countries, and institutions since 2005 (Hirsch, 2005).

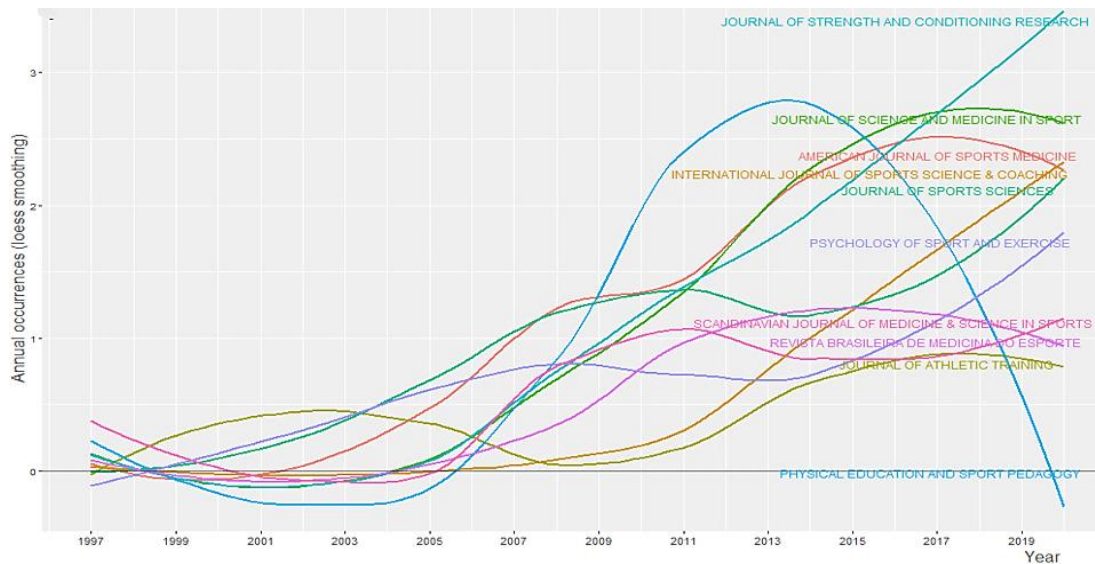
Table 2. Top sources

RANK	SOURCE	N.D	H-INDEX	TOTAL CITATION	Start year
1	AMERICAN JOURNAL OF SPORTS MEDICINE	29	19	1285	2006
2	JOURNAL OF SPORTS SCIENCES	22	14	732	2002
3	JOURNAL OF SCIENCE AND MEDICINE IN SPORT	27	13	707	2088
4	PHYSICAL EDUCATION AND SPORT PEDAGOGY	23	13	508	2010
5	PSYCHOLOGY OF SPORT AND EXERCISE	16	10	539	2003
6	BRITISH JOURNAL OF SPORTS MEDICINE	10	9	533	2002
7	JOURNAL OF STRENGTH AND CONDITIONING RESEARCH	27	8	363	2008

Table 2. Continued

RANK	SOURCE	N.D	H-INDEX	TOTAL CITATION	Start year
8	SCANDINAVIAN JOURNAL OF MEDICINE & SCIENCE IN SPORTS	13	8	188	1999
9	JOURNAL OF APPLIED SPORT PSYCHOLOGY	10	8	324	2004
10	JOURNAL OF ATHLETIC TRAINING	11	7	678	2001

As can be seen, the American Journal of Sports Science and Medicine and the Journal of Sports Sciences are the leaders in terms of number and H-index, respectively. Figure 5 represents the trend of articles over the last two decades. At first glance, it is evident that all journals have an upward trend, and the Journal of Strength & Conditioning Research has received more attention than other journals. In contrast, the Physical Education and Sport Pedagogy have experienced a sharp decline since 2013, so that in 2020 the number of Published articles has reached zero.

**Fig 5.** Source growth

3-4- Authors, country production

The top ten authors are introduced in figure 6. Three of them have published seven articles, three of them have each published six articles, and the other three have published five articles, while Cobley has published ten documents.

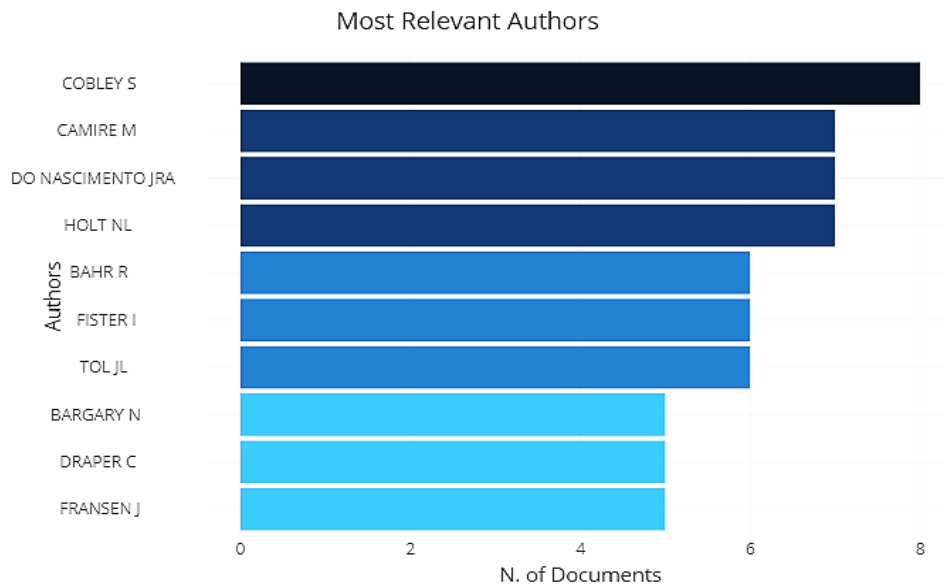


Fig 6. Top authors

As shown in

Fig figure 7 presents the authors in "sports data mining" have the highest cooperation level. The blue cluster has the highest cooperation level with five articles; the red and green have three and two articles, respectively. The more a box is bigger in each class, depicts more importance for each author.

