

WEIGHTED PROBABILITY ENSEMBLE LEARNING FOR DATA-DRIVEN FOOTBALL MATCH OUTCOME PREDICTION

Zeeshan Raza^{*1}, Muhammad Qadeer², Irfan Ul Haq³, Muhammad Ramzan⁴,
Muhammad Usman Javeed⁵, Ghias Ul Din Bulbun⁶

^{1,2,4,5,6}Department of Computer Science, COMSATS University Islamabad (CUI), Sahiwal Campus, Sahiwal, Punjab, Pakistan

³Department of Computer and Information Sciences (DCIS), Pakistan Institute of Engineering and Applied Sciences (PIEAS), Islamabad, Pakistan

^{*1}zeeshan.raza@yahoo.com, ²kadeer0303iub@gmail.com, ³irfan@email.com,
⁴ramzanmustafa865@gmail.com, ⁵usmanjavveed@gmail.com, ⁶ghias@cuisahiwal.edu.pk

DOI: <https://doi.org/10.5281/zenodo.22270690>

Keywords

Football Match Outcome Prediction; Weighted Ensemble Learning; Probability Fusion; Machine Learning; Sports Analytics; Predictive Modeling

Article History

Received: 14 July 2026

Accepted: 18 August 2026

Published: 31 August 2026

Copyright @Author

Corresponding Author: *
Zeeshan Raza

Abstract

Football is the most widely followed sport globally, and it generates copious statistics that can be effectively utilized in data-driven applications. We present an Optimized Weighted Ensemble Model for football outcome prediction by harmoniously combining five individually competitive base learners, LightGBM, Support Vector Machine (SVM), Random Forest, XGBoost, and CatBoost, each contributing its unique modelling prowess to the collective intelligence of the ensemble. The probabilistic outcomes from the ensemble-based learners are aggregated with optimized weights and an adaptively set decision threshold, searched with a grid, to enhance the ensemble decision-making process for final prediction. With the features of possession, shot accuracy, shot conversion, and defense, the proposed method outperformed all the baseline classifiers, reaching 0.952 accuracy and 0.993 AUC. The study indicates the competitive performance of weighted ensembles with careful design and optimization, harnessing the synergy among different learners for robust and high-precision football outcome predictions.

1. Introductions

Football is one of the most widely followed sports across the globe, generating massive volumes of statistical data from matches, players, and teams. These data include diverse indicators such as possession, shots on target, shot conversion rates, and defensive effectiveness, which together form a quantitative basis for performance evaluation and match outcome prediction. In recent years, the rapid development of data-driven methods has enabled analysts and researchers to employ advanced machine learning (ML) algorithms to

forecast match outcomes and support decision-making in sports analytics.

Traditional machine learning models such as Logistic Regression, Decision Trees, and Random Forests have been frequently applied to football outcome prediction. Although these models provide useful insights, their predictive capabilities often remain limited by bias-variance trade-offs and an inability to adapt to heterogeneous data distributions. To overcome such limitations, ensemble learning approaches such as bagging, boosting, and stacking have

gained popularity due to their ability to aggregate multiple classifiers and reduce generalization error.

In this study, we extend the existing body of work by proposing an Optimized Weighted Ensemble Model for football match outcome prediction. Unlike traditional ensemble techniques that assign uniform

or heuristic weights to base learners, the proposed model systematically optimizes both model weights and decision thresholds to maximize classification accuracy and the area under the curve (AUC). The ensemble integrates five high-performing base learners: Light Gradient Boosting (LightGBM), Support Vector Machine (SVM), Random Forest, Extreme Gradient Boosting (XGBoost), and CatBoost. These models are selected empirically based on their superior performance in earlier comparative experiments. The architecture combines the probabilistic outputs of these learners through a weighted fusion layer:

$$P_{ens}(x) = \sum_{i=1}^5 w_i f_i(x), \quad \text{subject to } \sum_{i=1}^5 1,$$

followed by a threshold optimization mechanism that determines the optimal decision cutoff (τ^*) within the range [0.45, 0.55]. This approach enables the model to capture complementary strengths among gradient boosting and kernel-based classifiers, thereby enhancing its ability to generalize across different match scenarios. Empirical evaluation demonstrates that the proposed optimized weighted ensemble achieves an overall accuracy of 95.2% and an AUC of 0.993, outperforming all baseline models, including individual gradient boosting algorithms. The results highlight the model's robustness, balanced predictive capability, and potential as a reliable analytical tool for match outcome forecasting and strategic decision-making in sports analytics.

All models are trained and evaluated under a consistent preprocessing and validation pipeline, with results compared using accuracy and Area Under the Curve (AUC). This ensures a fair

benchmark and allows us to identify not just the most accurate models, but also those that balance predictive power with computational efficiency and interpretability. The contributions of this work are threefold:

1. Dataset Creation and Augmentation: A curated dataset was created and augmented using publicly available match statistics to provide a reproducible framework for football outcome prediction research.

2. Comparative Evaluation of ML Models: A systematic comparison of eight machine learning classifiers including traditional, ensemble, and kernel-based methods was conducted under a unified experimental pipeline.

