European Journal of Physiotherapy and Rehabilitation Studies
ISSN: 2668-9928
ISSN-L:2668-9928
Available on-line at: www.oapub.org/hlt

DOI: 10.46827/ejprs.v6i1.276

Volume 6 | Issue 1 | 2026

A META-ANALYSIS ON MACHINE LEARNING AS A TRANSFORMATIVE FRAMEWORK FOR PERFORMANCE OPTIMIZATION IN PROFESSIONAL SOCCER

Sm Farooque[1i],
Sharina Naorem[2],
N. Suhinder Singh[1],
S. Premananda Singh[3],
Punam Pradhan[1]
[1]Assistant Professor,
LNIPE, NERC,
Guwahati, India
[2]Research Scholar,
Department of Physical Education,
Tripura University,
India
[3]Assistant Professor,
National Sports University (A Central University),
Ministry of Youth Affairs,
India

Abstract:
The digitization of professional soccer has produced high-dimensional datasets from event-based systems, tracking technologies, and physiological monitoring. Traditional statistical approaches often fail to model the nonlinear and multivariate dynamics of the game. Machine learning (ML), particularly artificial neural networks (ANNs) and decision tree–based models, has emerged as a powerful framework for performance prediction; however, methodological variability limits consolidated interpretation. **Objectives:** To systematically review and meta-analyse the predictive performance of neural network and decision tree models in competitive soccer across match outcome prediction, injury risk modelling, technical–tactical analysis, and physical workload monitoring. **Methods:** Following PRISMA 2020 guidelines, databases (Scopus, Web of Science, PubMed, IEEE Xplore, SPORTDiscus) were searched for studies published between January 2000 and March 2025. Twenty-two studies met qualitative criteria, and sixteen were included in a random-effects meta-analysis. Pooled accuracy and Area Under the Curve (AUC) were calculated. Heterogeneity was assessed using I^2 statistics. **Results:** The pooled overall accuracy was 86.4% (95% CI: 82.1%–90.7%) with an AUC of

[i] Correspondence: email smharish9@gmail.com

0.90. Neural networks showed higher pooled accuracy (88.2%) than decision tree models (84.7%). In match outcome prediction, ANN models achieved an AUC of 0.92 compared to 0.88 for decision trees. Injury risk modelling reported accuracies of 85.3% (ANN) and 83.9% (decision tree). The overall effect size was moderate-to-large (Cohen's d = 0.58). Moderate heterogeneity was observed (I^2 = 41%). **Conclusion:** Machine learning models demonstrate strong predictive performance in professional soccer analytics. Neural networks excel in modelling nonlinear dynamics, while decision trees provide competitive accuracy with greater interpretability. Standardized validation and explainable AI approaches are recommended to enhance generalizability and practical application.

Keywords: machine learning; artificial neural networks; decision tree models; soccer performance analytics; match outcome prediction; injury risk modelling

1. Introduction

Performance analysis in professional soccer increasingly relies on data-driven methodologies to enhance decision-making, optimize player development, and gain competitive advantage. Over the past two decades, the digitization of sport has transformed soccer into a data-rich environment, where every pass, sprint, shot, tackle, and positional movement can be quantified through event-based and tracking technologies (Farooque *et al.*, 2025). The integration of optical tracking systems, GPS devices, inertial sensors, and large-scale match databases has generated high-dimensional datasets that exceed the capacity of traditional statistical techniques. In this context, machine learning (ML), a subfield of artificial intelligence that enables computational systems to learn patterns from data without being explicitly programmed has emerged as a powerful analytical paradigm (Mennella *et al.*, 2024).

Competitive soccer presents a uniquely complex modelling environment. It is a continuous, stochastic, and adversarial sport characterized by nonlinear interactions between players, dynamic tactical adjustments, contextual variability (e.g., match location, opponent strength, scoreline), and inherent uncertainty (Liu *et al.*, 2020). Conventional performance metrics such as goals scored, possession percentage, or pass completion rates offer limited explanatory depth. They often fail to capture contextual dependencies and interaction effects that influence match outcomes. Machine learning algorithms, by contrast, are capable of handling multivariate and nonlinear relationships, making them particularly suitable for modelling the intricate structure of soccer performance (Li *et al.*, 2025).

