Original research article

The strategy tradeoff between risk and reward in soccer: Expected ball gain and expected possession value in Bundesliga

Leander Forcher[1,2] (iD)**, Leon Forcher[2]** (iD)**, Stefan Altmann[1,3]** (iD)**, and Alexander Woll[1]**

Abstract

Tactical behavior of soccer players always comes with both risk and reward depending on the outcome of the actions. This is especially true during the often chaotic playing phase of transition following a ball gain/loss, where player actions can highly impact a team's success. Therefore, this study analyses the risks and rewards of transitions, focusing on the objectives of both the defensive and offensive teams. Official tracking and event data of 612 matches played during two Bundesliga seasons (2022/23 and 2023/24) were analyzed. To assess transitions, two separate machine learning models were developed based on over 30 expert-driven features: (i) an expected possession value (EPV) model that predicts the probability of the ball-gaining team to score a goal in the following seconds of the possession, (ii) an expected ball gain (xBG) model that predicts the probability of the ball losing team to regain the ball in the following seconds of the possession. Overall, 220,226 match situations during the 58,868 transitions were analyzed. The resulting EPV (AUC: 0.88) and xBG (AUC: 0.64) models showed satisfactory prediction performance. Based on the combination of both models, the risk and reward of match situations for both teams can be quantified in detail which ultimately assists in analyzing the tactical strategy in offensive and defensive transition. This is illustrated in an in-depth individual player analysis to evaluate tactical player decision-making and determine the influence of the players on the success of a team.

Keywords

Association football, decision-making, performance analysis, tactics

Introduction

In elite soccer, a single decisive match situation can make the difference between success and defeat, which can have enormous economic consequences for the teams. This is why vast sums of money are spent to increase the chances of success.[1] To gain such an on-field advantage, the match performance of soccer players has become a main topic of interest[2] with the aim to assess players' performances objectively and ultimately assist them in reaching their potential.

For this purpose, the research on soccer match performance has focused on the assessment of four different performance facets which are mutually dependent but can be assessed individually: physical, technical, psychological, and tactical match performance.[3,4] In detail, tactical performance describes the effectiveness of a player's behavior in achieving a tactical objective (e.g. scoring[5]). Accordingly, the assessment of tactical behavior should always be placed in the context of its effectiveness in order to judge a performance reasonably. Since the tactical objectives differ significantly depending on the possession of the ball and the context of the match situation, six distinct playing phases

are differentiated in the analysis of tactical match performance: defensive & offensive play, defensive & offensive transition, defensive & offensive set plays.[6,7] Those playing phases proceed in a mutual interrelationship of both teams with opposing tactical objectives. For example, while one team is in offensive transition after a ball gain (i.e. objective to control the ball and eventually score), the opposing team is in defensive play after a ball loss (i.e. objective to regain the

Corresponding author:
Leander Forcher, Institute of Sports and Sports Science, Karlsruhe Institute of Technology, Engler-Bunte-Ring 15, 76133 Karlsruhe, Germany.
Email: leander.forcher@kit.edu

ball and save the own goal). In this context, most research on tactical match performance assesses one playing phase at a time, such as the individual analysis of offensive play[8] or the individual analysis of defensive play.[9] However, to gain a deeper insight into the complexities of tactical match performance it is arguably a powerful approach to analyze the opposing phases of the game simultaneously to get a complete picture of a match situation. Therefore, this study focuses on the analysis of the mutual opposing playing phases of transition, thereby considering the objectives of both the defensive and offensive teams.

