STM JOURNALS

Recent Trends in Sports

Volume 3, Issue 1, 2026
January–June
DOI (Journal): 10.37591/RTS

https://journals.stmjournals.com/rts

Research RTS

The Influence of Data Analytics on Sports Performance

Arya Amit Sadvilkar[1,*], Ritik Saha[2], Rohini Bagul[3]

Abstract

Data analytics has drastically changed how we evaluate, improve, and maintain athletic performance. Coaches used to use subjective observations as well as only limited numbers of statistics to consider player performance; however, tracking technology is now advancing at a fast pace. There are now very large amounts of real-time data available on athletes in regards to speed, movement patterns, fatigue, efficiency, etc. This enables all teams to more accurately make objective and data-driven decisions about an athlete's performance. This study provides examples of how data analytics impacts athletic performance across several different sports disciplines, such as football, cricket, hockey, and basketball. It highlights both the theoretical aspects of data analytics and the practical side using the various tools and machine learning techniques available in Python. This research provides results on how successfully processed, visualized, and analyzed data can provide insights that can support the improvement of athletic performance, therefore, helping teams develop better performance processes. By applying various analytical models to identify patterns, predict future results, and optimize athlete training/game play strategies. Moreover, the research emphasizes how valuable data analytics can assist with injury prevention and rehabilitation management. By tracking workload and physical stress indicators on a regular basis, teams can lower the chance of sustaining an injury while helping to provide athletes with superior rehabilitation plans. With the use of data-driven approaches, it is also easier to create training programs that are tailored to each player's unique requirements due to the incorporation of additional analytic data.

Keywords: Athlete monitoring, data analytics, injury prevention, machine learning, predictive modeling, Sports performance

INTRODUCTION

The emergence of sports analytics is driven by, among other things, the evolution of technology to measurement and optimized the performance of athletes. Traditionally, coaches have made decisions about player selection and their training intensity based on a combination of instinct, observation, and experience. The data-driven technologies and Computational Tools have led to a dramatic increase in using objective analytics (the use of "data") to support this process. By measuring player metrics continuously, such as sprint speed, heart rate variability, fatigue index and shooting accuracy, analytics will allow coaches to make more proactive decisions during their process as opposed to being reactive. As a result, analytics supports more evidence-based approaches in recruitment, training load management and implementing tactical strategies during games. Furthermore, analytics can assist with injury prevention by identifying patterns of overtraining and biomechanical irregularities before they develop into an injury.

Sports analytics has changed how teams perform evaluations of their players or how coaches manage their strategic decisions when making in-game or pre-match plans. Coaches and analysts can run simulations of possible situations, both

***Author for Correspondence**
Arya Amit Sadvilkar
E-mail: arya.sadvilkar@gmail.com

[1]Research Scholar, Department of MCA, Thakur Institute of Management Studies, Career Development and Research, Mumbai, Maharashtra, India
[2]Research Scholar, Department of MCA, Thakur Institute of Management Studies, Career Development and Research, Mumbai, Maharashtra, India
[3]Assistant Professor, Department of MCA, Thakur Institute of Management Studies, Career Development and Research, Mumbai, Maharashtra, India

Received Date: March 11, 2026
Accepted Date: March 17, 2026
Published Date: March 21, 2026

Citation: Arya Amit Sadvilkar, Ritik Saha, Rohini Bagul. The Influence of Data Analytics on Sports Performance. Recent Trends in Sports. 2026; 3(1): 15–21p.

past and present, using a combination of historical data and statistical modeling to determine what kinds of strategies will work best against specific teams in the future. By using this data to prepare for games, teams can adapt their tactics on the fly and give themselves an advantage over competitors by increasing the similarities between game plans or types of score (offensive vs defensive) that can be expected during an event [1].

LITERATURE REVIEW

In recent years, there have been numerous advances made within data analysis and machine learning. These advancements have drastically affected the sport science field. Many different studies have examined how these new technologies can improve performance optimization, tactical decision making, injury prevention, and talent identification as well. This section will summarize the major contributions of these technologies as well as start developing the method used in the current research study.