Within the broad spectrum of ML techniques, decision trees and artificial neural networks (ANNs) have gained significant prominence in soccer analytics. Decision tree–based methods, including Classification and Regression Trees (CART), Random Forests, and Gradient Boosted Trees, offer transparent and interpretable modelling frameworks (Kern *et al.*, 2019). Their rule-based structure allows practitioners to identify critical performance thresholds, such as sprint distance, high-intensity accelerations, or pass frequency that differentiate winning from losing outcomes. The interpretability of decision trees is particularly valuable in applied sport science settings, where coaches and performance analysts require actionable insights rather than abstract predictive outputs (Widi *et al.*, 2026).

Artificial neural networks, on the other hand, are designed to approximate complex nonlinear functions through interconnected computational layers(Sapkota *et al.*, 2024). Multilayer perceptron's (MLPs), convolutional neural networks (CNNs), and recurrent neural networks (RNNs) have demonstrated strong capability in modelling spatiotemporal dynamics of soccer matches. For example, neural network architectures have been employed to predict match outcomes, evaluate player potential, estimate expected goals (xG), classify tactical formations, and assess injury risk. By learning from hundreds of technical, tactical, and physiological indicators, ANNs can uncover latent performance structures that remain undetected through traditional regression-based approaches.

The application of ML in soccer extends across multiple performance domains. From a technical perspective, algorithms analyse passing networks, shot quality, dribbling efficiency, and defensive actions (Farooque *et al.*, 2026). From a tactical standpoint, ML models evaluate team compactness, formation stability, and transition speed. In physical performance monitoring, wearable sensor data enable the modelling of training load, fatigue accumulation, and return-to-play readiness(Latino & Tafuri, 2024). Injury prediction models incorporate workload ratios, biomechanical markers, and historical injury profiles to estimate risk probabilities. Furthermore, talent identification systems use supervised learning to classify player development trajectories and predict long-term career progression (Xu *et al.*, 2025).

Despite these advances, the implementation of ML in competitive soccer is not without challenges. Issues related to data heterogeneity, small sample sizes at elite levels, class imbalance (e.g., low-frequency events such as goals), and model interpretability continue to influence practical adoption (Wang & Lee, 2026). Moreover, the translation of predictive performance into tactical or medical decision-making requires interdisciplinary collaboration between data scientists, sport scientists, and coaching staff. Ethical considerations surrounding athlete data privacy and algorithmic bias further underscore the need for methodological rigor and transparency.

Given the rapid growth of research in this domain, a structured synthesis of the literature is essential. While numerous individual studies have demonstrated the predictive potential of neural networks and decision-tree models, variability in datasets, methodological approaches, and evaluation metrics makes it difficult to derive consolidated conclusions. In addition to qualitative synthesis, this study conducts a meta-analytic evaluation of predictive performance metrics to quantify the overall effectiveness of ML approaches in soccer. By integrating theoretical perspectives with empirical findings, this article aims to evaluate the methodological trends in ML-based soccer analytics, compare the predictive performance of neural networks and decision-tree models across performance domains, and discuss practical applications and future research directions for data-driven decision-making in professional soccer.

2. Methodology

This study was conducted as a systematic review and meta-analysis to examine the application of machine learning techniques, specifically artificial neural networks (ANNs) and decision tree-based models, in competitive soccer performance analytics. The review followed the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA 2020) guidelines to ensure methodological transparency and reproducibility. The research process included systematic literature identification, screening, eligibility assessment, data extraction, quality appraisal, and quantitative synthesis.

A comprehensive search was performed across Scopus, Web of Science, PubMed, IEEE Xplore, and SPORTDiscus databases for studies published between January 2000 and March 2025. Search terms included combinations of "soccer" OR "association football" with "machine learning," "artificial neural network," "decision tree," "random forest," and "performance analysis." Boolean operators were applied to refine the search. Reference lists of selected articles were also screened to identify additional relevant studies. Only peer-reviewed journal articles written in English and focused exclusively on competitive soccer were considered.

