

Article

An Expected Goals Model for Analyzing a 5-a-Side Soccer for the Blind Using Ten Machine Learning Algorithms with SHAP Interpretability

Boryi A. Becerra-Patiño ^{1,2,*} , Rodrigo Yáñez-Sepúlveda ^{3,4}  and José Pino-Ortega ⁵ 

¹ Programa de Doctorado en Ciencias de la Actividad Física y del Deporte, University of Murcia, San Javier, 30720 Murcia, Spain

² Faculty of Physical Education, National Pedagogical University, Bogota 111166, Colombia

³ Faculty Education and Humanities, School of Sport Sciences, Universidad Andres Bello, Viña del Mar 2520000, Chile; rodrigo.yanez.s@unab.cl

⁴ School of Medicine, Universidad Espíritu Santo, Samborondón 092301, Ecuador

⁵ Faculty of Sport Science, University of Murcia, 30100 Murcia, Spain; josepinoortega@um.es

* Correspondence: babecerrap@pedagogica.edu.co or boryialexander.becerrap@um.es

Abstract

Background: Currently, expected goal models are tools that enable quantitative analysis in the study of conventional sports, although they have seen very little application in the Paralympic context. **Objective:** To present a trained expected goals model for 5-a-side blind soccer games based on an analysis of 164 offensive plays by the national team that won first place at the 2022 IBSA Copa América. The novelty of this work lies in being, to our knowledge, the first expected goals (xG) model developed for Paralympic blind football (B1): conventional xG weights cannot be transferred directly because shooting in F5 is governed by auditory orientation, the absence of an offside rule, a smaller rebound-walled pitch, and fully blind executors, so a sport-specific, reproducible and SHAP-interpretable benchmark is required where none previously existed. **Materials and Methods:** The SHapley Additive exPlanations library was used to analyze the data via partial dependency plots, dependency scatter plots, waterfall plots, decision plots, and SHAP heatmaps. Additionally, ten machine learning algorithms were compared, including logistic regression, random forest, extra trees, gradient boosting, XGBoost, LightGBM, CatBoost, support vector machine, k-nearest neighbors, and multilayer perceptron, using a 70/30 stratification process with fivefold stratified cross-validation to define the main hyperparameters. **Results:** The most consistent model was CatBoost (F1 = 0.778; AUC-ROC = 0.913; AUC-PR = 0.828; MCC = 0.729; Brier = 0.072), which allowed for independent analysis and evaluation of the dataset. The five main offensive variables were determined to be (i) distance to the goal before the shot; (ii) lateral coordinate; (iii) absolute magnitude of the shooting angle; (iv) magnitude of the progression vector; (v) proximity to the side kickboard. However, none of these variables proved to be decisive in the tournament (n = 24), a characteristic that the model captured as a significant negative contribution from the opponent variable. **Conclusions:** The expected goals model considered for this study serves as a starting point for further analysis of tactical variables in 5-a-side soccer for the blind. Because the model was trained on a single team in a single tournament with few positive cases, these results should be read as preliminary, hypothesis-generating tactical insights rather than validated performance estimates, and require external validation before transfer to other teams or competitions.



Academic Editor: Oliver Ramos Álvarez

Received: 7 May 2026

Revised: 26 June 2026

Accepted: 1 July 2026

Published: 3 July 2026

Copyright: © 2026 by the authors.

Licensee MDPI, Basel, Switzerland.

This article is an open access article distributed under the terms and

conditions of the [Creative Commons](https://creativecommons.org/licenses/by/4.0/)

[Attribution \(CC BY\)](https://creativecommons.org/licenses/by/4.0/) license.

Keywords: sports analytics; adapted sports; sports science; technical–tactical performance; expected goals; explainable artificial intelligence; SHAP; blind football

1. Introduction

One of the main characteristics of 5-a-side blind soccer (F5) is that many variables influence player performance, including body composition [1], sociodemographic and psychological characteristics [2], and physical and technical–tactical variables [3,4].

F5 is characterized as a high-intensity sport in which tactical actions prevail as determining factors of athletic performance [4,5]. In this context, the main variables analyzed in this sport show that passes, dribbles, and shots on goal have emerged as the most frequently analyzed variables [5]. These technical–tactical characteristics are essential for players to meet the specific demands of competition [6,7]. Thus, the analysis of shots on goal has emerged as one of the variables influencing the outcome of the game [7].

Constant advances in research have revealed that technical and tactical demands have undergone significant changes, with a more elaborate style of play now prevailing, characterized by using passing sequences, ball control, and dribbling to create scoring opportunities [7]. However, although there are a growing number of studies analyzing technical-tactical variables, the scientific literature remains insufficient to identify new variables and analytical models for assessing game performance in F5 [4,8].

A systematic review of the study of technical–tactical actions in F5 determined that the effectiveness of shots on goals and the analysis of passing, control, and shooting actions are the most studied variables [4]; however, there is a limitation in the number of studies conducted using machine learning techniques to analyze the performance of F5 players on the basis of the tactical actions they execute in competition. Thus, one of the most widely used models in conventional soccer is the expected goals (xG) model, which seeks to estimate the probability that a shot-on goal will result in a goal on the basis of various variables derived from the shot; these probabilities can then be quantified to provide an independent description of the outcome and performance of the teams analyzed [9]. Consequently, the xG metric is one of the most representative models in soccer analysis [10,11], although no studies have been identified to date that apply this model to F5 players.

Advances in science have made it possible to develop models for the predictive quantification of specific game actions in team sports, particularly the xG model [12,13]. This xG model assigns an estimated goal probability to each analyzed offensive action based on characteristics such as angle, distance, specific game situation, and the player's body part, thereby analyzing the evaluation of the offensive structure according to technical execution [9,14]. The analyzed information is used to employ advanced statistical approaches through machine learning, aiming to optimize performance analysis and predict specific outcomes [15].

A review of the literature reveals that there are few studies on Paralympic sports and that this knowledge gap is particularly evident in a sport such as conventional soccer, especially among players with visual impairments (B1 category) [16]. This sport is played on a 40 × 20-m field and features four fully blind field players and a goalkeeper who may be partially or fully sighted (B2/B3) [16–18]. To the best of our knowledge, this is one of the first studies to consider the xG model in the Paralympic context, more specifically in the analysis of performance in F5. Therefore, the objective of this study was to use a trained xG model in F5 since the analysis of 164 offensive plays by the Argentine National Team, which won first place in the 2022 IBSA Copa América.

Conventional xG models cannot be transferred directly to blind football, and this is precisely the gap the present study addresses. First, the game is governed by different spatial and regulatory constraints: a 40×20 m pitch enclosed by rebound side boards (kickboards) that keep the ball in play, the absence of an offside rule, and a sound-emitting ball with verbal guidance from behind the goal mean that the geometry, location and decision context of shots differ fundamentally from sighted soccer. Second, the executors are fully blind (B1) and rely on auditory rather than visual orientation, so the relationship between distance, angle and scoring probability is not assumed to follow the calibrated weights learned from large professional datasets. Consequently, the marginal effects, feature hierarchy and base goal probability must be re-estimated within the sport rather than borrowed, and the proximity to the lateral kickboard becomes a candidate predictor with no equivalent in conventional xG. The main advantage of the proposed approach is therefore twofold: it provides the first reproducible probabilistic benchmark calibrated to F5, and it couples model comparison with explainable artificial intelligence (SHAP) so that each prediction is decomposed into transparent, sport-specific feature contributions that coaches and analysts can interpret directly [19,20]. Explainable artificial intelligence is increasingly used to make predictive sports models auditable rather than black boxes, both for xG in conventional football [21] and for broader performance and decision-support frameworks [22]; comparing multiple learners under a common interpretability layer is also a recommended practice to avoid over-reliance on a single architecture [23].

2. Materials and Methods

2.1. Study Design

A retrospective study [24] that analyzed specific preexisting game situations to develop an xG model [25] was applied to F5.

2.2. Dataset

We analyzed 164 offensive plays by the Argentine national team during the IBSA Blind Football 5 Copa América, which took place from 23 October to 30 October 2022. The data are available on the Kaggle repository (<https://www.kaggle.com/datasets/agustingermanrojas/blind-football> (accessed on 5 May 2026)). The tournament, held in Córdoba, Argentina, consisted of a group stage featuring six national teams: Argentina, Brazil, Chile, Colombia, Peru, and Mexico. The first phase consisted of a round-robin tournament in which the national teams faced each other in a single-round league format, with each team playing five matches, followed by a final phase in which the two teams with the most points competed for the gold medal. The data were obtained from the six matches played by the Argentine national team. The Copa América tournament rules were applied in accordance with the IBSA [16].

2.3. Variables and Characteristics

To analyze the variables, a vector comprising 14 predictive characteristics was designed in 4 domains. First, spatially continuous variables were analyzed based on the X and Y coordinates of the final action, distance to the goal before taking the shot, length of the progression vector, and absolute magnitude of the angle of the shot relative to the center of the goal. Second, temporal continuous variables were analyzed in absolute match minutes and minutes per half. A third domain involved categorical variables through analysis of the origin of the play (open play, high recovery, counterattack, set piece, penalty), type of combination (individual or collective), kicking leg (right, left, or body), and the opponent (Brazil, Chile, Colombia, Mexico, Peru). Finally, a fourth domain of binary variables is derived from proximity to the sideline kickboard ($Y < 3$ m or $Y > 17$ m), the area near

the opponent's goal ($X < 6$ m), and the second half. Each of the continuous variables was imputed using the median and a standardization process (mean = 0; standard deviation = 1). Categorical variables were encoded via one-hot encoding, with the first category omitted. Binary variables were retained without any transformation. The 14 predictors were defined a priori from the F5 performance-analysis literature and the spatial logic of the sport, rather than through data-driven feature selection; given the small sample and low event count, no stepwise or wrapper selection was applied, as such procedures would risk overfitting and unstable variable rankings. Before modelling, multicollinearity among the continuous predictors was screened using the pairwise Spearman correlation matrix and the variance inflation factor (VIF). This screen confirmed strong collinearity among the geometric predictors derived from the shot coordinates: because distance to goal and shot angle are deterministic functions of the X and Y coordinates, the X-coordinate's distance to goal and angle showed VIF values well above the conventional threshold of 10 (in our data, $VIF \approx 130$, ≈ 110 and ≈ 14 , respectively), whereas the remaining predictors (Y, progression and match time) showed VIF below 4; the two match-time variables (absolute and per-half minute) are likewise collinear by construction. This redundancy was anticipated and is the main reason the final model is a gradient-boosted decision-tree ensemble (CatBoost), which is robust to multicollinearity because correlated inputs do not destabilize recursive split selection in the way they inflate coefficients in linear models. As a consequence, the SHAP contributions of the correlated geometric features (distance, X-coordinate and angle) are interpreted jointly as a shared spatial signal, rather than as fully independent marginal effects.