This comprehensive way of examining opposing tactical objectives can also be interpreted as *risk-reward* analyses. For instance, Neuberg and Thiem[10] analyzed risk and reward in handball and showed, among other things, that underdogs can benefit more from taking higher risks than favorites do. In soccer, this type of analysis has been shown to grant deeper insights into the assessment of tactical match performance. For instance, it was used to assess the passing performance of players by modeling the success probabilities of passing options.[11,12] Following this idea applied to the present assessment of transitions, the team in defensive transition (i.e. after a ball loss) aims at regaining the ball and saving the own goal. In order to achieve those objectives the players counter-press to regain the ball as a potential *reward*[13] while minimizing the *risk* of conceding as a result of a dangerous counter-attack, e.g. by securing deep spaces behind the defending line using rest defense.[14] On the other hand, the team in offensive transition may directly drive a counter-attack with a high probability in scoring a goal taking advantage of an unorganized opposing defense after a ball loss as potential *reward*[15] without taking to great of a *risk* at losing the ball possession.

To assess those detailed behavioral patterns of players during a competitive match, the almost continuous and highly accurate tracking data has been established as the most promising data to evaluate in the last years.[16–18] The resulting Big-Data sets of tracking data and complementary data formats such as event data make them especially useful for the deployment of computer science methods such as machine learning.[19,20] Using those highly time-efficient procedures which are able to filter the most important information out of large data sets by detecting patterns such as success factors in players' behavior makes them especially attractive for soccer analytics.[21,22] Based on tracking data and machine learning (ML), offensive performance analytics have produced the most well-known performance metric in soccer, namely expected goals (xG).[23,24]

While xG estimates the likelihood of a shot resulting in a goal, recent approaches also assessed the scoring probability of all on-ball actions or entire possessions, enabling a more detailed evaluation of all actions on the pitch. In detail, various approaches have been developed to model the success of actions and attacking sequences, differing in both the input data and the methodologies applied.

For instance, xThreat relies solely on event locations and uses a Markov process to estimate scoring probabilities based on a scoring probability surface across pitch zones, without using machine learning methodologies.[25] In contrast, action based models, such as VAEP models,[26] use shallow machine learning methodologies and event data information to estimate the change in scoring and conceding probabilities resulting from each action. In contrast to both event data based approaches (xThreat & VAEP), expected possession value (EPV) models divide the match into distinct possession phases of consecutive actions where one team continuously controls the ball.[27,28] Based on those possession phases they use complex input data (e.g. raw tracking data) and deep neural networks to continuously predict the outcome of the playing sequence.[29–31] These presented EPV models also account for off-ball behavior, such as modeling pass option surfaces and space control, and their sophistication reflects the complexity of goal scoring in soccer.[30]

However, those recent EPV studies approached the problem of predicting the scoring probability of possessions by applying the ML model across all possessions without differentiating between playing phases.[30] However, organized playing sequences (offensive & defensive play) differ significantly in the match situation context of transition playing sequences (offensive & defensive transition). Consequently, research has already shown that the use of different models for different match situations, such as in differing playing phases, can improve the quality of models in soccer.[23] Therefore, this study solely focuses on match situations during transitions to gain deeper insights into the tactical match performance with this specific match context.

Furthermore, the recent EPV approaches mainly modeled two objective states: scoring a goal and conceding a goal. In defense, however, a team's most important tactical objective is not to score. Rather, the defending team aims at regaining the ball[32] or sometimes even concentrates exclusively on protecting its own goal (e.g. deep defending). Therefore, predicting the chance of scoring for a team that is not even in possession of the ball is of limited value. Therefore, on the defensive side, this paper focuses on the prediction of ball regains as a defensive objective.[33] Using ball regains to analyze defensive success has proven to be an effective approach for analyzing tactical behavior defensive play.[34,35] This idea of expected ball gains (xBG) enables a reasonable judgment of the tactical performance of the defending team by playing the defending behavior in the context of the true objectives of the defending team. This enables a more detailed analysis of match situations from both an offensive and defensive viewpoint.

So far, both approaches (EPV & xBG) have been deployed either without a differentiation according to playing phases[29–31] or solely on organized playing sequences of

offensive and defensive play.[33] Table 1 presents a concise overview of the first author's work in this area. This paper is the first to analyze both models solely for the playing phases of transition.