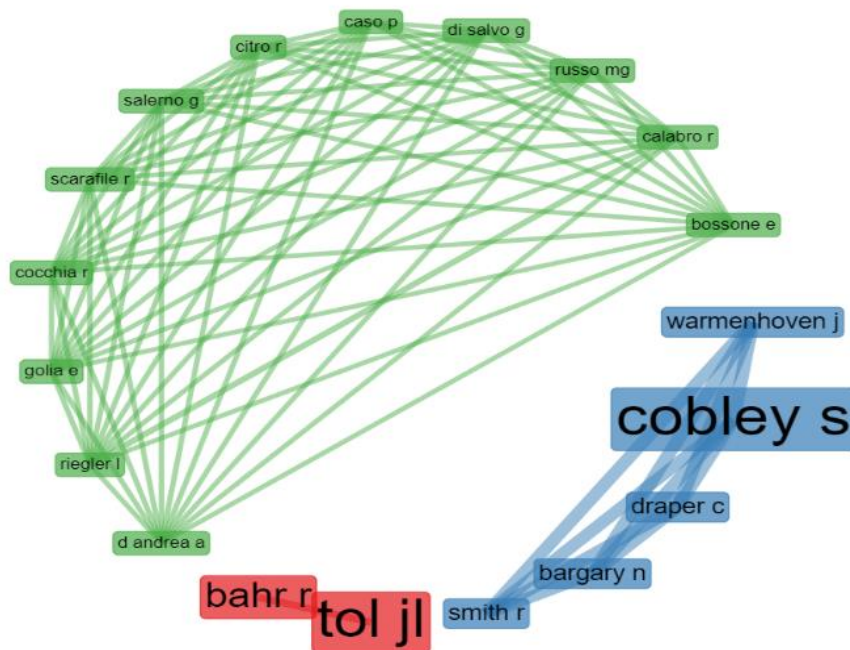


Fig 7. Co-authors

The important thing about the blue cluster's authors' cooperation was two significant similarities in 5 mutual articles in that the FDA² method was used in canoeing. They have focused on analyzing biomechanical and bivariate data with the help of analyzing techniques. They also conducted another research to evaluate whether the gender and the boat's orientation affect the angle-force curvature on 40 national athletes in canoeing. Afterward, by applying FPCA³ on the angle-force data, the curvature's primary state in the curves was recognized. As a result, gender is considered a dependent variable so that sportswomen have less paddling strength. In contrast, the boat's orientation is introduced as an independent variable. (John Warmenhoven, Cobley, et al., 2018; John Warmenhoven et al., 2019; J Warmenhoven et al., 2017; John Warmenhoven, Harrison, et al., 2018; John Warmenhoven, Smith, et al., 2018).

As a result of the article, applying FPCA and ACP⁴ as two FAD techniques on the propulsive pin force data in solo canoeing is asymmetrical paddling due to the sportsman's performance, a study is conducted.

It is shown in their article that the BFPCA technique has a high potential in analyzing sports biomechanical data. In another research by Warmenhoven, the force-angle data was collected from a group of elite athletes. Through BFPCA-based analysis, athletes' performances were inferred from their force-pressure curves (J Warmenhoven et al., 2017).

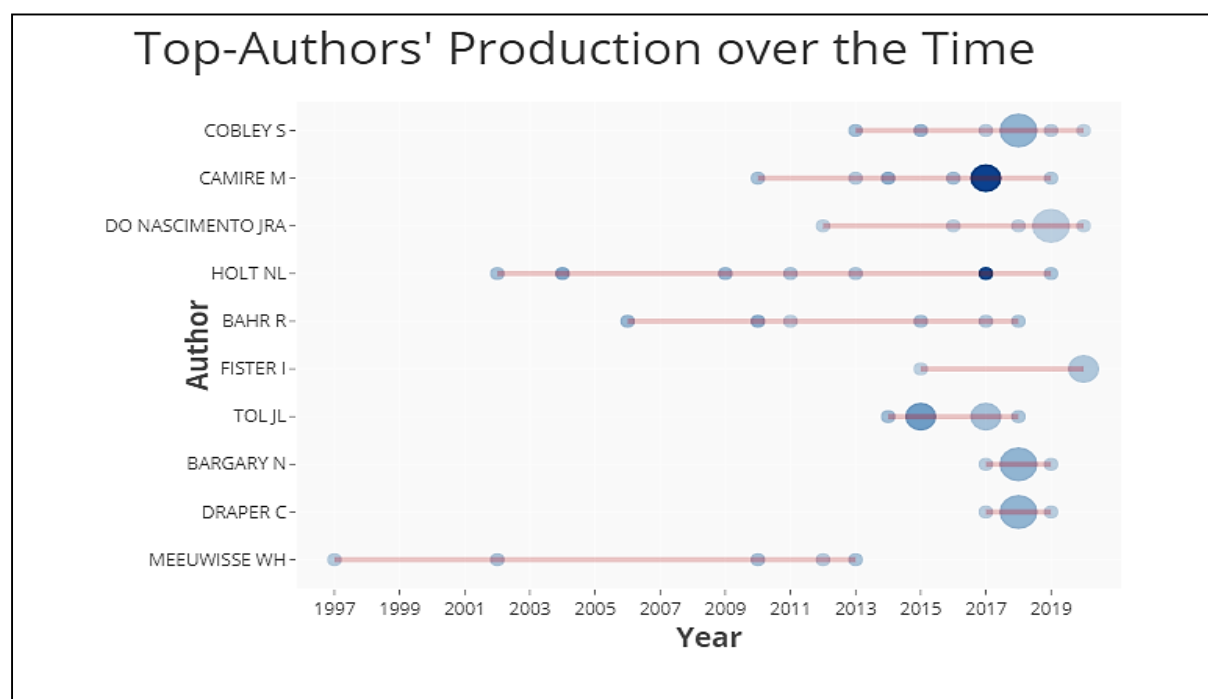


Fig 8. Top-Authors' production over the time

Figure 8 illustrates the scientific productions of the top 10 researchers over time. As a result of this research, it is possible to identify researchers who have continuously engaged in scientific production in data analysis in sports. Researchers who have been well focused in this field in recent years have also been shown that Camire m, Cobley s, Bahr r, holt nl, has continuously published scientific articles. It should be noted that the size of each node depends on the number of articles published in the year.

According to the United States, Britain, Australia, Brazil, Canada, Spain have the most participants in the field with 561, 187, 166, 134, 131, 117 articles. Toronto, Sydney, Calgary, Limerick have been leading universities in their respective fields. Figure 10 illustrates the top countries in terms of citation

² Functional Data Analysis

³ Functional Principal Components Analysis

⁴ Analysis of Characterizing Phases

per year: Lebanon, France, Belgium, and Norway, which show that emerging countries are in data mining in sports and have provided authoritative articles

