A Comprehensive Review of Soccer Match Simulation and Predictive Analytics

Allen Roman

✦

Abstract—This paper presents a comprehensive review of soccer match simulation and predictive analytics, examining both traditional statistical approaches and modern artificial intelligence methods. The integration of machine learning and sports analytics has created new opportunities for analyzing soccer matches, from player performance evaluation to match outcome prediction. We systematically review recent developments in soccer analytics, covering event data analysis, player tracking metrics, and virtual training environments. Special attention is given to the application of various machine learning models, including deep learning and reinforcement learning approaches, in predicting match outcomes and analyzing player performance. The review also examines available data sources and technologies, highlighting the challenges and opportunities in combining traditional sports statistics with AI-driven analytics. By synthesizing findings from multiple research streams, we provide insights into the current state of soccer analytics and identify promising directions for future research. Our analysis suggests that while significant progress has been made in predictive modeling for soccer, there remains substantial opportunity for improvement in areas such as real-time analysis and decision support systems. The findings of this review contribute to both the theoretical understanding of sports analytics and its practical applications in professional soccer.

Keywords: soccer analytics, artificial intelligence, machine learning, match prediction

1 INTRODUCTION

1.1 Background

Soccer is a complex team sport characterized by dynamic player interactions, tactical strategies, and real-time decision-making. As highlighted by [4], in recent years, automated soccer match analysis, evaluation, and decision-making have received increasing interest from both professional sports analytics and the academic artificial intelligence (AI) research community. The advancement in data collection technologies, including video capturing systems, optical tracking systems, and wearable devices, has enabled more sophisticated analysis of soccer matches. Moreover, the development of AI technologies, particularly in areas of computer vision and deep learning, has opened new possibilities for analyzing and predicting soccer match outcomes.

1.2 Motivation and Importance

The motivation for this research stems from the growing need for more accurate and comprehensive soccer match analysis tools. According to [7], predicting soccer match outcomes is a crucial task that impacts various stakeholders, from team managers to betting organizations. Traditional approaches to match analysis often rely on basic statistics and expert intuition. However, the complexity of soccer, with its numerous interacting variables and dynamic nature, makes it an ideal candidate for advanced analytical methods. As noted by [5], machine learning approaches can help uncover patterns and relationships in soccer data that might not be apparent through conventional analysis.

1.3 Research Objectives and Contributions

This research aims to achieve the following objectives:

- To develop a realistic simulation of the UEFA Champions League (UCL) using a combination of discrete-event simulation and machine learning to model matches
- To incorporate player statistics, team formations, and style of play in modeling match dynamics
- To simulate each second of a match, capturing major events like passes, tackles, shots, and substitutions
- To predict player actions and their likelihood of success based on player stats, game context, and fatigue levels

The key contributions of this work include:

- A novel approach combining discrete-event simulation with machine learning for soccer match modeling
- Integration of dynamic aspects like player fatigue, injuries, and strategic shifts

- Development of insights into player performance, substitution effects, and winning probabilities under various scenarios
- A framework for analyzing tactical formations and team strategies

1.4 Paper Structure

The remainder of this paper is organized as follows: Section II reviews related work in traditional sports analytics and AI applications in soccer. Section III presents our methodology, including problem formulation, data processing, and model development. Section IV describes the experimental setup and results analysis. Section V discusses key findings, practical implications, and limitations. Finally, Section VI concludes the paper with a summary of contributions and future research directions.

2 LITERATURE REVIEW AND RELATED WORK

2.1 Traditional Sports Analytics in Soccer

2.1.1 Match Statistics and Event Data Analysis

Traditional soccer analytics primarily focuses on recording and analyzing key match events and statistics. According to [5], event data provides a chronological record of match events with structured information about event types (passes, shots, fouls, tackles), timestamps, involved players, and player positions. [4] notes that event data is mainly obtained through manual annotation, though recent efforts have explored automatic annotation methods using machine learning approaches.

The granularity of event data varies by provider, with mainstream vendors like Opta, WyScout, and StatsBomb offering different levels of detail [5]. These datasets typically contain 1000-2000 events per match, providing a comprehensive view of match dynamics. [7] emphasizes that while event data is valuable for analyzing key moments, it may miss important contextual information between recorded events.