2.4. Data Split and Protection Against Overfitting

The data split was performed via a 70/30 stratified approach for the 164 actions analyzed. These were divided into a first phase ($n = 115$) and a final phase ($n = 49$), ensuring the overall proportion of the positive class (goal = 17.1%). Moreover, the test set was isolated from all phases of the current model and hyperparameter tuning and was later incorporated into the final evaluation. In the first-phase dataset, it was necessary to perform a grid search via fivefold stratified cross-validation (5-fold inner CV) on the AUC-ROC metric to select the hyperparameters. Similarly, a second fivefold stratified cross-validation process was performed on the final selected model to ensure the stability of each algorithm. Random seeds were fixed (random-state = 42) to promote reproducibility. Class imbalances were addressed using `class_weight = balanced` or the equivalent setting in each of the models that supported it. A single stratified hold-out (70/30) was deliberately preferred over repeated whole-sample cross-validation. The partition was stratified by outcome so that the goal prevalence (17.1%) was preserved in both subsets ($n = 115$ for training and $n = 49$ for the held-out test set), and the random seed was fixed (random_state = 42) for exact reproducibility. The rationale is statistical rather than temporal: with only 28 positive events, repeated k-fold cross-validation over the entire dataset would reuse the same scarce goals across folds for both hyperparameter tuning and performance reporting, producing optimistically biased and unstable estimates, whereas keeping the test partition completely untouched during imputation, scaling, encoding and model selection guarantees a leakage-free estimate of generalization. Internal model selection and hyperparameter tuning were nonetheless performed with 5-fold stratified cross-validation restricted to the training partition, and a second 5-fold stratified cross-validation on the training partition was used only to quantify the stability (mean $\pm$ SD of AUC-ROC) of each algorithm. We acknowledge that a single hold-out yields a higher-variance point estimate than repeated cross-validation, and this trade-off is stated explicitly as a limitation.

2.5. Compared with Algorithms

Ten representative algorithms from supervised machine learning were compared: (1) logistic regression with L2 regularization; (2) random forest; (3) extra trees; (4) classic gradient boosting; (5) XGBoost; (6) LightGBM; (7) CatBoost; (8) support vector machine with an RBF kernel; (9) k-nearest neighbors; (10) multilayer perceptron. Each of the selected algorithms was integrated into a scikit-learn pipeline to enable the addition of missing data through imputation, as well as the scaling and encoding of categorical variables [26]. This was done to prevent contamination from the cross-validation folds. The hyperparameter grids and the most stable selected values are reported in Table 1 below. The best hyperparameter configurations were selected via 5-fold internal cross-validation on the training set, with the goal of optimizing the AUC-ROC to ensure replicability. These configurations, along with `random_state = 42`, allow for the complete replication of all presented results. The hyperparameters were selected through an exhaustive grid search (scikit-learn `GridSearchCV`) nested inside the training partition: for every candidate combination, the mean AUC-ROC across the five inner stratified folds was computed, and the combination maximizing this inner-CV AUC-ROC was retained. The objective metric was AUC-ROC because it is threshold-independent and appropriate under class imbalance, and the same scoring rule was applied to all ten algorithms to keep the comparison fair. The search spaces were intentionally compact to limit the risk of optimization overfitting in a small sample and covered, for the tree ensembles, the number of estimators/iterations, maximum depth (or number of leaves), and learning rate; for logistic regression and SVM, the regularization strength (and kernel coefficient); for KNN, the number of neighbours and weighting scheme; for the MLP, the hidden-layer architecture, the L2 penalty (alpha), and the initial learning rate. The complete per-model grids, the full `GridSearchCV` object, and the resulting cross-validation tables are provided in the openly available reproducibility repository together with the random seed (`random_state = 42`), so the entire selection procedure can be re-executed and audited (Table 1 reports the winning configurations).

Table 1. The best hyperparameters for each algorithm for replication based on the chosen model.

Model	Best Hyperparameters (Selected by 5-Fold Inner CV on AUC)
Logistic Regression	<code>C = 10.0</code> ; <code>penalty = l2</code> ; <code>solver = lbfgs</code>
Random Forest	<code>max_depth = 5</code> ; <code>min_samples_split = 2</code> ; <code>n_estimators = 300</code>
Extra Trees	<code>max_depth = 5</code> ; <code>n_estimators = 300</code>
Gradient Boosting	<code>learning_rate = 0.1</code> ; <code>max_depth = 2</code> ; <code>n_estimators = 200</code>
XGBoost	<code>learning_rate = 0.05</code> ; <code>max_depth = 3</code> ; <code>n_estimators = 200</code>
LightGBM	<code>learning_rate = 0.1</code> ; <code>n_estimators = 200</code> ; <code>num_leaves = 15</code>
CatBoost	<code>depth = 3</code> ; <code>iterations = 100</code> ; <code>learning_rate = 0.1</code>
SVM (RBF)	<code>C = 0.5</code> ; <code>gamma = scale</code>
KNN	<code>n_neighbors = 7</code> ; <code>weights = uniform</code>
MLP	<code>alpha = 0.01</code> ; <code>hidden_layer_sizes = (64, 32)</code> ; <code>learning_rate_init = 0.01</code>

Note: XGBoost: eXtreme Gradient Boosting; LightGBM: Light Gradient Boosting Machine; CatBoost: Category Boosting; SVM (RBF): Support Vector Machine—Radial Basis Function; KNN: K-Nearest Neighbors; MLP: Multilayer Perceptron.

2.6. Evaluation Metrics

The evaluation metrics were calculated via the following measures: accuracy, balanced accuracy, precision, recall, F1 score, area under the ROC curve (AUC-ROC), area under the precision–sensitivity curve (AUC-PR), the Matthews correlation coefficient (MCC) [27], the Brier score, and log-loss. Furthermore, the combined use of F1, MCC, and the Brier score enables a robust evaluation to analyze the low-class imbalance and the model’s probabilistic calibration [28]. These metrics are defined as follows: Accuracy is the proportion of

correctly classified actions over all actions, $(TP + TN)/(TP + TN + FP + FN)$; because it is misleading under class imbalance, balanced accuracy, the mean of sensitivity and specificity, is also reported. Precision (positive predictive value) is $TP/(TP + FP)$, the proportion of predicted goals that were true goals; recall (sensitivity) is $TP/(TP + FN)$, the proportion of true goals correctly identified; the F1 score is their harmonic mean, $2 \times (\text{precision} \times \text{recall})/(\text{precision} + \text{recall})$. AUC-ROC is the area under the receiver operating characteristic curve (sensitivity versus $1 - \text{specificity}$) and quantifies threshold-independent discrimination, with 0.5 indicating chance and 1.0 perfect separation. AUC-PR is the area under the precision–recall curve and is more informative than AUC-ROC under low prevalence, with the no-skill baseline equal to the positive-class prevalence (here 0.171). The MCC is a correlation coefficient between observed and predicted classes, $(TP \times TN - FP \times FN)/\sqrt{[(TP + FP)(TP + FN)(TN + FP)(TN + FN)]}$, ranging from -1 to $+1$ and robust to imbalance. The Brier score is the mean squared difference between predicted probabilities and outcomes, measuring probabilistic calibration (lower is better), and log-loss (cross-entropy) penalizes confident wrong predictions, which also involves lower is better. To assess whether the differences in discrimination among the competing algorithms were statistically meaningful, the per-model 5-fold cross-validated AUC-ROC distributions were compared, and pairwise differences in test-set ROC-AUC were tested with the DeLong test for two correlated ROC curves; the corresponding statistics are reported in Section 3.1 and the reproducibility repository.

2.7. SHAP Analysis

To enhance the interpretability of the best identified model (CatBoost), it was necessary to use the Shapley Additive exPlanations (SHAP) library, which breaks down each prediction individually through additive contributions attributed to each input variable [19,20]. SHAP TreeExplainer values (an exact algorithm for tree-based models) were calculated across the 50 instances in the dataset. Additionally, the following 7-step visualizations were generated: (1) global importance bar chart; (2) Beeswarm plot of impact by instance; (3) SHAP heatmap by instance; (4) Scatter/dependency plots for the four most important predictors; (5) partial dependency plots; (6) Waterfall plots for a representative positive prediction and a representative negative prediction showing the trajectory from the base value $E[f(X)]$ to the individual output $f(x)$; (7) decision plots showing the cumulative trajectories of actions that resulted in a goal and those that did not.

2.8. Implementation

Each analysis was performed via the Python 3.12 programming language (NumPy 1.26, pandas 2.2, scikit-learn 1.5, XGBoost, LightGBM, CatBoost, SHAP 0.46) [29]. The global random seed number was set to 42. The final code for the training pipeline, the socialized trained models, the calculated SHAP values and the scripts for generating figures and tables.

2.9. Ethical Considerations

All the data used in this study were anonymized and sourced from official IBFA reports and public records. This research adhered to the principles established by the Declaration of Helsinki [30] regarding privacy and data handling.