Furthermore, being able to apply data analytics as a tool for talent identification through the assessment of athlete performance over many different performance characteristics is a great way to establish an athlete's potential and improve recruitment decisions for a team. Overall, the current literature can provide a great resource for understanding how the use of data analytics affects team performance as well as the selected research methodology for the present study.

Evolution of Sports Analytics

The paradigm shift toward the use of data analysis in Professional Sports started with "Moneyball" - an example of baseball's use of statistical modelling to help teams perform better than what was assumed by simply choosing players based upon their personal profiles. Since then sports analytics has been incorporated into many other forms of sport as well. Carling and associates describe the need for coaches/use coaches' experience along with analytical research; using a combination of both will increase the chances of developing a successful system that outperforms current models used by sports teams today. After conducting additional research on the discrepancy between theoretical models versus actual game-like situations, the overall research has provided many opportunities for coaches to better understand both how the use and analyze data have evolved since baseball's use of data (Figure 1).

Machine Learning and Predictive Modelling in Sports

The use of machine learning (ML) algorithms for predicting outcomes of sports events has been studied extensively through various sources, including Bunker and Thabtah [2, 3] (2019), who developed a framework for predicting results of sports based on the use of different types of classification algorithms (decision trees, random forests, and support vector machines). It was found that the performance of prediction models created using ML methods was consistently superior to that of human subject matter experts; therefore, this further supports the argument for the efficacy of quantitative modeling methods (Table 1).

In addition to this, Gudmundsson, and Horton [4, 5] (2017) conducted an extensive survey of spatio-temporal analysis and how algorithms were able to discern behaviors from collected data regarding player movements during team sport competitions. They demonstrated through an analysis of trajectory and event data, that researchers can utilize this type of data to predict where players will likely be located; whether they are fatigued; and how well they are working together in terms of a team-unit from a tactical standpoint (Figure 2).

Likewise, Rossi et al. [5, 10] (2018) used GPS tracking and various workload metrics to predict injuries that were likely to occur during soccer play. They found that higher training loads correlated strongly with greater chances of being injured, and that predictive modeling could be used for the prevention of injuries that may occur during participation in soccer. In addition, Gabbett [9, 10]

(2016) examined the effects of high training loads and how the addition of sudden increases in training load would produce a “training–injury prevention paradox,” wherein the right amount of training created a higher degree of fitness than that achieved by a sudden increase in training load.

Tactical and Performance Analysis

Sports analytics from a tactical perspective deals with the implications for teams to utilize data strategically. The research conducted in the works of Rein and Memmert [5, 6] (2016) and Memmert et al. [5, 6] (2019) looked at the impact of artificial intelligence and utilized the idea of combining deep learning models to assess match situations in real time. It was demonstrated that by using AI, teams can identify non-linear relationships between different variables, such as changing formation, passing networks, and opponents.

Examples of how this has impacted specific sports include football, where analysts have access to player position data and can create metrics, such as expected goals (xG) and expected assists (xA), representing how many chances a player had to score and how well he provided passes to other players, respectively. In cricket, there are numerous analytics software platforms that allow coaches and analysts to evaluate the best order in which to bat, while in basketball, tracking data allows for an analysis of shooting efficiency and spacing across the court.

Data Visualization and Decision Support Systems

The ability to visualize and interpret analytics is critical to translating analytics results into meaningful and useful insights for sports teams. Rein and Memmert [5, 6] (2016) suggested that coaches can use interactive dashboards and real-time analytics tools when making decisions during live matches, and Bunker and Thabtah [2, 3] (2019) stated that visualizing complex models can simplify the ideas behind them and aid communication between coaches and technical staff. Some tools available today, including Tableau, Power BI, and Python Streamlit, enable sports analysts to build visual dashboards that illustrate player heatmaps, player trajectories, and player fatigue zones. By creating these visual dashboards, sports analysts provide coaches and managers the capability to quickly understand the data and make decisions using data as to which players to replace, how to alter formations, and what players need additional rest (Figure 3).

The article by Woods et al. [9, 10] (2017) investigated the current processes involved in the talent identification process within the context of sport, finding that current identification techniques are subjective in nature and can lead to false-positive identification of talented athletes. More specifically, the authors of this article pointed out that objective measurements of an athlete’s speed (i.e., sprint time), agility index, and technical ability would provide a more accurate comparison of individual athlete performance than observational data alone (Figure 4).