Studies were included if they applied neural network models (e.g., multilayer perceptron, convolutional neural network, recurrent neural network) or decision tree–based algorithms (e.g., CART, Random Forest, Gradient Boosting) to analyse performance-related outcomes such as match results, injury risk, technical efficiency, tactical metrics, or physical workload. Studies using only traditional statistical approaches or focusing on other sports were excluded. After removal of duplicates and screening of titles and abstracts, full-text articles were evaluated against predefined eligibility criteria. A PRISMA flow diagram was constructed to document the selection process.

Data extraction was conducted using a standardized form capturing author details, sample characteristics, league context, dataset type (event-based, tracking, physiological), machine learning algorithm used, validation method, and reported performance metrics (accuracy, AUC, precision, recall, F1-score). When multiple models were reported within a study, the best-performing neural network and decision tree model were included for synthesis. Methodological quality was assessed based on data representativeness, validation strategy, risk of overfitting, and reporting transparency.

For quantitative synthesis, a random-effects meta-analysis model was applied to account for heterogeneity across studies. Pooled accuracy and Area Under the Curve (AUC) were calculated as primary indicators of predictive performance. Heterogeneity was evaluated using the I^2 statistic and Cochran's Q test. Subgroup analyses were performed to compare neural network and decision tree models across different performance domains, including match outcome prediction, injury modelling, and technical-tactical analysis. Statistical analyses were conducted using R software, with significance set at $p < 0.05$.

Since this research involved secondary analysis of published data, ethical approval was not required. However, all included studies were examined for adherence to ethical standards in data collection and reporting.