However, besides the described approaches, there have been several other studies that analyzed transitions in soccer which are of value for the present study. In defensive transition, it has been shown that the time to regain the ball is the most important KPI indicating successful performance.[36,37] Further, two group tactics in defensive transition, namely counter-pressing[13] and rest defense,[14] have been assessed in detail. On the offensive side, several overarching studies have analyzed the differences between organized and counter-attacks,[15] and counter-attacks have been assessed in more detail.[38] Still, to the best of the authors' knowledge, there is no study that analyzed the *risk* and *reward* of both offensive and defensive transitions exploiting the possibilities of tracking data and machine learning.

Therefore, the aim of this study is to analyze playing sequences of (offensive & defensive) transition by developing two separate machine learning models that analyze tactical match performance. First, an EPV model that predicts the probability of the attacking team after a ball gain to score a goal in the following seconds of the possession. Second, an expected ball gain (xBG) model that predicts the probability of the defending team after a ball loss to regain the ball in the following seconds of the possession. Those models are applied in a practice-oriented *risk-reward* assessment for the attacking team's transition (EPV: *reward* to score, xBG: *risk* to lose possession of the ball) and the defending team's transition (EPV: *risk* to concede, xBG: *reward* to regain the ball).

Methods

This study was conducted according to the guidelines of the Declaration of Helsinki and approved by the local ethics committee (Human and Business Sciences Institute, Saarland University, Germany, identification number: 22–02, 10 January 2022).

Data

The presented study analyzed two seasons of official tracking and event data (2022/23 and 2023/24) of the German Bundesliga including a total of 612 matches.

The tracking data of the ball and the on-field players were measured using a semi-automatic multi-camera tracking system with a measuring frequency of 25 Hz (TRACAB, ChyronHego, Melville, NY, USA). This system has been validated for soccer-specific performance assessment.[41]

The notational event data was collected by Sportec Solutions (Sportec Solutions AG, Ismaning, Germany) consisting of about 30 events with over 100 attributes defined by the official match data catalog of the German Soccer League (DFL).[42]

Both data types were synchronized by Sportec Solutions following the methodology of Anzer & Bauer demonstrating high robustness (i.e. shots were synchronized with an average error of 0.23 ± 0.49 s).[23]

All steps of data preprocessing, analysis, modeling, and visualization were completed using python 3.11.8 with the Pandas, NumPy, Math, SciPy, Matplotlib, SHAP, and scikit-learn libraries.

To ensure the traceability of the presented study the main processing steps are described below. This includes the specification of the playing phases, the derivation of features, and the detailed description of the sophisticated features deep runs and expected pass reception.

Features

Every on-ball action of the ball-possessing team during its own playing sequence of the playing phase of offensive transition was analyzed as an individual match situation. Accordingly, the observation level was event-based, while success as target variable was defined by the outcome of the same playing sequence (see Statistics). the level of observation of was event based while the success was defined using the outcome of the playing sequence of the same possession. For every match situation, the features were derived from the identified matching time frame of tracking data. This approach was selected to ensure a selective analysis of match situations where the attacking team is in control of the ball. For instance, analyzing the defending team's coverage of passing options is only meaningful if an attacker has ball control and can take advantage of a passing opportunity.

The features were carefully selected and specified by two professional match analysts of a Bundesliga team, who are also the first two authors of this paper (LeaF & LeoF), and were only considered when they reflected a technical or tactical element of match performance.[33,43] This procedure was chosen to ensure the validity of the presented approach in specifically examining the technical and tactical behavior of the players to eventually prevent false conclusions.[2]

The features were selected to provide detailed information on the match context (e.g. pitch position), the on-ball and off-ball behavior of the offense (e.g. availability of close pass options), and the off-ball behavior of the defending team (e.g. defensive pressure on the ball-leader).

All detailed feature specifications can be found in Table 2. Due to their importance and complexity, the identification of the *playing phase of transition*, *tactical formations*, *deep runs* and *expected pass reception* are presented in more detail below.

