

# Impact of Big Data in Sports

Aanchal Yadav  
University Institute Of Engineering  
Chandigarh University  
Mohali, India  
[22bda70136@cuchd.in](mailto:22bda70136@cuchd.in)

Ashish Sharma  
Electrical Engineering Department  
Chandigarh University  
Mohali, India  
[ashish.e10096@cumail.in](mailto:ashish.e10096@cumail.in)

Rashmvi Kaur Grang  
University Institute Of Engineering  
Chandigarh University  
Mohali, India  
[22bda70163@cuchd.in](mailto:22bda70163@cuchd.in)

Saksham Sharma  
University Institute Of engineering  
Chandigarh University  
Mohali, India  
[22bda70136@cuchd.in](mailto:22bda70136@cuchd.in)

Saksham Chandna  
University Institute Of engineering  
Chandigarh University  
Mohali, India  
[22bda70188@cuchd.in](mailto:22bda70188@cuchd.in)

Manudev Kashyap  
University Institute Of engineering  
Chandigarh University  
Mohali, India  
[22bda70193@cuchd.in](mailto:22bda70193@cuchd.in)

**Abstract** — Sports big data has become a challenging issue due to the rapid growth of information technology and sports. These data are easily accessible and contain valuable information on athletes, coaches, and sports events. To extract the value from this data, amazing data analysis technologies have been developed. This article provides an overview of sports big data, including its background, management, analysis methods, and applications such as performance evaluation and prediction. It also highlights research issues, including predicting athlete performance, identifying rising stars in sports, creating a unified sports big data platform, open data access, and privacy protection. This paper aims to assist researchers in gaining a broader understanding of sports big data and its potential applications. Due to the advancement of the internet, cloud computing, and Internet of Things technology, coupled with the increased use of medical and health information technology, the amount of medical and health-related data is increasing rapidly. The development of genomic technology and wearable mobile medical and mobile health technology has also led to the era of big data in the field of healthcare. Traditional methods of sports risk assessment are time-consuming and inaccurate, which limits their use in large-scale sports assessment. This paper promotes the transformation and upgrading of sports health management and advancing the industrialization of big data in sports analytics.

**Keywords**— Big Data, sports analytics, machine learning, data visualization, performance tracking, knowledge graph

## I. INTRODUCTION

Sports and technology have always gone hand in hand, with new advancements continuously revolutionizing the way athletes train, perform, and even how fans engage with their favourite sports. In recent years, big data has emerged as a game-changer in the sports industry, providing coaches, athletes, and sports organizations with valuable insights and analytics to improve performance and enhance the overall fan experience. Big data analytics enables the collection, analysis, and interpretation of vast amounts of data, providing real-time insights that can help make more informed decisions. This review paper aims to explore the impact of big data on sports, highlighting its applications, benefits, and challenges, and providing insights into how big data can continue to transform the sports industry. Big data usage in sports refers to the collection, analysis, and utilization of large and complex data

sets to gain insights, make informed decisions, and optimize performance in various aspects of sports.

### A. Key Considerations:

- 1) *Performance analysis*: Big data can be used to analyze player performance data such as player statistics, biometric data, and match data to identify strengths, weaknesses, and patterns of play. This information can be used for tactical planning, player development, and game preparation.
- 2) *Injury Prevention*: Big data can help in monitoring and analysing player health data to identify injury risks, track recovery progress, and implement injury prevention strategies, such as workload management and personalized training programs.
- 3) *Fan Engagement*: Big data can be used to analyze fan data, including social media interactions, ticket sales, and merchandise purchases, to better understand fan preferences, behavior, and engagement levels. This information can be used to create personalized fan experiences, targeted marketing campaigns, and revenue generation opportunities.

### B. Methodology of Analysing Data:

- 1) *Data Collection*: Data can be collected from various sources such as wearables, sensors, video feeds, social media, and other relevant databases. The data should be clean, accurate, and reliable.
- 2) *Data Integration*: Data from different sources should be integrated and stored in a centralized database for analysis.
- 3) *Data Analysis*: Advanced analytics techniques such as machine learning, data mining, and predictive modeling can be used to analyze the data and derive meaningful insights.
- 4) *Data Visualization*: Data can be presented in visually appealing and easy-to-understand formats such as charts, graphs, and dashboards to aid decision-making.

