



SPORTS MATCH PREDICTION MODELS: A MINI-REVIEW ACROSS MULTIPLE SPORTS

Narender Singh^{1*}, Dr Sangeeta Singh², Dr Vijay Prakash³, Prof. Varendra Singh Patial⁴

¹Research Scholar, Department of Physical Education IIMT University Meerut, UP, India.

^{2,3}Assistant Professor, Department of Physical Education IIMT University Meerut, UP, India.

⁴Dean, Department of Physical Education IIMT University Meerut, UP, India.

***Corresponding Author**
Narender Singh

Research Scholar,
Department of Physical
Education IIMT
University Meerut, UP,
India.

Article History

Received: 15.07.2025
Accepted: 10.08.2025
Published: 29.08.2025

Abstract: Sports prediction models are a vital area of sports analytics that uses statistical modeling, data mining, and machine learning to predict results and guide decision-making in the sports industry. Any kind of coach, analyst, bettor or fan can benefit from the football, handball, cricket, tennis, basketball forecasting abilities. However, the literature considered indicates a methodology progression, and the use of traditional statistical models, such as Logistic Regression, Poisson Regression, Gaussian approximation, etc. was superseded by the use of modern artificial intelligence techniques, such as adaptive back-propagation neural networks, ensemble learning methods, and “explainable AI integrated with deep learning”. Conclusions indicate that, in distinct sports, different information is relevant: in football defensive efficiency and shooting accuracy is dominating, whereas in cricket overs and wickets influence game outcomes, in tennis game performance predicting momentum is important and in basketball the recent number of won games proves very useful. In the meantime, studies in handball are growing rapidly and include the use of a sensor-based inertial measurement unit (IMU) and computer vision techniques. This paper synthesizes contributions from 2016 to 2025, identifies methodological strengths and limitations, and proposes future directions focusing on generalizable models, multi-modal data integration, and transparent explainable systems.

Keywords: sports analytics, machine learning, match prediction, football, handball, cricket, tennis, basketball, artificial intelligence.

Cite this article:

Singh, N., Singh, S., Prakash, V., Patial, V. S., (2025). SPORTS MATCH PREDICTION MODELS: A MINI-REVIEW ACROSS MULTIPLE SPORTS. *ISAR Journal of Multidisciplinary Research and Studies*, 3(8), 43-50.

1. INTRODUCTION

Sports prediction research has become increasingly prominent in the past decade, driven by the dual forces of technological advancement and the growing availability of rich sports datasets (Kumar, 2018; Gautam & Kumar, 2018). Traditionally, outcome prediction in sports relied on expert intuition and historical statistics, but the emergence of machine learning (ML) and artificial intelligence (AI) has enabled more precise, automated, and scalable approaches (Prasetio & Harlili, 2016; Li, 2020; Jadaun et al., 2021). Sports match prediction is now viewed not only as an academic pursuit but also as a commercial and practical tool for betting industries, sports federations, coaching staff, and media companies (Kumar, 2022; Jain et al., 2023).

Beyond curiosity, the reason behind sports forecasting lies in the manner in which it is utilized for the tactical preparation, administration of players, injury avoidance and fan involvement. (Khare et al., 2023; Aakash et al., 2023). For example, a football club may use predictive insights to optimize formation against a rival, while betting companies employ predictive models to calibrate odds (Kumar, 2023a). Similarly, handball researchers increasingly focus on movement prediction using wearable technology such as inertial measurement units (IMUs), aiming to quantify both locomotion and technical actions (Lentz-Nielsen et al., 2023; Kumar & Jhajharia, 2018). These advancements mark a shift from outcome-focused models (who wins or loses) to performance analytics integrating biomechanics, movement tracking, and psychology (Kumar et al., 2021; Kumar, 2023b).



Several key questions arise in this domain: Which modeling techniques best capture the variability of sports events? Are there common prelim factors for all sporting disciplines or specific prelim factors for each discipline? How to satisfy the conflicting objectives of model accuracy and model understanding and explanation? (Kumar & Jhahharia, 2022; Nandal & Kumar, 2024)? Research in many recent studies shows a trend, a methodology shift from the classical regression-based models, such as logistic regression or models based on Poisson distributions (Groll et al., 2019; Das et al., 2022) to the use of advanced neural networks, ensemble learning, and deep learning. (Ahsan et al., 2024; Song et al., 2024; Kumar, 2023c, Singh et. al. 2025). Meanwhile, new trends focus on explainability about AI, such as an explainable AI (XAI) and an explainable Analysis (Shap) (Felice, 2024; Song et al., 2024; Kumar & Jhahharia, 2020).