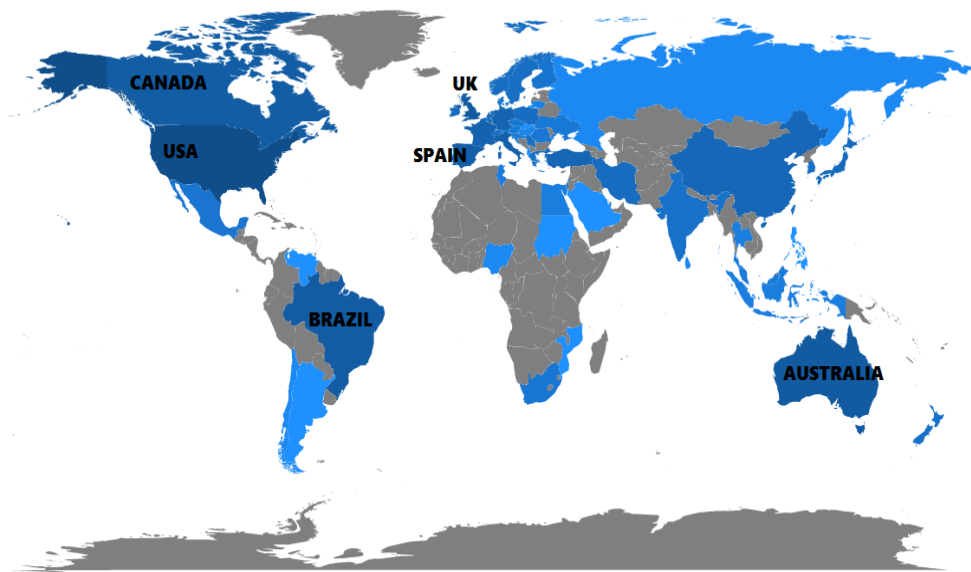


Fig 9. Top countries

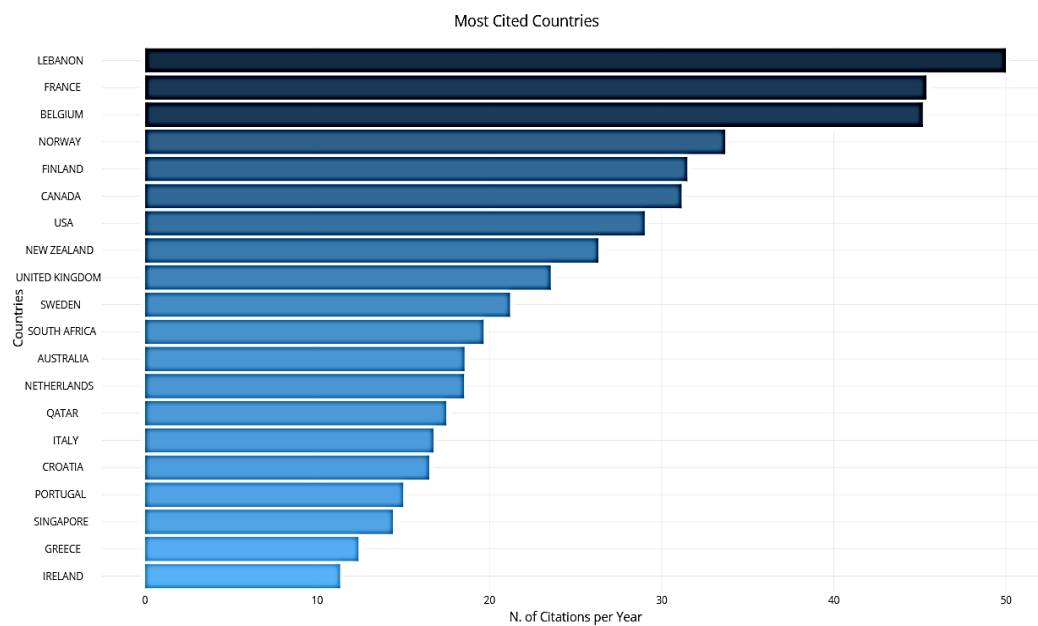


Fig 10. Top Countries

The best index for article verification is citation per year. The top 10 articles are introduced in table 3. The most cited articles were related to "Sports Injuries," most of which focused on predicting athletes' sports injuries and recovery time.

Table 3. List of top 10 articles

Author	Title	Year	TP	TC	Sub –Topic	<u>Topic</u>
Werthner et al.	A New Theoretical Perspective for Understanding How Coaches Learn to Coach	2003	51.8	830	Analysis of reported data from brain injuries in football	sports injury
BOILEAU et al.	Risk Factors for Recurrence of Shoulder Instability after Arthroscopic Bankart Repair	2006	33.9	441	Evaluation of risk factors in osteoarthritis of the knee (arthroscopic Bankart)	sports injury
Holt et al.	A grounded theory of positive youth development through sport based on results from a qualitative meta-study	2016	26.5	53	Review of qualitative studies conducted in the field of data analysis in sports (63 articles reviewed)	Qualitative articles
Gall et al.	Anthropometric and fitness characteristics of international, professional and amateur male graduate soccer players from an elite youth academy	2010	16.5	149	Anthropometric identification (physical dimensions) influential in the development of talent in the sport of football	Sports talent search
Pelvis et al.	Trunk, Pelvis, Hip, and Knee Kinematics, Hip Strength, and Gluteal Muscle Activation during a Single-Leg Squat in Males and Females With and Without Patellofemoral Pain Syndrome [29]	2012	19	139	Reported data from athletes (including men and women) with patellofemoral pain syndrome were studied.	sports injury
Lefevre et al	Return to Sport After Primary and Revision Anterior Cruciate Ligament Reconstruction [30]	2017	17	34	Predicting the time to return to exercise after a cruciate ligament injury	sports injury
Milewski et al.	Chronic Lack of Sleep is Associated with Increased Sports Injuries in Adolescent Athletes	2014	15.2	76	Investigating the relationship between sleep deprivation and sports injuries	sports injury
DWYER et al	Motivations Associated With Nondisclosure of Self-Reported Concussions in Former Collegiate Athletes[32]	2016	15	45	A study was performed at the university level on people with brain injuries during exercise. The data were extracted through a questionnaire. The Multivariate binomial regression technique was used for analysis.	sports injury
Kerr et al	Global positioning system data analysis: velocity ranges and a new definition of sprinting for field sport athletes	2012	13.8	97	GPS-recorded time-motion data in several sports (including men and women) as input data for motion analysis to detect and define the exact range of speed in each sport	sport physiology
Krane et al	Living the Paradox: Female Athletes Negotiate Femininity and Muscularity[26]	2004	10.8	163	Data from interviews with female athletes were analyzed to find out women's social expectations	sport psychology