2.1.2 Player Tracking and Performance Metrics

Player tracking systems have revolutionized performance analysis in soccer. As detailed by [2], modern tracking systems can record players' positions at rates of up to 10 frames per second, enabling detailed analysis of movement patterns and physical performance. Performance metrics commonly tracked include total distance covered (TDC) and distance covered in high-intensity running (HIR, speed > 19.8 km/h).

Research by [1] shows that during official matches, players are exposed to varied physical demands, including running, jumping, sprints, accelerations, decelerations, and constant directional changes. [2] found that professional soccer players cover between 9000 and 12,500 meters per match (100-130 m/min), with significant variations based on playing position and match context.

2.1.3 Team Tactical Analysis

Team tactical analysis has evolved significantly with the availability of detailed tracking data. [4] describes how tactical analysis now incorporates both individual player actions and collective team behavior. Modern analysis examines formations, pressing patterns, and spatial relationships between players.

A significant advancement in tactical analysis is the consideration of match status phases. [2] identified seven distinct phases that affect team performance: Drawing to Winning (DW), Losing to Drawing (LD), Winning to Winning (WW), Drawing to Drawing (DD), Losing to Losing (LL), Drawing to Losing (DL), and Winning to Drawing (WD). Their research showed that these phases significantly influence players' physical performance and tactical decisions.

2.1.4 Existing Prediction Models

Early prediction models in soccer relied heavily on statistical approaches. [6] pioneered the use of Bayesian models for match prediction, incorporating team abilities and match-specific factors. Their work demonstrated the possibility of predicting match outcomes using historical data and statistical inference.

More recent statistical models have become increasingly sophisticated. [7] describes various approaches including:

- Poisson models for goal prediction
- Bayesian networks for team performance analysis
- Time series models for form prediction
- Regression models for specific outcome predictions

However, these traditional models face limitations. As noted by [5], they often struggle to capture the complex, non-linear relationships in soccer data and may oversimplify the dynamic nature of match play. [4] suggests that these limitations have driven the recent shift toward machine learning approaches, which can better handle the complexity and dimensionality of modern soccer data.

2.2 AI and Machine Learning in Soccer

2.2.1 Recent AI Applications

The integration of AI in soccer analysis has grown significantly in recent years. As [4] notes, "AI brings game-changing approaches for soccer analytics where soccer has been a typical benchmark for AI research." Recent applications span multiple domains, from performance analysis to injury prediction and tactical decision-making.

A key development has been in computer vision and NLP technologies. [5] describes how "the development of technology, and subsequently the large amount of data available, has become ML in an important strategy to help team staff members in decision-making predicting dose-response relationship reducing the chaotic nature of this team sport." These technologies enable automated extraction of match events and player movements from video data, significantly improving data collection efficiency.

[3] highlights that machine learning is being increasingly used in medicine and health sciences for soccer, particularly in injury prevention and performance optimization. The authors emphasize that "machine learning use is based on the assumption that the computer and the algorithms will learn as we feed them with more data."

2.2.2   Machine Learning Models

Various machine learning models have been applied to soccer analysis. [7] presents a comprehensive framework incorporating multiple algorithms including:

- Logistic Regression (LR)
- Random Forest (RF)
- Support Vector Machine (SVM)
- XGBoost (XGB)
- Light Gradient Boosting Machine (LightGBM)

The authors found that "ensemble techniques such as stacking or voting are also explored to improve the accuracy of basic machine learning models." Their research demonstrated prediction accuracies ranging from 54 to 58 percent, surpassing traditional betting odds providers.

2.2.3   Deep Learning Approaches

Deep learning has shown particular promise in soccer analysis. [4] describes how "successful applications of computer vision (CV) and natural language processing (NLP) technologies represent the maturity of perception intelligence, especially with recent large-scale models." These approaches have enabled more sophisticated analysis of player movements and team tactics.

Convolutional Neural Networks (CNNs) have become increasingly popular. As [5] notes, CNNs have been successfully applied to:

- Player detection and tracking
- Event recognition
- Tactical pattern analysis
- Performance prediction

2.2.4   Reinforcement Learning in Soccer

Reinforcement learning (RL) represents a particularly promising direction for soccer analytics. [4] highlights that "the development of deep reinforcement learning (DRL), such as Alpha-x of DeepMind and the series OpenAI algorithms, have contributed a large amount to decision-making intelligence."

