OPEN ACCESS

ISAR Journal of Multidisciplinary Research and Studies

Volume 3, Issue 8, 2025 | pp: 43-49

Abbriviate Title- ISAR J Mul Res Stud

ISSN *(Online)*- 2583-9705

https://isarpublisher.com/journal/isarjmrs

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

Narender Singh[1*], Dr Sangeeta Singh[2], Dr Vijay Prakash[3], Prof. Varender Singh Patial[4]

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

[2,3]Assistant Professor, Department of Physical Education IIMT University Meerut, UP, India.

[4]Dean, 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 have become an essential field within sports analytics, combining statistical reasoning, data mining, and machine learning to forecast outcomes and inform decision-making. The ability to forecast results in sports such as football, handball, cricket, tennis, and basketball is valuable for coaches, analysts, bettors, and fans. The reviewed literature demonstrates a methodological evolution from traditional statistical models—such as logistic regression, Poisson regression, and Gaussian approximations—to modern artificial intelligence techniques including adaptive back-propagation neural networks, ensemble learning methods, and explainable AI integrated with deep learning. Findings highlight that predictors vary across sports: defensive efficiency and shooting accuracy are dominant in football, overs and wickets significantly influence cricket outcomes, momentum predicts tennis performance, and recent win rate is crucial in basketball. Meanwhile, handball research is expanding rapidly, integrating sensor-based inertial measurement units (IMUs) and computer vision approaches. 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), 43-49.

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).

The rationale behind sports forecasting extends beyond curiosity—it encompasses decision-making in tactical preparation, athlete management, injury prevention, and fan engagement (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 predictive factors universal across sports or specific to each discipline? How can models balance accuracy and interpretability

(Kumar & Jhajharia, 2022; Nandal & Kumar, 2024)? Recent studies illustrate a methodological transition: from regression-based models like logistic regression and Poisson models (Groll et al., 2019; Das et al., 2022) toward advanced neural networks, ensemble techniques, and deep learning approaches (Ahsan et al., 2024; Song et al., 2024; Kumar, 2023c, Singh et. al. 2025). At the same time, new trends emphasize explainability, particularly through SHAP analysis and explainable AI (Felice, 2024; Song et al., 2024; Kumar & Jhajharia, 2020).

Thus, the objective of this review is to synthesize contributions across multiple sports—football, handball, cricket, tennis, and basketball—to examine methodological progress, identify sport-specific predictors, and highlight emerging trends (Jain et al., 2023; Kumar, 2023b).

2. Literature Review

2.1 Handball Prediction Models

Handball prediction research has grown steadily, with different strands of inquiry. Groll et al. (2019) introduced an under dispersed Poisson and Gaussian modeling approach to predict outcomes of the 2019 IHF World Men's Handball Championship. Using historical championship data (2011–2017), they compared Poisson, Gaussian response, and negative binomial regression. The Gaussian response model emerged as the most robust predictor, enabling simulations of tournament outcomes and winning probabilities for all teams. This model identified Denmark as the most likely winner, followed by France.

In an extended version, Groll et al. (2020) applied a sparse Gaussian approximation model, reinforcing the superiority of Gaussian regression for handball match scores. Such statistical frameworks provide probabilities of progression through stages, thus supporting both coaching staff and analysts.

More recent studies expanded the methodological toolkit. Lentz-Nielsen et al. (2023) used IMUs and extreme gradient boosting to classify locomotion and throwing events. Their model achieved excellent classification for throw events (F1 = 0.95) and running (F1 = 0.86) but struggled with dynamic movement (F1 = 0.66). This highlighted the role of biomechanical features such as interquartile range (IQR) and zero-crossing frequency as significant predictors.

Felice (2024) introduced deep learning combined with explainable AI (xAI) for the Paris 2024 Olympic Games. This model not only forecasted match outcomes but also provided interpretable explanations using Large Language Models (LLMs). Such integration of xAI ensures that predictions are actionable for coaches, addressing the common "black box" problem in AI-based analytics.

Finally, Kobayashi et al. (2023) demonstrated the use of computer vision in predicting outcomes in 2-vs-2 handball games by applying multiple object tracking and logistic regression. Their approach achieved 60% accuracy without sensors, offering practical solutions for non-professional settings.

2.2 Football (Soccer) Prediction Models

Football has historically dominated sports analytics, given its global following and extensive match data. Early works relied heavily on regression methods. Prasetio and Harlili (2016) built a logistic regression model on Barclays Premier League data, achieving 69.5% accuracy. Their model emphasized the importance of defensive statistics such as "home defense" and "away defense" as key predictors.