3. Results

The 164 offensive actions analyzed had a goal rate of 17.1% (28/164) distributed across the six matches played by eight players on the Argentine national team. Figure 1 shows the distributions for the most important continuous and categorical variables, which, in turn, are stratified by outcome.



Figure 1. Dataset characterization for the xG-B1 model. (a) Class distribution, (b–d) distributions of distance to the goal, progression magnitude and shot angle by outcome, (e) outcome by executing the foot and (f) goal rate by attack origin.

Table 2 presents the various descriptive statistics analyzed. Three key descriptive patterns that serve as a reference for developing the predictive model are summarized. First, plays that resulted in a goal occurred at a shorter-than-average distance (10 ± 3.1 m versus 12.2 ± 4.2 m). Second, the left foot demonstrated greater effectiveness, accounting for approximately 30.0% of the goal rate, than the right foot (13.0%). Finally, none of the 24 plays recorded in the final against Brazil proved decisive in resulting in a goal, a pattern that highlights one of the most representative features of predictive models. In contrast, the goal rate for Peru reached 24.4%.

Table 2. Descriptive statistics are presented as the means and standard deviations for each type of variable analyzed.

Variable	Type	No-Goal Mean $\pm$ SD	Goal Mean $\pm$ SD	Overall Mean $\pm$ SD
Final X-coordinate (m)	Continuous	10.32 $\pm$ 4.33	8.57 $\pm$ 2.93	10.18 $\pm$ 4.26
Final Y-coordinate (m)	Continuous	5.39 $\pm$ 2.88	5.52 $\pm$ 1.98	5.40 $\pm$ 2.81
Distance to goal (m)	Continuous	11.73 $\pm$ 4.15	10.02 $\pm$ 2.25	11.59 $\pm$ 4.05
Progression vector (m)	Continuous	14.19 $\pm$ 7.18	15.19 $\pm$ 6.96	14.27 $\pm$ 7.14
Shot angle ($^{\circ}$)	Continuous	26.52 $\pm$ 15.13	29.49 $\pm$ 16.43	26.76 $\pm$ 15.20
Match minute (absolute)	Continuous	23.41 $\pm$ 18.64	20.79 $\pm$ 18.85	22.96 $\pm$ 18.65
Match minute (per-half)	Continuous	13.85 $\pm$ 9.31	13.64 $\pm$ 9.59	13.82 $\pm$ 9.33
Origin: Open play	Categorical	57 (41.9%)	12 (42.9%)	69 (42.1%)
Origin: Set piece	Categorical	43 (31.6%)	6 (21.4%)	49 (29.9%)
Origin: Counterattack	Categorical	17 (12.5%)	4 (14.3%)	21 (12.8%)
Origin: High recovery	Categorical	19 (14.0%)	4 (14.3%)	23 (14.0%)
Origin: Penalty	Categorical	0 (0.0%)	2 (7.1%)	2 (1.2%)
Combination: Individual	Categorical	88 (64.7%)	24 (85.7%)	112 (68.3%)
Combination: team	Categorical	48 (35.3%)	4 (14.3%)	52 (31.7%)

Table 2. Cont.

Variable	Type	No-Goal Mean $\pm$ SD	Goal Mean $\pm$ SD	Overall Mean $\pm$ SD
Leg: Right	Categorical	107 (78.7%)	16 (57.1%)	123 (75.0%)
Leg: Left	Categorical	28 (20.6%)	12 (42.9%)	40 (24.4%)
Leg: Body	Categorical	1 (0.7%)	0 (0.0%)	1 (0.6%)
Rival: PER	Categorical	31 (22.8%)	10 (35.7%)	41 (25.0%)
Rival: COL	Categorical	28 (20.6%)	8 (28.6%)	36 (22.0%)
Rival: CHL	Categorical	26 (19.1%)	4 (14.3%)	30 (18.3%)
Rival: MEX	Categorical	27 (19.9%)	6 (21.4%)	33 (20.1%)
Rival: BRA	Categorical	24 (17.6%)	0 (0.0%)	24 (14.6%)
Near kickboard ($Y < 3$ m or > 17 m)	Binary	34 (25.0%)	0 (0.0%)	34 (20.7%)
Close zone ($X < 6$ m)	Binary	31 (22.8%)	4 (14.3%)	35 (21.3%)
Second half	Binary	65 (47.8%)	10 (35.7%)	75 (45.7%)

Note: m: meters; (°): angle; SD: standard deviation. Descriptive statistics for the 14 input characteristics, stratified by class. Continuous variables are presented as the means $\pm$ SDs; categorical and binary variables are presented as counts and percentages. $n = 164$ actions in total ($n = 136$ without a target; $n = 28$ with a target).

3.1. Comparison of the Ten Algorithms

Figure 2 shows the ROC curves and precision–sensitivity metrics for the ten algorithms analyzed. The tree-based boosting models (CatBoost, gradient boosting, XGBoost, LightGBM, and random forest) achieved the highest scores, with AUC-ROC (0.840–0.927) and AUC-PR (0.796–0.829) values. Moreover, logistic regression, extra trees, SVM, and MLP achieved moderate performance, with an AUC-ROC of 0.875–0.913, whereas the K-nearest neighbors algorithm achieved a high AUC-ROC of 0.915 and a low AUC-PR of 0.603, indicating low sensitivity for analyzing minority actions. Figure 2 shows the discrimination performance of each of the 10 candidate xG-B1 models. (a) ROC curves. (b) Precision–recall curves, where the dashed line in panel (b) marks the prevalence-based reference line.

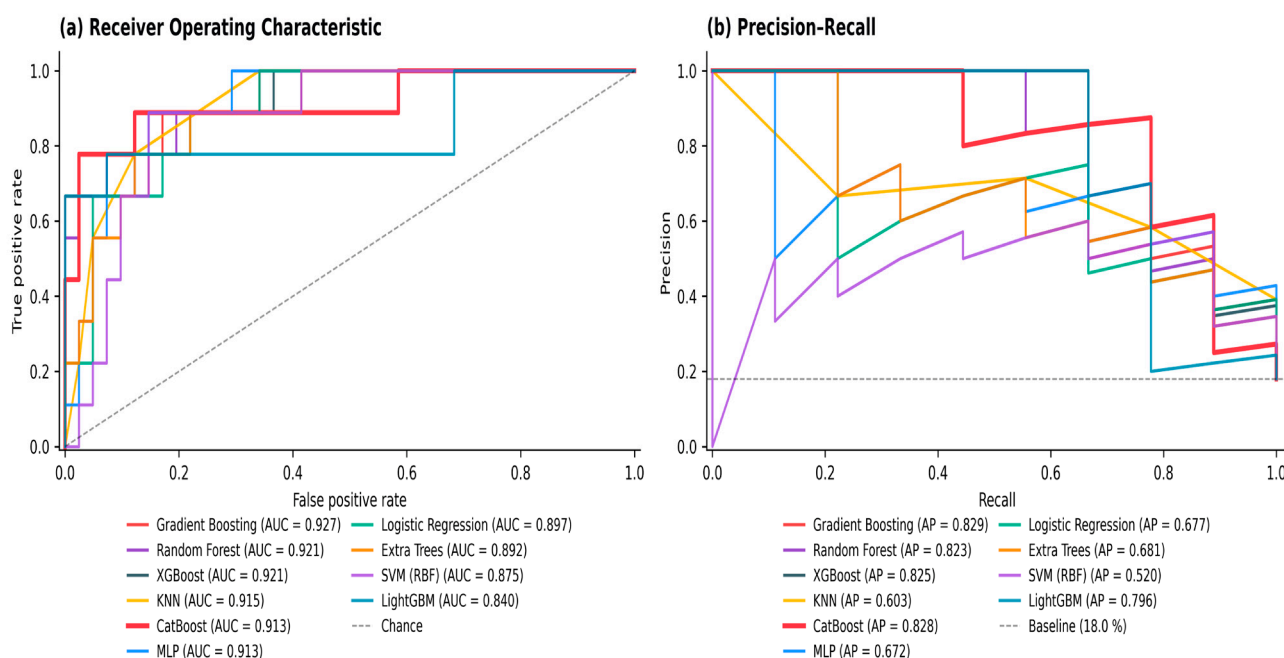


Figure 2. Discrimination performance of the 10 candidate xG models.

Figure 3a illustrates the stability of each analyzed algorithm following a fivefold cross-validation process, and Figure 3b provides a detailed view of the multidimensionality of each of the metrics obtained. Four of the analyzed models achieved high cross-validation and AUC-ROC values exceeding 0.86 (KNN, CatBoost, random forest, and gradient boost-

ing). Furthermore, the chosen model was CatBoost for achieving high levels ($F1 > 0.778$), accuracy under class imbalance, high MCC (0.729), a low Brier score (0.072), consistency in the cross-validation process ($AUC = 0.871 \pm 0.139$), and performance across the entire test set ($AUC = 0.913$) (Table 3). Importantly, the differences in discrimination among the leading models were not statistically distinguishable at this sample size. The 5-fold cross-validated AUC-ROC confidence intervals overlapped substantially across the top performers (CatBoost 0.871 ± 0.139 , KNN 0.880 ± 0.075 , Random Forest 0.869 ± 0.139 , Gradient Boosting 0.863 ± 0.172), and the wide standard deviations reflect the scarcity of positive events per fold. Consistently, the DeLong test for paired test-set ROC AUCs found no significant pairwise difference between CatBoost and the other tree-based ensembles (all $p > 0.05$). CatBoost was therefore selected, not as a statistically superior model but as the most balanced and best-calibrated option (highest F1 and MCC with the lowest Brier score) among a set of statistically comparable candidates, a nuance that should temper any ranking interpretation.