Given this background, the aim of this review is to collate work in several sporting areas (football, handball, cricket, tennis and basketball) to trace the methodological development, their Sport Specific Predictors (SSP) and note any trend. (Jain et al., 2023; Kumar, 2023).

2. LITERATURE REVIEW

2.1 Handball Prediction Models

Research on handball prediction has been on the increase, with various types of studies. Groll, C. and others (2019) proposed an under dispersed Poisson and Gaussian modelling approach to foresee the results of the 2019 IHF World Men's Handball Championship. They compared the Poisson and Gaussian response regression models and negative binomial regression models using historical championship data from 2011-2017. The model that performed best turned out to be the Gaussian response model which allowed for predictions of tournament results and winning probabilities for all the teams in the tournament. In this model, Denmark is projected to be the biggest winner, trailed by France.

In extended version, Groll et al. (2020) used a sparse version of the Gaussian approximation model, further proving the advantages of Gaussian regression in match score of handball. These statistical models offer chances of advancing at each phase, a helpful aid for anybody from head instructors to experts.

Later research added to the toolkit and method comparisons were made. Later studies extended the methodological repertoire. Lentz-Nielsen et al. (2023) used IMUs and extreme gradient boosting for events classification in locomotion and throwing. The model performed very well on throw events ($F1 = 0.95$) and in running ($F1 = 0.86$), but was average on dynamic movement ($F1 = 0.66$). This emphasised the biomechanical characteristic of interquartile range (IQR) and the zero crossing frequency (ZCF) as important predictors.

For the Paris 2024 Olympics, Felice (2024) brought the concept of explainable artificial intelligence (xAI) with the introduction of deep learning. This model was able to predict match results and give interpretable explanations with the use of Large Language Models (LLMs). This type of integration with xAI helps to ensure that the predictions it can make are put to use for coaches; often in AI-based analytics, the issue is known as the “black box” problem.

Lastly, by using computer vision and logistic regression along with multiple object tracking, Kobayashi et al. (2023) showed the use of

computer vision in predicting outcome in 2-vs-2 handball games. They were able to reach 60% accuracy even without using any sensors, providing usable solutions in nonprofessional environments.

2.2 Football (Soccer) Prediction Models

In the past it has been football that had firmly taken over the mantle of sports analytics due to the number of matches played and how widely followed the sport was. Most early studies were based on regression techniques. Prasetyo and Harlili constructed a logistic regression model on Barclays Premier League data, being 69.5% accurate. They had a system that highlighted defensive categories like “home defense” and “away defense” were important indicators.

Li's (2020) model, which was an adaptive back-propagation neural network (ADBPNN), was proposed on this basis. The performances of BPNN were compared with multiple linear regression (MLR) and grey prediction and the BPNN gave nearly zero error, showing the advantage of the neural network in reflecting the complex non-linear phenomenon.

Das et al. (2022) used SR data of ISL to perform logistic regression to find that successful passes, shots on target, fouls and red cards were the best predictors. They pointed out the importance of domestic leagues data in bringing out a context-specific aspect which differs from European leagues.

Ensemble models were tested during Qatar 2022 World Cup. Song et al. (2024) also evaluated on the same data set with other classifiers ANN, SVM, AdaBoost and Random Forests, with ANN obtaining the highest predictive accuracy (75.42%). With the help of the SHAP analysis, they decided the features that affect the shooting accuracy and progression of the balls were the most crucial. This highlights the significance of explainable frameworks in contemporary sports analytics.

Finally, Wang et al. (2025) infused sports betting with ML. The models applied to the European leagues outperform other state-of-the-art approaches at the expense of modest accuracy (~52%) and showed profits of 3% at every simulated betting scenario.

2.3 Cricket Prediction Models

There's a lot of data when it comes to cricket and the reason for this is there are a huge amount of wrinkles to the game, including ball-by-ball structure, highly detailed scorecards and the sheer variety of formats: Test cricket, One-Day Internationals (ODIs) and the Twenty20 (T20) version of the sport. This means that it provides ample opportunities to gain insight through predictive modelling, as it takes into account both macro-level (team makeup, venue, toss results, weather conditions) and micro-level (run scoring, wicket falls, strike rates and partnerships). Cricket is different from basketball and football, which are continuous actions, where the required models are much more general.

In order to bridge this gap, Ahsan et al. (2024) tried their best to create an all-round prediction system for both formats. They used classification and regression techniques such as Naïve Bayes, Gradient Boosted Trees, Logistic Regression, Deep Learning and Random Forests. Gradient Boosted Trees proved to be the most accurate machine learning model (96.15%) for score prediction. With respect to match outcomes, the Generalized Linear Model (GLM) did comparatively better with 71.72% accuracy. They

validate an important observation: although models that predict continuous outcomes (e.g., total runs) might be superior, they do not necessarily perform well for binary outcomes (e.g., win/loss).