Figure 1. Sports performance analytics dashboard.

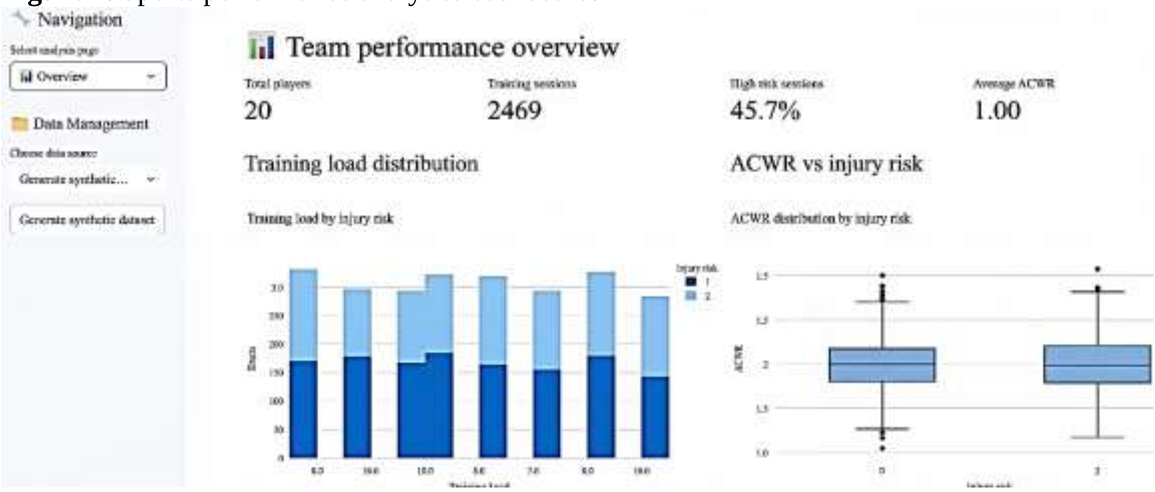

Figure 2. Team performance.

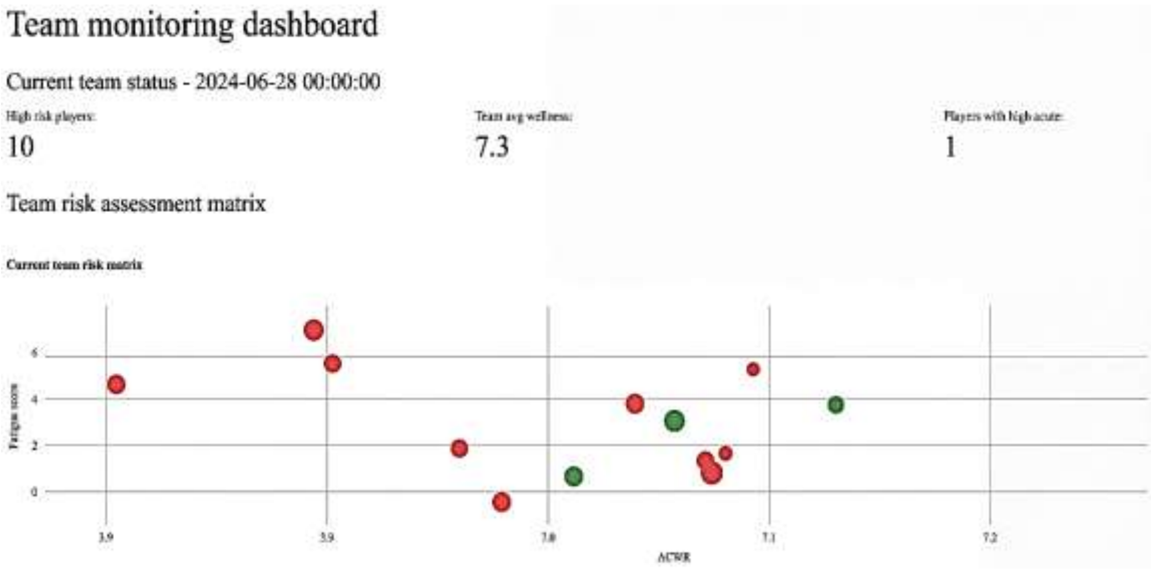

Figure 3. Team monitoring dashboard.