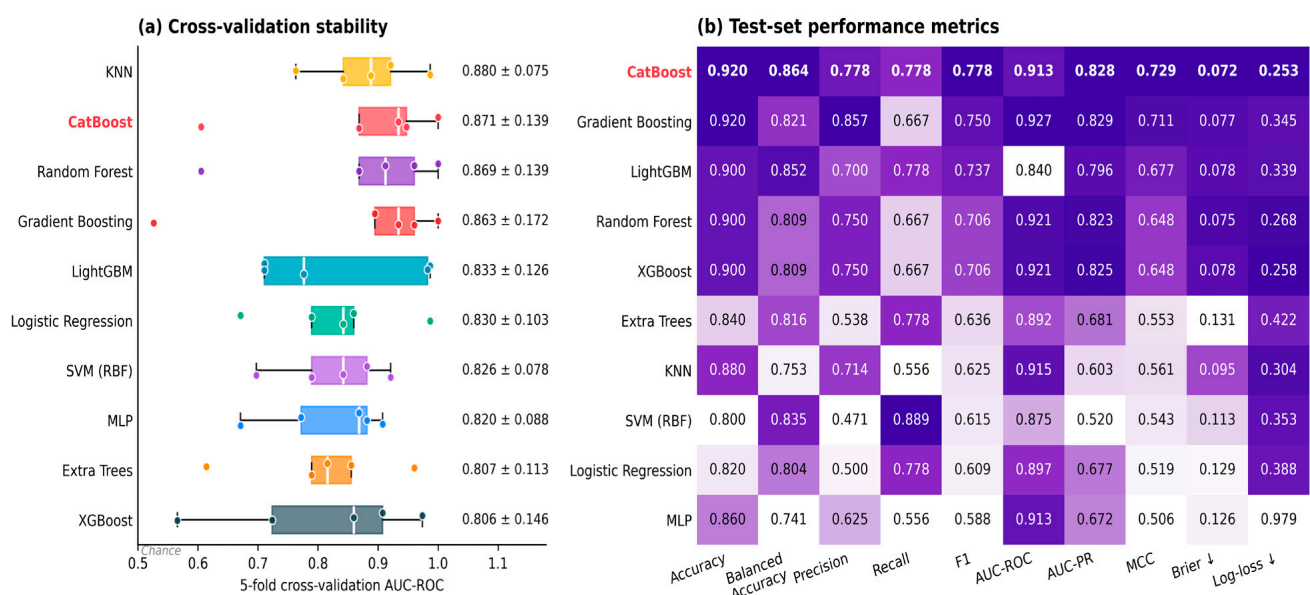


Figure 3. Stability and performance of the 10 candidate xG-B1 models. (a) 5-fold CV AUC distribution per model. (b) Heatmap of test-set performance metrics ordered by F1 (best model on top). Note: In the heat map (panel (b)), the color intensity represents each model's performance for each metric. Darker shades of purple indicate better performance when the metric should be maximized (for example, Accuracy, Balanced Accuracy, Precision, Recall, F1, AUC-ROC, AUC-PR, and MCC). In contrast, lighter shades represent lower values for these metrics. For error metrics (Brier Score and Log-loss), the interpretation is reversed: lighter colors correspond to lower errors and, therefore, better performance, while darker shades indicate higher errors.

Table 3. Test-set performance metrics for all 10 models.

Model	Accuracy	Bal. Acc.	Precision	Recall	F1	AUC-ROC	AUC-PR	MCC	Brier ↓	Log-Loss ↓	CV-AUC (5-Fold)
CatBoost	0.920	0.864	0.778	0.778	0.778	0.913	0.828	0.729	0.072	0.253	0.871 ± 0.139
Gradient Boosting	0.920	0.821	0.857	0.667	0.750	0.927	0.829	0.711	0.077	0.345	0.863 ± 0.172
LightGBM	0.900	0.852	0.700	0.778	0.737	0.840	0.796	0.677	0.078	0.339	0.833 ± 0.126
Random Forest	0.900	0.809	0.750	0.667	0.706	0.921	0.823	0.648	0.075	0.268	0.869 ± 0.139
XGBoost	0.900	0.809	0.750	0.667	0.706	0.921	0.825	0.648	0.078	0.258	0.806 ± 0.146
Extra Trees	0.840	0.816	0.538	0.778	0.636	0.892	0.681	0.553	0.131	0.422	0.807 ± 0.113
KNN	0.880	0.753	0.714	0.556	0.625	0.915	0.603	0.561	0.095	0.304	0.880 ± 0.075
SVM (RBF)	0.800	0.835	0.471	0.889	0.615	0.875	0.520	0.543	0.113	0.353	0.826 ± 0.078
Logistic Regression	0.820	0.804	0.500	0.778	0.609	0.897	0.677	0.519	0.129	0.388	0.830 ± 0.103
MLP	0.860	0.741	0.625	0.556	0.588	0.913	0.672	0.506	0.126	0.979	0.820 ± 0.088

Metrics for which a lower value is better are indicated with ↓. The cross-validation AUC (CV-AUC) was calculated via cross-validation stratified into 5 groups on the training set with the best hyperparameter configuration (Table 3).

3.2. SHAP Interpretability Based on the Model's Overall Importance

Figure 4 shows the hierarchy of each variable based on the SHAP values computed in the test model via CatBoost. Thus, the five main predictors are identified on the basis of the mean absolute value of the SHAP score: distance to the goal at the moment of the shot (mean SHAP = 1.00), lateral coordinate on the Y-axis (0.78), total magnitude of the shooting angle (0.51), magnitude of the forward momentum vector (0.48), and proximity to the side kickboard (0.43). On this basis, the lateral Y coordinate becomes the second most important predictor in the model (Figure 4a). Moreover, the Beeswarm plot (Figure 4b) reveals the pattern by which each selected variable is directed. For distance to the goal, low values (blue points) tend to generate positive SHAP values (higher probability of a goal), whereas high values (yellow points) generate negative SHAP values. The Y coordinate exhibits a bidirectional pattern where extreme values (low Y on the right or high Y on the left) tend to push the prediction in opposite directions.

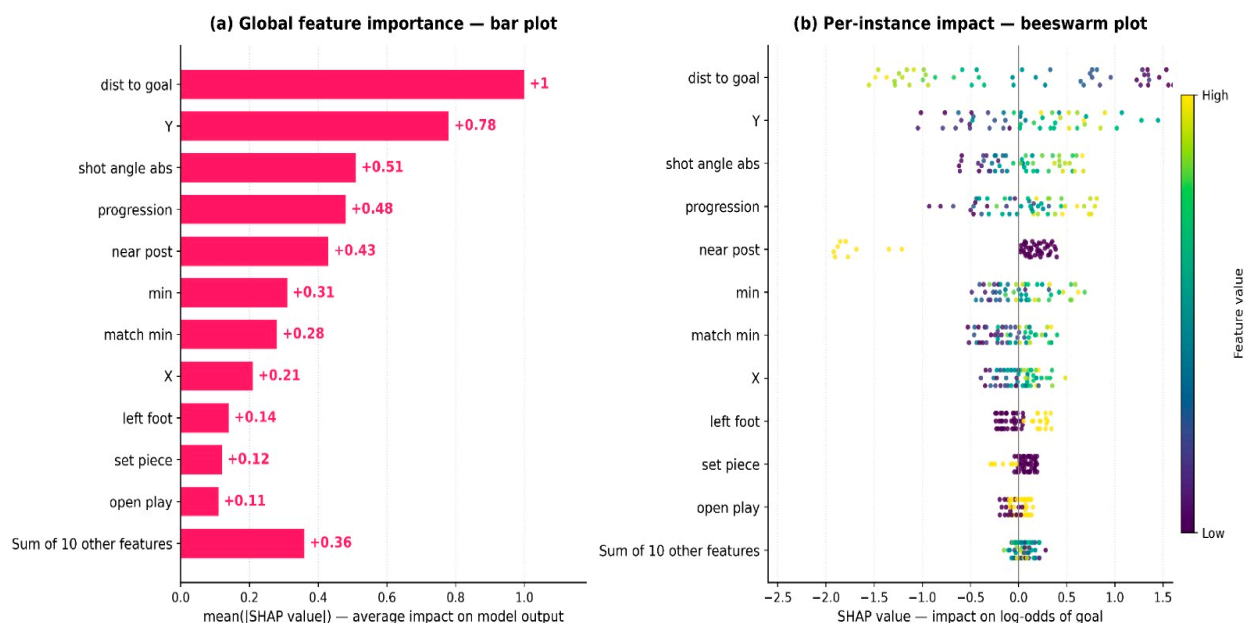


Figure 4. Global SHAP interpretability of the CatBoost xG-B1 model. (a) Bar plot of mean absolute SHAP values per feature and (b) Beeswarm plot showing the distribution of SHAP values per instance, with feature values encoded by color.

3.3. SHAP Heatmap

Figure 5 shows how the top curve represents the model's output $f(x)$ on a log-odds scale for each instance, allowing us to visually identify the actions with the highest probability. This figure reveals three subgroup patterns: (i) demands dominated by positive contributions from distance and Y, following the instances on the far right with higher $f(x)$ values; (ii) proximity to the kickboard, which plays a distinguishing role; (iii) demands with low $f(x)$ values concentrated on the far left, converging with negative contributions.

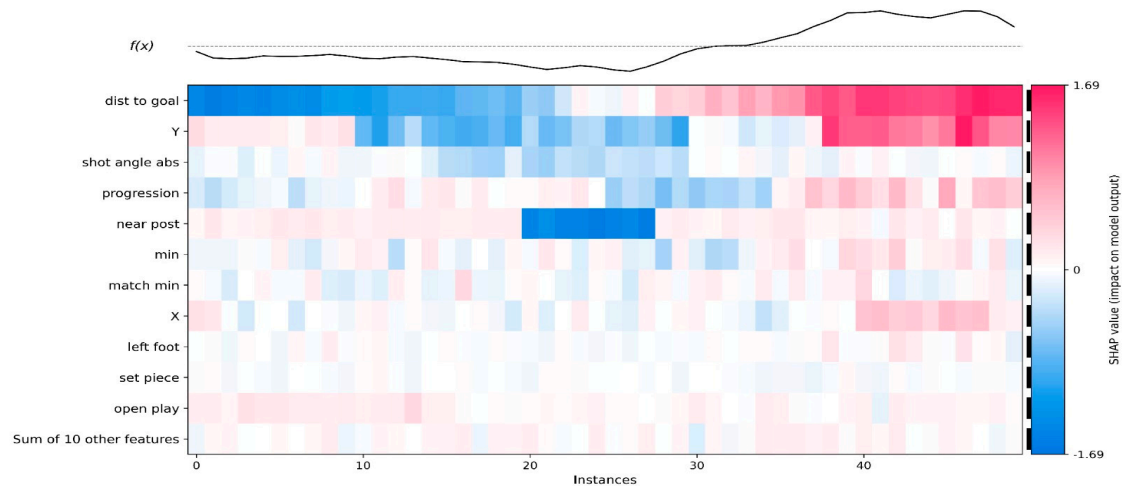


Figure 5. SHAP heatmap of test-set instances. Each column is a single test instance ordered by hierarchical similarity. The black curve at the top traces $f(x)$, the model output (log-odds of goal). The cell color encodes the SHAP value of each feature for that instance.