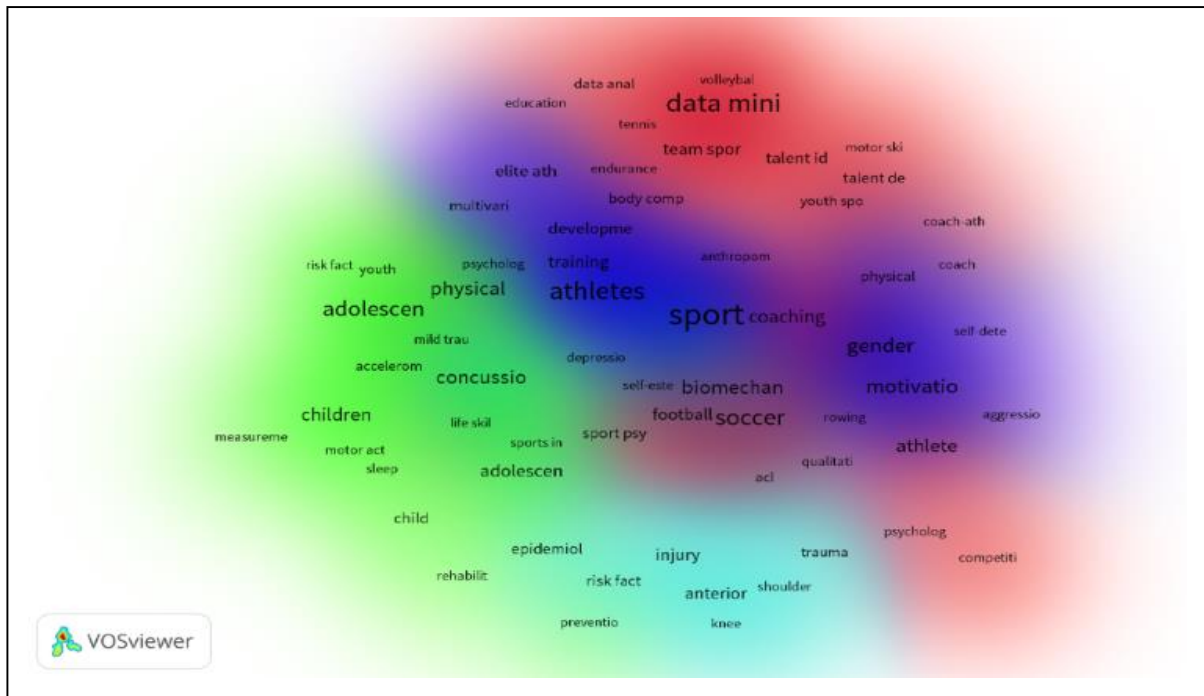


Fig 11. Top keywords

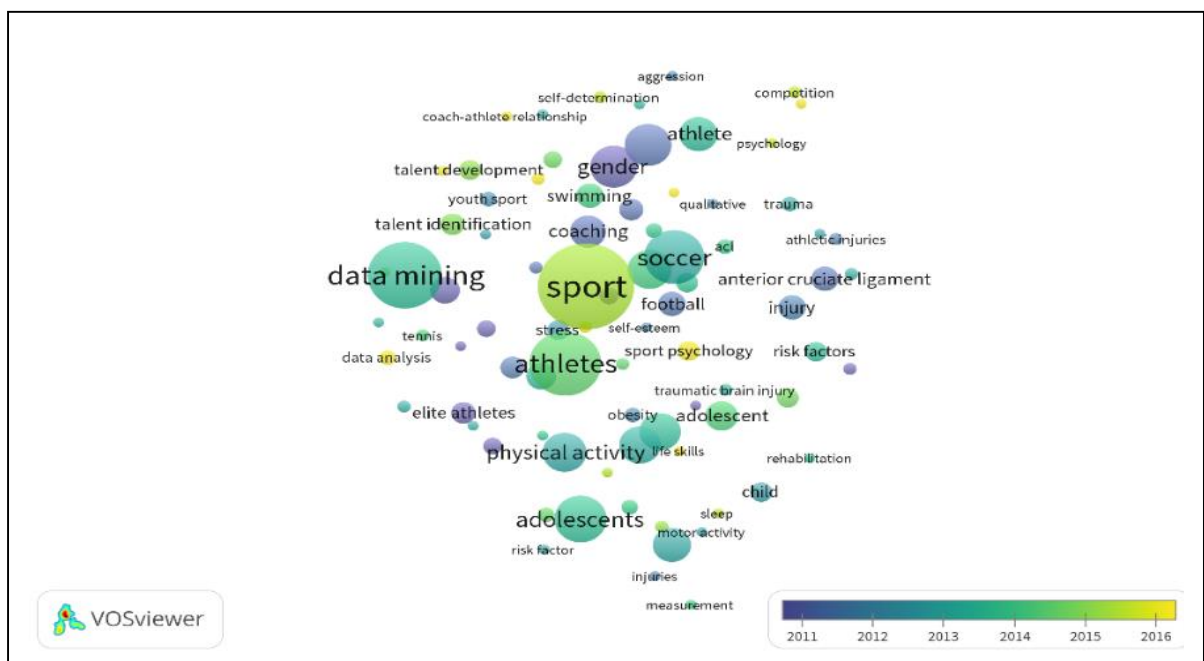


Fig 12. Top keywords-times

3-6-Co-citation analysis

Co-citation is defined as the new literature that mentions two articles from previous literature. (Davoodi et al., 2010; Lemyre et al., 2007). Different types of co-citations can be used based on the unit of analysis: Document co-citation analysis, author co-citation analysis, and journal co-citation analysis. In this study, Document- Co-citation analysis was performed using CiteSpace II, an application that combines visualization, bibliometrics, and data mining methods in an interactive visualization tool (Synnvestedt et al., 2005). Additionally, using Kleinberg's burst detection algorithm (Kleinberg,

2003), co-citation clusters are labeled by extracting burst terms within titles and abstracts.

3-7-Document co-citation

Document analysis lets us know which documents define the mental structure in sustainability and socially responsible mutual funds. According to one hundred ninety-eight articles with more than five citations were classified into six clusters. Each node represents an article, and the color of each node represents the year of publication. The yellow node represents new articles from 2015 to 2020. Each node's size represents the importance of that article based on the number of citations. Information about each cluster, including the number of articles, year, and name of clusters, is given in table 4. Further, we will briefly explain each cluster, including its purposes and the results from their most important articles.