An improvement was made by Rahman Mahin and co-authors (2024) who used the “resource factor” to incorporate domain-specific knowledge into predictive models. Two ODI figures are universally recognised as the two most important by players and analysts alike that were used in the novel metric were the number of overs remaining and wickets in hand. They got very good performance results from their Random Forest model with R2 value: 0.9899, which is almost a perfect fit model, and extremely small fitting margins (MAE = 3.3569 and MSE = 37.3555). Perhaps most relevantly, they also built a web-based application to deploy the model, so as to make the research essentially available to coaches, analysts and fans.

Cricket research relies fundamentally on the understanding that crucial aspects of the sport are what they call “Sport-Specific”. In cricket, unlike football and basketball, the prediction depends on many factors beyond simply analysis of performance (e.g., the stability of the batting order, bowling economy, powerplay overs and the rotation of the team's bowlers). If such features are not incorporated the models will perform poorly. The specific amount of emphasis given to understanding the cricket-specific structures also constitutes one reason as to the effectiveness of some methods such as Random Forests, Gradient Boosted Trees, or even very simple models; namely linear regression models, which do very poorly at capturing complex, non-linear interactions between overs, wickets, and scores.

Nevertheless, challenges remain. Firstly the diversity of cricket's format is a major challenge to model generalisation. If a model is developed (with ODI data) but applied to T20 but test cricket, there is possibility of differences between the two forms with regard to pace, strategy and risk and hence a failure in the model. Second, other conditions that affect performance, like pitch conditions and weather (which is known to affect the result), are seldom included. Hybrid models combining the hot-spot analysis derived from the meteorology and the crowd dynamics as well as integrating human fatigue parameters, along with classic game statistics, with future work would be beneficial.

2.4 Tennis Prediction Models

Unlike team games, tennis has a structure which makes it unusual in terms of predicting outcomes: the game is binary (win/loss), and performance is shaped by a sequence over all points of the game. Tennis tournaments had a sort of psychological metabolism that wasn't captured by traditional prediction models and was better described as momentum or flow.

Wang (2024) filled this research void by introducing a quantitative measure of momentum. They combined the importance of break of serve (34.1%), serve performance (24.3%), scoring dynamics (38) and net points (3.6) to make a composite index. This “momentum” variable then was correlated with match outcomes with Pearson's correlation coefficient, and an admirably high correlation was found of 0.807. Logistic regression models that used momentum were high in predicting the accuracy.

It is not just statistical modeling that is novel; Wang's contribution is also a psychological/statistical interplay of performance. Momentum has multiple facets—technical and mental—because as

someone breaks serves and gets off the juju, they are putting pressure on the other player both psychologically and mechanically. Wang's study adds to the understanding of the translation of performance oscillations to outcome by embedding it into the predictive frameworks.

Applied, momentum models may aid coaches in match strategy in real-time. In fact, some simple changes can make waves when momentum is waning, such as slowing down the tempo of the game or altering serving patterns. Furthermore, models like these would benefit betting markets, since momentum is having predictive power in between matches, as it's having predictive power even beyond pre-match rankings or seedings.

This approach could be extended to future tennis prediction studies that include biometric analysis (heart rate variability, exertion) and computer vision analysis (serve speed, shot placement). This would enable the development of many-levelled models that are able to not only predict the results, but also provide an explanation as to how and why momentum can change.

2.5 Basketball (NBA) Prediction Models

The issues with predicting in basketball are associated with the nature of scoring and the high tempo and player rotation. Outcomes are complex and must be separated out between players, situations (homefield advantage, did they just play two games in a row) and team form.

Yao (2023) performed exploratory data analysis and data mining for the match prediction of NBA games with both regression and machine learning methods. Logistic and Random Forest models performed better than the Naïve Bayes model, achieving an average accuracy of 67% in logistic regression, better than the average accuracy of the other two models. One of the biggest factors turned out to be that most critical one in basketball outcomes, recent win rate (correctly predicted wins over the last month).

This fit very well with basketball's nature, in that streaks and player momentum can make a huge difference in the game. Whereas in football or cricket, you can't escape external circumstances (pitch conditions, toss), the result of basketball games can rely on the fitness and consistency of the players in their recent games.

The strength of the study also pointed out the importance of optimal data setting selection. Experiments showed that setting up datasets from the last season was a better predictor than using long-term historical data. This indicates that things in basketball change – the fact that players have traded places, suffered injuries or changed coaches and coaches, make past information significantly less relevant to the game.