3.4. Functional Dependence of the Main Predictors

Figure 6 shows the four most important predictive variables in the model. The scatter plots (dependence) indicate that the distance from the goal has a nonlinear relationship with its SHAP value, particularly because shots taken from very long or very short distances yield negative SHAP values, whereas shots from medium distances (~12 m) contribute positively to the model. Similarly, the Y-coordinate has the greatest positive effect when it takes on values close to the sideline of the field, extending toward both flanks. The absolute magnitude of the shooting angle also tends to have a concentrated positive effect when moderate angles ($0\text{--}30^\circ$) occur. Finally, the magnitude of the run-up shows a positive trend that decreases as the run-up length increases.

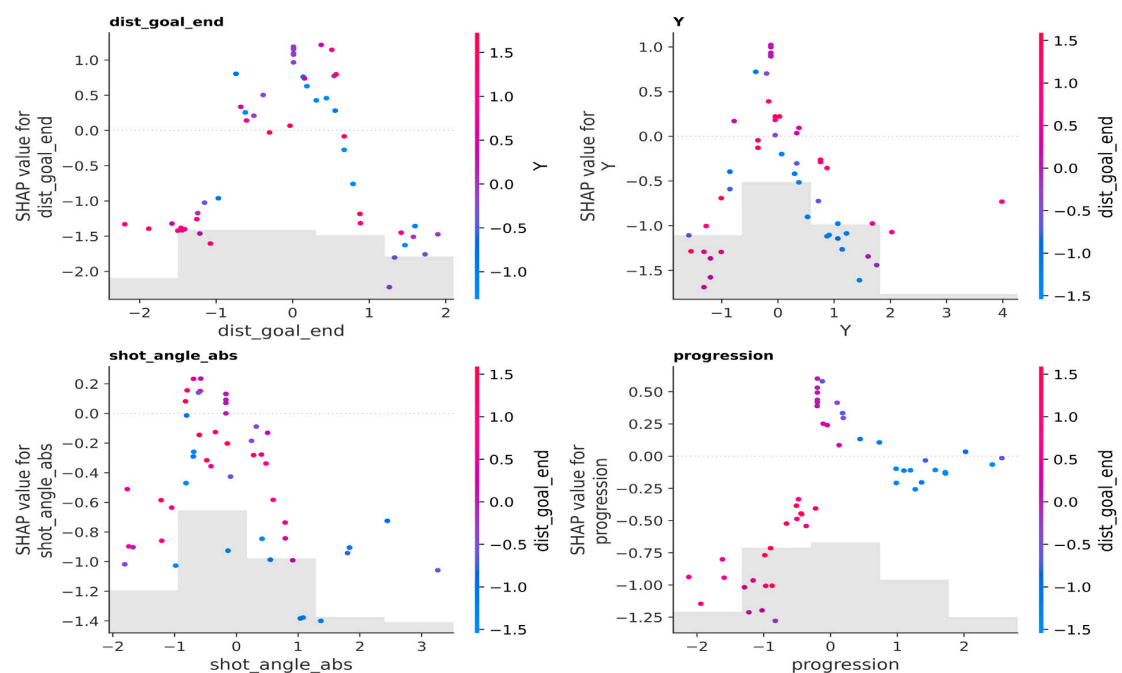


Figure 6. SHAP scatter/dependence plots for the top 4 predictors. Each panel plots the SHAP value against the feature value; color encodes the most strongly interacting secondary feature, revealing two-way effects on the predicted log-odds of the goal.

The expected probability of a goal peaks at approximately 0.37 for distances close to the average, whereas it decreases to approximately 0.08 for very long distances. It also exhibited a similar pattern of decline for runs longer than the average (Figure 7).

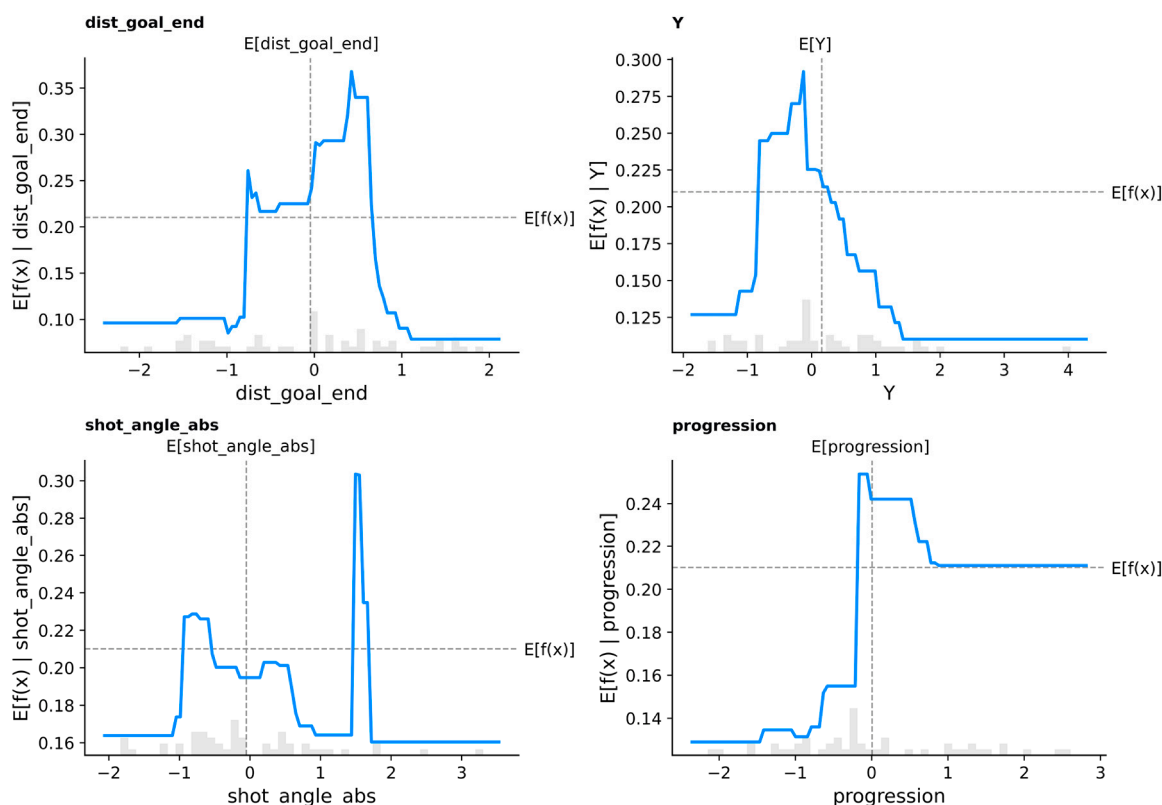


Figure 7. SHAP partial dependence plots for the top 4 predictors. Each panel shows the marginal effect of one feature on the predicted probability of the goal, integrating all other features. The vertical and horizontal dashed lines mark the mean feature value and the model expected value, respectively. Note: This figure shows Accumulated Local Effects (ALE) plots, which illustrate how each variable influences the model's prediction while holding the effects of the other variables constant. The blue line represents the cumulative effect of each variable on the model's prediction. Rises in the curve indicate that certain values of the variable increase the probability of the event of interest, while declines reflect a decrease in that probability. The dotted horizontal line ($E[f(x)]$) corresponds to the mean value of the model's prediction, and the gray bars at the bottom show the distribution of the data used to estimate the effect.

3.5. Individual Decomposition: Waterfall and Decision Plots

Figure 8 shows how the model constructs two representative predictions with the base value $E[f(X)] = -0.226$. For prediction (a), an action correctly classified as a goal (probability of 0.982) has positive contributions from Y, distance to the goal, progression, and X, with a total $f(x)$ value of (+4.002 on the log-odds scale). For prediction (b), an action correctly classified as a nongoal (probability of 0.002) has negative contributions, resulting in $f(x)$ of -6.063 .

Figure 9 shows the decision plots, which extend to different instances simultaneously, illustrating the various trajectories that accumulate from the base value to the model's final output. Contribution (a) shows the nine main actions that resulted in a goal, in which their trajectories converge toward high values and heterogeneous paths for prediction. Contribution (b) shows 15 actions that did not result in a goal, where their trajectories primarily exhibit negative values and a more similar pattern.