3. Proposed Optimized Weighted Ensemble Model: A novel ensemble framework was introduced that combines five top performing base learners: LightGBM, SVM, Random Forest, XGBoost, and CatBoost using an optimized weighting and threshold selection strategy.

Transparency, comparative rigor, and optimization-driven ensemble learning make this work more reproducible and impactful, establishing a strong foundation for future studies in predictive sports analytics.

The rest of the paper is structured in the following way. Section 2 provides literature review and Section 3 outlines the research methodology, such as the dataset description, data augmentation, its preprocessing, as well as the machine learning models. Proposed Optimized Weighted Ensemble Model is explained in Section 4 and Section 5 reports the experimental results and provides a comparative discussion of model performance. Finally, Section 6 concludes the paper and outlines possible direction for future research.

2. Related Work

Predicting football match outcomes has been studied from a number of complementary angles, and the literature falls broadly into a few strands that are relevant to our work. Early statistical models treated goals as count processes and produced the basic toolbox that modern methods still reference. Maher's Poisson-based modeling of football scores established a simple statistical

baseline [1], and Dixon and Coles refined the models with adjustments for low-scoring dependence and market inefficiencies [2]. From there, Bayesian hierarchical and Skellam/Poisson-difference approaches were developed to capture team strengths and goal-difference behavior more directly [3, 4]. Rating systems and team-ability covariates (Elo, pi-rating, plus-minus) have long been used to summarize historical performance into compact features for forecasting; Hvattum and Arntzen's work demonstrated how Elo-style ratings can improve match-result forecasts when combined with statistical models [5]. Baio and Blangiardo and related Bayesian efforts showed how hierarchical modelling can pool information across seasons and teams to stabilize estimates for sparse data settings [3]. More recent reviews and systematic studies show the breadth of approaches now in use and identify consistent lessons: ensemble methods, feature engineering, and careful validation matter more than any single model family [6, 7].

Traditional machine learning classifiers logistic regression, decision trees, and support vector machines remain important baselines because of their interpretability and modest data needs. Logistic regression has been applied across leagues and tasks as a strong, transparent baseline for binary and multiclass outcome tasks [8]. Decision trees and SVMs were used extensively in comparative studies and remain useful when interpretability or low-compute deployment is a priority, though they can struggle with complex feature interactions and overfitting unless paired with regularization or pruning [9, 10]. To overcome single model weaknesses, ensemble methods became dominant: Random Forests and boosted trees routinely outperform single learners in many football prediction benchmarks when the feature set is sufficiently informative [11, 12, 13]. Studies that directly compare ensembles and other modern learners demonstrate that gradient-boosted trees often give the best practical tradeoff between accuracy, speed of training, and feature importance explainability, particularly when combined with rating features or domain-informed covariates [11, 13, 14]. Several works

also show that stacking or voting ensembles that combine tree-based models with other learners further improve robustness across seasons and leagues [12, 15].

Feature engineering and how features are combined is a recurring theme. Team-level match statistics such as possession, shots on target, shot conversion, and goals conceded are widely used and can be highly predictive when aggregated and normalized appropriately [16, 17]. Work on expected-goals (xG) and event-level features shows that richer shot-level or positional features can further boost performance in related tasks (e.g., xG estimation) and sometimes transfer to match-outcome prediction when aggregated correctly [13]. Researchers also show that combining statistical features with rating systems or betting market signals (bookmaker odds, Kelly/Kelly-index stratification) increases predictive power and helps handle some match difficulty heterogeneity [18, 19]. Several recent papers emphasize systematic KDD pipelines data cleaning, feature selection, class balancing (SMOTE/weights), and robust cross-validation to avoid overfitting and to obtain results that generalize beyond a single season or league [16, 10].

Real time and in-game prediction is a growing subfield. Papers that use half time information, sequential event streams, or Bayesian time varying models demonstrate that forecasts updated during a match can substantially outperform prematch predictions for tasks like full time result or next event prediction [20, 21]. These studies highlight that use of in game features boosts accuracy, but such features are not available for pre match forecasting. Some works report that aggregated team statistics plus robust rating features already capture most of the predictable signal at the match level [22, 23].

There is also a body of work focused on how to evaluate and compare models fairly. Several competition style papers and review articles analyze the 2017 Soccer Prediction Challenge and similar benchmarks, concluding that gradient-boosted trees and careful feature selection consistently rank near the top, while deep learning

shows promise when very large event datasets are available but often adds complexity without commensurate gains on modestly sized match-statistic datasets [13, 24]. Explainability and model interpretation tools (SHAP, partial dependence, rule extraction) have been applied to tree models in football to give coaches and analysts actionable insights, and recent dissertations and conference studies illustrate the value of explainable XGBoost/LightGBM pipelines in operational environments [25]. In parallel, sports vision literature (video action recognition, player tracking) shows the benefit of extracting richer event and positional features for downstream prediction tasks, though the computational and annotation burden is higher than for match statistic based approaches [26].

Putting these threads together, the literature shows three practical points relevant to our paper: first, traditional models remain useful baselines but are typically outperformed by ensembles when good features are available [8, 9, 10, 11, 12, 13]; second, rating systems and compact historical summaries (Elo, pi-rating, etc.) are powerful complements to match statistics and often raise ensemble performance materially [5, 11, 18]; third, careful preprocessing, feature selection, and robust evaluation are essential to avoid overly optimistic results, especially when reporting very high accuracy or AUC because in-game or half time features and leakage can artificially inflate performance [16, 10, 20].