Building on this, Li (2020) proposed an adaptive back-propagation neural network (BPNN) model using UEFA Champions League data. By comparing with multiple linear regression (MLR) and grey prediction, the BPNN achieved almost zero error, demonstrating the value of neural networks in capturing complex non-linear dynamics.

Das et al. (2022) applied logistic regression to Indian Super League data, identifying successful passes, shots on target, fouls, and red cards as predictors. Their study highlighted how domestic league data can reveal context-specific factors differing from European leagues.

The Qatar 2022 World Cup provided a testbed for ensemble models. Song et al. (2024) compared ANN, SVM, AdaBoost, and Random Forests, finding that ANN achieved the highest predictive accuracy (75.42%). Using SHAP analysis, they identified shooting accuracy and ball progression as the most influential features. This emphasizes the role of explainable frameworks in modern sports analytics.

Finally, Wang et al. (2025) integrated ML into sports betting. By testing LightGBM and AdaBoost on European leagues, their models achieved modest accuracy (~52%) but delivered consistent 3% profit margins in simulated betting strategies.

2.3 Cricket Prediction Models

Cricket is one of the most data-rich sports in the world, largely due to its ball-by-ball structure, detailed scorecards, and the multiplicity of formats—Test matches, One-Day Internationals (ODIs), and Twenty20 (T20) cricket. This naturally makes it a fertile ground for predictive modeling, especially since outcomes depend on both **macro factors** (team composition, venue, toss results, weather conditions) and **micro events** (runs scored, wickets lost, strike rates, partnerships). Unlike football or basketball, where performance is continuous, cricket allows researchers to capture granular discrete events, leading to highly specialized predictive models.

Ahsan et al. (2024) attempted to bridge the gap between formats by designing a comprehensive prediction framework for both T20 and Test matches. Their approach incorporated **classification and regression models** including Naïve Bayes, Gradient Boosted Trees, Logistic Regression, Deep Learning, and Random Forests. For score prediction, Gradient Boosted Trees emerged as the most reliable with an accuracy of 96.15%. For match outcomes, the Generalized Linear Model (GLM) performed comparatively better with 71.72% accuracy. These findings underline an important trend: models that excel in predicting **continuous outcomes** (like total runs) do not necessarily perform best in predicting **binary outcomes** (win/loss).

Rahman Mahin et al. (2024) went a step further by integrating **domain-specific knowledge** into predictive frameworks through the introduction of the "resource factor." This novel metric combined the weight of **overs remaining** and **wickets in hand**—two variables universally recognized by players and analysts as the most crucial in ODI cricket. Their Random Forest model achieved exceptional results with an R^2 value of 0.9899, an almost perfect fit, and impressively low error margins (MAE = 3.3569, MSE = 37.3555). Importantly, they also developed a **web-based application** to deploy the model, making the research practically accessible to coaches, analysts, and fans.

What distinguishes cricket research is the recognition that **sport-specific features drive predictive accuracy**. While football and basketball predictions often rely on general performance metrics (e.g., passes, shooting accuracy, recent form), cricket models must account for factors like batting order stability, bowling economy, powerplay overs, and strike rotation. Without integrating these features, models tend to underperform. The emphasis on cricket-specific structures also explains why methods like Random Forest and Gradient Boosted Trees are particularly successful: they handle complex, non-linear interactions between overs, wickets, and scores better than linear regression models.

Nevertheless, challenges remain. First, cricket's **format diversity** complicates model generalization. A model trained on ODI data may fail when applied to T20 or Test cricket due to differences in pacing, strategy, and risk tolerance. Second, external factors such as **pitch conditions** and **weather** are rarely integrated despite their known influence on outcomes. Future work could benefit from **hybrid models** that merge meteorological data, crowd dynamics, and player fatigue with traditional game statistics.

2.4 Tennis Prediction Models

Compared to team sports, tennis presents a unique structure for prediction because outcomes are **binary (win/loss)** at the match level but influenced by **point-by-point sequences**. Traditional prediction frameworks struggled to capture the psychological swings that define tennis matches—often referred to as momentum or flow.

Wang (2024) addressed this gap by operationalizing **momentum** into a quantifiable metric. By weighting factors such as break of serve (34.1%), serve performance (24.3%), scoring dynamics (38%), and net points (3.6%), they created a composite index. This "momentum" variable was then correlated with match outcomes using Pearson's coefficient, revealing a remarkably strong correlation of 0.807. Logistic regression models incorporating momentum demonstrated high predictive accuracy.