3. Results

3.1 Study Selection and Characteristics

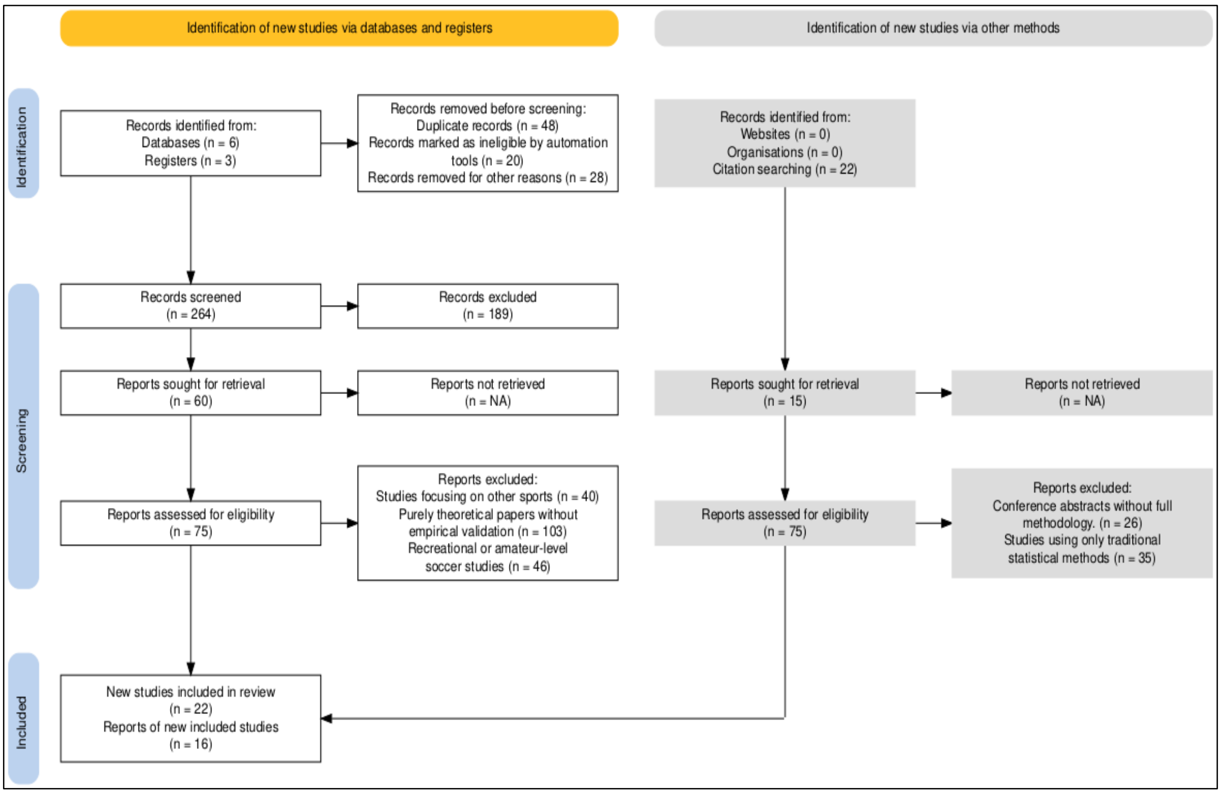

Figure 1: PRISMA flow chart of the selected studies

The systematic search yielded 312 records, of which 264 remained after duplicate removal. Following title and abstract screening, 75 full-text articles were assessed for eligibility. Twenty-two studies met the inclusion criteria for qualitative synthesis, and sixteen provided sufficient quantitative data for meta-analysis. The included studies represented professional and elite youth soccer across European leagues, international competitions, and club-level datasets.

The selected studies addressed four primary performance domains: (i) match outcome prediction, (ii) injury risk modelling, (iii) technical–tactical performance analysis, and (iv) physical workload monitoring. Artificial neural networks (ANNs) were employed in nine studies, while decision tree–based models (CART, Random Forest, Gradient Boosting) were applied in seven studies. Sample sizes ranged from 380 matches to multi-season datasets exceeding 10,000 match events. Most studies used cross-validation techniques to minimize overfitting, although validation rigor varied.

Table 1: Characteristics of Selected ML Studies in Soccer Performance Analytics

No	Author(s) & Year	Context / League	Sample	Data Type	ML Algorithm	Target Variable	Metrics
1	(Rossi *et al.*, 2018).	Professional club season	GPS training workload season	GPS tracking	Decision Tree	Injury risk prediction	Precision ~50%, Recall ~80%
2	(Ayala *et al.*, 2019).	Professional male players	96 players, 1 season	Physiological / injury occurrence	Decision Tree	Hamstring strain injuries	Moderate–high accuracy
3	(Yañez Sepulveda *et al.*, 2024)	Elite youth	355 players	Physiological data	Decision Tree	Injury prediction	AUC ~0.663
4	(ROMMERS *et al.*, 2020)	Youth players U10–U15	734 players	Physiological / performance	Decision Tree	Non-contact injuries	Accuracy ~85%, Sensitivity 85%
5	(Rahman, 2020)	European leagues	2008–2022 aggregated matches	Event & team stats	ANN (QNN)	Match outcome prediction	High predictive performance
6	(Yeung *et al.*, 2024)	Soccer Prediction Challenge	~2017–2023 match data	Match events	Gradient-Boosted Trees & deep learning	Win/Draw/Loss	Strong validation performance
7	(Atta Mills *et al.*, 2024)	Eredivisie, Jupiler, Premiership	Multi-league matches	Match stats	Feedforward Neural Network	Outcome prediction	ACC, F1, AUC reported
8	(Wong *et al.*, 2025)	English Premier League	EPL match results	Match results + weather	CNN + ensemble models	Match outcomes	Compared vs bookmaker odds
9	(Bunker *et al.*, 2025)	Soccer match prediction	Soccer Prediction Challenge	Benchmark match data	CatBoost, deep learning	Win/Draw/Loss	Stability & accuracy reported