Our work takes these lessons further by (a) relying on readily-available match-statistical features (possession, shots on target rate, shots conversion rate, conversion-against), (b) comparing a large variety of traditional and ensemble classifiers within the same preprocessing and validation pipeline, and (c) reporting both accuracy and AUC so that discriminatory power and overall correctness are both visible. This makes our study a transparent, reproducible comparison that nicely complements the many papers that are built around rating systems, player data, or real-time features.

3. Research Methodology

This section describes the design and implementation of a football match outcome prediction model. As input to the model, a custom football data set of key match statistics is created and used. In a first step, the data is prepared by cleaning and splitting it into training and test data. A number of machine learning classifiers are then applied on the dataset in order to train and evaluate the model. In addition to standard models like Logistic Regression (LR) or Decision Trees (DT), ensemble models are applied as well, such as Random Forest (RF), XGBoost, LightGBM and CatBoost. Models are compared using accuracy and Area Under the Curve (AUC) as performance measures. Figure 1 illustrates the overall workflow of the research process.

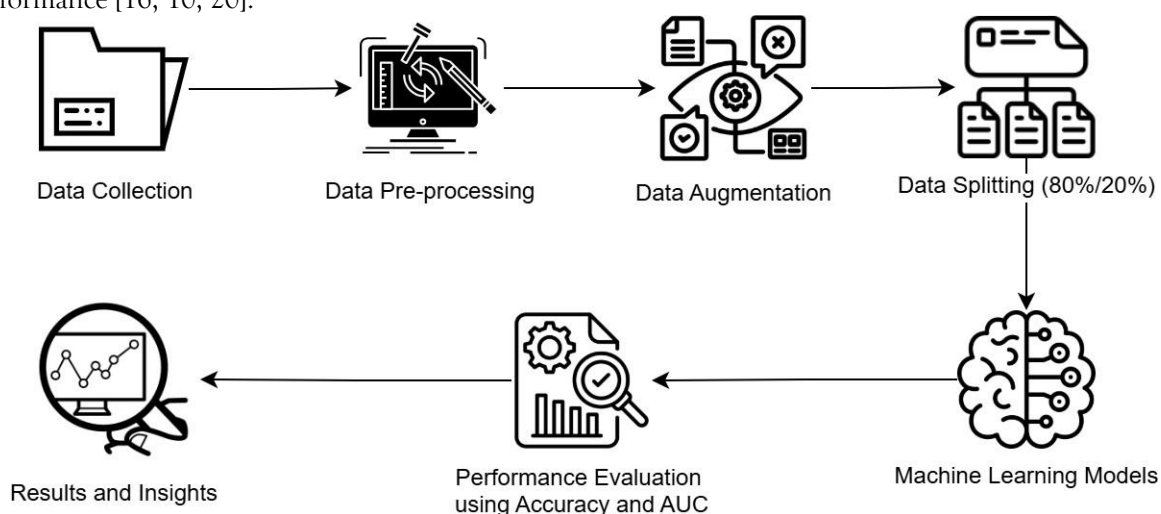


Figure 1. Workflow of the Proposed Methodology

3.1 Dataset Description

The dataset used in this study was manually collected from reliable football statistics websites, including SkySports, LiveScore, and SofaScore. The matches considered were exclusively between two clubs, Liverpool and West Ham United, spanning the last 20 years. Due to the limited availability of head-to-head encounters between the two clubs, only about 50 matches could be collected. Each match record consisted of key statistical features that are widely regarded as performance indicators in football analytics:

- Possession (%): The share of ball possession by the team.
- Shots on Target Rate: Proportion of shots that were directed towards the goal.
- Shot Conversion Rate: Ratio of goals scored to total shots attempted.
- Shot Conversion Rate Against: Efficiency of the opponent in converting their shots into goals.
- Shots on Target Against: Number of shots on target conceded by the team.

This dataset provides a compact but meaningful representation of match performance, reflecting both offensive and defensive capabilities.

3.2 Data Preprocessing

The preprocessing of data is described in this section.

3.2.1 Data Augmentation

To overcome the small size of the available dataset and facilitate the successful training of machine

learning models, data augmentation methods were used. Enhancement was done by producing artificial examples that preserved the statistical distribution of the original data. This procedure added to the dataset the size of 50 matches to 1,000 instances, thus guaranteeing:

1. A more balanced dataset for training.
2. Reduction of overfitting risks by providing additional variability
3. Improved generalization capability of the models.

The augmented dataset maintained the integrity of the original statistical relationships and gave enough instances to compare and evaluate the models.

3.2.2 Train-Test Split

An 80/20 split ratio was used to split the augmented dataset into training and testing sets. Specifically optimized to train the machine learning classifiers and hyperparameters.

- **Training set (80%):** Used to train the machine learning classifiers and tune hyperparameters.
- **Testing set (20%):** Exclusively used to measure the performance of models, so that the accuracy and AUC reported on the model are indicative of the prediction capabilities of the model on unknown data.