The novelty of Wang's contribution lies not only in statistical modeling but also in bridging the **psychological and statistical dimensions** of performance. Momentum reflects both technical execution and mental resilience—a player who secures multiple breaks of serve not only gains a scoring advantage but also exerts psychological pressure on the opponent. By embedding this dynamic into predictive frameworks, Wang's study enriches the understanding of how **performance oscillations translate into outcomes**.

From an applied perspective, momentum-based models could assist coaches in designing **real-time match strategies**. For example, if analytics detect declining momentum, interventions such as slowing down the pace of play or changing serving patterns could help reverse trends. Moreover, betting markets stand to benefit from such models, as momentum provides **intra-match predictive power** beyond pre-match rankings or seedings.

Future tennis prediction studies could expand this line of work by integrating **biometric tracking** (heart rate variability, player exertion) and **computer vision analysis** (serve speed, shot placement). Such approaches would create multi-layered models capable of not just predicting outcomes but also explaining *why* momentum shifts occur.

2.5 Basketball (NBA) Prediction Models

Basketball presents unique predictive challenges due to its high-scoring nature, fast tempo, and player rotations. Predicting outcomes requires disentangling the contributions of **team form, player efficiency, and situational factors** (home court advantage, fatigue from back-to-back games).

Yao (2023) conducted exploratory data analysis and data mining on NBA match prediction using both regression and machine learning methods. Logistic regression proved surprisingly effective, achieving an average accuracy of 67%, outperforming more complex models such as Random Forest and Naïve Bayes. Among predictive features, **recent win rate** was identified as the single most significant factor, highlighting the importance of **short-term form** in basketball outcomes.

This finding aligns with basketball's unique rhythm, where streaks and player momentum significantly shape results. Unlike football or cricket, where external factors (pitch conditions, toss) dominate, basketball outcomes often hinge on immediate player fitness and consistency across recent games.

The study also emphasized the need for **optimal data setting selection**. Experiments revealed that generating datasets from the most recent season improved predictive accuracy compared to long-term historical data. This suggests that basketball outcomes are **time-sensitive**, with player trades, injuries, and coaching changes rendering older data less relevant.

Moving forward, the integration of **player-tracking data** (shot efficiency by location, defensive matchups, rebound positioning) could enhance predictive frameworks. Combining such spatial-temporal data with logistic regression or ensemble ML may yield models both accurate and interpretable.

3. Methodology

3.1 Research Design

This study adopts a **narrative mini-review design**, which is particularly suited for synthesizing findings from a relatively diverse set of studies across multiple sports. Unlike systematic reviews that follow rigid protocols such as PRISMA, narrative reviews allow for greater flexibility in drawing cross-disciplinary insights. The purpose here is not to statistically meta-analyze effect sizes but to identify **methodological progressions, sport-specific predictors, and thematic patterns** across football, handball, cricket, tennis, and basketball.

3.2 Source Identification and Selection

A total of 14 peer-reviewed studies published between 2016 and 2025 were included in this review. These studies were identified from the dataset provided (conference papers, journal articles, and arXiv preprints), representing a blend of both theoretical modeling and applied research. Each study was analyzed 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

Each study was read carefully, and key elements were extracted into 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).

Thematic grouping was then applied. Studies were clustered by sport, but also by **methodological approach**. For example, regression-based models were compared across football, tennis, and basketball, while ensemble ML approaches were compared across cricket and football.

3.5 Limitations of Methodology

While this review synthesizes a broad set of findings, some limitations exist:

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. **Scope bias:** Only five sports were represented, potentially excluding emerging areas (e.g., rugby, baseball, eSports).
3. **Comparability issues:** Different studies reported accuracy in different forms (e.g., percentage accuracy, F1-score, R^2), making direct comparison challenging.
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.

Despite these limitations, the methodology is appropriate for identifying **trends, gaps, and cross-sport insights**—the primary objectives 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	69.5% accuracy; defensive stats most predictive
	Li (2020)	UEFA Champions League	Back-propagation neural network (BPNN), MLR, Grey models	BPNN ≈ minimal prediction error; superior to regression
	Das et al. (2022)	Indian Super League	Logistic regression	Passes, shots, fouls significant; contextual domestic insights
	Song et al. (2024)	FIFA World Cup 2022	ANN, SVM, AdaBoost, Random Forest + SHAP	ANN 75.42% accuracy; shots on target & ball progression most influential
	Wang et al. (2025)	European leagues (betting context)	LightGBM, AdaBoost	~52% accuracy but consistent 3% profit margin
Handball	Groll et al. (2019)	IHF Championship (2011–2017)	Poisson, Gaussian regression	Gaussian response most robust predictor
	Groll et al. (2020)	IHF Championship	Sparse Gaussian approximation	Improved 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 96.15% accuracy for score prediction; GLM best for win/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 of win/loss
Basketball	Yao (2023)	NBA (recent season)	Logistic regression, RF, Naïve Bayes	Logistic regression = 67% accuracy; recent win rate most decisive