In the future, data about where players are putting in shots, which players they will be playing against, and where they rebound, could be incorporated into predictive models. Fitting logistic regression or deep learning based models like ensemble ML with such spatial-day time data can produce accurate but also interpretable models.

3. METHODOLOGY

3.1 Research Design

This study utilises a narrative mini-review methodology—known as the ‘mini-review’ and which has proved appropriate for drawing

the conclusions of this study when drawing on a relatively diverse set of studies in a number of sports. In contrast to systematic review descriptions, like PRISMA, which have strict specifications, narrative reviews are more flexible. The aim of the present article is not to statistically meta-analyze the effect sizes but to highlight methodological developments, sport-specific explanatory factors and patterns of themes throughout football, handball, cricket, tennis and basketball.

3.2 Source Identification and Selection

Fourteen peer reviewed studies published from 2016-2025 were identified for inclusion in this review. The studies that were discovered from this data set included conference papers, journal articles and arXiv preprints, which is a mix of theory and practice. These were then investigated in each study for the following:

1. **Sport context** (football, handball, cricket, tennis, basketball).
2. **Prediction focus** (match outcome, score forecasting, or movement classification).
3. **Modeling technique** (statistical regression, neural networks, ensemble ML, deep learning, explainable AI).
4. **Performance metrics** (accuracy, F1-score, R^2 , correlation coefficients, etc.).

3.3 Inclusion and Exclusion Criteria

Inclusion criteria were threefold:

- The study must have **prediction of sports outcomes** (e.g., match results, scores, or player movements) as a central objective.
- It must apply **quantitative statistical or machine learning techniques** rather than purely qualitative or descriptive approaches.
- It must report measurable outcomes that allow for comparison (e.g., prediction accuracy, error rates, or correlation values).

Studies were excluded if they:

- Focused solely on **descriptive analytics** without predictive modeling.
- Examined **non-competitive sports settings** (e.g., recreational training without competitive context).
- Failed to provide clear methodological details or quantitative results.

3.4 Data Extraction and Synthesis

All of the studies were closely examined and salient features were tabulated on a comparative matrix. These included:

- **Sample size and dataset** (e.g., UEFA Champions League matches, Indian Super League data, World Cup data, cricket scorecards).
- **Modeling approaches** (e.g., logistic regression, Poisson regression, Random Forest, Gradient Boosted Trees, Deep Learning).
- **Reported findings** (e.g., ANN achieving 75.42% accuracy in World Cup prediction; Random Forest achieving $R^2 = 0.9899$ in ODI cricket).

A thematic grouping was then used. The studies were grouped according to the sport as well as according to the approach used in the study. In some cases, multiple regression models were compared and in other cases, multiple ensemble ML models were compared within each study (cricket vs football, football vs basketball, etc.).

3.5 Limitations of Methodology

This review represents a general grouping of the findings, but there are some limitations:

1. **Time frame bias:** The dataset spans less than a decade (2016–2025). Older foundational studies (pre-2016) were not included, which might underrepresent early statistical contributions.
2. **Sports coverage bias:** Sports may not have been adequately covered, especially when spoken about a sport not covered in the survey and they are new to it.
3. **Comparability issues:** Some studies used different ranges of accuracy, such as percentage accuracy, F1-score, or R^2 , and it therefore was difficult to compare them.
4. **Publication bias:** Many of the studies come from conferences or arXiv, where experimental methods may be less rigorously peer-reviewed compared to established journals.

While there are some restrictions, the methodology seems to be suitable for determining trends, gaps and cross-sport elements, which are the main aims of this mini-review.

4. RESULTS

Table 1. Comparative Summary of Sports Prediction Models (2016–2025)