The application of RL in soccer faces unique challenges. According to [4], these include:

"i) Modelling challenge: Soccer features more on-pitch players (22), a longer match period (90 min for a regular match), and a larger pitch size compared with other sports.
ii) Cooperation challenge: Soccer matches rarely involve goals scored, which is a typical sparse reward problem.
iii) Interpretability challenge: Unlike other video game problems such as Starcraft or Dota, the analysis and decision-making models for soccer should be more interpretable."

To address these challenges, researchers have developed various approaches. [7] describes the use of "hybrid machine learning frameworks that combine multiple algorithms to improve prediction accuracy." These frameworks often integrate traditional statistical methods with modern AI approaches to leverage the strengths of both.

Virtual environments have become increasingly important for RL research in soccer. [4] notes that "in 2019, Google released a RL training environment, i.e., Google Research Football (GRF). Until now, GRF has been a fundamental benchmark for multi-agent RL (MARL) research." These environments enable researchers to train and test AI models without the constraints and costs associated with real-world data collection.

2.3   Discrete-Event Simulation

2.3.1   Discrete-Event Simulation Framework

Discrete-event simulation (DES) provides a powerful framework for modeling soccer matches by breaking down the continuous game into discrete events and decision points. According to [7], the simulation can model individual actions and their outcomes while accounting for the dynamic state of the game. Rue and Salvesen [6] pioneered this approach, demonstrating that soccer matches can be modeled as a series of connected probabilistic events.

The simulation framework typically divides match progression into key events such as passes, shots, tackles, and strategic decisions. As noted by [4], "Soccer matches require numerous event-based decisions including running, jumping, sprints, accelerations, decelerations, and constant changes in direction." Each event can be modeled individually while maintaining the temporal and tactical relationships between events.

2.3.2 Event Modeling and Prediction

Individual events within the match can be modeled using various predictive approaches. [7] describes a comprehensive framework where "models were optimized through a grid-search cross-validation process" to predict specific match events. The authors found that different types of events require different modeling approaches:

- Pass events: Modeled using spatial-temporal features and player positioning
- Shot events: Incorporating expected goals (xG) models and player skill metrics
- Tactical decisions: Using formation analysis and team strategy patterns
- Physical events: Based on player stamina and fatigue models

2.3.3 Dynamic State Representation

A critical aspect of discrete simulation is maintaining an accurate representation of the game state. [2] emphasizes the importance of tracking match status phases, identifying seven distinct phases that affect team performance and decision-making. The authors note that "on average, phases resulting in the change of the match status occur during the first half, while all phases maintaining the result in the second half."

The state representation must include multiple factors:

- Player positions and movements
- Team formations and tactical setups
- Physical condition of players
- Match context (score, time, importance)
- Environmental factors

2.3.4 Integration with Machine Learning

Modern discrete simulation approaches increasingly integrate machine learning models to improve prediction accuracy. [4] notes that "the simulation engine achieves a functional state, capable of simulating full matches with dynamic player behavior, tactical variability, and event-driven outcomes."

The integration typically occurs at multiple levels:

- Event outcome prediction: ML models predict the success probability of individual actions
- State transition modeling: Neural networks model how game states evolve
- Decision modeling: RL agents learn optimal strategies for different situations
- Performance adaptation: Models account for fatigue and other dynamic factors

Critically, [7] found that "combining elements of modern portfolio theory yielded positive cumulative profits in experiments." This suggests that discrete simulation combined with ML can provide not just accurate modeling but also practical utility for decision-making.

The simulation environment must also account for what [3] calls the "multi-factorial phenomena" of soccer, where outcomes depend on complex interactions between players, tactics, and match conditions. The authors emphasize that "well-developed tactical, technical, physical, and cognitive skills are increasingly being demanded from players."

3 KEY CONCEPTS AND RECAP

3.1 Discrete-Event Simulation for Soccer

Drawing from [6], who pioneered statistical match modeling, we examine their Bayesian dynamic generalized linear model approach. The authors note that "it is more intricate than we might think to estimate the properties like the strengths of attack and defence for each team." Their framework considers:

- Time-dependent skills estimation
- Attack and defense strengths as primary variables
- Markov chain Monte Carlo techniques for inference
- Home-ground advantage effects

3.2 Match Status and Performance

[2] They provide crucial insights into how match status affects performance. They identify seven distinct phases:

- DW (Drawing to Winning)
- LD (Losing to Drawing)
- WW (Winning to Winning)
- DD (Drawing to Drawing)
- LL (Losing to Losing)
- DL (Drawing to Losing)
- WD (Winning to Drawing)

Their research shows that "players participating in the UEFA Champions League matches cover the longest TDC in DW, DL and DD phases" and that "TDC in these stages was between 111 and 123 m min1." This understanding of how match status affects performance is crucial for realistic simulation.