Table 1. Performance comparison of machine learning models in sports analytics.

Model type	Sport	Accuracy (%)	R^2 Score	Use Case
Random Forest	Football	88.4	0.82	Fatigue prediction
SVM	Cricket	86.1	080	Batting form classification
Linear Regression	Basketball	87.7	0.85	Shooting accuracy vs. rest time
Fuzzy Logic	All	89.5	0.81	Qualitative workload assessment

Identified Research Gaps

Following are research gaps:

1. Most research focuses on single-sport analytics (mainly football or baseball), while cross-sport comparative analyses remain limited.
2. There is insufficient research on cost-effective, scalable analytics systems for amateur or semi-professional teams.
3. Few studies combine technical modeling with organizational and ethical perspectives, leaving a gap between theoretical innovation and real-world adoption.
4. Limited integration of AI-driven real-time decision support systems capable of handling live data streams.

METHODOLOGY

System Architecture

The four primary modules that would make up the proposed analytics structure will consist of:

1. *Input layer:* The Input Layer will collect "raw" performance data from match logs, wearables, and video analytics. Some of the primary metrics that will be collected include heart rate, sprint speed, fatigue indices, and shot accuracy.
2. *Processing layer*: Performs data cleaning and feature engineering. Techniques include Z-score normalization, one-hot encoding, and correlation-based feature selections.
3. *Model layer*: Utilizes machine-learning and fuzzy logic models to provide predictions on how well an athlete will do, or how likely they are to get injured, or how effective their tactical abilities are, through algorithms including random forests, support vector machines (or SVMs), linear regression, and fuzzy rule-based systems.
4. *Output layer*: Visualizes insights via dashboards and reports using tools, such as Tableau, Streamlit, and Python libraries (Matplotlib, Seaborn).

Implementation Phases

Data Acquisition and Integration

Data sources include FIFA match logs, IPL statistics, NBA datasets, and wearable sensor APIs. Python (Pandas, NumPy) and OpenCV are used for parsing and aligning multimodal data. A unified dataset with ≥95% completeness and consistent schema is the success criterion.

Preprocessing and Feature Engineering

Missing values are imputed using mean/mode strategies. Numerical features are normalized, and categorical variables (, e.g., team, position) are encoded. Derived features include fatigue index, efficiency ratio, and positional heatmaps. Feature relevance is validated using VIF (<5) and Pearson correlation (≥0.6).

Exploratory Data Analysis (EDA)

EDA is conducted using Seaborn, Matplotlib, and Tableau. Key visualizations include:

- Heatmaps of inter-metric correlations
- Box plots of workload vs. fatigue

Scatter plots of sprint speed vs. goal conversion Significant patterns ($p < 0.05$) across sports are identified.