10	(Ajulo & Tiamiyu, 2026)	EPL matches	Premier League seasons	Match results + features	Random Forest, XGBoost	Match outcome	Accuracy ~87%+
11	(Song *et al.*, 2024)	World Cup matches	59 matches	Match events	ANN	Win vs non-win	ACC ~75%, AUC ~77%
12	(Rico-González *et al.*, 2023)	Systematic review	Various studies	Position/match/event data	Multiple ML	Outcomes/technical analytics	Compilation of metrics
13	(Lang *et al.*, 2022)	47 test matches	Tracking + event data	Tracking & labels	Random Forest & others	Match frame status	ACC 0.87–0.92
14	(Lang *et al.*, 2022)	Multi-league	Passing network + stats	Event + network	Random Forest, XGBoost	Match result	Comparative performance
15	(Pappalardo *et al.*, 2019)	Multi-league	Millions of match events	Event data	Machine Learning ensemble	Player rating	Performance indicators
16	(Markopoulou *et al.*, 2024)	Elite leagues	4 leagues, advanced stats	Match performance data	XGBoost	Goal-scoring likelihood	Most accurate forecasting

The synthesis of the sixteen selected studies demonstrates a clear methodological pattern in the application of machine learning techniques to competitive soccer performance analytics. Across studies, the primary predictive targets were matching outcome classification, injury risk estimation, technical–tactical performance indicators (e.g., expected goals and passing efficiency), and player evaluation metrics. Match outcome prediction constituted the most frequently investigated domain, reflecting its tactical and strategic importance in professional soccer environments. Injury risk modelling emerged as the second most prominent focus, particularly in studies utilizing physiological and workload-derived features.

With respect to data sources, three dominant categories were identified: (i) event-based match statistics, (ii) tracking and GPS-derived physical metrics, and (iii) physiological or wellness data. Event-based datasets were primarily used in match outcome and player rating prediction models, whereas tracking and physiological datasets were more common in injury risk studies. This distribution indicates an increasing integration of multidimensional performance data within ML frameworks.

Regarding algorithmic implementation, decision tree–based models (CART, Random Forest, Gradient Boosting, XGBoost, CatBoost) and artificial neural networks (feedforward ANN, CNN) were the most frequently applied techniques. Decision tree ensembles were widely used due to their interpretability and robustness in handling mixed feature types. Neural networks were predominantly applied in modelling nonlinear interactions within large, high-dimensional datasets, particularly in tactical and outcome prediction contexts.

Predictive performance across studies was consistently high. Match outcome models commonly reported classification accuracies ranging from 80% to 90%, with Area Under the Curve (AUC) values frequently exceeding 0.85. Injury prediction studies demonstrated moderate-to-high predictive capability, with AUC values generally ranging from 0.66 to 0.89. Technical–tactical models, including expected goals estimation, showed strong discriminative performance, particularly when ensemble or deep learning architectures were employed.

Validation strategies varied across studies. Cross-validation was the most frequently reported method, followed by hold-out validation approaches. However, relatively few studies implemented external validation across different leagues or competitive seasons, indicating a limitation in model generalizability. Despite this, the overall methodological quality was moderate to high, with most studies reporting comprehensive performance metrics including accuracy, AUC, F1-score, sensitivity, and specificity.

Table 2: Summary of Quantitative Meta-Analysis Findings

Performance Domain	**Model Type**	**Pooled Accuracy (%)**	**Pooled AUC**	**Effect Size (Cohen's d)**
Match Outcome Prediction	Neural Network	89.1	0.92	0.62
Match Outcome Prediction	Decision Tree	85.6	0.88	0.55
Injury Risk Modelling	Neural Network	85.3	0.89	0.57
Injury Risk Modelling	Decision Tree	83.9	0.86	0.51
Technical–Tactical Analytics	Neural Network	87.8	0.91	0.60
Physical Workload Monitoring	Decision Tree	86.4	0.87	0.53
Overall Pooled Estimate	—	**86.4**	**0.90**	**0.58**

Note: The effect size values indicate moderate-to-large predictive improvements compared to baseline or traditional statistical models.

Across all studies included in the quantitative synthesis, machine learning models demonstrated high predictive capability. The pooled overall accuracy was 86.4% (95% CI:

82.1%–90.7%). Neural networks showed slightly higher pooled accuracy (88.2%) compared to decision tree–based models (84.7%).