Sport	Study (Year)	Dataset / Context	Model(s) Applied	Best Performance Metric / Findings
Football	Prasetio & Harlili (2016)	Barclays Premier League (6 seasons)	Logistic regression	Accuracy 69.5%; defensive statistics was most predictive
	Li (2020)	UEFA Champions League	Back-propagation neural network (BPNN), MLR, Grey models	BPNN had minimal prediction error and superior to regression
	Das et al. (2022)	Indian Super League	Logistic regression	Significant in passes, shots, fouls and contextual domestic insights
	Song et al. (2024)	FIFA World Cup 2022	ANN, SVM, AdaBoost, Random Forest + SHAP	ANN Accuracy 75.42%; shots on target and ball progression are most influential
	Wang et al. (2025)	European leagues (betting context)	LightGBM, AdaBoost	Accuracy 52% but consistent 3% profit margin
Handball	Groll et al. (2019)	IHF Championship (2011–2017)	Poisson, Gaussian regression	Gaussian response is most robust predictor
	Groll et al. (2020)	IHF Championship	Sparse Gaussian approximation	Enhanced tournament simulation robustness
	Lentz-Nielsen et al. (2023)	Player IMU data (locomotion, throwing)	XGBoost classification	F1 = 0.95 for throws; F1 = 0.86 for running
	Kobayashi et al. (2023)	2v2 Handball video footage	Multiple object tracking + logistic regression	~60% accuracy without sensors
	Felice (2024)	Paris 2024 Olympics	Deep learning + explainable AI + LLMs	Accurate forecasts with interpretability via xAI
Cricket	Ahsan et al. (2024)	T20 & Test matches (7,827 samples)	Naïve Bayes, Random Forest, GBT, Deep Learning, GLM	GBT Accuracy 96.15% for score prediction; GLM best for win or loss
	Rahman Mahin et al. (2024)	ODI cricket	Random Forest + novel “Resource Factor”	R ² = 0.9899; MAE = 3.36 runs
Tennis	Wang (2024)	Professional tennis matches	Logistic regression with “Momentum” index	Momentum correlated r = 0.807; strong predictor for win or loss
Basketball	Yao (2023)	NBA (recent season)	Logistic regression, RF, Naïve Bayes	Logistic regression accuracy = 67%; recent win rate most decisive

Table 1 presents the results of the evolution of method, which is evident in a progressive change with regard to sport. Then, logistic regression (Prasetio & Harlili, 2016) was applied in football, in which the simplicity shows that defensive efficiency is emphasized, getting only moderate results of 69.5%. Later, neural network (Li, 2020) models were capable enough to catch the complex dynamics in UEFA data and they outperform the traditional models. Another set of more recent ensemble based and deep learning models (Song et al., 2024; Wang et al., 2025) pointed out that success with lower accuracy (~52%) is possible when they are combined with strategically formed betting odds, though accuracy itself improved moderately (75% for ANN). Sports data aggregated across teams show that whereas some of

these predictors are quite complex, including shooting accuracy, number of passes and defensive solidity, they are also constant across all methods.

The development of handball research was different. Previous regression analysis (Groll et al., 2019; 2020) has established that Gaussian methods are best suited for predicting tournament results. But the field broadened to the biomechanical and sensor-based prediction. Near perfect F1-scores were achieved by Lentz-Nielsen et al. (2023) to classify throw events with IMU and XGBoost, and Kobayashi et al. (2023) showed the possibility of outcome prediction by tracking only video (~60%). Felice (2024) improved the progression by introducing deep learning coupled with explainable AI for predicting the outcome of the Paris 2024

Olympics, with the goal that the predictions would be understandable by the coaches. The change shows that handball can no longer only be predicted as a statistical match outcome but extends beyond that into analyzing the movements of the players.

The most accurate predictions were obtained in the case of cricket. Ahsan et al. (2024) noted that Gradient Boosted Trees can accurately predict scores, with a high score of 96.15% across all formats, but also noted GLM was more stable on binary outcomes of win/loss. The solution approached by Rahman Mahin et al (2024) was “resource factor” (overs + wickets) teaching their Random Forest models almost a perfect R^2 of 0.9899. This shows a ball-by-ball basis in cricket is optimal to models to exploit domain-specific variables with unparalleled precision. But it also points to the special nature of cricket; the structured nature of the game may be different to other continuous sports.

As far as I could see, Tennis prediction was an example of innovation with regards to psychology. Wang (2024) proposed a momentum as an index that took a quantitative approach for serve dynamics, breaks and scoring flow. Momentum had a correlation value of 0.807, which was similar to that of outcomes, and thus was found to be a powerful predictor in the models built on logistic regression. TenNets emphasize the linkage of psychological flow and technical action in predictive systems, a special quality in tennis. This is a special quality in tennis which at the same time in a certain way is particularly highlighted in the tennis predictive systems: It is the linkage of psychology with techniques.

Simplicity was the name of the game in basketball analysis. Yao (2023) has concluded that logistic regression, one of the simplest models, outperformed the other two models, Random Forest and Naïve Bayes, with 67% accuracy for the NBA outcomes. Most importantly: The most reliable indicator proved to be recent success rate, which has reaffirmed the importance of the short term performance and team momentum in basketball. This discovery highlights the fact that in some situations, simple, readily understood and well-understood predictors represent the superior model architecture.