Table 4. Details of each cluster

cluster	number	Mean year	Cluster name
Cluster1	47	2007	Coach education
Cluster2	45	2010	Concussion pain
Cluster3	41	2010	Parental involvement
Cluster4	26	2009	Talent selection
Cluster5	24	2007	Biomechanical Measures
Cluster6	15	2009	Hamstring injuries

Cluster 1 is the most significant cluster with 47 articles, and most of the articles in this cluster have focused on expanding the coach's knowledge. The articles in this cluster are qualitative; therefore, the data are collected through questionnaires and analyzed through descriptive statistics to discover the pattern and framework for expanding the coach's knowledge and talents. In 2006, Penny Werthner and Pierre Trudel conducted conceptual research on an experienced coach to discover a new theoretical pattern and how a coach becomes an elite coach. The data were extracted through questionnaires and analyzed through methods of descriptive statistics. They concluded that cognitive learning parameters and psychological issues are critical factors in developing a coach (Werthner & Trudel, 2006). Afterward, in 2007, Francois Lemyre et al. recognize the effective pattern that helps a coach in the growth stage to develop their talents and become an elite coach by collecting data from 36 coaches in the growth stage in three prevalent fields of sport in the United States and analyzing the data based on descriptive statistics. They discovered three important practical factors in developing the coaching talent. These factors are education, experience as a player, and, most importantly, strengthening relationships and interaction with others (Werthner & Trudel, 2006). In 2007, Trevor Wright et al. collected data from 35 volunteer ice hockey coaches through interviews and analyzed them using descriptive statistics tests. This study aimed to evaluate the different situations that promote and develop coaching knowledge. The results indicate that educational seminars, official councils, books, videos, individual experience as a player, family, interactions with other coaches, and the internet play an important role in developing coaching talent. Each situation also plays a unique role; therefore, it can be concluded that educating coaches must be a combination of the mentioned learning situations (Wright et al., 2007). In 2010, Nicholas Holt published an article to recognize organizational stress-inducing factors in athletes and classify them. Their results showed that the stress-inducing factors could be divided into four main categories: leadership, cultural and team, logistic and environmental, and personal problems. These categories can be considered the essential stress-inducing factors in sports organizations (Holt & Dunn, 2004).

Cluster 2 has 45 articles that have focused on sports injuries in different fields and predicting them. Most articles focus on head injuries, which shows the most common sports injuries are to the head. This category of articles usually focuses on collecting data in a period and obtaining results through quantitative analysis methods. The most important articles in this cluster are presented based on the link strength and citation.

Daniel Gould and Kristen Dieffenbach published a scientific article named “Psychological Characteristics and Their Development in Olympic Champions” about using doctors and health experts to focus on injured athletes (Gould et al., 2002). In 2003, Guskiewicz KM et al. conducted a case study on athletes with head injuries across the United States university to evaluate and find the relation between the experience of past injury and the probability of future injuries to estimate the recovery time. The results showed that the athlete with an experience of past injuries are more prone to future injuries. In other words, there is a relation between the number of previous injuries and the possibility of future injuries. Also, headaches are common signs of head injuries (Guskiewicz et al., 2003). Mallika Marar et al. published the next most significant article in 2012. They analyzed the data from head injury reports from 2008 to 2010 related to 20 fields of sports across high schools with SPSS software to evaluate the occurrence method and the epidemiology of head injuries in athletes. The results showed that the possibility of head injuries in competitions, compared to training, is higher. Most head injury reports were related to football, men’s wrestling, and women’s basketball. They also showed that the severity of injuries is higher in women’s sports. The most usual injury mechanism is player-to-player contact, in a way that 50 percent of the injured athletes would return in 1 to 3 weeks (Marar et al., 2012). Steven P. et al. published research in 2014 to propose the best clinical instructions based on previous studies’ results for managing head injuries in athletes. The results indicated that for better managing the injured athletes, recovery must be the forerunner in developing a detailed plan for head injury management strategy and sharing it with sports executives. This plan must include an essential evaluation of athletes, such as necessary let’s neural and motor criteria (Broglio et al., 2014).

Cluster 3, with 41 articles, evaluates functional parameters in terms of psychology and in developing sports talents for becoming an elite athlete. The data in most of the articles were collected through questionnaires and analyzed through descriptive statistics. This cluster’s articles show that family is the most crucial parameter in developing youth sports talents. It is also observed that clusters 3 and 1 have similarities in some parts because the articles in these two clusters are qualitative types based on designing questionnaires. Table 4 illustrates more details about this cluster. Jean Côté, in 1999, examined the role of family, in terms of dynamic, in developing athletes’ sports talents. The results were the discovery of a constant and mutual pattern of the athletes’ family dynamics and their significant effect on developing the youth sports talents, in such a way that in the first years, the family plays the role of the leader and during an investment and sports maturity, the role of the supporter (Côté, 1999)

Daniel Gould et al. conducted a study on 10 American Olympic champions to evaluate their psychological characteristics and determine the effective parameters in their success. They realized that six psychological characteristics, including controlling anxiety, self-confidence, endurance, purposefulness, optimism, and perfectionism, were the most critical psychological parameters in the athletes’ success. Moreover, the results showed that family, society, and individuals significantly affect the development of these parameters (Gould et al., 2002). Further, in 2004, S. Wuerth A. et al. evaluated the parents’ participation from the athletes’ perspectives. They examined the parents’ participation pattern in youth sports during a career phase and professional transmission. They concluded that families’ support plays a significant role in helping athletes become professional; they can tolerate and control different severities during their sports careers with support from the families (Wuerth et al., 2004). Then, in 2005, Nicholas L. Holt and Laura E. Wolfenden A. conducted a study to evaluate the effectiveness of the three factors, athlete’s perceptive intelligence, family, and coach, in developing tennis talent in elite youths. The results generally showed that parents play the most prominent role in providing emotional support; however, the coach’s role is to provide technical advice. This article’s main concept was that players, parents, and coaches play unique roles in developing a professional team. (Wolfenden & Holt, 2005). In 2010, Nicholas L. Holt, using questionnaires and interviews, utilized the data from a group of young professional football players and first-class coaches in England. They evaluated and recognized the psychological abilities among professional players and achieved a

constant success factor with football theory. The results of this research led to the discovery of the fourth crucial psychological-social parameter. If a player can obtain and harness these factors, he/she is more likely to be chosen for professional football. These factors tend towards selflessness, commitment, endurance, and social support (Holt & Dunn, 2004). Brett Smith and Kerry R. McGannon published another important article in this cluster in 2007. They believe that qualitative research in sports psychology is increasing daily; therefore, they focused on evaluating the standard methods of conducting and refereeing qualitative research in sports psychology. This research concluded that conducting and referring to this article type must be changed and then proposed some suggestions concerning this topic (Smith & McGannon, 2017).