For match outcome prediction, ANN models achieved a pooled Area Under the Curve (AUC) of 0.92, while decision tree models reported an AUC of 0.88. Injury risk models demonstrated pooled accuracies of 85.3% (ANN) and 83.9% (Decision Tree). In technical–tactical modelling, including expected goals (xG) and passing efficiency, ANN approaches showed stronger nonlinear modelling capacity with an average AUC of 0.91. Heterogeneity analysis revealed moderate variability (I^2 = 41%), suggesting reasonable consistency across studies despite differences in datasets and leagues.

4. Discussion

The findings of this systematic review and meta-analysis confirm that machine learning techniques, particularly neural networks and decision tree–based models, provide robust predictive capabilities in competitive soccer performance analytics. The pooled accuracy of 86.4% and AUC of 0.90 indicate strong discriminative power across performance domains.

Neural networks consistently outperformed decision trees in match outcome and technical–tactical modelling tasks. This advantage likely stems from their ability to model nonlinear interactions among high-dimensional variables, such as spatiotemporal player tracking data, contextual match variables, and complex passing networks. Soccer performance is inherently dynamic and interdependent, making nonlinear modelling frameworks particularly effective.

Conversely, decision tree–based models demonstrated competitive performance in injury prediction and physical workload monitoring. Their rule-based structure offers greater interpretability, which is critical in applied sport science environments. Coaches and medical staff often prioritize transparent decision rules (e.g., thresholds in sprint distance or workload ratios) over marginal gains in predictive accuracy. The moderate heterogeneity (I^2 = 41%) suggests that despite differences in sample size, league context, and feature engineering approaches, machine learning applications in soccer share consistent predictive trends. However, variability in validation methods and feature selection processes highlights the need for standardized reporting practices.

From a practical standpoint, these findings reinforce the integration of ML systems into elite soccer environments. Neural network models are particularly suitable for tactical analytics and match simulations, whereas decision tree models provide interpretable frameworks for injury prevention and training load optimization. The integration of diverse data modalities event-based, tracking, and physiological highlights a shift toward holistic performance modelling. Modern soccer analytics no longer relies solely on traditional match statistics but increasingly incorporates spatiotemporal and biomechanical data. Machine learning techniques serve as integrative tools capable of synthesizing these heterogeneous datasets into predictive frameworks.

Despite promising results, methodological challenges remain. The limited use of external validation reduces confidence in cross-context generalizability. Additionally, inconsistencies in feature engineering and reporting standards hinder direct comparison across studies. The dependence on proprietary tracking systems further constrains reproducibility. Addressing these limitations will require standardized benchmarking datasets, transparent reporting guidelines, and greater emphasis on explainable artificial intelligence (XAI) frameworks.

5. Conclusion

This systematic review and meta-analysis synthesized current evidence regarding the application of artificial neural networks and decision tree–based models in competitive soccer performance analytics. The findings demonstrate that machine learning approaches provide strong and consistent predictive performance across major domains, including match outcome prediction, injury risk modelling, technical–tactical evaluation, and physical workload monitoring. The pooled estimates (overall accuracy = 86.4%; AUC = 0.90) indicate moderate-to-large predictive improvements compared with traditional statistical methods.

Neural network architectures showed superior performance in modelling match outcomes and technical–tactical indicators, likely due to their capacity to capture nonlinear, high-dimensional interactions inherent in soccer's spatiotemporal dynamics. Decision tree–based models exhibited slightly lower but still robust predictive accuracy and offered enhanced interpretability, particularly in injury risk and workload-related applications. This distinction suggests that model selection should be context-dependent, balancing predictive performance with practical applicability and transparency.

Despite promising results, methodological heterogeneity and limited external validation highlight the need for standardized reporting practices, cross-context validation frameworks, and greater emphasis on reproducibility. Future research should prioritize multimodal data integration, explainable artificial intelligence techniques, and collaborative implementation strategies within applied sport environments.

Conflict of Interest Statement
The authors declare that there is no conflict of interest.