The results in table 1 show a **progressive methodological evolution** across sports. In football, initial reliance on logistic regression (Prasetio & Harlili, 2016) yielded moderate accuracy of 69.5%, emphasizing defensive efficiency. Later, neural networks (Li, 2020) captured complex dynamics in UEFA data, outperforming traditional models. More recent ensemble and deep learning approaches (Song et al., 2024; Wang et al., 2025) highlight that accuracy improved moderately (75% for ANN), but profitability in betting applications could be achieved even at lower accuracy (≈52%) when coupled with strategic odds modeling. Collectively, football results indicate that *predictors such as shooting accuracy, passes, and defensive solidity consistently emerge as decisive*, irrespective of methodological sophistication.

Handball research evolved differently. Early regression models (Groll et al., 2019; 2020) confirmed Gaussian approaches as superior for forecasting tournament outcomes. However, the field expanded into biomechanical and sensor-based prediction. Lentz-Nielsen et al. (2023) achieved near-perfect F1-scores in classifying throw events with IMU and XGBoost, while Kobayashi et al. (2023) demonstrated the feasibility of outcome prediction using only video-based tracking (~60% accuracy). Felice (2024) advanced this trajectory by embedding deep learning with explainable AI for the Paris 2024 Olympics, ensuring predictions were interpretable to coaches. This shift illustrates that handball prediction increasingly merges *match outcome forecasting with player movement analytics*, extending beyond pure statistics.

Cricket yielded the **highest predictive accuracies** among all sports reviewed. Ahsan et al. (2024) showed that Gradient Boosted Trees could predict scores with 96.15% accuracy across formats, though GLM remained more stable for binary win/loss outcomes. The breakthrough came from Rahman Mahin et al. (2024), whose "resource factor" (overs + wickets) enabled Random Forest models to achieve near-perfect R^2 (0.9899). This demonstrates that cricket's ball-by-ball structure allows models to exploit domain-specific variables with unparalleled precision. However, it also highlights cricket's uniqueness—its structured nature may not generalize to other continuous sports.

Tennis prediction demonstrated innovation in capturing psychological dimensions. Wang (2024) introduced "momentum" as a quantifiable index combining serve dynamics, breaks, and scoring flow. With a correlation of 0.807 to match outcomes, momentum proved a powerful predictor when embedded in logistic regression models. Unlike other sports, tennis highlights the integration of *psychological flow and technical execution* in predictive frameworks.

Basketball analysis emphasized simplicity. Yao (2023) found that logistic regression, a relatively simple model, outperformed Random Forest and Naïve Bayes, achieving 67% accuracy in NBA outcomes. Crucially, *recent win rate* emerged as the strongest predictor, confirming basketball's reliance on short-term form and team rhythm rather than long historical trends. This finding underscores that in certain contexts, simple, interpretable models with strong predictors may outperform complex architectures.

5. Discussion

The reviewed literature illustrates a **decade-long methodological journey**, beginning with traditional regression approaches, advancing to ensemble machine learning, and culminating in deep learning integrated with explainable AI. This progression mirrors broader trends in data science but also reveals that predictive success in sports analytics is *context-dependent and sport-specific*. In this discussion, three key dimensions are examined: methodological evolution, sport-specific predictors, and practical implications, followed by a critical evaluation of limitations and future pathways.

Between 2016 and 2019, statistical regression models dominated sports prediction research. These models, such as logistic regression and Poisson regression, offered high interpretability and were computationally efficient. For example, Prasetio & Harlili (2016) used logistic regression to analyze Premier League matches, identifying home and away defense as significant predictors. Similarly, Groll et al. (2019) demonstrated the effectiveness of Gaussian regression for handball championship outcomes. While transparent, these models often achieved only modest predictive power (≈65–70%), as they struggled to capture nonlinear interactions inherent in team sports.

From 2020 to 2023, the landscape shifted toward **ensemble learning and neural networks**. Neural models like Li's (2020) BPNN minimized error in UEFA predictions, while Random Forests and Gradient Boosted Trees (Ahsan et al., 2024) excelled in cricket. These models effectively captured non-linear interactions between multiple variables (passes, wickets, overs, player rotations). Yet, ensemble methods carried risks of overfitting, particularly in sports with limited datasets such as handball or tennis. The key trade-off in this period was accuracy versus generalizability.