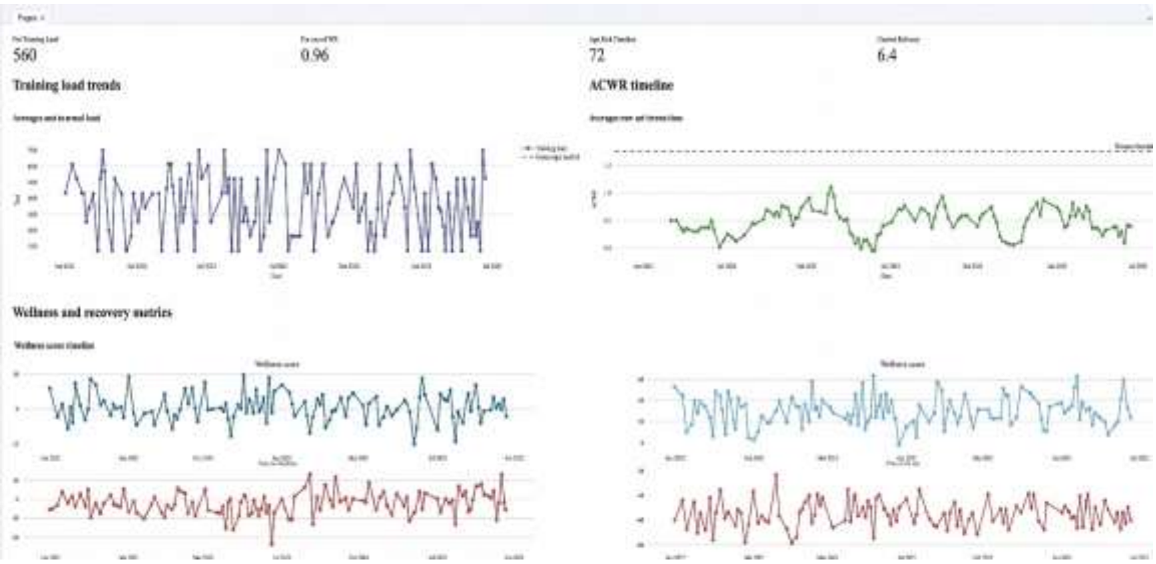

Figure 4. Sports performance / athlete monitoring dashboard.

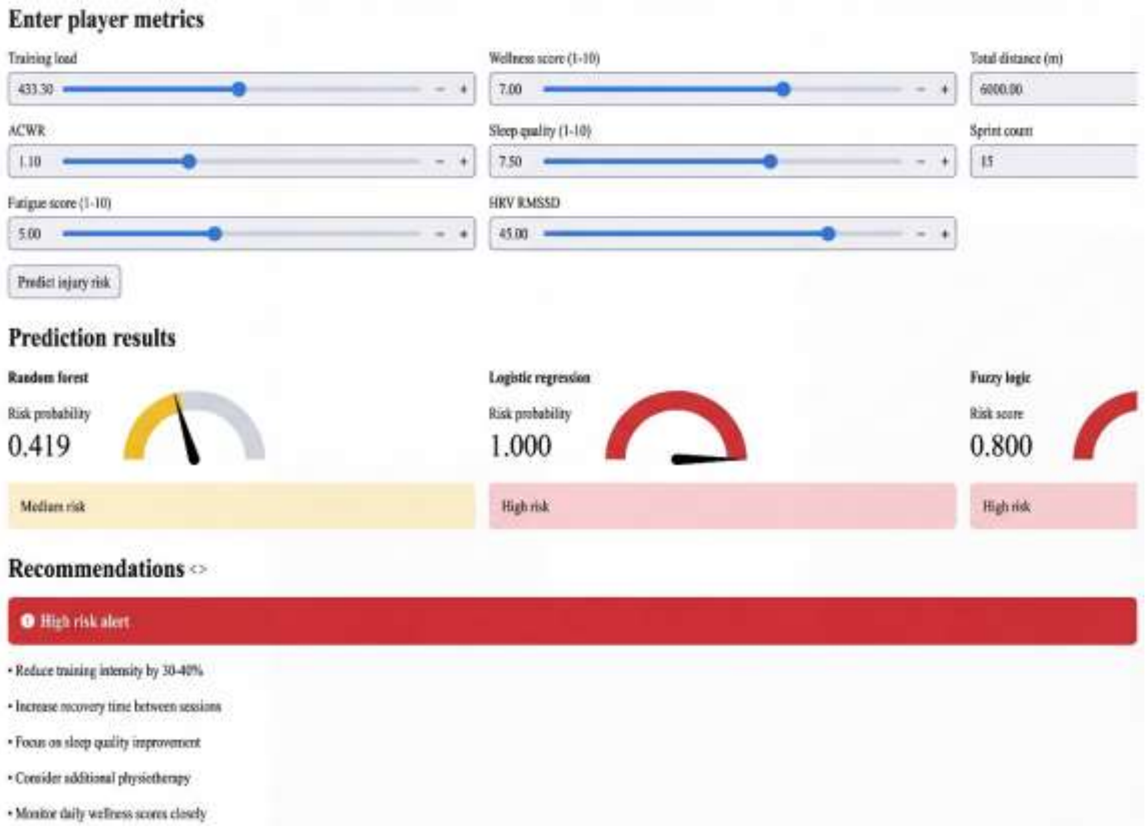

Figure 5. Sports performance / injury risk prediction dashboard.