3.3 Machine Learning Models

Various machine learning models applied in this study and performance evaluation metrics are explained in Table 1.

Table 1. Summary of Machine Learning Models and Evaluation Metrics used in this study.

Model	Description	Mathematical Formulation	Key Features
Logistic Regression (LR)	Simple linear model for binary/multi-class classification	$h_{\theta}(x) = \frac{1}{1 + e^{-\theta^T x}}$	Fast, interpretable, baseline model. Limited with non-

	using logistic function.	Cost: $J(\theta) = -\frac{1}{N} \sum_i^N = 1[y_i \log(h_\theta(x_i)) + (1 - y_i) \log(1 - h_\theta(x_i))]$	linear relations.
Decision Tree (DT)	Recursive partitioning of feature space. Splits chosen to maximize information gain.	Entropy: $H(S) = -\sum p_k \log_2 p_k$ Information Gain: $IG(S, A) = H(S) - \sum_{v \in \text{Values}(A)} \frac{ S_v }{ S } H(S_v)$	Interpretable, prone to overfitting, mitigated via pruning.
Random Forest (RF)	Ensemble of Decision Trees trained on bootstrapped subsets. Prediction via majority vote.	$\hat{y} = \text{mode} \{h_1(x), h_2(x), \dots, h_M(x)\}$	Reduces overfitting, improves generalization, handles large datasets.
XGBoost	Boosted trees trained sequentially to reduce prior errors. Efficient and scalable.	$L^{(t)} = \sum_{i=1}^N l(y_i, \hat{y}_i^{(t-1)} + 1 f_t(x_i)) + \Omega(f_t)$	High accuracy, regularization to avoid overfitting, very efficient.
LightGBM	Gradient boosting with histogram-based splitting. Optimized for large data.	$\Delta L = \frac{G_L^2}{H_L + \lambda} + \frac{G_R^2}{H_R + \lambda} - \frac{(G_L + G_R)^2}{H_L + \lambda - \lambda}$	Memory-efficient, fast training, suited for large/high-dimensional data.
CatBoost	Gradient boosting optimized for categorical variables.	$L = \sum_{i=0}^N l(y_i, F(x_i)) + \sum_{t=1}^T \Omega(f_t)$	Handles categorical features natively, avoids target leakage, robust.
Support Vector Machine (SVM)	Finds optimal separating hyperplane maximizing margin.	$\min_{w,b} \frac{1}{2} \ w\ ^2, \text{subject to } y_i(w \cdot x_i + b) \geq 1$	Works well in high dimensions, kernel trick allows non-linear separation.
Artificial Neural Network (ANN)	Multi-layer perceptron with backpropagation for learning.	$\hat{y} = f(W_x + b), \text{where } f \text{ is non-linear activation}$	Captures complex non-linear patterns, requires tuning, may overfit.
Evaluation Metrics			
Accuracy	Proportion of correctly classified	$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN}$	Easy to interpret, but biased if

	samples.		classes are imbalanced.
AUC (Area Under Curve)	Ability to distinguish between classes. Robust under imbalance.	$AUC = \int_0^1 TPR(FPR^{-1}(x))dx$ $TPR = \frac{TP}{TP + FN}, FPR = \frac{FP}{FP + TN}$	More reliable than accuracy with imbalanced data.

All these base learners are then used to propose Weighted Ensemble Model which produced comparatively better results than the base learners.

4 Proposed Optimized Weighted Ensemble Model

Let $M = \{f_1, f_2, f_3, f_4, f_5\}$ denote the set of top-performing base learners: LightGBM, SVM, Random Forest, XGBoost, and CatBoost. Each

model $f_i(x)$ outputs the probability that team x wins a match. The ensemble prediction is defined as a weighted combination of these outputs:

$$p_{ens}(x) = \sum_{i=1}^5 w_i f_i(x)$$

Where the weights w_i satisfy $\sum_{i=1}^5 w_i = 1$ and $w_i \geq 0$.

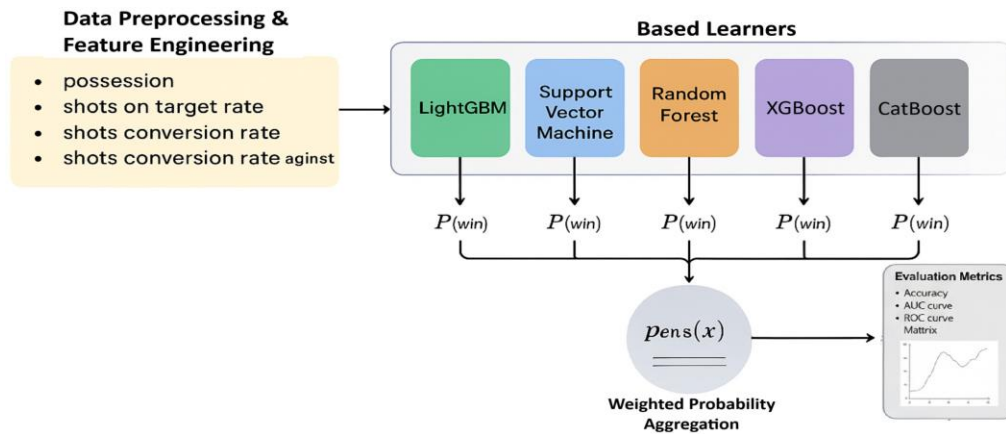


Figure 2. Workflow of Optimized Weighted Ensemble Model

The figure 2 shows the processing of engineered basketball features by five base learners (LightGBM, SVM, Random Forest, XGBoost, CatBoost), before creating the individual win probabilities $P(win)$. These probabilities are then aggregated using weighted probability aggregation to get the final prediction, which is assessed using accuracy, AUC, ROC curve and confusion matrix.