Cluster 4 includes 26 articles regarding the number of citations and link strength—the articles in this cluster focus on finding sports talents, especially in youth football players. In most articles, the data used to find a model for discovering top talents is anthropometric data or aspects of the samples. Also, the results of these articles significantly help coaches consider the most critical factors for choosing players. One of the fundamental articles in this cluster was published by M. Williams A. and T. Reilly in 2000. They reviewed different articles for predicting sports talents and concluded that sports sciences play an important role in recognizing, observing, and training sports talents (Williams & Reilly, 2000). T. Reilly et al. examined the anthropometric, psychological, and physiological data of 30 athletes, including 15 professional and 15 semi-professional athletes, to design a model for evaluating and recognizing their talents. The results showed that the talent search procedure should be done by combining different types of tests, which must be flexible according to age and athletes' educations (Reilly et al., 2000). In 2006, Vaeyens et al. researched youth football players to determine the relationship between physical features and players' skills. It was done using analyzed anthropometric data. The results obtained from this multivariate analysis showed that top players were better than other players in the strength, speed, flexibility, and aerobic parameters. Also, the most critical variables in each group were specified. Finally, it was concluded that recognizing talents is a dynamic process, and opportunities for long-term talent development must be prepared (Vaeyens et al., 2006). In 2007, Renaat M. Philippaerts et al. conducted a study on 33 male football players in the same time interval. The changes in their heights and weights, and functional variables were obtained to discover their relationship. For example, when the height reaches its peak, other body variables such as speed and strength reach their peaks, too. They suggested that coaches must be aware of each compensatory growth feature to discover talents (Philippaerts et al., 2006). Afterward, in 2008, Roel Vaeyens et al. conducted a study focusing on top talent problems, given that talent discovery was a popular topic in sports. However, there was no accepted framework or pattern for discovering sports talents. This study considers the current talent discovery models ineffective due to unidimensionality based on the literature in this field. It states that genetic and physical features must be included (Vaeyens et al., 2008).

Cluster 5, with 24 articles, is called Biomechanical Measures. These articles focus on musculoskeletal injuries in the lower body. In most of these studies, the data are extracted with biomechanical measurements in controlled ambients and usually focus on predicting future injuries and essential factors in muscular injuries and predicting the recovery time with regression tools. One of the primary articles in this cluster was published in 2005 by Timothy E. Hewett et al. They conducted a study to discover and predict injuries in an ACL⁵ through variance analysis and linear regression on the data during training and exercise, usually in descent, in female athletes. Motion and knee angles are essential in predicting ACL injuries (Hewett et al., 2005). Mark V. Paterno et al., in 2010, conducted a study to recognize essential variables in predicting future injuries through multivariate logistic regression on the collected data from people who have once had ACL injuries in the past. Those four crucial variables are torque on the hip joint, asymmetry on the knee's surface during descent and stationary body position, and knee motion range. The results also illustrated that a person with previous ACL injury experience, lack of control on thigh and knee joint, and asymmetry in the body posture are signs of ACL injury for the second time (Paterno et al., 2010). Patellofemoral pain syndrome (PFP) is one of the most common musculoskeletal injuries among athletes. In 2011, Kimberly L. Dolak et al. conducted a study to determine a group with the most improvement. They used a mixed-model analysis of variance on two

⁵ Anterior Cruciate Ligament

groups of females with PFP⁶. Group one had strengthened the thigh joint in a specific period, and group two strengthened the quadriceps muscle. The results showed that the group, who had strengthened the thigh joint after the specific period, had improved more (Dolak et al., 2011).

Cluster 6, as the smallest cluster, includes 15 articles that analyze the data from athletes' muscle injury reports, especially hamstring muscle. One of the most important articles in this cluster was published by Jan Ekstrand et al. in 2011. Given that muscle injuries are the most common injuries among athletes, they conducted a study and analyzed the recorded data by the medical team of 51 European football teams. The results showed that, on average, a team of 25 people records 15 muscle injury reports in each season. Also, 37 percent of all the injuries are muscle injuries, and 27 percent of all absences are due to muscle injuries. Generally, lower-body muscles are more prone to injuries, and injuries to the thigh's posterior muscles are the most common muscle injuries (Ekstrand et al., 2011). Afterward, in 2012, Jan Ekstrand examined the MRI data from players with injuries in the hamstring in professional European teams to find a relationship between the MRI outcome and the type of injury. The results showed that MRI data could be a precise tool in determining the injury's extent and predicting the recovery time. They divided the outcomes into four clusters, 0, 1, 2, and 3, based on the extent of the injury and treatment duration, from low to high. Most of the reports belong to clusters 0 and 1 (Ekstrand et al., 2012)

4- Conclusion and discussion

In some studies, the bibliometric approach has been used in various fields and the number of articles that apply this method to the study of literature is increasing. (Bartolini et al., 2019; Fabregat-Aibar et al., 2019; Gao et al., 2016; Geng et al., 2017; Vishwakarma & Mukherjee, 2019). It is seen that mostly they did not deal with all the analyses and relied only on software in the analyses. In contrast, this article has covered all the details entirely through bibliometric analyses. This article provides an overview of published articles that have analyzed quantitative and qualitative data using data science tools. Compared to other review articles, this research used bibliometric analysis to identify articles, journals, authors, countries, and institutions that have played a significant role in developing this science. Meaningful communication networks between authors, articles were drawn, and their knowledge was extracted. The recently published articles 790articles were also categorized based on the topic and method.

In contrary to other studies that used only one software, for obtaining the best results, three valid bibliometric software (R, Cite space, and VOSviewer) were used simultaneously in this study. Overall, research on sports data mining has increased significantly among the authors (Fig). Football was also one of the most analyzed sports, and many studies have focused on it. The United States, Australia, and the United Kingdom were the top three countries, and Toronto, Sydney, and Calgary were the top three universities that researched this area. A significant majority of articles focused on the prediction of sports injuries and recovery time. Research on Sports medicine and talent searches have increased significantly in recent years, and it is noteworthy that most of these articles focused only on analyzed qualitative data. In contrast, quantitative data is beneficial in the analysis of sports talent searches. Therefore, it can be an essential area for future research.

This study tried to comprise all the articles from the emergence of this area. However, it should be considered that the results presented were dynamic. With the increase in the number of citations and the new emergence of authors, institutions, and countries in sports data mining may change. It should also be noted that there may be articles in WOS and SCOPUS without a keyword that is not in the search domain.