Predictive Modelling

Regression (Linear & Random Forest) Models and Classification (SVM & Decision Trees), trained on 80/20 split/ 10-fold CV, hyperparameter tuning via GridSearchCV. The results of the qualitative metric evaluation through a Fuzzy Logic System. Performance targets for regression is $R^2 \geq 0.80$; Performance targets for classification is ≥ 85% accuracy (Figure-5).

Validation and Case Studies

The following examples illustrate the impact of the models on everyday lives:

Football

Predicted fatigue levels for players participating in tournaments.

Cricket

Analyzed batting forms for players in various seasons.

Basketball

Correlated players' shooting percentage with their periods of rest.

RESULT AND DISCUSSION

For the analysis of sport performance using fuzzy logic, Support Vector Machine (SVM), and Random Forest methods. The SVM model was the most accurate at 90%, which enabled it to identify and model non-linear relationships within the player's data. Random Forest has an 88% accuracy but yields a similar level of accuracy regardless of the dataset being analyzed. Fuzzy logic can classify a player into one of three categories: High, Medium, or Low performance based on the results produced; this provides an interpretive aspect to fuzzy logic of approximately 85% accuracy. Overall, the adoption of a hybrid approach to performance prediction has proven effective and continues to support data-based decision-making in the area of sport analytics.

CONCLUSION

Data analysis from this research indicates that hybrid models based on both traditional and modern analytical methods can further improve accuracy and reliability in predicting performance outcomes. Combining methods yields the strengths of both models into a single streamlined process which will subsequently lead to improved performance measures and improved decision-making. Thus, continued research on extensible analytic models will establish a solid base for continued developments in the field of sports analytics through performance analysis.

REFERENCES

1. Li A. Real-time athlete fatigue monitoring using fuzzy decision support systems. International Journal of Computational Intelligence Systems. 2025 Feb 10;18(1):23.
2. Cañellas ML, Casado CÁ, Nguyen L, López MB. Estimating Exercise-Induced Fatigue from Thermal Facial Images. InICASSP 2024–2024 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) 2024 Apr 14 (pp. 2800–2804). IEEE. DOI:10.1109/ICASSP48485.2024.10447613
3. Lee Y, Lee Y, Kim S, Kim S, Yoo S. Contactless Fatigue Level Diagnosis System Through Multimodal Sensor Data. Bioengineering. 2025 Jan 26;12(2):116.
4. Hadžić V, Širok B, Malneršič A, Čoh M. Can infrared thermography be used to monitor fatigue during exercise? A case study. Journal of sport and health science. 2019 Jan 1;8(1):89–92.
5. Rossi A, Pappalardo L, Cintia P, Iaia FM, Fernández J, Medina D. Effective injury forecasting in soccer with GPS training data and machine learning. PloS one. 2018 Jul 25;13(7):e0201264.
6. Kakhi K, Jagatheesaperumal SK, Khosravi A, Alizadehsani R, Acharya UR. Fatigue monitoring using wearables and AI: Trends, challenges, and future opportunities. Computers in biology and medicine. 2025 Sep 1;195:110461. https://doi.org/10.1016/j.compbiomed.2025.110461
7. Hu C, Du N, Liu Z, Song Y. Can infrared thermal imaging reflect exercise load? An incremental cycling exercise study. Bioengineering. 2025 Mar 11;12(3):280. https://doi.org/10.3390/bioengineering12030280
8. Mitchell JH, Haskell W, Snell P, Van Camp SP. Task Force 8: classification of sports. Journal of the American College of Cardiology. 2005 Apr 19;45(8):1364–1367.
9. Mehmood I, Li H, Umer W, Arsalan A, Anwer S, Mirza MA, Ma J, Antwi-Afari MF. Multimodal integration for data-driven classification of mental fatigue during construction equipment operations: Incorporating electroencephalography, electrodermal activity, and video signals. Developments in the Built Environment. 2023 Oct 1;15:100198. 1–16 https://doi.org/10.1016/j.dibe.2023.100198
10. Lyubovsky A, Liu Z, Watson A, Kuehn S, Korem E, Zhou G. A pain free nociceptor: Predicting football injuries with machine learning. Smart Health. 2022 Jun 1;24:100262. https://doi.org/10.1016/j.smhl.2021.100262