The most recent years (2024–2025) highlight **deep learning coupled with explainable AI (XAI)**. Song et al. (2024) used ANN for World Cup football predictions, supplementing results with SHAP analysis to identify critical features like shots on target. Felice (2024) advanced handball analytics by integrating deep learning with large language models to produce interpretable forecasts for the Paris Olympics. These advances mark a new paradigm: prediction accuracy is no longer sufficient—interpretability and human usability have become equally essential. Wang et al. (2025) further extended this by demonstrating that ML could generate profits in betting applications despite modest predictive accuracy, underscoring the commercial applicability of predictive models.

A recurring theme across all sports is that **sport-specific variables consistently outperform generic indicators**. In football, passing accuracy, ball progression, and shooting efficiency were repeatedly identified as decisive features (Das et al., 2022; Song et al., 2024). Handball, in contrast, demonstrated the predictive value of biomechanical data, where IMU-based features such as throw frequency and locomotion intensity achieved near-perfect classification scores (Lentz-Nielsen et al., 2023). Cricket confirmed the dominance of overs and wickets as predictors, a conclusion quantitatively reinforced by Rahman Mahin et al. (2024) through the resource factor. Tennis broke new ground by embedding psychological constructs into prediction models, as Wang (2024) showed that momentum, a variable reflecting both technical and mental resilience, strongly predicted outcomes. Basketball highlighted the dominance of short-term form, with recent win rate outstripping more complex indicators (Yao, 2023). Together, these findings demonstrate that *predictive accuracy improves dramatically when domain-specific knowledge is embedded into models.*

The practical implications of these models are extensive. For **coaching and training**, predictive insights can shape tactical preparation—football managers can prioritize passing and shooting drills, cricket coaches can simulate overs-and-wickets scenarios, and tennis coaches can develop strategies for momentum recovery. In **sports science and injury prevention**, wearable and vision-based models in handball can identify biomechanical stressors, offering early warning systems for overuse injuries. In the **betting industry**, Wang et al. (2025) showed that modestly accurate ML models could produce profitable margins, though ethical concerns regarding gambling responsibility must be acknowledged. Finally, in **fan engagement and broadcasting**, prediction probabilities combined with explainable AI have the potential to enrich live commentary, offering spectators real-time tactical insights.

Despite remarkable progress, significant limitations remain. Many studies rely on **small sample sizes** (especially in handball and tennis), which limits generalizability and heightens the risk of overfitting. Prediction metrics are **heterogeneous**—accuracy, F1-score, R^2—making cross-sport comparisons imprecise. External factors known to influence outcomes, such as weather in cricket, injuries in football, referee decisions in basketball, and crowd influence in tennis, remain largely absent from existing models. Additionally, the **format diversity of cricket** (Test, ODI, T20) exemplifies the challenge of building generalizable models across contexts. Without addressing these gaps, predictive models risk being highly accurate in controlled datasets but less effective in real-world applications.

The future of sports prediction lies in **multi-modal data integration**. Combining match statistics with biometric monitoring, wearable sensors, and video-tracking data will create holistic prediction systems. **Real-time predictive analytics** represent another frontier, enabling in-game tactical decision-making—for instance, adjusting strategies mid-match when momentum shifts in tennis. **Generalizability** across contexts must also be prioritized; cricket models should adapt across formats, and basketball models across seasons with frequent trades and injuries. Finally, **explainability and ethics** remain central. Coaches and fans need interpretable predictions, and the betting applications of sports prediction demand responsible deployment. Techniques such as SHAP, XAI, and natural language explanations by large language models will be vital in bridging the gap between algorithmic complexity and human usability.

6. CONCLUSION

This review demonstrates that sports match prediction has advanced remarkably over the last decade. Football and cricket dominate research, but handball, tennis, and basketball contribute unique insights. The evolution from regression to neural networks and explainable AI mirrors broader data science trends.

The most significant lesson is that **predictors must be sport-specific**: overs and wickets in cricket, momentum in tennis, defensive efficiency in football, and recent form in basketball. Models that integrate such tailored variables achieve the highest accuracies.

Moving forward, the field must balance **accuracy, interpretability, and real-world applicability**. Multi-modal data integration, real-time prediction, and coach-friendly interpretability will define the next generation of research.

Sports prediction is not just an academic exercise—it is an applied science shaping how athletes train, how fans engage, and how sports are experienced globally.

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., Schauberger, 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., Schauberger, 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/tojqi.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