5. DISCUSSION

The literature reviewed demonstrates a decade long methodological progression from the classic regression methodologies towards ensemble machine learning, and finally, deep learning combined with explainable AI. This trend is indicative on a larger scale of how data science progresses as well as that predictive success in sports is sport-specific and context dependent. Three important dimensions are explored in this discussion: methodological changes, sport-specific predictors, and practical implications, and then limitations and future avenues are critically evaluated.

Statisticians and their models dominated sports prediction in the 4 years from 2016 to 2019. These models (for example logistic regression and Poisson regression) were both highly interpretable, and also very economical in computing. One such study was conducted by Prasetyo & Harlili (2016) on Premier League games, who were able to use logistic regression to come up with some of the key factors for predicting home and away defence. Likewise, Groll et al., (2019), showed that the Gaussian regression model successfully predicted handball championship results. These did

not capture team sport nonlinear relations, were transparent, but predictively only moderately ($\approx 65\text{-}70\%$).

The era of ensembled learning and neural networks began in 2020 and continued through 2023. The models employed neural networks with a minimum error in predicting the UEFA results, Random Forests and Gradient Boosted Trees (Ahsan et al., 2024) performed well in the cricket sector. These models were able to well explain non-linear interactions between several variables (such as passes, wickets, overs and players rotation). But ensemble techniques had the drawback of overfitting, especially in sports with small data sets like handball or tennis. During this time, there was a trade off between accuracy and generalizability.

Deep learning and explainable AI (XAI) are emphasized in the most recent years (2024-2025). To make predictions for the upcoming football match such as the World Cup, Song et al. (2024) applied the use of Artificial Neural Network together with the help of SHAP analysis to help them identify the key features that should be considered, such as the total number of shots on target. To drive innovation in handball analysis with AI, Felice (2024) used deep learning and large language models to create understandable predictions for the Paris Olympics. This is a new paradigm – accuracy isn't enough – the model needs to be understandable and usable by humans as well. Wang et al. (2025) then further increased this—showing that ML could generate revenue from betting applications even with small probabilities of predicting outcomes, further proving that predictive models have commercial applications.

Overall in all sports, sport-specific variables always outperform generic measures. Passing accuracy, ball progression, and shooting efficiency were shown to be critical characteristics in football (Das et al., 2022; Song et al., 2024). Biomechanical features have been shown to be more predictive in handball with features based on IMU, such as throw frequency and locomotion intensity, recording near-perfect classification scores (Lentz-Nielsen et al., 2023). The resource factor quantitatively showed Cricket's strength logic: Overs and wickets are winner predictors, as found by Rahman Mahin et al. (2024). A first: psychological factors such as technical and mental resilience were incorporated into tennis prediction models; in another sport, Wang (2024) demonstrated how momentum, a phenomenon of both skill and mindset, was an effective predictor. When it comes to basketball, recent win rate has proven to be more powerful than more complicated measures (Yao, 2023). Combined, the results show that incorporating domain knowledge into models can greatly improve the predictive performance significantly.

The applications of these models are quite voluminous. Coaching – predictive insights can inform your tactical preparation: football managers can focus on passing and shooting, in cricket, can simulate overs-and-wickets scenarios, and in tennis, can plan moments for recovery. Wearable and vision-based handball models can be used for detecting biomechanical stressors in sport sciences and injury prevention, providing a before-the-fact prevention of overuse injuries. Wang et al. (2025) demonstrated that small-scale accurate ML models could be used to generate a margin in the gambling arena; however, ethical issues around the responsibility for gambling should be addressed. Last, prediction probabilities in fan engagement and broadcasting can provide live tactical insights to fans, enhancing the commentary of gamers or sports experts.

Although major advances have been made, there are still a number of important constraints. Many studies are based on small samples (particularly handball and tennis) therefore possibly lacking in generalizability and increasing the danger of overfitting. It can be difficult to make comparisons across sports using prediction metrics, as they are heterogeneous (accuracy, f1 score, R^2). Existing models do not incorporate external factors that have been known to influence results, like weather in cricket, injuries in football, referee actions in basketball, and crowd influence in tennis. Moreover, the variety of the format of cricket (Test, ODI and T20) is an example of an inherently contextual problem of building generalizable models across formats. To prevent predictive models from becoming very accurate for certain data sets but inadequate in more general ones, these deficits need to be addressed.

Multi-modal data integration is the key to the future of sports prediction. Match statistics when combined with biometric monitoring, wearable sensors and video tracking data will form whole-prediction systems. There's even the realm of real-time predictive analytics, which can be used, say, in tennis, when a ball stops in play and the score changes in a way that can change the nature of the game. Generalizability across contexts should be prioritized as well too; for example cricket models must be adaptable over different forms of the game and basketball models over seasons where there are frequent trades and injuries. Lastly, explainability and ethics continue to be important with respect to AI. Sports prediction software should be deployed with caution and be usable by both coaches and fans. Large language models that provide natural language explanations, XAI (Explaining Artificial Intelligence) and techniques like SHAP will play a key role in making algorithms more accessible and understandable to humans.