### C. Methods to Shun:

- 1) *Privacy Violation:* Care should be taken to ensure that data collected and used comply with relevant privacy laws and regulations. Personal and sensitive information should be protected and used only for legitimate purposes.
- 2) *Over-Reliance on Data:* While big data can provide valuable insights, it should not be the sole basis for decision-making. Human expertise and intuition should also be considered.
- 3) *Lack of Data Quality Assurance:* Ensuring the quality and accuracy of data is crucial. Data should be validated, cleaned, and verified to avoid erroneous conclusions or decisions.
- 4) *Neglecting Ethical Considerations:* Ethical considerations such as fairness, bias, and transparency should be taken into account when using big data in sports. Decisionmaking should be transparent, and biases in data or algorithms should be minimized.

## II. BIOMETRIC ANALYSIS

Biometric analysis is one of the areas where big data analytics is having a significant impact on sports. Biometric data refers to physiological and behavioural measurements such as heart rate, muscle activity, movement patterns, and sleep quality. By collecting and analysing biometric data, sports organizations can gain insights into the health, fitness, and performance of athletes, as well as monitor their recovery and injury risk.

Some specific examples of biometric analysis in sports include:

1. *Wearable sensors:* Athletes can wear sensors that capture data on their heart rate, body temperature, and movement patterns. This data can be used to monitor the intensity of their training, identify areas for improvement, and prevent injuries.
2. *Video analysis:* Video footage can be analysed using computer vision techniques to track the movements of athletes and identify patterns and trends. For example, analysts can track the movements of basketball players to identify their shooting patterns and tendencies.
3. *Sleep tracking:* Athletes' sleep patterns can be tracked using sensors that monitor brain activity, eye movement, and heart rate. This data can be used to optimize athletes' sleep schedules and improve their recovery time.
4. *Nutrition tracking:* Biometric data can also be used to track athletes' nutrition and hydration levels, which can impact their performance and recovery. For example, sensors can track athletes' hydration levels during training and games to ensure they are properly hydrated. All the headings in the main body of your paper are numbered (automatically).

## III. EXISTING SOLUTIONS

It's important to note that the usage of big data in sports is a rapidly evolving field, and new solutions and technologies continue to emerge as data analytics and technology advancements progress.

1. *Data integration and management:* One of the biggest challenges in using big data analytics in sports is integrating and managing different types of data from various sources such as sensors, video footage, social media, and historical data. To address this, sports organizations are adopting data management platforms and tools that can collect, store, process, and analyze data in a centralized and standardized manner.

2. *Machine learning and AI:* Another solution is to apply machine learning and artificial intelligence techniques to sports data to extract insights and generate predictions. These techniques can help teams and organizations make data-driven decisions about player performance, game strategy, and fan engagement. For example, predictive models can be used to identify high-potential players or predict game outcomes based on historical data.

3. *Collaborative approaches:* Successful implementation of big data analytics in sports requires collaboration among different stakeholders, including teams, leagues, players, coaches, and fans. Sports organizations are adopting collaborative approaches such as open data initiatives, data sharing agreements, and hackathons to encourage innovation and knowledge exchange.

4. *Ethics and governance frameworks:* As big data analytics becomes more pervasive in sports, there is a growing need for ethical and governance frameworks to ensure transparency, fairness, and accountability in the use of data. Sports organizations are developing policies and guidelines for data privacy, data security, data quality, and data ethics to address the risks and challenges associated with big data analytics in sports.

5. *Social media analytics:* Sports organizations can use social media analytics to monitor and analyze fan engagement, sentiment, and preferences on social media platforms, which can inform marketing and fan engagement strategies.

6. *Athlete Recruitment Decisions:* The Case of the University of Virginia: Athletes who participate in sports at colleges and universities frequently benefit from scholarships that cover a portion of their tuition costs. Not to mention that successful athletes can improve universities' reputations. This has resulted in several universities concentrating on increasing their chances of discovering future athletic stars through the power of big data algorithms. One such model is that of the College of Virginia. Algorithms have been used by the university to predict whether a football player will go to them. school instead of another one or enter the NFL. The recruiters can use this information to stay up to date on where to focus their efforts and time.