3.3 AI Framework Integration

From [4], we adopt the OODA (Observation-Orientation-Decision-Action) loop model for match analysis. The authors explain that "a soccer match is essentially a game about command and control (C2), i.e., controlling the right player so that they come up in the right place at the right time with the right action."

Key elements of their framework include:

- Observation: Data collection through various technologies
- Orientation: Analysis and prediction based on current status
- Decision: Strategy selection and tactical adjustments
- Action: Implementation of decisions in real-time

3.4 Machine Learning Model Integration

[7] provides a comprehensive predictive analytics framework incorporating multiple algorithms. Their research found that "using weather data alongside team-based box score data, as well as constructs such as fatigue and momentum" improved prediction accuracy. Their key contributions include:

- Ensemble techniques for prediction improvement
- Integration of environmental factors
- Consideration of team dynamics and momentum
- Novel evaluation metrics for model performance

3.5 Physical Performance Factors

From [1], we incorporate understanding of how physical fatigue affects performance. Their systematic review reveals that "physical fatigue leads to:

- Lower frequency of movements such as accelerations and sprints
- Decreased distance covered and velocity of actions
- Reduced efficiency of passes and shooting
- Increased incidence of goals in final minutes"

3.6 ML Applications Framework

[5] provides insights into machine learning applications in soccer. They emphasize that "ML becomes more relevant in soccer due to its chaotic nature and the unpredictability of players' behavior." Their framework suggests considering:

- Injury prediction and prevention
- Performance improvements
- Player potential prediction
- Team tactics optimization

3.7 Injury Risk and Performance

[3] contributes understanding of injury risk assessment through machine learning. They note that "machine learning does not seem to have a high predictive ability at the moment" but can help identify early signs of elevated injury risk. Key concepts include:

- Integration of multiple data sources
- Real-time risk assessment
- Personalized injury prevention strategies
- Performance optimization within safety constraints

4 DISCUSSION

4.1 Key Findings and Insights

Our review of soccer match simulation and predictive analytics reveals several key findings. First, the integration of traditional sports analytics with AI approaches has significantly improved our ability to model and predict soccer match outcomes. As demonstrated by [7], ensemble machine learning techniques can achieve prediction accuracies surpassing traditional bookmakers.

The importance of match status phases, as identified by [2], provides crucial insights into how game dynamics affect player performance. Their finding that "phases resulting in the change of the match status occur during the first half, while all phases maintaining the result in the second half" has significant implications for both simulation and prediction models.

The OODA loop framework proposed by [4] offers a comprehensive approach to soccer match analysis and decision-making. This framework successfully bridges the gap between real-world observation and AI-driven decision support systems, providing a structured approach to match analysis and prediction.

4.2 Practical Implications

The practical implications of our findings are substantial for various stakeholders in soccer:

For Coaches and Teams:

- Enhanced ability to analyze opponent strategies and prepare tactical responses
- Better understanding of how match status affects player performance
- Improved injury risk assessment and prevention strategies
- Data-driven decision support for player rotation and substitutions

For Analysts and Data Scientists:

- Framework for integrating multiple data sources and analysis approaches
- Methods for combining discrete-event simulation with machine learning
- Improved models for predicting match outcomes and player performance
- Tools for real-time analysis and decision support

For Sports Organizations:

- Better understanding of factors affecting match outcomes
- Improved player development and scouting systems
- Enhanced fan engagement through detailed analytics
- More efficient resource allocation based on predictive insights

4.3 Limitations and Challenges

Several significant challenges and limitations exist in current approaches:

Technical Challenges:

- As noted by [3], "machine learning does not seem to have a high predictive ability at the moment" for certain aspects like injury prediction
- The complexity of soccer makes it difficult to capture all relevant variables in simulation models
- Real-time data processing and analysis remain computationally challenging
- Integration of different data sources and formats poses technical difficulties

Methodological Limitations:

- [1] points out the limited availability of high-quality training data
- Difficulty in quantifying subjective factors like team morale and player psychology
- Challenge of modeling rare but significant events
- Balancing model complexity with interpretability

4.4 Future Research Directions

Based on our analysis, several promising directions for future research emerge:

Advanced Modeling Approaches:

- Development of more sophisticated multi-agent reinforcement learning models
- Integration of psychological and tactical factors into simulation models
- Improved methods for real-time analysis and prediction
- Enhanced frameworks for modeling player interactions and team dynamics

Data Integration and Analysis:

- Better methods for combining different types of match data
- Improved approaches for handling missing or noisy data

- Development of standardized data formats and analysis protocols
- Enhanced techniques for real-time data processing

Practical Applications:

- Development of more user-friendly analysis tools for coaches and analysts
- Creation of automated tactical recommendation systems
- Improved injury prediction and prevention systems
- Enhanced methods for player development and scouting

As suggested by [4], bridging the gap between real-world and virtual domains represents a particularly promising direction for future research. This includes developing better simulation environments and improving the transfer of learned strategies from virtual to real-world scenarios.

5 CONCLUSION

5.1 Summary of Contributions

This comprehensive review has examined the key components needed for developing a realistic simulation of the UEFA Champions League using AI and sports analytics. Our analysis has identified several important contributions:

First, we have established the theoretical foundation for combining discrete-event simulation with machine learning for soccer match modeling. This integration enables the simulation of individual match events while accounting for dynamic factors such as player fatigue, tactical adjustments, and match context [4].

Second, we have identified the critical components needed for modeling player performance and decision-making within matches. This includes methods for simulating key events like passes, tackles, shots, and substitutions, while incorporating player statistics, team formations, and playing styles [2], [7].

Third, we have outlined frameworks for predicting player actions and their success probabilities based on multiple factors including player stats, game context, and fatigue levels [1]. This multi-factorial approach provides a more realistic basis for simulation than traditional statistical methods alone.

5.2 Concluding Remarks

The development of a comprehensive soccer simulation system represents a significant step forward in sports analytics. Our review suggests that by combining discrete-event simulation with machine learning techniques, it is possible to create a realistic model

of soccer matches that accounts for both individual player actions and team-level dynamics.

The proposed simulation framework, focusing on the UEFA Champions League, provides a valuable tool for:

- Testing tactical concepts and refining strategic principles
- Analyzing player performance under various match conditions
- Evaluating the impact of substitutions and formation changes
- Predicting match outcomes with consideration of multiple variables

Moving forward, this work lays the groundwork for developing more sophisticated simulation tools that can serve both analytical and practical purposes in professional soccer. The integration of AI with traditional sports analytics offers promising opportunities for advancing our understanding of soccer dynamics and improving decision-making in competitive environments.

REFERENCES

[1] F. Dambroz, F. M. Clemente, and I. Teoldo, "The effect of physical fatigue on the performance of soccer players: A systematic review," *PLOS ONE*, vol. 17, no. 7, pp. e0270099, 2022.

[2] M. Konefał, Ł. Radzimiński, J. Chmura, T. Modrić, M. Zacharko, A. Padrón-Cabo, D. Sekulic, S. Versic, and P. Chmura, "The seven phases of match status differentiate the running performance of soccer players in UEFA Champions League," *Scientific Reports*, vol. 13, pp. 6675, 2023.

[3] G. P. Nassis, E. Verhagen, J. Brito, P. Figueiredo, and P. Krustrup, "A review of machine learning applications in soccer with an emphasis on injury risk," *Biology of Sport*, vol. 40, no. 1, pp. 233–239, 2023.

[4] Z. Pu, Y. Pan, S. Wang, B. Liu, M. Chen, H. Ma, and Y. Cui, "Orientation and decision-making for soccer based on sports analytics and AI: A systematic review," *IEEE/CAA Journal of Automatica Sinica*, vol. 11, no. 1, pp. 37–57, 2024.

[5] M. Rico-González, J. Pino-Ortega, A. Méndez, F. M. Clemente, and A. Baca, "Machine learning application in soccer: a systematic review," *Biology of Sport*, vol. 40, no. 1, pp. 249–263, 2023.

[6] H. Rue and Ø. Salvesen, "Prediction and retrospective analysis of soccer matches in a league," *Journal of the Royal Statistical Society: Series D (The Statistician)*, vol. 49, no. 3, pp. 399–418, 2000.

[7] A. Wong, E. Li, H. Le, G. Bhangu, and S. Bhatia, "A predictive analytics framework for forecasting soccer match outcomes using machine learning models," *Decision Analytics Journal*, vol. 14, pp. 100537, 2025.