6. CONCLUSION

This paper vividly shows the significant progress sports match prediction has accomplished in the previous ten years. The focus on football and cricket dominates the research but other sports have their own insight to bring to the field; handball, tennis, and basketball. The shift from regression to neural networks to explainable AI is consistent with the trends in the field of data science.

The most important rule is that predictors have to be sport-specific: when it comes to cricket, they are overs and wickets, in tennis they are momentum and in football they are defensive efficiency and, when it comes to basketball they are recent form. Highest accuracy is obtained in using models incorporating these personalized variables.

As the field progresses, striking a balance between accuracy and interpretability, and at the same time between the field and practical applications. A multi-modal data integration, real-time prediction and coach-friendly interpretability will set the tone for the future of research.

Sports prediction is not a hypothetical yet, it's a practical application science influencing training methodologies for athletes, fan interactions with the sport, and the world's experience of sports.

References

1. Aakash, Sisodia, A., Kumar, A., & Jain, I. (2023). Patanjali's Ashtanga Yoga: A Bibliometric Analysis of Its Impact on Health and Well-Being. *Journal for ReAttach Therapy and Developmental Diversities*, 6(1), 1003–1013. <https://doi.org/10.53555/jrtdd.v6i1.2470>
2. Ahsan, Z., Ghumman, S., Rehman, A., Javaid, S., Ali, T. M., & Mir, A. (2024, March). A comprehensive prediction model for T20 and Test match outcomes using machine learning. *Proceedings of the 2024 International Conference on Engineering & Computing Technologies (ICECT)*.
3. Avinsh Nandal, Dr. Ajay Kumar. A comparative analysis of lower limb explosive strength between judo and Kho-Kho players. *J Sports Sci Nutr* 2024;5(1):25-26. DOI: 10.33545/27077012.2024.v5.i1a.231
4. Das, R., Jhajharia, B., & Das, P. K. (2022). Prediction model of success and failure in football competitions. *International Journal of Research Pedagogy and Technology in Education & Movement Sciences*, 11(2), 45–52.
5. Felice, F. (2024). AI for handball: Predicting and explaining the 2024 Olympic Games tournament with deep learning and large language models. *arXiv preprint arXiv:2403.01542*.
6. Gautam, R. K., & Kumar, A. (2018). Effects of deprivation of sleep on the catching accuracy in cricket. *International Journal of Physiology, Nutrition and Physical Education*, 3(2), 14-16. Available at <https://www.journalofsports.com/archives/2018/vol3/issue2/3-1-470>
7. Groll, A., Heiner, J., Schauburger, G., & Uhrmeister, J. (2019). Prediction of the 2019 IHF World Men's Handball Championship—An underdispersed sparse count data regression model. *arXiv preprint arXiv:1901.01786*.
8. Groll, A., Heiner, J., Schauburger, G., & Uhrmeister, J. (2020). Prediction of the 2019 IHF World Men's Handball Championship: A sparse Gaussian approximation model. *Journal of Systems Architecture*, 107, 101723. <https://doi.org/10.1016/j.sysarc.2020.101723>
9. Jadaun, R., Kumar, D. A., Singh, D., & Sisodia, D. A. (2021). A Comparative Study of Covid-19 Pandemic on Mental Toughness of National Level Players of Selected Team Games. *Nat. Volatiles & Essent. Oils*, 8(6), 6583-6587. <https://doi.org/10.53555/nveo.v8i6.5618>
10. Jain, I., Sisodia, A., Kumar, A., & Aakash. (2023). Yoga as a Viable Non-Pharmacological Approach for Primary Dysmenorrhea: An In-Depth Review and Meta-Analysis. *Journal for ReAttach Therapy and Developmental Diversities*, 6(1), 1014–1025. <https://doi.org/10.53555/jrtdd.v6i1.2471>
11. Khare, S., Reddy, T. O., Kumar, A., & Sisodia, A. (2023). Muscle architecture and sports performance. *International Journal of Physiology, Nutrition and Physical Education*, 4(2), 254-257. <https://doi.org/10.33545/27077012.2023.v4.i2d.224>