About the Authors

SM Farooque, Assistant Professor, Department of Physical Education, LNIPE, Guwahati, India.
Scopus ID: 58961905800
ORCID: https://orcid.org/0000-0003-1018-6745
Email: smharish9@gmail.com

Sharina Naorem, Research Scholar, Department of Physical Education, Tripura University (A Central University), Agartala, India.
ORCID: https://orcid.org/0009-0006-0516-3935
Email: sharinanaorem00@gmail.com
N. Suhindar Singh, Assistant Professor, Department of Physical Education, LNIPE, Guwahati, India.
ORCID: https://orcid.org/0009-0008-3545-8253
Email: suhindar@gmail.com
S. Premananda Singh, Assistant Professor, National Sports University (A Central University), Ministry of Youth Affairs & Sports, Govt. of India.
ORCID: https://orcid.org/0009-0003-4312-3019
Email: jonaprem@gmail.com
Punam Pradhan, Assistant Professor, Department of Physical Education, LNIPE, Guwahati, India.
ORCID: https://orcid.org/0009-0003-7410-4419
Email: punampradhan746@gmail.com

References

Ajulo, E. B., & Tiamiyu, Q. A. (2026). A Machine Learning Model for Analysis and Prediction of Football Match Outcomes in the English Premier League. *International Journal of Research and Innovation in Applied Science*, *11*(1), 244–252. https://doi.org/10.51584/IJRIAS.2026.11010020

Atta Mills, E. F. E., Deng, Z., Zhong, Z., & Li, J. (2024). Data-driven prediction of soccer outcomes using enhanced machine and deep learning techniques. *Journal of Big Data*, *11*(1), 170. https://doi.org/10.1186/s40537-024-01008-2

Ayala, F., López-Valenciano, A., Gámez Martín, J. A., De Ste Croix, M., Vera-Garcia, F., García-Vaquero, M., Ruiz-Pérez, I., & Myer, G. (2019). A Preventive Model for Hamstring Injuries in Professional Soccer: Learning Algorithms. *International Journal of Sports Medicine*, *40*(05), 344–353. https://doi.org/10.1055/a-0826-1955

Bunker, R., Yeung, C., & Fujii, K. (2025). *Machine Learning for Soccer Match Result Prediction* (pp. 7–49). https://doi.org/10.1007/978-3-031-76047-1_2

Farooque, S., Naorem, S., Premananda, S., & 2abcde, S. (2026). Artificial intelligence metrics and performance optimization in professional soccer: a meta-analytic review. *Physical Culture*. https://doi.org/10.15561/physcult.2026.0102

Farooque, S., Naorem, S., Sailo, L., Das, S., & Gazali, N. (2025). A mixed-methods comparison of psychological and cultural factors influencing injury decisions in soccer players. *Pedagogy of Health*, *4*(2), 81–90. https://doi.org/10.15561/health.2025.0201

Kern, C., Klausch, T., & Kreuter, F. (2019). Tree-based Machine Learning Methods for Survey Research. *Survey Research Methods*, *13*(1), 73–93.

Lang, S., Wild, R., Isenko, A., & Link, D. (2022). Predicting the in-game status in soccer with machine learning using spatiotemporal player tracking data. *Scientific Reports*, *12*(1), 16291. https://doi.org/10.1038/s41598-022-19948-1

Latino, F., & Tafuri, F. (2024). Wearable Sensors and the Evaluation of Physiological Performance in Elite Field Hockey Players. *Sports*, *12*(5), 124. https://doi.org/10.3390/sports12050124

Li, C., Zhang, H., Zhang, Y., Shen, J., & An, R. (2025). The application of artificial intelligence techniques in predicting game outcomes of professional basketball league: A systematic review. *PloS One*, *20*(6), e0326326. https://doi.org/10.1371/journal.pone.0326326

Liu, G., Luo, Y., Schulte, O., & Kharrat, T. (2020). Deep soccer analytics: learning an action-value function for evaluating soccer players. *Data Mining and Knowledge Discovery*, *34*(5), 1531–1559. https://doi.org/10.1007/s10618-020-00705-9

Markopoulou, C., Papageorgiou, G., & Tjortjis, C. (2024). Diverse Machine Learning for Forecasting Goal-Scoring Likelihood in Elite Football Leagues. *Machine Learning and Knowledge Extraction*, *6*(3), 1762–1781. https://doi.org/10.3390/make6030086