Playing phase transition

To analyze the match situations during the specific playing phase of transition, playing phases were defined in order to assign each match situation to one of the six playing phases

Table 1. Summarizing information on the recent work of the first author related to the presented study.

	Paper	Topic	Data	Methods
Detailed analyses of the models featured in this work	Forcher, L., Beckmann, T., Wohak, O., Romeike, C., Graf, F., & Altmann, S. (2024). Prediction of defensive success in elite soccer using machine learning—Tactical analysis of defensive play using tracking data and explainable AI. Science and Medicine in Football, 1–12. https://doi.org/10.1080/24733938.2023.2239766[33]	Analysis of tactical match performance in defense	Tracking & Event data of Bundesliga season 2022/23	Machine learning classifier (methods similar to xBG in the presented study)
General information on performance analysis in soccer	Forcher, L. (2024). Success Factors in Soccer Defense – Match Analysis in Soccer based on Positional Tracking Data [Karlsruher Institute of Technology (KIT)]. https://doi.org/10.5445/IR/1000173445[7]	General analysis of tactical match performance in defensive play	Tracking & Event data of Bundesliga season 2022/23	Not applicable
	Forcher, L. (2024). The Beautiful Game Unveiled – The Influence of Tactical Factors on Soccer Match Performance [Karlsruher Institute of Technology (KIT)]. https://doi.org/10.5445/IR/1000169718[3]	General analysis of tactical factors and their influence on match performance	Tracking & Event data of Bundesliga season 2022/23	Not applicable
	Forcher, L., Forcher, L., & Altmann, S. (2024). How Soccer Coaches Can Use Data to Better Develop Their Players and Be More Successful. In P. Düking & B. Sperlich (Eds.), Individualizing Training Procedures with Wearable Technology (pp. 99–123). Springer International Publishing. https://doi.org/10.1007/978-3-031-45113-3_7[4]	Summarizing Article on the analysis of match performance in soccer	Not applicable	Not applicable
Detailed analysis of defensive KPIs	Forcher, L., Forcher, L., Altmann, S., Jekauc, D., & Kempe, M. (2024). The keys of pressing to gain the ball—Characteristics of defensive pressure in elite soccer using tracking data. Science & Medicine in Football, 8(2), 161–169. https://doi.org/10.1080/24733938.2022.2158213[34]	Analysis of tactical match performance in defense	Tracking & Event data of Bundesliga season 2022/23	Detailed analysis of KPI "defensive pressure"
	Forcher, L., Forcher, L., Altmann, S., Jekauc, D., & Kempe, M. (2024). Is a compact organization important for defensive success in elite soccer? – Analysis based on player tracking data. International Journal of Sports Science & Coaching, 19(2), 757–76. https://doi.org/10.1177/17479541231172695[39]	Analysis of tactical match performance in defense	Tracking & Event data of Bundesliga season 2022/23	Detailed analysis of KPIs "spread" and "surface area"
	Forcher, L., Forcher, L., Altmann, S., Jekauc, D., & Kempe, M. (2023). The Success Factors of Rest Defense in Soccer – A Mixed-Methods Approach of Expert Interviews, Tracking Data, and Machine Learning. Journal of Sports Science and Medicine, 22(4), 707–725. https://doi.org/10.52082/jssm.2023.707[14]	Analysis of tactical match performance in defense	Tracking & Event data of Bundesliga season 2022/23	Detailed analysis of rest defense match situations in defensive transition
Formation changes	Forcher, L., Preine, L., Forcher, L., Wäsche, H., Jekauc, D., Woll, A., Gross, T., & Altmann, S. (2023). Shedding some light on in-game formation changes in the German Bundesliga: Frequency, contextual factors, and differences between offensive and defensive formations. International Journal of Sports Science & Coaching, 18(6), 2051–2060. https://doi.org/10.1177/17479541221130054[40]	Analysis of tactical formation changes	Bundesliga season 2020/21	Detailed analysis on formation changes in Bundesliga

Table 2. Features of tactical & technical match performance, colored according to the three main feature categories (match situation: gray, offense: orange, defense: blue).