TABLE I

Sport	Data collection	Use of data
Baseball	Track player movements, pitch velocity, spin rate, and trajectory using cameras and sensors	Analyse player performance, optimize game plans, evaluate pitch velocity, spin rate, and trajectory
Basketball	Track player movements, shooting accuracy, and defensive coverage using cameras and sensors	Analyse player performance, optimize game plans, evaluate shooting accuracy, and defensive coverage
Football	Track player movements on the field, evaluate offensive and defensive strategies using cameras and sensors and wearable tech	Analyse player performance, optimize game plans, evaluate offensive and defensive strategies
Soccer	Track player movements on the field, evaluate pass accuracy and shot selection using cameras and sensors	Analyse player performance, optimize game plans, evaluate pass accuracy, shot selection, and player positioning
Tennis	Track player movements on the court, evaluate shot selection using cameras and sensors	Analyse player performance, optimize training programs, evaluate shot selection and player positioning

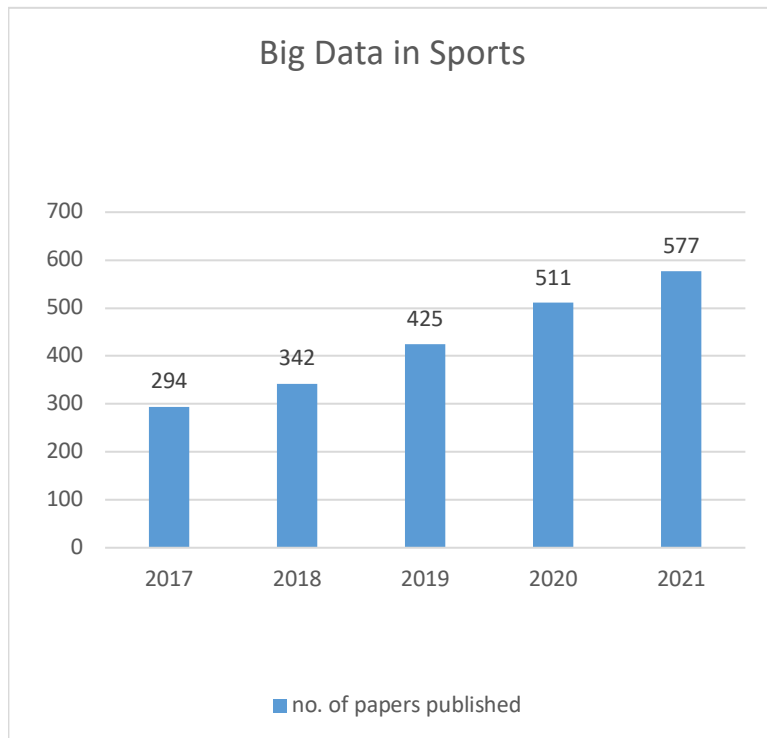


Fig.1 total no. of papers published in recent years.

#### IV. AUTHORS' PERSPECTIVE

- A. *Sports and Health Management Using Big Data Based on Voice Feature Processing and Internet of Things*  
 BY- Lina Sun and Mingzhi Li.  
 Published 26 August 2021

Sports and Health Management Using Big Data Based on Voice Feature Processing and Internet of Things This paper has used MFCCs which are a widely used feature extraction technique in speech and audio processing. MFCCs are derived from the short-term power spectrum of a signal and are used to represent the spectral envelope of the signal. The process of computing MFCCs involves several steps. The first step is to divide the signal into short overlapping frames, typically 20-40 milliseconds in length. The next step is to convert the frequency scale from Hertz to the Mel scale, which is a nonlinear scale that better approximates the human auditory system's perception of frequency. The logarithm of the power spectrum is then taken, and the resulting values are transformed using the discrete cosine transform (DCT) to obtain the MFCCs. MFCCs have been shown to be effective in

a wide range of speech and audio processing tasks, such as speaker identification, speech recognition, and music genre classification. MFCCs can have several applications in the sports industry. For instance, they can be used to extract features from audio data collected during training or performance sessions, which can help coaches and trainers to evaluate and improve athletes' technique, form, and overall performance. In sports like swimming or rowing, underwater microphones can be used to capture the sound of an athlete's stroke, which can be analyzed using MFCCs to identify patterns and variations in their movements. This information can then be used to adjust training techniques and provide feedback on technique and form. MFCCs can also be used in sports broadcasting to enhance the viewing experience for fans. For instance, they can be used to identify specific sound events, such as the sound of a basketball hitting the rim or the crack of a baseball bat and enhance the audio to make it more immersive and engaging for viewers. Overall, MFCCs can help the sports industry to extract meaningful insights from audio data, leading to improved training techniques, enhanced viewing experiences, and better performance outcomes.