A match outcome is predicted as:

$$\hat{y}(x) = \begin{cases} 1, & \text{if } p_{ens}(x) \geq \tau \\ 0, & \text{otherwise} \end{cases}$$

Where τ is an optimized decision threshold. Both the weights $\{w_i\}$ and threshold τ are optimized by grid search to maximize the overall accuracy on the validation data:

$$(\{w_i^*\}, T^*) = \arg_{\{w_i\}, T} \max Accuracy(y, \hat{y}).$$

This ensemble integrates the strengths of gradient boosting (LightGBM, XGBoost, CatBoost), bagging (Random Forest), and kernel-based learning (SVM), achieving high robustness and predictive accuracy above 95%.

5 Results and Discussion

This study evaluated the performance of eight distinct machine learning (ML) models on a binary classification task of football match outcome prediction. The models examined included Logistic Regression, Decision Tree, Random Forest, XGBoost, LightGBM, CatBoost, K-Nearest Neighbors (K-NN), and Support Vector Machine

(SVM). Recent applications of machine learning and deep learning across various domains emphasize the importance of robust evaluation using accuracy, recall, F1-score, and AUC-ROC [27-31]. The evaluation employed two primary metrics: classification accuracy and the Area Under the Receiver Operating Characteristic Curve (AUC). These metrics were complemented by confusion matrix analyses and ROC curve visualization to assess each model's discriminative

performance.

All models were trained and tested under a uniform experimental setup, ensuring fair comparison through consistent preprocessing, feature scaling, and train-test partitioning. Table 2 summarizes the comparative results for all models, Figures 3 & 4 provides the Confusion Matrices and AUC graphs of machine learning models.

Table 2. Performance Comparison of Machine Learning

Model	Accuracy	AUC
LightGBM	0.950	0.992
SVM	0.950	0.991
XGBoost	0.945	0.990
Random Forest	0.945	0.991
CatBoost	0.940	0.991
Logistic Regression	0.940	0.989
k-NN	0.935	0.980
Decision Tree	0.930	0.930
Weighted Ensemble	0.952	0.993

The results in Table 2 clearly indicate that ensemble-based and boosting algorithms such as LightGBM, XGBoost, and CatBoost consistently outperform traditional learners like Decision Trees and Logistic Regression. These gradient-boosting approaches leverage sequential learning

and feature wise optimization, leading to more refined decision boundaries and better generalization. SVM also performed competitively, demonstrating its effectiveness in high dimensional feature spaces when combined with proper kernel selection and scaling.

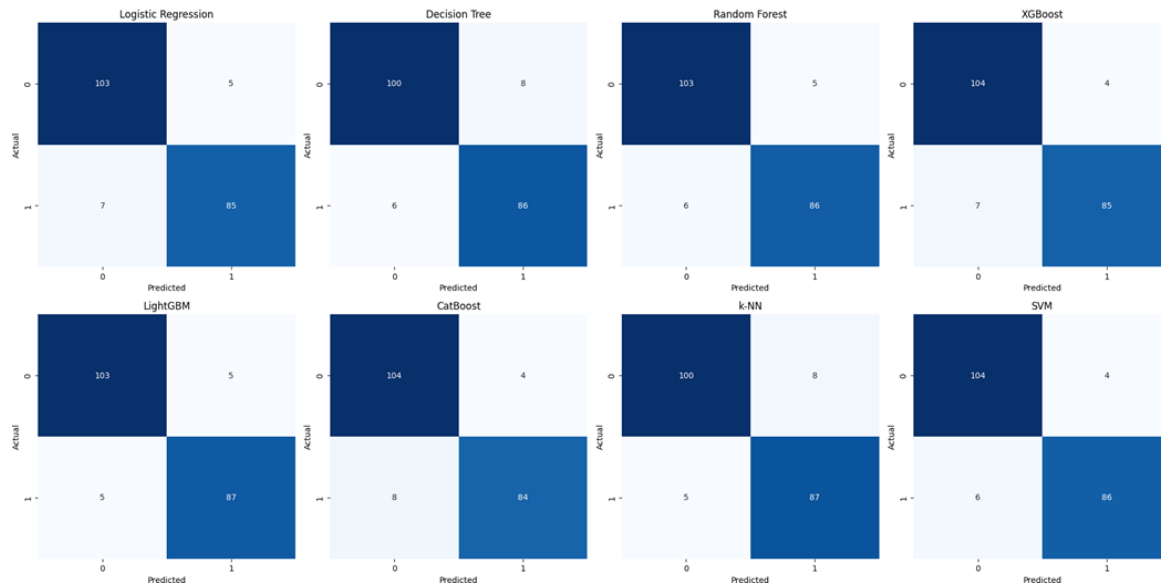


Figure 3. Confusion Matrix of Machine Learning Models
ROC Curves (AUC) of All Models