Category	Subcategory	Feature	Exact features and expression	Calculation	Technical & Tactical Idea
Match situation	Absolute pitch position	absolute_pitch_position_lane	Categorical: outer lane ($\hat{=}1$), half lane ($\hat{=}2$), middle lane ($\hat{=}3$)	Ball position on the pitch, 5 vertical zones	The description of the absolute pitch position in thirds and lanes assists in roughly subdividing match situations
		absolute_pitch_position_third	Categorical: first third ($\hat{=}1$), middle third ($\hat{=}2$), final third ($\hat{=}3$)	Ball position on the pitch, 3 horizontal zones	In match analysis practice this subdivision of the pitch is used to describe different situational context in which different tactics are deployed Forcher et al. [33] could show that successful behavior in defense differs depending on the pitch position of the outer lane or middle lane EPV: It is expected that bringing the ball in the final third and the middle lane may increase the chance of scoring in the following seconds of the possession xBG: It is expected that channelizing the opposing ball possession towards the outer lane may increase the chance of regaining the ball in the following seconds
		absolute_pitch_position_distance_to_goal	Numerical: [m]	Distance of the ball to the center of the attacked goal	Several studies showed that the distance to the goal is the most important variable in predicting the goal probability of shots [23] or possessions EPV: It is expected that bringing the ball closer to the goal may increase the chance of scoring in the following seconds of the possession xBG: It is expected that keeping the ball possession team away from the own goal may increase the chance of regaining the ball in the following seconds
		absolute_pitch_position_angle_to_goal	Numerical: [°] (0°–90°) 0° $\hat{=}$ ball is directly beneath goal, 90° $\hat{=}$ ball is in the middle of the pitch directly in front of goal	angle of the ball towards the center of the attacked goal	Anzer & Bauer [23] showed that the angle towards the goal is an important variable in predicting the goal probability of shots EPV: It is expected that bringing the ball in a greater angle in front of the goal may

(continued)

Table 2. Continued.

Category	Subcategory	Feature	Exact features and expression	Calculation	Technical & Tactical Idea
					increase the chance of scoring in the following seconds of the possession
	Relative pitch position	relative_pitch_position_horizontal_zones	Categorical: in front of attackers of defending team (≙1), between attackers and midfielders of defending team (≙2), between midfielders and defenders of defending team (≙3), behind defenders of defending team (≙4)	Based on the formation lines of the tactical formation of the defending team in playing phase of defensive transition (see methods section, Features) 4 horizontal zones	The relative pitch position is based on the tactical principle of play to outplay opposing formation lines to disrupt the opponent's defensive organization and bring the ball closer to the opponent's goal. Based on the idea of Fernandez et al.[29] this feature was shown to be highly important in predicting successful possessions of offensive play. EPV: It is expected that outplaying opposing formation lines and thus bringing the ball behind opposing formation lines may increase the chance of scoring in the following seconds of the possession. xBG: It is expected that the chance of regaining the ball in the following seconds is increased if there are more defensive formation lines between the ball and the own goal
	History of possession	match_situation_number_actions	Ordinal: number of actions	Number of previous actions of current possession (including passes, tacklings, other ball actions, etc.)	One tactical objective during offensive transition is to move the ball quickly toward the opposing goal to make use of the unordered defensive organization. Furthermore, quick passes after a ball gain can effectively assist in outplaying possible counter-pressing by the opposing team. EPV: It is expected that more actions of the attacking team in less time may increase the chance of scoring in the following seconds of the possession. xBG: It is expected that a shorter time of the opposing possession increases the
		match_situation_possession_duration	Numerical: [sec]	Time in which the attacking team is already in ball possession up to the current match situation	

Table 2. Continued.