The field of education and teaching reforms is currently facing a challenge due to the absence of a reliable and efficient information management system for physical health that is based on scientific evidence. This lack of a platform could result in difficulties in monitoring and evaluating students' physical health, identifying potential health risks, and providing targeted interventions to promote well-being. To address this issue, there is a need to develop a comprehensive information management platform that integrates data on various aspects of physical health and offers tools for tracking, monitoring, and providing personalized interventions. Such a platform could help to foster a culture of health and wellness within schools and promote positive outcomes for students' academic performance and overall health. The goal of creating a sports health management system is to perform a thorough assessment of an athlete's physical test results, which would enable a precise understanding of their overall health level. By leveraging the information stored in the database, the system can categorize an individual's physical state in a detailed manner. This, in turn, facilitates the generation of a more comprehensive evaluation of their physical condition and enables the system to provide better fitness guidance programs to help them achieve optimal performance. Ultimately, this system aims to help athletes gain a more profound understanding of their health and fitness status, which could enhance their athletic capabilities and well-being.

How Sports & Health industry plays role in economic development The sports and health industry can play a significant role in economic development in several ways:

1. Job Creation: The sports and health industry can create jobs in various sectors such as sports management, coaching,

fitness instruction, sports medicine, and sports marketing. These jobs not only provide employment opportunities but also generate income for the employees and contribute to the growth of the economy.
- 2. Tourism: Sports events and competitions attract visitors from all over the world, which leads to increased tourism and revenue for the local economy. For example, hosting a major sporting event can boost the hospitality industry, transportation, and retail sectors.
- 3. Infrastructure Development: Sports facilities such as stadiums, arenas, and training centers require significant investment in infrastructure. This investment creates jobs and stimulates economic activity in the construction industry, providing a boost to the local economy.
- 4. Health and Wellness: The sports and health industry promote healthy living, which can lead to a decrease in healthcare costs, increased productivity, and higher quality of life. A healthy population also results in lower absenteeism and higher worker productivity, contributing to economic growth.
- 5. Sponsorship and Advertising: The sports industry attracts significant sponsorship and advertising revenue from companies seeking to reach a large audience. This revenue can contribute to economic growth through increased sales and brand awareness for the sponsoring companies.

Overall, the sports and health industry have the potential to drive economic growth through job creation, tourism, infrastructure development, health and wellness, and advertising revenue.

Speech Feature Processing and Big Data Enabled Sports Health Management Using Internet of Things (IoT) is an innovative approach to sports health management that utilizes cutting-edge technologies to improve accuracy, personalization, timeliness, and efficiency in data processing and presentation. The article highlights the use of big data in physique and health management, which allows for comprehensive evaluation of physical test results and a more accurate understanding of an athlete's health level. Physical education departments are encouraged to conduct research into body composition and health management using voice feature processing and IoT technologies. By implementing personalized management and services, physical education departments can improve health outcomes and provide more appropriate fitness guidance programs. The article emphasizes the importance of tailoring sports programs to enhance excitement and scientific of participation in sports and exercise, ultimately leading to better health outcomes. AI technology can track and evaluate physical quality in stages, allowing for improved fitness programs and promoting efficient improvement of physique and health levels. Future includes integrating fog computing to improve efficiency and accuracy even further. By incorporating the latest technologies and utilizing big data, Speech Feature Processing and Big Data-Enabled Sports Health Management Using Internet of Things is poised to transform the way

physical education departments manage and promote sports and exercise, leading to healthier and more active lifestyles. The article suggests that physical education departments should modernize by researching body composition and health management using IoT technologies and voice feature processing and implementing personalized management and services. Big data can improve accuracy, personalization, timeliness, and efficiency in data processing and presentation. The sports program should enhance participation and scientific to improve health outcomes. AI can track and evaluate physical quality and provide improved fitness programs to achieve physical education goals. Future includes integrating fog computing for improved efficiency and accuracy.