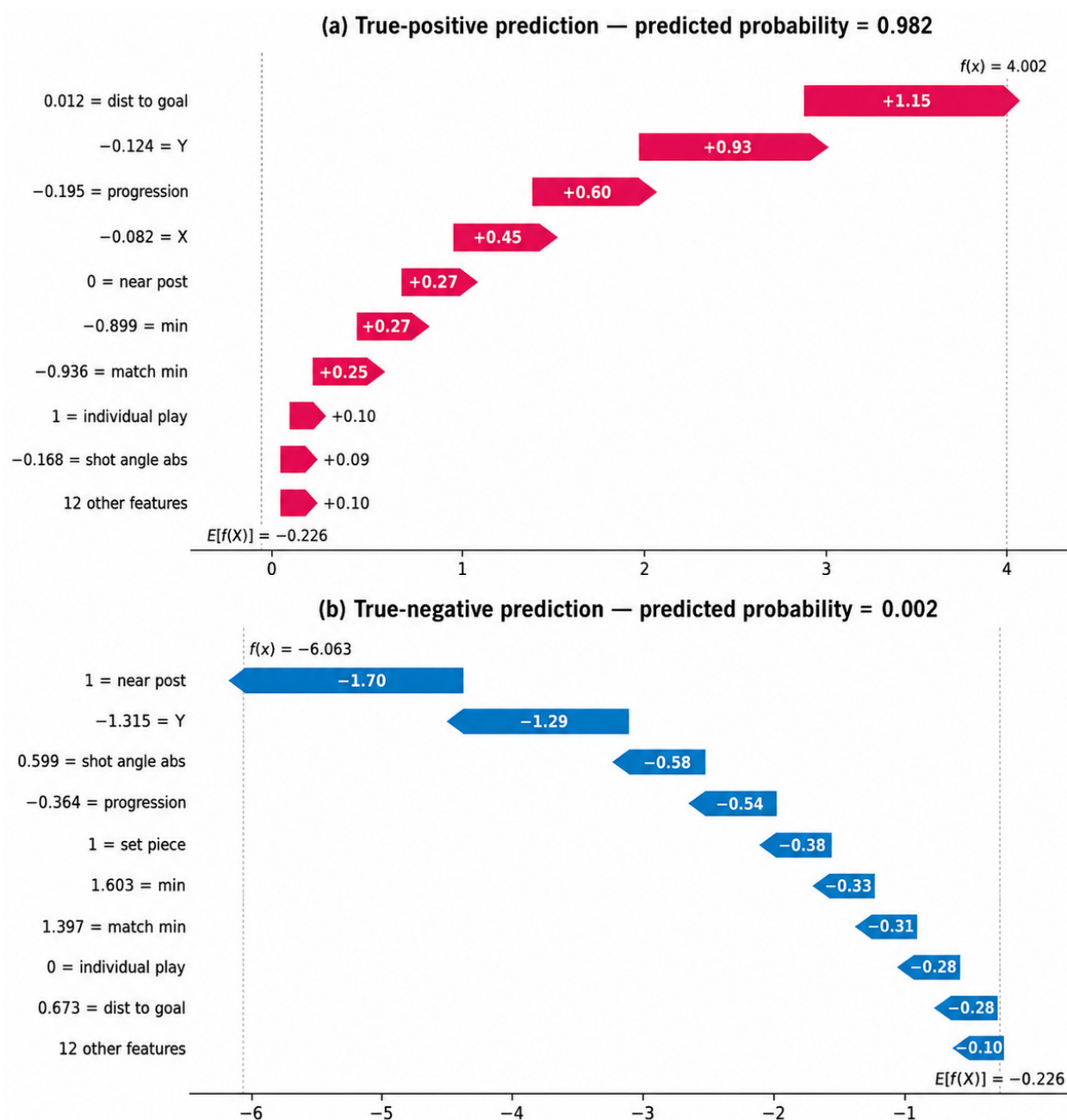


Figure 8. SHAP waterfall plots for two representative test instances. **(a)** True-positive prediction with a predicted probability of 0.982. **(b)** True-negative prediction with a predicted probability of 0.002. Each plot decomposes the prediction from the model expected value $E[f(X)]$ to the individual output $f(x)$.

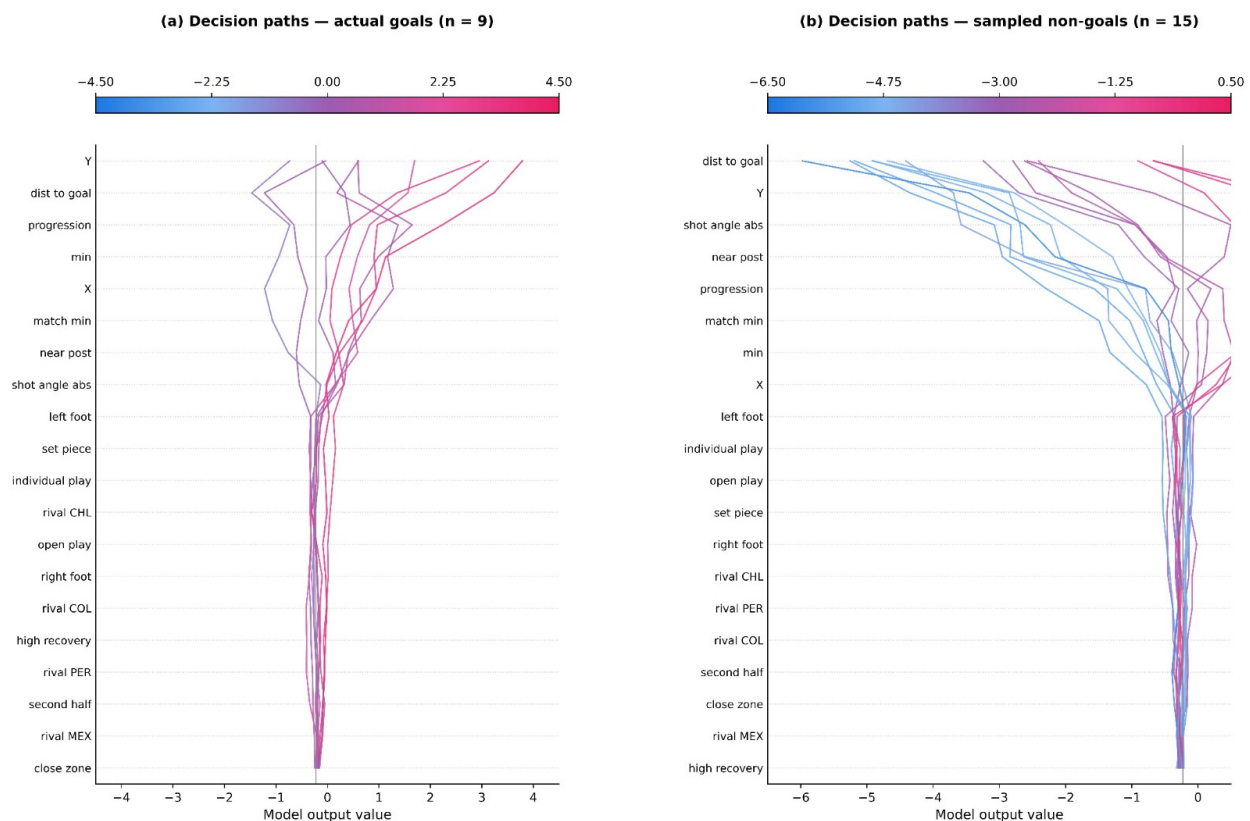


Figure 9. SHAP decision plots showing cumulative feature contributions for (a) the $n = 9$ test-set actions that ended in goal and (b) a random sample of $n = 15$ nongoal actions. Each line traces one prediction path from the base value (bottom) to its final model output (top).

4. Discussion

This study presents an xG model that was specifically calibrated for F5 players and a national team that competed in 2022 IBSA Copa América. Four major findings are identified. First, three-based models, specifically CatBoost, appeared to demonstrate greater discrimination and calibration ($F1 = 0.778$; $AUC-ROC = 0.913$; $Brier = 0.072$). Second, the structural goal predictors defined by SHAP indicate that the distance from the goal, angle, and lateral position are key variables. The third finding is that the magnitude of progression has a nonlinear effect when short and medium distances ranging from 5 to 15 m are analyzed. The fourth finding indicates that the opponent plays an important role in the predictive signal of the model to be used.

Several studies have used the xG model to evaluate performance and analyze offensive plays. For example, a study by Bandara et al. [12] used a random forest model and concluded that opportunities arising from the flanks of the penalty area, as well as shots taken in front of the opponent's 18-yard box and shots following successful passes to the far post, were the primary predictors of goal probability. Similarly, an analysis of shots in the Bundesliga reported that an extreme gradient boosting algorithm using data synchronization yields a classified probability ($RPS = 0.197$) higher than any other model previously analyzed, where SHAP values reveal that the most important variables were distance to goal, angle, and goalkeeper position [9]. These results, which employ the xG model in professional soccer, align with the findings of the present study involving F5 players, where SHAP values demonstrate that distance from the goal, angle, and lateral position are determining variables, with the latter being particularly important owing to the specific demands of the sport, which features a protective barrier to safeguard athletes and promote greater continuity of play.

The results of this study fall within the upper range of values reported by other xG models in soccer. For this study, values of $F1 = 0.778$ and $AUC-ROC = 0.913$ are reported via fivefold cross-validation, findings similar to those reported in analyses of matches from international leagues and elite European clubs, where fivefold cross-validation yielded AUC-ROC values of 0.883, whereas separate analysis of the test data showed AUC-ROC values of 0.826 [12]. Other studies report high AUC-ROC values of 0.822 when analyzing data from professional Bundesliga players with many variables associated with passes, ball control prior to a shot (direct, two-touch, by dribbling > 10 m and < 10 m, or free kicks), shots with the foot and head, fouls, and saves [9], whereas another study reported AUC-ROC values of 0.800 when analyzing data related to the distance measured via the shot coordinates (x, y) to the center of the opposing goal, angle, body parts used (head, dominant and nondominant foot), and the specific game situation (open play, counterattack, free kicks, penalty) [31]. Thus, the xG model serves as a valuable tool for analyzing the collective and individual performance of players in both soccer and F5 in highly competitive settings [15]. This comparison must nonetheless be interpreted cautiously: the benchmark studies cited above are based on tens of thousands of shots, whereas the present AUC-ROC of 0.913 was obtained from a test set of only 49 actions containing 9 goals. With so few positive cases, a single hold-out AUC is a high-variance estimate and is prone to optimistic bias, so the fact that our value sits at the upper end of the reported range should be read as a property of a small, favourable sample rather than as evidence that the F5 model discriminates better than well-powered professional models. The overlapping cross-validation intervals reported in Section 3.1 reinforce this caution.