Category	Subcategory	Feature	Exact features and expression	Calculation	Technical & Tactical Idea
					chance of regaining the ball in the following seconds
tactical match performance offense	On-ball behavior: Ball control ball-leading player	offense_onball_ball_control_distance_to_ball	Numerical: [m]	Distance of ball-leading player to the ball	Ball control of the ball-leading player is central to deliberately performing actions on the ball and controlling the ball during the possession to eventually create a scoring opportunity
		offense_onball_ball_control_speed_ballleader	Numerical: [km/h]	Speed of ball-leading player	A small distance of the ball-leading player to the ball, a small difference of the speed of the ball may indicate a high ball control of the ball-leading player Furthermore, it is assumed that a ball with higher speed is harder to control for the ball-leading player
		offense_onball_ball_control_speed_ball	Numerical: [km/h]	Speed of the ball	EPV: It is expected that a higher ball control of the ball-leading player may increase the chance of scoring in the following seconds of the possession
		offense_onball_ball_control_speed_difference_player_ball	Numerical: [km/h]	Difference in speed of ball-leading player and ball	xBG: It is expected that a lower ball control of the ball-leading player may increase the chance of regaining the ball in the following seconds
	Off-ball behavior: Pass options	offense_offball_pass_options_close_to_ball_5_mean	Numerical: [%] (0–100%) mean of expected pass reception value for 5 attacking players closest to the ball	Expected pass reception prediction (see methods section, expected pass reception) for the 5 closest pass options to the ball	One tactical objective during offensive transition is to move the ball quickly from the origin of the ball gain to outplay possible counter-pressing of the opposing team
		offense_offball_pass_options_close_to_ball_5_higher50	Ordinal: number of players (0–5) 5 attacking players closest to the ball whit an expected pass reception value above 50% (0–2)		Several studies showed that the coverage of close pass options is crucial to regain the ball in defense [33,34] Interpreted conversely, this shows the importance of open pass options to maintain the ball control EPV: It is expected that a higher number of open close pass options may increase the chance of scoring in the following seconds of the possession xBG: It is expected that high coverage of close pass options resulting in a low

(continued)

Table 2. Continued.

Category	Subcategory	Feature	Exact features and expression	Calculation	Technical & Tactical Idea
					number of open pass options may increase the chance of regaining the ball in the following seconds
		offense_offball_pass_options_dangerous_2_mean	Numerical: [%] (0–100%) mean of expected pass reception value for 2 attacking players closest to the goal	Expected pass reception prediction (see methods section, expected pass reception) for the 2 pass options closest to the attacked goal	The main tactical objective during offensive transition is to move the ball quickly toward the opposing goal by taking advantage of the opposing unbalanced defense after a ball loss. Therefore, pass options closest to the opposing goal are expected to be the most dangerous pass receivers to create a scoring opportunity in the following actions. EPV: It is expected that if those dangerous pass options are not covered it may increase the chance of scoring in the following seconds of the possession. xBG: It is expected that if those dangerous pass options are covered by the defending team it may increase the chance of regaining the ball in the following seconds
		offense_offball_pass_options_dangerous_2_higher50	Ordinal: number of players (0–2) 2 attacking players closest to the goal whith an expected pass reception value above 50 [%] (0–2)		
	Off-ball behavior: Space control	offense_offball_space_control_final_third_offense_percentage	Numerical: [%] (0–100)	Percentage of space control of attacking team in the final third of the pitch based on voronoi diagrams	One tactical principle of attacking is to create space. This controlled space is especially valuable in the final third of the pitch in front of the goal and was used as a performance metric in several other approaches [8]. EPV: It is expected that a higher space control of the attacking team in the final third may increase the chance of scoring in the following seconds of the possession
	Off-ball behavior: Box occupation	offense_offball_box_occupation_offense_total	Ordinal: number of players (0–11)	Number of players in the penalty box in front of the attacked goal	One tactical principle of offensive play is to bring and position attackers in the opposing penalty box and in the best
		offense_offball_box_occupation_numerical_superiority_offense	Ordinal: difference in number of attackers and		

(continued)

Table 2. Continued.