*B. Big data in sport analytics: applications and risks*  
*BY- Euodia Vermeulen and Sarma Venkata*  
*Yadavalli*  
*(October 29 – November 1, 2018)*

The use of physical activity trackers has led to the generation of large amounts of data, which has the potential to revolutionize various aspects of sports,

including training monitoring, rehabilitation, talent acquisition, and urban planning. However, there are also risks associated with using big data in sports, such as concerns over user privacy, data accuracy, interpretation of information, and athlete autonomy. It is important to consider ethical considerations when implementing these advances in sports, which may slow down the process. While big data has great potential in sports, it is important to proceed with caution and ensure that ethical considerations are addressed. Now big data is characterized in terms of the three V's that are: Volume: enormous data set sizes measured in terabytes or petabytes; Velocity: data is being created and transmitted in near-real time which results in an extremely fast arrival rate. Variety: the organization of the data is diverse and presented as structured, semi-structured and unstructured. Introduction: This paper tells us about the recent uses of big data generated using fitness trackers and sporting websites. It uses wearable microelectronicmechanical (MEM) systems that are used to collect biometric (physiological, kinematic, and kinetic data) and geo-spatial tracking data as athletes physically move through space during sporting activities. Wearable devices also called as wearables include pedometer anklets, chest straps, running or activity watches and smart phones with monitoring applications.

TABLE II

Title	Author	Publication Year	Explanation
"Predicting performance in professional soccer: Big Data analytics approaches"	Martín-González et al.	2021	The paper describes an approach for using big data analytics to predict player performance in professional soccer, using a combination of machine learning algorithms and traditional statistical methods.
"Big Data analytics in professional sports: A systematic review"	Abella-García et al.	2020	The paper provides a systematic review of the literature on the use of big data analytics in professional sports, highlighting the main applications, challenges, and opportunities for future research in the field.
"Big data analytics in sports: A review"	Arjunan et al.	2020	The paper presents a review of the different types of big data analytics techniques used in sports, including descriptive, predictive, and prescriptive analytics, and provides examples of their application in various sports.
"Big data analytics in sports: A survey"	He et al.	2019	The paper surveys the current state of the art in big data analytics in sports, focusing on the key challenges and opportunities for using big data to improve sports performance, prevent injuries, and enhance fan engagement.
"A big data approach to predicting soccer matches"	Kostovski et al.	2018	The paper proposes a big data approach for predicting soccer match outcomes based on historical data on team performance, player injuries, weather conditions, and other factors.
"A big data approach to understanding elite soccer strategy"	Liu et al.	2018	The paper describes a big data approach for analysing the strategy of elite soccer teams based on data on player positioning, movement, and passing patterns, and provides insights into the factors that contribute to team success.
"Big Data in Basketball: The impact of advanced analytics on team success"	Arora et al.	2017	The paper examines the impact of advanced analytics on team success in basketball, using data on team performance, player salaries, and other factors to identify the key factors that contribute to winning games.

## V. CONCLUSION

In this paper, a detailed overview of sports big data, with a specific emphasis on the management, analysis methods, and applications of sports big data has been reviewed. It can be concluded that big data is transforming the world of sports in various ways. With the help of advanced analytics and machine learning techniques, big data is enabling sports teams to make better decisions, optimize performance, and gain a competitive edge. It is also being used to improve fan engagement and create new revenue streams through personalized content and targeted advertising.

For example, in the competitive sports area, big data analysis technology can not only help coaches and athletes to analyze the previous training and competition sports behavior but also can pin the athletes' movement and physical condition and adjust the athletes' training activities to improve their competition performance.

Our recommendation is to use big data for tasks such as predicting athletes' performance using knowledge graphs, identifying rising stars in sports, creating a unified sports big data platform, promoting open access to sports big data, and ensuring adequate privacy protections in sports events on an international level.