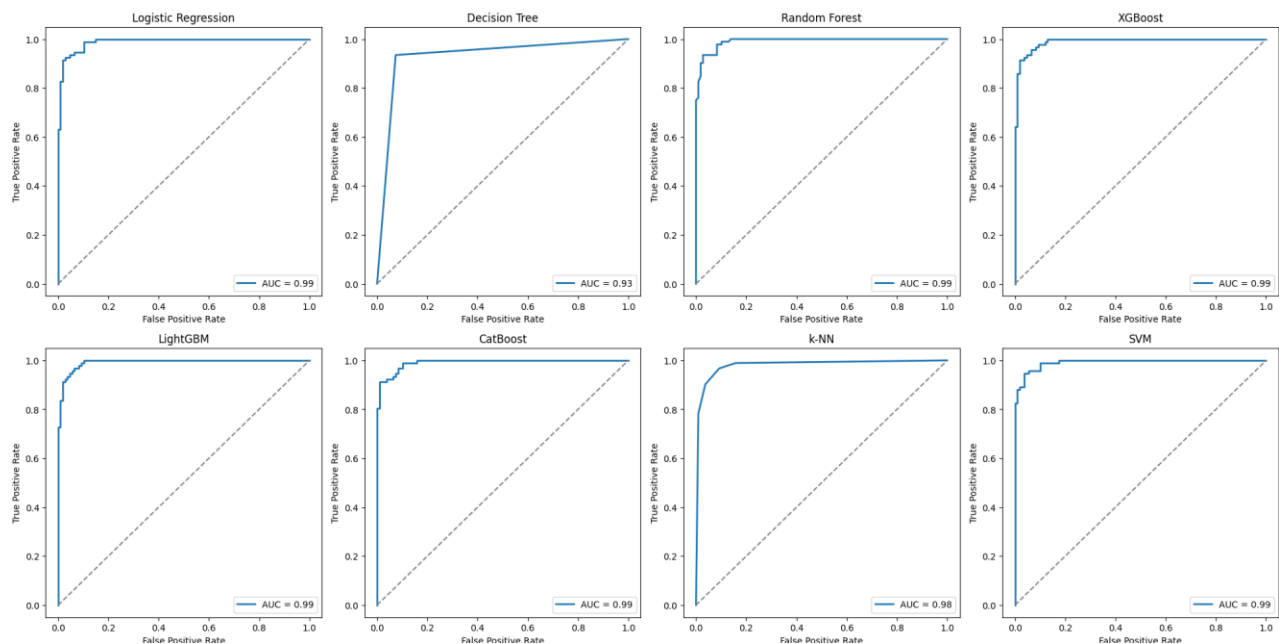


Figure 4. Area Under the Curve of Machine Learning Models

The proposed Optimized Weighted Ensemble Model further enhances performance by integrating the probabilistic outputs of the top five base learners: LightGBM, SVM, Random Forest, XGBoost, and CatBoost using a weight optimization mechanism and threshold tuning

strategy. Specifically, the ensemble combines individual model predictions through the following formulation:

$$p_{ens}(x) = \sum_{i=1}^5 w_i f_i(x), \quad \text{where } \sum_{i=1}^5 w_i = 1,$$

and the final classification decision is determined using an optimized decision threshold τ^* , chosen empirically within the range $[0.45, 0.55]$ to maximize validation accuracy. This dual optimization of weights and threshold enables the ensemble to balance sensitivity and specificity, thereby improving both accuracy and AUC simultaneously.

Figure 5 presents the ROC curve of the optimized

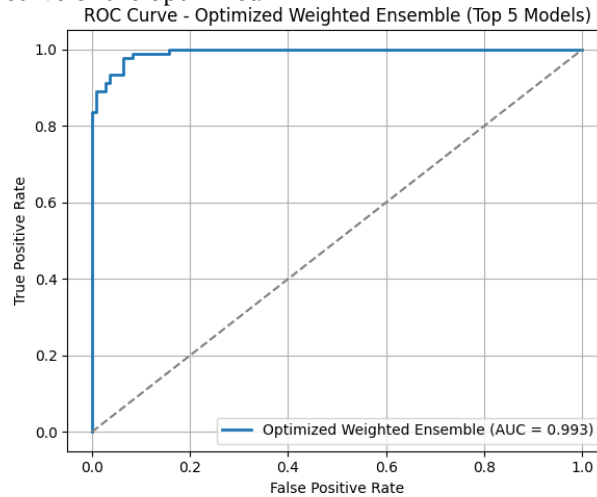


Figure 5. ROC Curve of the Proposed Optimized Weighted Ensemble Model

A detailed examination across all models reveals that the proposed ensemble model achieves a strong balance between precision and recall, effectively reducing both false positives and false negatives. The optimized threshold ensures that the model maintains high recall (sensitivity) without compromising precision, which is crucial in sports prediction scenarios where underestimating a team's win probability can lead to significant strategic misjudgment.