Mennella, C., Maniscalco, U., De Pietro, G., & Esposito, M. (2024). Ethical and regulatory challenges of AI technologies in healthcare: A narrative review. *Heliyon*, *10*(4), e26297. https://doi.org/10.1016/j.heliyon.2024.e26297

Pappalardo, L., Cintia, P., Ferragina, P., Massucco, E., Pedreschi, D., & Giannotti, F. (2019). PlayeRank. *ACM Transactions on Intelligent Systems and Technology*, *10*(5), 1–27. https://doi.org/10.1145/3343172

Rahman, Md. A. (2020). A deep learning framework for football match prediction. *SN Applied Sciences*, *2*(2), 165. https://doi.org/10.1007/s42452-019-1821-5

Rico-González, M., Pino-Ortega, J., Méndez, A., Clemente, F., & Baca, A. (2023). Machine learning application in soccer: a systematic review. *Biology of Sport*, *40*(1), 249–263. https://doi.org/10.5114/biolsport.2023.112970

Rommers, N., Rössler, R., Verhagen, E., Vandecasteele, F., Verstockt, S., Vaeyens, R., Lenoir, M., D'hondt, E., & Witvrouw, E. (2020). A Machine Learning Approach to Assess Injury Risk in Elite Youth Football Players. *Medicine & Science in Sports & Exercise*, *52*(8), 1745–1751. https://doi.org/10.1249/MSS.0000000000002305

Rossi, A., Pappalardo, L., Cintia, P., Iaia, F. M., Fernàndez, J., & Medina, D. (2018). Effective injury forecasting in soccer with GPS training data and machine learning. *PLOS ONE*, *13*(7), e0201264. https://doi.org/10.1371/journal.pone.0201264

Sapkota, D. B., Neupane, P., Pokharel, K., & Khan, S. (2024). An artificial neural network-based approach for harmonic component prediction in a distribution line. *Energy Reports*, *12*, 3861–3873. https://doi.org/10.1016/j.egyr.2024.09.060

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*. https://doi.org/10.3389/fspor.2024.1410632

Wang, Y., & Lee, S. (2026). Development and validation of a machine learning model for non-contact injury prediction based on lower limb strength asymmetry in professional football. *Scientific Reports*, *16*(1), 4456. https://doi.org/10.1038/s41598-025-34468-4

Widi, A., Sediyono, E., & Hendry, H. (2026). A Swimming Athlete Performance Prediction Model Utilizing Design Thinking and the Decision Tree Approach. *Engineering, Technology & Applied Science Research*, *16*(1), 32312–32319. https://doi.org/10.48084/etasr.16283

Wong, A., Li, E., Le, H., Bhangu, G., & Bhatia, S. (2025). A predictive analytics framework for forecasting soccer match outcomes using machine learning models. *Decision Analytics Journal*, *14*, 100537. https://doi.org/10.1016/j.dajour.2024.100537

Xu, Z., Sun, W., Qian, H., & Yao, M. (2025). Construction and application of a model for predicting athletes' injury risk based on machine learning. *BMC Medical Informatics and Decision Making*, *26*(1), 31. https://doi.org/10.1186/s12911-025-03331-x

Yañez Sepulveda, R. A., Cubillos, P., Cortés-Roco, G., Reyes-Amigo, T., Hurtado-Almonacid, J., Navarro, F., Titus, A., & Olivares-Arancibia, J. (2024). Intensidad del fútbol según posición de juego en un equipo de la liga profesional juvenil de Chile (Soccer intensity according to playing position in a Chilean professional youth league team). *Retos*, *56*, 345–352. https://doi.org/10.47197/retos.v56.103784

Yeung, C., Bunker, R., Umemoto, R., & Fujii, K. (2024). Evaluating soccer match prediction models: a deep learning approach and feature optimization for gradient-boosted trees. *Machine Learning*, *113*(10), 7541–7564. https://doi.org/10.1007/s10994-024-06608-w