Another important consideration when analyzing the usefulness of xG models in sports is the difficulty of comparing these models with less explored fields, given that in F5, there are smaller sample sizes ($n = 164$), lower prevalence rates (17%), and a less represented Paralympic field compared with more robust studies in conventional soccer that analyzed 105,627 shots [9], manual notational analysis of 18,000 shots over a season in the Bundesliga and the Premier League [32], or the prediction of outcomes from shots by analyzing 10,000 shots [33].

Furthermore, the findings of this study provide an initial approach to compare xG models, particularly when tree-based models are used, indicating that the CatBoost model represents a first step in analyzing xG in F5, whereas for conventional soccer, it has been reported that the most commonly used models are logistic regression, gradient boosting, neural networks, support vector machines, and tree-based classification algorithms [34–36]. Therefore, a wide variety of models can be used in xG analysis; consequently, incorporating machine learning into F5 analysis—as a less-explored field—should be done with caution [26] and by considering larger samples as well as longer-term follow-ups. Our head-to-head evaluation of ten learners is consistent with recent benchmarking work in conventional football: comparative studies across the major European leagues likewise report that gradient-boosted and tree-based ensembles tend to lead the field while remaining close to logistic and neural baselines [37], and that explicitly modelling player- or position-level effects further refines shot-level probabilities [38]. Extending the present comparison to the newest architectures (e.g., attention-based tabular networks or Bayesian hierarchical xG [15]) was not pursued here because such models are data-hungry and would be unidentifiable on 164 actions; we therefore restricted the benchmark to algorithms that are stable in small, imbalanced samples and flag richer architectures as a priority once larger multi-tournament F5 datasets become available.

Finally, a key finding of this study relates to the effects exerted by the opponent; this is because the opponent's identity and collective organization could serve as inputs for the model's predictive signal to capture the contribution of the opponent variable, so

future studies in F5 should take this variable into account. These results are consistent with findings from other studies that reported that not only does the opponent variable condition the model, but other factors, such as context and type of competition, also vary when evaluating a team's offensive performance via xG ($r = 0.488$, $p < 0.001$) [13].

4.1. Practical Implications

The model used in this study initially has three specific applications. The first of these relates to information that could be relevant to coaches and the technical staff of F5 national teams, as it considers the estimated probability that can be used to quantify the quality of offensive actions to understand offensive performance based on specific metrics. A second application could be related to training planning processes, which seek to identify optimal SHAP zones and ranges to promote specific offensive actions because of specific training sessions and exercises. Finally, a third application could be associated with professionals involved in adapted sports, technical staff, and researchers seeking to understand the interaction between offensive actions and individualized physical and physiological variables to customize training loads, addressing the specific requirements of competition [39,40].

4.2. Limitations and Future Directions

Although this is an initial study exploring machine learning models to analyze offensive plays in F5, it is important to note that it has several limitations that future research could address. First, the sample size ($n = 164$) of plays analyzed represents a limitation associated with a modest model compared with other studies that employ these same techniques in professional soccer. Second, since the data were drawn from a single tournament and a single national team, the study is not generalizable, which means that conclusions derived from it should be interpreted with caution. Nevertheless, it serves as an initial contribution that could serve as a reference. Third, there is an inherent limitation in analyzing competition when only offensive variables are considered without accounting for physiological metrics such as distances, speeds, accelerations, decelerations, heart rate, and others, which could substantially help improve the predictive capacity of the model used. A fourth limitation relates to the effect of the opponent, which was captured by the model; this may lead to errors in the observed correlations, which do not necessarily reflect causal relationships applicable to other contexts. These limitations deserve stronger emphasis than a generic caveat. Because every analyzed action belongs to one national team in one tournament (164 offensive plays, only 28 goals), the model unavoidably encodes that team's style of play and its specific set of opponents; the learned feature weights and the prominence of the "opponent" signal may therefore reflect team-specific and tournament-specific bias rather than transferable structural determinants of goal probability in F5, and the model is not expected to generalize to the playing styles of other teams. The combination of a small sample, low event prevalence, and flexible learners (XGBoost, MLP, gradient-boosted trees) also entails a real risk of overfitting and of optimistically inflated discrimination; although a fully isolated final-phase test set, compact hyperparameter grids, class weighting and a fixed seed were used to mitigate this, they reduced but could not eliminate it. Accordingly, the high AUC-ROC reported here should be regarded as an upper-bound, sample-dependent estimate, and the present work should be read strictly as a proof-of-concept benchmark requiring external, multi-team, and multi-tournament validation before any tactical conclusion is transferred to practice.

Consequently, it is suggested that this type of analysis be expanded to include larger tournament formats with a greater number of teams; however, this would require making public data available. Similarly, it would be necessary to expand the use of these

types of models to other tournaments, categories, and F5 competitions, thereby seeking to estimate population and contextual uncertainty regarding the various xG estimates and thus strengthen predictors of performance in F5. Bayesian hierarchical models with random effects for each player, match, and tournament should be incorporated. This would allow for the determination of interindividual variability. Finally, it would be necessary to integrate longitudinal data that consider injury incidence, physiological markers, and physical and anthropometric variables [41], as well as reports of subjective training load, to construct predictive models adapted to the multidimensional approach of an invasion sport such as F5. The aim is to bridge the gap between academic research and practice to transfer knowledge that can be useful in different training contexts [42].

5. Conclusions

This is the first study to develop an xG model specifically for F5 players. By comparing 10 machine learning algorithms through a rigorous and systematic process using stratified splitting and cross-validation with multiple metrics, it was determined that CatBoost is the model with the best integration of discrimination and calibration ($F1 = 0.778$; $AUC-ROC = 0.913$; $Brier = 0.072$). For the analysis of the main variables, the SHAP analysis revealed that distance to the goal, lateral coordinate, shooting angle, magnitude of progression, and proximity to the lateral kickboard emerged as the five main structural predictors of goals in F5.

These data represent an initial contribution by supporting predictive frameworks that guide the analysis of Paralympic sports, demonstrating that nonlinear models with SHAP interpretability can serve as a useful tool at the theoretical level when developed using rigorous protocols against overfitting. These conclusions are deliberately framed as preliminary: derived from a single team and a single tournament with few goals, the reported performance is likely an optimistic, sample-dependent estimate, and the model is not assumed to generalize to other teams or competitions. External validation on larger, multi-team and multi-tournament F5 datasets is therefore a prerequisite before the identified predictors are used to inform tactical practice.

Author Contributions: Introduction: B.A.B.-P. and J.P.-O.; method: B.A.B.-P. and R.Y.-S.; analysis; B.A.B.-P. and R.Y.-S.; critical review and editing: B.A.B.-P., J.P.-O. and R.Y.-S.; discussion and conclusions: B.A.B.-P., J.P.-O. and R.Y.-S.; writing and preparation of the paper: B.A.B.-P., J.P.-O. and R.Y.-S.; revision and editing: B.A.B.-P. and J.P.-O. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The analyzed data is publicly available on the Kaggle repository (<https://www.kaggle.com/datasets/agustingermanrojas/blind-football>) (accessed on 5 May 2026). The complete training pipeline (feature engineering, the full GridSearchCV hyperparameter grids, the 5-fold cross-validation tables, the multicollinearity/VIF screen, the DeLong pairwise tests), the serialized models, the calculated SHAP values and the figure/table scripts, all with the fixed random seed ($random_state = 42$), are openly available at <https://www.kaggle.com/datasets/agustingermanrojas/blind-football>.

Conflicts of Interest: The authors declare no conflicts of interest.