⁶ Patellofemoral pain

References

- Agopyan, A. (2020). An analysis of movements with or without back bend of the trunk or large hip extension in 1st Juniors' Rhythmic Gymnastics World Championship-2019. Is there injury risk for gymnasts? *International Journal of Performance Analysis in Sport*, 1-18. doi:10.1080/24748668.2020.1850038.
- Akarçeşme, C. J. C. E. J. o. S. S., & Medicine. (2017). Is it possible to estimate match result in volleyball: A new prediction model. *19*(3), 5-17.
- Arabzad, S. M., Tayebi Araghi, M., Sadi-Nezhad, S., & Ghofrani, N. J. J. o. A. R. o. I. E. (2014). Football match results prediction using artificial neural networks; the case of Iran Pro League. *1*(3), 159-179.
- Araghi Niknam, F., Ghousi, R., Masoumi, A., Atashi, A., & Makui, A. (2021). Hybrid Medical Data Mining Model for Identifying Tumor Severity in Breast Cancer Diagnosis. *Advances in Industrial Engineering*, 55(2), 151-164. doi:10.22059/ijieng.2021.326775.1789
- Bartolini, M., Bottani, E., & Eric, H. J. J. o. C. P. (2019). Green warehousing: Systematic literature review and bibliometric analysis.
- Bonidia, R. P., Brancher, J. D., & Busto, R. M. J. I. L. A. T. (2018). Data mining in sports: A systematic review. *16*(1), 232-239.
- Broglio, S. P., Cantu, R. C., Gioia, G. A., Guskiewicz, K. M., Kutcher, J., Palm, M., & McLeod, T. C. V. (2014). National Athletic Trainers' Association position statement: management of sport concussion. *Journal of athletic training*, 49(2), 245-265.
- Cao, C. (2012). Sports data mining technology used in basketball outcome prediction.
- Côté, J. (1999). The influence of the family in the development of talent in sport. *The sport psychologist*, 13(4), 395-417.
- Davoodi, E., Khanteymoori, A. R. J. R. A. i. N. N., Fuzzy Systems, & Computing, E. (2010). Horse racing prediction using artificial neural networks. *2010*, 155-160.
- Dolak, K. L., Silkman, C., McKeon, J. M., Hosey, R. G., Lattermann, C., & Uhl, T. L. (2011). Hip strengthening prior to functional exercises reduces pain sooner than quadriceps strengthening in females with patellofemoral pain syndrome: a randomized clinical trial. *journal of orthopaedic & sports physical therapy*, 41(8), 560-570.
- Dzikowski, P. J. J. o. B. R. (2018). A bibliometric analysis of born global firms. *85*, 281-294.
- Ekstrand, J., Häggglund, M., & Waldén, M. (2011). Epidemiology of muscle injuries in professional football (soccer). *The American journal of sports medicine*, 39(6), 1226-1232.
- Ekstrand, J., Healy, J. C., Waldén, M., Lee, J. C., English, B., & Häggglund, M. (2012). Hamstring muscle injuries in professional football: the correlation of MRI findings with return to play. *Br J Sports Med*, 46(2), 112-117.
- Fabregat-Aibar, L., Barberà-Mariné, M. G., Terceño, A., & Pié, L. J. S. (2019). A bibliometric and visualization analysis of socially responsible funds. *11*(9), 2526.
- Gao, C., Sun, M., Geng, Y., Wu, R., & Chen, W. J. A. e. (2016). A bibliometric analysis based review on wind power price. *182*, 602-612.

- Geng, S., Wang, Y., Zuo, J., Zhou, Z., Du, H., Mao, G. J. R., & Reviews, S. E. (2017). Building life cycle assessment research: A review by bibliometric analysis. *76*, 176-184.
- Gould, D., Dieffenbach, K., & Moffett, A. (2002). Psychological characteristics and their development in Olympic champions. *Journal of applied sport psychology*, *14*(3), 172-204.
- Guskiewicz, K. M., McCrea, M., Marshall, S. W., Cantu, R. C., Randolph, C., Barr, W., . . . Kelly, J. P. (2003). Cumulative effects associated with recurrent concussion in collegiate football players: the NCAA Concussion Study. *Jama*, *290*(19), 2549-2555.
- Haghighat, M., Rastegari, H., & Nourafza, N. J. A. i. C. S. a. I. J. (2013). A review of data mining techniques for result prediction in sports. *2*(5), 7-12.
- Hewett, T. E., Myer, G. D., Ford, K. R., Heidt Jr, R. S., Colosimo, A. J., McLean, S. G., . . . Succop, P. (2005). Biomechanical measures of neuromuscular control and valgus loading of the knee predict anterior cruciate ligament injury risk in female athletes: a prospective study. *The American journal of sports medicine*, *33*(4), 492-501.
- Hirsch, J. E. J. P. o. t. N. a. o. S. (2005). An index to quantify an individual's scientific research output. *102*(46), 16569-16572.
- Hjørland, B. J. I. P., & Management. (2013). Citation analysis: A social and dynamic approach to knowledge organization. *49*(6), 1313-1325.
- Holt, N. L., & Dunn, J. G. J. J. o. a. s. p. (2004). Toward a grounded theory of the psychosocial competencies and environmental conditions associated with soccer success. *16*(3), 199-219.
- Ivanković, Z., Racković, M., Markoski, B., Radosav, D., & Ivković, M. (2010). *Analysis of basketball games using neural networks*. Paper presented at the 2010 11th International Symposium on Computational Intelligence and Informatics (CINTI).
- Jiwai, H., & Kamber, P. (2012). *Data Mining concepts and techniques third edition*: Morgan Kaufmann publishers.
- Kahn, J. J. W. W. W. e. p. (2003). Neural network prediction of NFL football games. 9-15.
- Kerr, Z. Y., Register-Mihalik, J. K., Kroshus, E., Baugh, C. M., & Marshall, S. W. J. T. A. j. o. s. m. (2016). Motivations associated with nondisclosure of self-reported concussions in former collegiate athletes. *44*(1), 220-225.
- Kleinberg, J. (2003). Bursty and hierarchical structure in streams. *Data mining and knowledge discovery*, *7*(4), 373-397.
- Kos, A., Wei, Y., Tomažič, S., & Umek, A. J. P. C. S. (2018). The role of science and technology in sport. *129*, 489-495.
- Lefevre, N., Klouche, S., Mirouse, G., Herman, S., Gerometta, A., & Bohu, Y. J. T. A. j. o. s. m. (2017). Return to sport after primary and revision anterior cruciate ligament reconstruction: a prospective comparative study of 552 patients from the FAST cohort. *45*(1), 34-41.
- Lemyre, F., Trudel, P., & Durand-Bush, N. J. T. s. p. (2007). How youth-sport coaches learn to coach. *21*(2), 191-209.