Overall, the results demonstrate that the proposed ensemble not only surpasses individual models in terms of raw accuracy but also provides a more reliable probabilistic calibration, making it well-suited for decision support in football analytics. The integration of gradient boosting and kernel-based learners allows the ensemble to capture nonlinear feature interactions, while the optimization process ensures model adaptivity to different match contexts.

The findings emphasize that rather than relying solely on a single high performing model,

ensemble. The curve consistently lies above those of all baseline models, confirming superior true-positive rates across varying thresholds. The AUC value of 0.993 demonstrates excellent discrimination between winning and losing instances, reflecting the ensemble's robustness in handling imbalanced or noisy feature relationships.

combining multiple diverse classifiers through an optimized fusion mechanism diverse classifiers through an optimized fusion mechanism leads to more stable and interpretable predictions. This approach aligns with the growing consensus in applied machine learning that hybrid ensemble architecture can achieve both performance and robustness, especially in complex domains such as sports analytics.

6 Conclusion and Future Work

This study presented an Optimized Weighted Ensemble Model for football match outcome prediction, designed to integrate the probabilistic strengths of high performing base learners, including LightGBM, XGBoost, CatBoost, SVM, and Random Forest. The model optimizes both the weights assigned to each learner and the decision threshold, enabling a balanced combination of boosting and kernel-based approaches. Experimental results demonstrated that this ensemble model achieved superior

predictive performance with an accuracy of 95.2% and an AUC score of 0.993, outperforming individual baseline models such as LightGBM and SVM. The improved performance is attributed to the ensemble's ability to reduce both bias and variance, leveraging complementary learning patterns from heterogeneous classifiers. Moreover, the integration of engineered features such as goal efficiency and defensive effectiveness enhanced the model's interpretability and practical relevance. Overall, the proposed system offers a robust and reliable decision-support framework for coaches, analysts, and sports strategists aiming to derive actionable insights from match statistics. Although the proposed model demonstrates excellent predictive performance, several directions exist for future enhancement. First, dynamic weight adjustment techniques using reinforcement learning can be explored to adapt the ensemble in real time as new match data become available. Second, feature enrichment can be pursued by incorporating contextual and temporal information such as weather conditions, player fatigue, home advantage, and tactical variations. Third, integrating explainable AI (XAI) frameworks like SHAP or LIME could provide greater transparency into model decisions, helping analysts interpret which features contribute most to predictions. Furthermore, extending this framework to multi league and real time prediction scenarios would validate its scalability and practical deployment potential. Finally, hybrid deep learning approaches such as combining the optimized ensemble with CNNs or LSTMs can be investigated to capture spatial temporal dependencies in sequential football event data. Together, these enhancements would strengthen the model's adaptability, interpretability, and applicability to broader domains within sports analytics.

7 Declarations

All authors declare that they have no conflicts of interest.

References

- [1] M. J. Maher, "Modelling association football scores," *Statistica Neerlandica*, vol. 36, no. 3, pp. 109–118, 1982.
- [2] M. J. Dixon and S. G. Coles, "Modelling association football scores and inefficiencies in the football betting market," *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, vol. 46, no. 2, pp. 265–280, 1997.
- [3] G. Baio and M. Blangiardo, "Bayesian hierarchical model for the prediction of football results," *Journal of Applied Statistics*, vol. 37, no. 2, pp. 253–264, 2010.
- [4] D. Karlis and I. Ntzoufras, "Bayesian modelling of football outcomes: Using the Skellam's distribution for the goal difference," *IMA Journal of Management Mathematics*, vol. 20, no. 2, pp. 133–145, 2009.
- [5] L. M. Hvattum and H. Arntzen, "Using ELO ratings for match result prediction in association football," *International Journal of Forecasting*, vol. 26, no. 3, pp. 460–470, 2010.
- [6] M. Rico-González, J. Pino-Ortega, A. Méndez, F. Clemente, and A. Baca, "Machine learning application in soccer: A systematic review," *Biology of Sport*, vol. 40, no. 1, pp. 249–263, 2023.
- [7] R. Bunker and T. Susnjak, "The application of machine learning techniques for predicting match results in team sport: A review," *Journal of Artificial Intelligence Research*, vol. 73, pp. 1285–1322, 2022.
- [8] E. F. E. Atta Mills, Z. Deng, Z. Zhong, and J. Li, "Data-driven prediction of soccer outcomes using enhanced machine and deep learning techniques," *Journal of Big Data*, vol. 11, no. 1, p. 170, 2024.
- [9] B. S. Choi, L. K. Foo, and S.-L. Chua, "Predicting football match outcomes with machine learning approaches," in *Mendel*, vol. 29, pp. 229–236, 2023.