References

- Hernández-Beltrán, V.; Gámez-Calvo, L.; Castelli Correia de Campos, L.F.; Bertu, F.; Gamonales, J.M. Analysis of the Body Composition of the Players of the Spanish Blind Football Team. *Rev. Andal. Med. Deporte* **2023**, *16*, 100–106. [\[CrossRef\]](#)
- Huertas-Pineda, L.V.; Zapata-Piratova, C.D.; Santos-Tavera, J.E.; Becerra Patiño, B.A. Análisis de variables sociodemográficas, psicológicas, físicas y técnicas de los futbolistas ciegos en relación con su rol de juego: Estudio exploratorio. *Sport. Sci. J. Sch. Sport Phys. Educ. Psychomot.* **2026**, *12*, 1–25. [\[CrossRef\]](#)
- Gamonales, J.M.; Muñoz-Jiménez, J.; León-Guzmán, K.; Ibáñez, S.J. 5-A-Side Football for Individuals with Visual Impairments: A review of the Literature. *Eur. J. Adapt. Phys. Act.* **2018**, *11*, 4. [\[CrossRef\]](#)
- Becerra-Patiño, B.A.; Montenegro-Bonilla, A.D.; Valencia-Sánchez, W.G.; Olivares-Arancibia, J.; Yáñez-Sepúlveda, R.; Pino-Ortega, J. Identification of Performance Variables in Blind 5-A-Side Football: Physical Fitness, Physiological Responses, Technical-Tactical Actions and Recovery Variables: A Systematic Review. *Sports* **2026**, *14*, 3. [\[CrossRef\]](#) [\[PubMed\]](#)
- Souza, R.P.; Alves, J.M.V.M.; Gorla, J.I.; Novaes, G.; Cabral, S.I.C.; Neves, E.B.; Nogueira, C.D. Characterization of the intensity of effort of blind athletes from the Brazilian Football 5-A-Side national team. *J. Health Biol. Sci.* **2016**, *4*, 218–226. [\[CrossRef\]](#)
- Finocchietti, S.; Gori, M.; Souza Oliveira, A. Kinematic Profile of Visually Impaired Football Players During Specific Sports Actions. *Sci. Rep.* **2019**, *9*, 10660. [\[CrossRef\]](#) [\[PubMed\]](#)
- Gamonales, J.M.; Hernández Beltrán, V.; León, K.; Espada, M.; Sanabria Jiménez, M.; Alemán Ramírez, C.; Castelli Correia de Campos, L.F.; Muñoz Jiménez, J. Analysis of the shots in Football for blind people in the 2021 World Grand Prix. *Cult. Cienc. Deporte* **2023**, *18*, 81–89. [\[CrossRef\]](#)
- Becerra Patiño, B.A.; Pino-Ortega, J.; Olivares-Arancibia, J. Exploratory review of the scientific production of 5-blind soccer with the Bibliometrix tool. *Retos* **2025**, *68*, 1025–1047. [\[CrossRef\]](#)
- Anzer, G.; Bauer, P. A Goal Scoring Probability Model for Shots Based on Synchronized Positional and Event Data in Football (Soccer). *Front. Sports Act. Living* **2021**, *3*, 624475. [\[CrossRef\]](#) [\[PubMed\]](#)
- Belloso, J.; Gómez-Ruano, M.Á.; Lago-Peñas, C. Unveiling the xG metric in soccer: How the adjusted xG improves match predictions. *Int. J. Perform. Anal. Sport* **2026**, 1–11. [\[CrossRef\]](#)
- Ruiz-de-Alarcón-Quintero, A.; Dela-Cruz-Torres, B. An Expected Goals on Target (xGOT) Metric as a New Metric for Analyzing Elite Soccer Player Performance. *Data* **2024**, *9*, 102. [\[CrossRef\]](#)
- Bandara, I.; Shelyag, S.; Rajasegarar, S.; Dwyer, D.; Kim, E.J.; Angelova, M. Predicting goal probabilities with improved xG models using event sequences in association football. *PLoS ONE* **2024**, *19*, e0312278. [\[CrossRef\]](#) [\[PubMed\]](#)
- Murillo García, C. The XG and their association with goals scored in elite football. *Rev. Iberoam. Cienc. Act. Fís. Deporte* **2025**, *14*, 85–93. [\[CrossRef\]](#)
- Davis, J.; Bransen, L.; Devos, L.; Jaspers, A.; Meert, W.; Robberechts, P.; Van Haaren, J.; Van Roy, M. Methodology and evaluation in sports analytics: Challenges, approaches, and lessons learned. *Mach. Learn.* **2024**, *113*, 6977–7010. [\[CrossRef\]](#)
- Iapteff, L.; Le Coz, S.; Rioland, M.; Houde, T.; Carling, C.; Imbach, F. Toward interpretable expected goals modeling using Bayesian mixed models. *Front. Sports Act. Living* **2025**, *7*, 1504362. [\[CrossRef\]](#) [\[PubMed\]](#)
- International Blind Sports Federation. *IBSA Blind Football Laws of the Game 2022–2025*; International Blind Sports Federation: Bonn, Germany, 2022; Available online: <https://ibsasport.org/sports/football/about/rules-and-downloads/> (accessed on 5 May 2026).
- Sakuma, T.; Kobayashi, M.; Kinoshita, H.; Matsui, Y.; Kobayashi, Y.; Watanabe, M. Three-dimensional kinematics analysis of blind football kicking. *Sports Biomech.* **2023**, *22*, 1136–1152. [\[CrossRef\]](#) [\[PubMed\]](#)
- Comité Paralímpico Español. *Fútbol-5. Recuperado el 3 de Julio de 2025*; Comité Paralímpico Español: Madrid, Spain, 2025; Available online: <https://www.paralimpicos.es/deportes-paralimpicos/futbol-5> (accessed on 1 May 2026).
- Lundberg, S.M.; Lee, S.I. A unified approach to interpreting model predictions. *Adv. Neural Inf. Process. Syst.* **2017**, *30*, 4765–4774.
- Lundberg, S.M.; Erion, G.; Chen, H.; DeGrave, A.; Prutkin, J.M.; Nair, B.; Katz, R.; Himmelfarb, J.; Bansal, N.; Lee, S.I. From Local Explanations to Global Understanding with Explainable AI for Trees. *Nat. Mach. Intell.* **2020**, *2*, 56–67. [\[CrossRef\]](#) [\[PubMed\]](#)
- Chen, S.; Li, X.; Ouyang, Y.; Hong, W.; Jiang, W.; Li, F.; Zhao, Y.; Liu, Y.; Zhao, Y.; Zhou, T. An explainable machine learning analysis of technical and tactical indicators associated with CSL match outcomes. *Front. Psychol.* **2026**, *17*, 1854812. [\[CrossRef\]](#)
- Cavus, M.; Biecek, P. Explainable expected goal models for performance analysis in football analytics. In Proceedings of the 2022 IEEE 9th International Conference on Data Science and Advanced Analytics (DSAA), Shenzhen, China, 13–16 October 2022; pp. 1–9. [\[CrossRef\]](#)
- Mănescu, D.C. Big Data Analytics Framework for Decision-Making in Sports Performance Optimization. *Data* **2025**, *10*, 116. [\[CrossRef\]](#)
- O'Donoghue, P. *Research Methods for Sports Performance Analysis*; Routledge: London, UK, 2010.
- Green, S. *Assessing the Performance of Premier League Goalscorers*; OptaPro Blog: London, UK, 2012.
- Rodu, J.; DeJong Lempke, A.F.; Kupperman, N.; Hertel, J. On Leveraging Machine Learning in Sport Science in the Hypothetico-deductive Framework. *Sports Med.-Open* **2024**, *10*, 124. [\[CrossRef\]](#) [\[PubMed\]](#)

27. Chicco, D.; Jurman, G. The Matthews correlation coefficient (MCC) should replace the ROC AUC as the standard metric for assessing binary classification. *BioData Min.* **2023**, *16*, 4. [\[CrossRef\]](#) [\[PubMed\]](#)
28. Shirdel, M.; Di Mauro, M.; Liotta, A. Worthiness Benchmark: A novel concept for analyzing binary classification evaluation metrics. *Inf. Sci.* **2024**, *678*, 120882. [\[CrossRef\]](#)
29. Sundaram, S.; Gowri, K.; Devaraju, S.; Gokuldev, S.; Jayaprakash, S.; Anandaram, H.; Manivasagan, C.; Thenmozhi, M. An Exploration of Python Libraries in Machine Learning Models for Data Science. In *Advanced Interdisciplinary Applications of Machine Learning Python Libraries for Data Science*; IGI Global Scientific Publishing: Hershey, PA, USA, 2023. [\[CrossRef\]](#)
30. World Medical Association. World Medical Association Declaration of Helsinki: Ethical Principles for Medical Research Involving Human Participants. *JAMA* **2025**, *333*, 71–74. [\[CrossRef\]](#) [\[PubMed\]](#)
31. Mead, J.; O'Hare, A.; McMenemy, P. Expected goals in football: Improving model performance and demonstrating value. *PLoS ONE* **2023**, *18*, e0282295. [\[CrossRef\]](#) [\[PubMed\]](#)
32. Rathke, A. An examination of expected goals and shot efficiency in soccer. *J. Hum. Sport Exerc.* **2017**, *12*, S514–S529. [\[CrossRef\]](#)
33. Ruiz, H.; Power, P.; Wei, X.; Lucey, P. "The Leicester City Fairytale?": Utilizing new soccer analytics tools to compare performance in the 15/16 & 16/17 EPL seasons. In Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Halifax, NS, Canada, 13–17 August 2017; pp. 1991–2000. [\[CrossRef\]](#)
34. Herbinet, C. *Predicting Football Results Using Machine Learning Techniques*; Technical Report; Imperial College: London, UK, 2018.
35. Moya, D.; Tipantuña, C.; Villa, G.; Calderón-Hinojosa, X.; Rivadeneira, B.; Álvarez, R. Machine Learning Applied to Professional Football: Performance Improvement and Results Prediction. *Mach. Learn. Knowl. Extr.* **2025**, *7*, 85. [\[CrossRef\]](#)
36. Bunker, R.; Yeung, C.; Fujii, K. Machine Learning for Soccer Match Result Prediction. In *Artificial Intelligence, Optimization, and Data Sciences in Sports*; Blondin, M.J., Fister, I., Jr., Pardalos, P.M., Eds.; Springer Optimization and Its Applications, 218; Springer: Cham, Switzerland, 2025. [\[CrossRef\]](#)
37. Gamonales, J.M.; Muñoz-Jiménez, J.; León, K.; Ibáñez, S.J. Differences between Championships of Football 5-a-Side for Blind People. *Appl. Sci.* **2021**, *11*, 8933. [\[CrossRef\]](#)
38. Gamonales Puerto, J.M.; Muñoz Jiménez, J.; León Guzmán, K.; Ibáñez Godoy, S.J. Efficacy of shots on goal in football for the visually impaired. *Int. J. Perform. Anal. Sport* **2018**, *18*, 393–409. [\[CrossRef\]](#)
39. Markopoulou, C.; Papageorgiou, G.; Tjortjis, C. Diverse Machine Learning for Forecasting Goal-Scoring Likelihood in Elite Football Leagues. *Mach. Learn. Knowl. Extr.* **2024**, *6*, 1762–1781. [\[CrossRef\]](#)
40. Hewitt, J.H.; Karakuş, O. A machine learning approach for player and position adjusted expected goals in football (soccer). *Frankl. Open* **2023**, *4*, 100034. [\[CrossRef\]](#)
41. Becerra-Patiño, B.A.; Monterrosa-Quintero, A.; Olivares-Arancibia, J.; López-Gil, J.F.; Pino-Ortega, J. Differences in Anthropometric and Body Composition Factors of Blind 5-a-Side Soccer Players in Response to Playing Position: A Systematic Review. *J. Funct. Morphol. Kinesiol.* **2025**, *10*, 238. [\[CrossRef\]](#) [\[PubMed\]](#)
42. Becerra Patiño, B.A.; Escorcia-Clavijo, J.B. The transfer and dissemination of knowledge in sports training: A scoping review. *Retos* **2023**, *50*, 79–90. [\[CrossRef\]](#)

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.