- Marar, M., McIlvain, N. M., Fields, S. K., & Comstock, R. D. (2012). Epidemiology of concussions among United States high school athletes in 20 sports. *The American journal of sports medicine*, 40(4), 747-755.
- Masoumi, A., Ghousi, R., Vazifehdoost, M., & Araghi Niknam, F. (2021). A quantitative scoring system to compare the degree of COVID-19 infection in patients' lungs during the three peaks of the pandemic in Iran. *Journal of Industrial and Systems Engineering*, 13(3), 61-69.
- McCabe, A., & Trevathan, J. (2008). *Artificial intelligence in sports prediction*. Paper presented at the Fifth International Conference on Information Technology: New Generations (itng 2008).
- McCulloh, I., Armstrong, H., & Johnson, A. (2013). *Social network analysis with applications*: John Wiley & Sons.
- Morris, S., DeYong, C., Wu, Z., Salman, S., Yemenu, D. J. C., & engineering, i. (2002). DIVA: a visualization system for exploring document databases for technology forecasting. 43(4), 841-862.
- Niu, B., Loaiciga, H. A., Wang, Z., Zhan, F. B., & Hong, S. J. J. o. H. (2014). Twenty years of global groundwater research: A Science Citation Index Expanded-based bibliometric survey (1993–2012). 519, 966-975.
- Norton, M. (2000). *Introductory concepts in information science*: Information Today, Inc.
- Paterno, M. V., Schmitt, L. C., Ford, K. R., Rauh, M. J., Myer, G. D., Huang, B., & Hewett, T. E. (2010). Biomechanical measures during landing and postural stability predict second anterior cruciate ligament injury after anterior cruciate ligament reconstruction and return to sport. *The American journal of sports medicine*, 38(10), 1968-1978.
- Philippaerts, R. M., Vaeyens, R., Janssens, M., Van Renterghem, B., Matthys, D., Craen, R., . . . Malina, R. M. (2006). The relationship between peak height velocity and physical performance in youth soccer players. *Journal of sports sciences*, 24(3), 221-230.
- Reilly, T., Williams, A. M., Nevill, A., & Franks, A. (2000). A multidisciplinary approach to talent identification in soccer. *Journal of sports sciences*, 18(9), 695-702.
- Saavedra, J. M., Pic, M., Lozano, D., Tella, V., & Madera, J. (2020). The predictive power of game-related statistics for the final result under the rule changes introduced in the men's world water polo championship: a classification-tree approach. *International Journal of Performance Analysis in Sport*, 20(1), 31-41.
- Schumaker, R. P., Solieman, O. K., & Chen, H. (2010a). *Sports data mining* (Vol. 26): Springer Science & Business Media.
- Schumaker, R. P., Solieman, O. K., & Chen, H. J. A. (2010b). Sports knowledge management and data mining. 44(1), 115-157.
- Shi, Z., Moorthy, S., & Zimmermann, A. (2013). *Predicting NCAAAB match outcomes using ML techniques—some results and lessons learned*. Paper presented at the ECML/PKDD 2013 Workshop on Machine Learning and Data Mining for Sports Analytics.
- Small, H. J. J. o. t. A. S. f. i. S. (1973). Co-citation in the scientific literature: A new measure of the relationship between two documents. 24(4), 265-269.

- Smith, B., & McGannon, K. R. (2017). Developing rigor in qualitative research: problems and opportunities within sport and exercise psychology. *International Review of Sport and Exercise Psychology*, 11(1), 101-121. doi:10.1080/1750984x.2017.1317357.
- Sujatha, K., Godhavari, T., Bhavani, N. P. J. I. J. o. M., & Methods, C. (2018). Football match statistics prediction using artificial neural networks. 3.
- Synnestvedt, M. B., Chen, C., & Holmes, J. H. (2005). *CiteSpace II: visualization and knowledge discovery in bibliographic databases*. Paper presented at the AMIA Annual Symposium Proceedings.
- Tax, N., Joustra, Y. J. T. o. k., & engineering, d. (2015). Predicting the Dutch football competition using public data: A machine learning approach. 10(10), 1-13.
- Thabtah, F., Zhang, L., & Abdelhamid, N. (2019). NBA Game Result Prediction Using Feature Analysis and Machine Learning. *Annals of Data Science*, 6(1), 103-116. doi:10.1007/s40745-018-00189-x
- Tümer, A. E., & Koçer, S. (2017). Prediction of team league's rankings in volleyball by artificial neural network method. *International Journal of Performance Analysis in Sport*, 17(3), 202-211.
- Vaeyens, R., Lenoir, M., Williams, A. M., & Philippaerts, R. M. (2008). Talent identification and development programmes in sport. *Sports medicine*, 38(9), 703-714.
- Vaeyens, R., Malina, R. M., Janssens, M., Van Renterghem, B., Bourgois, J., Vrijens, J., & Philippaerts, R. M. (2006). A multidisciplinary selection model for youth soccer: the Ghent Youth Soccer Project. *British journal of sports medicine*, 40(11), 928-934.
- Vishwakarma, P., & Mukherjee, S. J. T. R. R. (2019). Forty-three years journey of Tourism Recreation Research: a bibliometric analysis. 44(4), 403-418.
- Warmenhoven, J., Cogley, S., Draper, C., Harrison, A., Bargary, N., Smith, R. J. J. o. s., & sport, m. i. (2018). How gender and boat-side affect shape characteristics of force–angle profiles in single sculling: Insights from functional data analysis. 21(5), 533-537.
- Warmenhoven, J., Cogley, S., Draper, C., Harrison, A., Bargary, N., & Smith, R. J. S. b. (2019). Bivariate functional principal components analysis: considerations for use with multivariate movement signatures in sports biomechanics. 18(1), 10-27.
- Warmenhoven, J., Cogley, S., Draper, C., Harrison, A., Bargary, N., Smith, R. J. S. j. o. m., & sports, s. i. (2017). Assessment of propulsive pin force and oar angle time-series using functional data analysis in on-water rowing. 27(12), 1688-1696.
- Warmenhoven, J., Harrison, A., Robinson, M. A., Vanrenterghem, J., Bargary, N., Smith, R., . . . sport, m. i. (2018). A force profile analysis comparison between functional data analysis, statistical parametric mapping and statistical non-parametric mapping in on-water single sculling. 21(10), 1100-1105.
- Warmenhoven, J., Smith, R., Draper, C., Harrison, A., Bargary, N., Cogley, S. J. S. j. o. m., & sports, s. i. (2018). Force coordination strategies in on-water single sculling: Are asymmetries related to better rowing performance? , 28(4), 1379-1388.
- Werthner, P., & Trudel, P. J. T. s. p. (2006). A new theoretical perspective for understanding how coaches learn to coach. 20(2), 198-212.

Williams, A. M., & Reilly, T. (2000). Talent identification and development in soccer. *Journal of sports sciences*, 18(9), 657-667.

Wiseman, O. (2016). *Using Machine Learning to Predict the Winning Score of Professional Golf Events on the PGA Tour*. Dublin, National College of Ireland.

Wolfenden, L. E., & Holt, N. L. (2005). Talent development in elite junior tennis: Perceptions of players, parents, and coaches. *Journal of applied sport psychology*, 17(2), 108-126.

Wright, T., Trudel, P., & Culver, D. (2007). Learning how to coach: the different learning situations reported by youth ice hockey coaches. *Physical education and sport pedagogy*, 12(2), 127-144.

Wuerth, S., Lee, M. J., & Alfermann, D. (2004). Parental involvement and athletes' career in youth sport. *Psychology of sport and Exercise*, 5(1), 21-33.

Zdravevski, E., & Kulakov, A. (2009). *System for Prediction of the Winner in a Sports Game*. Paper presented at the International Conference on ICT Innovations.

Zupic, I., & Čater, T. J. O. R. M. (2015). Bibliometric methods in management and organization. 18(3), 429-472.