## REFERENCES

- [1] L. Cai, C. Zhao and X. Wang, "Situation and lessons of application of NBA big data technology," 2021 International Conference on Information Technology and Contemporary Sports (TCS), Guangzhou, China, 2021, pp. 228-231, doi: 10.1109/TCS52929.2021.00054.
- [2] J. Zeng and J. Jia, "The impact of big data on school sports and competitive sports," 2017 Chinese Automation Congress (CAC), Jinan, China, 2017, pp. 596-599, doi: 10.1109/CAC.2017.8242837.
- [3] C. Wangming, "Analysis and Application: Sports Education Information Management under the Background of Big data," 2021 13th International Conference on Measuring Technology and Mechatronics Automation (ICMTMA), Beihai, China, 2021, pp. 543-546, doi: 10.1109/ICMTMA52658.2021.00125.
- [4] G. Kaur and G. Jagdev, "Analyzing and Exploring the Impact of Big Data Analytics in Sports Science," 2020 Indo – Taiwan 2nd International Conference on Computing, Analytics and Networks (Indo-Taiwan ICAN), Rajpura, India, 2020, pp. 218-224, doi: 10.1109/Indo-TaiwanICAN48429.2020.9181320.
- [5] V. S. Tseng, C. -H. Chou, K. -Q. Yang and J. C. C. Tseng, "A Big Data Analytical Framework for Sports Behavior Mining and Personalized Health Services," 2017 Conference on Technologies and Applications of Artificial Intelligence (TAAI), Taipei, Taiwan, 2017, pp. 178-183, doi: 10.1109/TAAI.2017.47.
- [6] Vermeulen, Euodia & Sarma, Venkata. (2018). Big data in sport analytics: applications and risks.
- [7] H., Bai, Z., & Bai, X. (2021, January 30). *Sports Big Data: Management, Analysis, Applications, and Challenges*. Sports Big Data: Management, Analysis, Applications, and Challenges. <https://doi.org/10.1155/2021/6676297>
- [8] H. -C. Shih, "A Survey of Content-Aware Video Analysis for Sports," in IEEE Transactions on Circuits and Systems for Video Technology, vol. 28, no. 5, pp. 1212-1231, May 2018, doi: 10.1109/TCSVT.2017.2655624.
- [9] N. J. Kim and J. K. Park, "Sports analytics & risk monitoring based on hana platform: Sports related big data & IoT trends by using HANA In-memory platform," 2015 International SoC Design Conference (ISOC), Gyeongju, Korea (South), 2015, pp. 221-222, doi: 10.1109/ISOC.2015.7401795.
- [10] M. Izadi, A. Chia, B. Cheng and S. Wu, "NetClips: A Framework for Video Analytics in Sports Broadcast," 2018 IEEE International Conference on Big Data (Big Data), Seattle, WA, USA, 2018, pp. 1593-1597, doi: 10.1109/BigData.2018.8622333.
- [11] G. Capobianco, U. D. Giacomo, F. Mercaldo and A. Santone, "A Formal Methodology for Notational Analysis and Real-Time Decision Support in Sport Environment," 2018 IEEE International Conference on Big Data (Big Data), Seattle, WA, USA, 2018, pp. 5305-5307, doi: 10.1109/BigData.2018.8621892.
- [12] M. Ahsan, S. T. Hon and A. Albarbar, "Development of Novel Big Data Analytics Framework for Smart Clothing," in IEEE Access, vol. 8, pp. 146376-146394, 2020, doi: 10.1109/ACCESS.2020.3015152.
- [13] A. Srivastava, A. Chaudhary, D. Gupta and A. Rana, "Usage of Analytics in the World of Sports," 2021 9th International Conference on Reliability, Infocom Technologies and Optimization (Trends and Future Directions) (ICRITO), Noida, India, 2021, pp. 1-7, doi: 10.1109/ICRITO51393.2021.9596466.
- [14] J. A. Brown, A. Cuzzocrea, M. Kresta, K. D. L. Kristjanson, C. K. Leung and T. W. Tebinka, "A Machine Learning Tool for Supporting Advanced Knowledge Discovery from Chess Game Data," 2017 16th IEEE International Conference on Machine Learning and Applications (ICMLA), Cancun, Mexico, 2017, pp. 649-654, doi: 10.1109/ICMLA.2017.00-87.
- [15] M. Kempe, F. R. Goes and K. A. P. M. Lemmink, "Smart Data Scouting in Professional Soccer: Evaluating Passing Performance Based on Position Tracking Data," 2018 IEEE 14th International Conference on e-Science (e-Science), Amsterdam, Netherlands, 2018, pp. 409-410, doi: 10.1109/eScience.2018.00126.