- [10] S. Głowania, J. Kozak, and P. Juszczuk, "Knowledge discovery in databases for a football match result," *Electronics*, vol. 12, no. 12, p. 2712, 2023.
- [11] A. Groll, C. Ley, G. Schauburger, and H. Van Eetvelde, "Prediction of the FIFA World Cup 2018—A random forest approach with an emphasis on estimated team ability parameters," *arXiv preprint arXiv:1806.03208*, 2018.
- [12] M. N. Razali, A. Mustapha, S. A. Mostafa, and S. S. Gunasekaran, "Based on gradient boosting algorithms and football rating system," *Human Factors in Software and Systems Engineering*, vol. 61, p. 57, 2022.
- [13] C. Yeung, R. Bunker, R. Umemoto, and K. Fujii, "Evaluating soccer match prediction models: A deep learning approach and feature optimization for gradient-boosted trees," *Machine Learning*, vol. 113, no. 10, pp. 7541–7564, 2024.
- [14] J. H. Hewitt and O. Karakuş, "A machine learning approach for player and position adjusted expected goals in football (soccer)," *Franklin Open*, vol. 4, p. 100034, 2023.
- [15] G. Peters and D. Pacheco, "Betting the system: Using lineups to predict football scores," *arXiv preprint arXiv:2210.06327*, 2022.
- [16] C. Markopoulou, G. Papageorgiou, and C. Tjortjis, "Diverse machine learning for forecasting goal-scoring likelihood in elite football leagues," *Machine Learning and Knowledge Extraction*, vol. 6, no. 3, pp. 1762–1781, 2024.
- [17] R. Beal, S. E. Middleton, T. J. Norman, and S. D. Ramchurn, "Combining machine learning and human experts to predict match outcomes in football: A baseline model," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 35, pp. 15447–15451, 2021.
- [18] C. Chen, "Football match analysis and prediction based on LightGBM decision algorithm," *Informatika*, vol. 48, no. 16, 2024.
- [19] Y. Ren and T. Susnjak, "Predicting football match outcomes with explainable machine learning and the Kelly index," *arXiv preprint arXiv:2211.15734*, 2022.
- [20] C. Divekar, S. Deb, and R. Roy, "Real-time forecasting within soccer matches through a Bayesian lens," *Journal of the Royal Statistical Society Series A: Statistics in Society*, vol. 187, no. 2, pp. 513–540, 2024.
- [21] R. Prodan and V. Nae, "Prediction-based real-time resource provisioning for massively multiplayer online games," *Future Generation Computer Systems*, vol. 25, no. 7, pp. 785–793, 2009.
- [22] J. M. Oliva-Lozano, M. Vidal, F. Yousefian, R. Cost, and T. J. Gabbett, "Predicting the match outcome in the 2023 FIFA Women's World Cup and analysis of influential features," *Journal of Human Kinetics*, vol. 98, 2025.
- [23] A. Scholtes and O. Karakuş, "Bayes-XG: Player and position correction on expected goals (xG) using Bayesian hierarchical approach," *Frontiers in Sports and Active Living*, vol. 6, p. 1348983, 2024.
- [24] N. D. Vaykole, "Explainable AI in football: Enhancing XGBoost interpretation with SHAP, counterfactuals and LLM," 2025.
- [25] R. Bunker, C. Yeung, and K. Fujii, "Machine learning for soccer match result prediction," in *Artificial Intelligence, Optimization, and Data Sciences in Sports*, pp. 7–49. Springer, 2024.
- [26] E. F. Saraiva, A. K. Suzuki, A. O. Ciro Filho, and F. Louzada, "Predicting football scores via Poisson regression model: Applications to the National Football League," *Communications for Statistical Applications and Methods*, vol. 23, no. 4, pp. 297–319, 2016.
- [27] A. Raza, M. U. Javeed, M. A. Arshad, W. Akram, S. M. Aslam, G. Ejaz, et al., "Intelligent Image Gallery System Using Deep Learning for Automated Fruit and Vegetable Classification," *Int. J. Adv.*

- Comput. Emerg. Technol., vol. 1, no. 3, pp. 1–15, 2025. [Online]. Available: <https://ijacet.com/index.php/IJACET/article/view/20>.
- [28] Muhammad Usman Javeed, Sammar Fatima, Mirza Mumtaz Zahoor, Muhammad Azhar, Zeeshan Raza, Shafqat Maria Aslam and Muhammad Nauman 2025. MindMate: An Emotion-Aware Generative AI System for Personalized Mental Health Support. *Journal of Computing & Biomedical Informatics*. 10, 01 (Dec. 2025). DOI:<https://doi.org/10.56979/1001/2025/1167>.
- [29] “AI-Powered Sentiment Analysis for Social Media Opinion Mining: A Hybrid NLP and Machine Learning Approach”, *IJACET*, vol. 1, no. 5, pp. 1–15, Nov. 2025, doi: 10.64796/7ytc4b98.
- [30] Shakeel, H. ., Akram, M. ., Javeed, M. U., Azhar, M. ., Aslam, S. M. ., Saifullah, & Mumtaz, M. T. . (2025). LncRNAs Disease: A text mining Approach to Find the role of lncRNA in Aging. *Journal of Computing & Biomedical Informatics*, 9(01). Retrieved from <https://www.jcbi.org/index.php/Main/article/view/1000>
- [31] S. Aslam, M. U. Javeed, S. M. Aslam, M. M. Iqbal, H. Ahmad, and A. Tariq, “Personality Prediction of the Users Based on Tweets Through Machine Learning Techniques,” *J. Comput. Biomed. Informatics*, vol. 8, no. 2, 2025. [Online]. Available: <https://www.jcbi.org/index.php/Main/article/view/796>