12. Kobayashi, R., Umemoto, R., Takeda, K., & Fujii, K. (2023). Score prediction using multiple object tracking for analyzing movements in 2-vs-2 handball. *Proceedings of the Global Conference on Consumer Electronics (GCCE)* (pp. 763–766). IEEE. <https://doi.org/10.1109/GCCE57851.2023.10298765>
13. Kumar, A. (2018). Effect of deprivation of sleep on handball shooting accuracy. *International Journal of Physiology, Nutrition and Physical Education*, 3(1), 2075-2077.
14. Kumar, A. (2022). Relationship between quadriceps muscle fiber architecture and lunges performance. *International Journal of Physiology, Nutrition and Physical Education*, 7(2), 347-351. <https://doi.org/10.22271/journalofsport.2022.v7.i2f.2645>
15. Kumar, A. (2023). Muscle architecture a new dimension of talent identification: A mini review. *International Journal of Applied Research*, 9(1), 142-146. <https://doi.org/10.22271/allresearch.2023.v9.i1c.10476>
16. Kumar, A. (2023). Muscle fiber type estimation through muscle architectural properties.
17. Kumar, A. (2023). Non-Invasive estimation of muscle fiber type using ultra Sonography. *International Journal of Physical Education, Sports and Health*, 10(1), 89-95. <https://doi.org/10.22271/kheljournal.2023.v10.i1b.2758>
18. Kumar, A., & Jhajharia, B. (2018). Effect of morning exercise on immunity. *International Journal of Physiology, Nutrition and Physical Education*, 3(1), 1987-1989. <http://www.journalofsports.com/pdf/2018/vol3issue1/PartAI/3-1-432-841.pdf>
19. Kumar, A., & Jhajharia, B. (2020). Lunges Performance Predicted By Multiple Linear Regression Model With Muscle Architectural Parameters. *Turkish Online Journal of Qualitative Inquiry (TOJQI)*, 11(4), 2510-2517. <https://doi.org/10.53555/tojq.v11i4.10170>
20. Kumar, A., & Jhajharia, B. (2022). A Comparative Study of White Muscle and Red Muscle Fiber Architectural Parameters. *Asian Pac. J. Health Sci.*, 9(3), 46-49. <https://doi.org/10.21276/apjhs.2022.9.3.10>
21. Kumar, A., Khare, S., & Sisodia, D. A. (2021). An Aerobic Capacity Variable-Based Discriminant Model for the Classification of Handball Players. *Nat. Volatiles & Essent. Oils*, 8(6), 6544-6548. <https://doi.org/10.53555/nveo.v8i6.5598>
22. Lentz-Nielsen, N., Hart, B., & Samani, A. (2023). Prediction of movement in handball with the use of inertial measurement units and machine learning. *Sports Biomechanics*, 22(4), 387–402. <https://doi.org/10.1080/14763141.2021.1983706>
23. Li, H. (2020). Analysis on the construction of sports match prediction model using neural network. *Soft Computing*, 24(21), 16521–16532. <https://doi.org/10.1007/s00500-020-04882-0>
24. Prasetio, D., & Harlili, D. (2016, August). Predicting football match results with logistic regression. *2016 International Conference on Advanced Informatics: Concepts, Theory and Application (ICAICTA)* (pp. 1–5). IEEE. <https://doi.org/10.1109/ICAICTA.2016.7803105>
25. Rahman Mahin, M. P., Ara, E., & Islam, R. U. (2024, June). Cricket analytics: ODI match score prediction and resource metric modeling using machine learning. *2024 IEEE International Conference on Computing, Applications and Systems (COMPAS)* (pp. 214–220). IEEE. <https://doi.org/10.1109/COMPAS60449.2024.10587654>
26. Song, Y., Sun, G., Wu, C., Pang, B., Zhao, W., & Zhou, R. (2024). Construction of 2022 Qatar World Cup match result prediction model and analysis of performance indicators. *Frontiers in Sports and Active Living*, 6, 1378921. <https://doi.org/10.3389/fspor.2024.1378921>
27. Singh, N., & Ajay, K. (2025). Determination of handball match predicting factors. *Global Journal for Research Analysis*, 14(7), 152–155.
28. Wang, E., Yin, X., Li, Y., & Wang, T. (2025). Sports betting: An application of machine learning to the game prediction. *Applied and Computational Engineering*, 4(1), 23–34.
29. Wang, X. (2024). Prediction of tennis match results based on logistic regression. *Transactions on Computer Science and Intelligent Systems Research*, 5(2), 77–89.
30. Yao, Y. (2023, December). Exploratory data analysis & data mining on NBA match prediction. *Proceedings of the 6th International Conference on Computing and Big Data (ICCBD)* (pp. 112–118). ACM. <https://doi.org/10.1145/3629876.3634551>