Category	Subcategory	Feature	Exact features and expression	Calculation	Technical & Tactical Idea
			defenders (−11–11) positive values $\hat{=}$ attackers are in numerical superiority negative values $\hat{=}$ defenders are in numerical superiority		case outnumber the opposing defensive players This behavior increases the controlled space in front of the goal and thereby increases the chance of successfully receiving a pass and with it, creating a scoring opportunity since most goals are scored within the box EPV: It is expected that a higher box occupation of the attackers may increase the chance of scoring in the following seconds of the possession
	Off-ball behavior: Deep runs	offense_offball_deep_runs	Ordinal: Number of offensive deep runs in the moment of the match situation (0–10)	Deep runs (see methods section, Deep Runs)	One tactical principle in attacking is deep runs to threaten the spaces behind the defending line and in front of the goal By performing a deep run, the attacking player may create a scoring opportunity by receiving a pass behind the defending line On the other hand, the player disrupts the defense and creates open spaces for other teammates to receive a pass Further, the presence of deep runs may indicate a high dynamic of the attack which in turn can contribute to creating a scoring opportunity EPV: It is expected that performing deep runs during the match situation may increase the chance of scoring a goal
tactical match performance defense	Off-ball behavior: Overall defensive behavior	defense_offball_defensive_pressure_on_ballleader	Numerical: [%] (0–100)	Defensive pressure on the ball-leading player is quantified using the defensive pressure metric that was used to quantify pass options (compare methods section, Pass Options & Appendix I)	One tactical objective of defensive play is to regain the ball. Several studies showed that defensive pressure on the ball-leading player is the most important performance indicator to deny the actions of the ball-leader (e.g. passes or dribblings) to eventually regain the ball [34] EPV: It is expected that a smaller defensive pressure on the ball-leading

Table 2. Continued.

Category	Subcategory	Feature	Exact features and expression	Calculation	Technical & Tactical Idea
					player may increase the chance of scoring in the following seconds of the possession xBG: It is expected that a higher pressure on the ball-leading player may increase the chance of regaining the ball in the following seconds
		defense_offball_block_efficiency	Numerical: [%] (0–100)	Space control of defending team in the defensive block	One tactical principle in defense is to build a defensive compact organization [39] to deny possible playing space for the opposing attacking team within this organization EPV: It is expected that a low block efficiency indicating a poor defensive organization may increase the chance of scoring in the following seconds of the possession xBG: It is expected that a higher block efficiency within the defensive organization may increase the chance of regaining the ball in the following seconds
		defense_offball_dis_def_mid defense_offball_dis_mid_att	Numerical: [m] distance between the mean longitudinal position of the players of the respective formation line		Creating small inter-line distances within the defensive team the defensive principle of a compact organization is applied [33] EPV: It is expected that high inter-line distances of the defending team indicating a poor defensive organization may increase the chance of scoring in the following seconds of the possession xBG: It is expected that small inter-line distances (10–13 [m],[39]) may increase the chance of regaining the ball in the following seconds
		defense_offball_defenders_in_ line_of_shot	Ordinal: number of players (0–11)	Defending players in shot area the shot area stretches between the ball position	One of the tactical objectives in defensive play is to protect the own goal In the final stages of an opposing attack

(continued)

Table 2. Continued.

Category	Subcategory	Feature	Exact features and expression	Calculation	Technical & Tactical Idea
				and the two goalposts of the attacked goal	defenders, therefore, place themselves between the ball-leading player and the own goal to block a possible shot [23] EPV: It is expected that a smaller number of defenders in line of the shot may increase the chance of scoring in the following seconds of the possession
		defense_offball_goalkeeper_in_goal	Categorical: [True, False]	Goalkeeper in goal is defined by the position of the goalkeeper (goalkeeper in line of shot) and his distance to the goal (closer than 11 meters to the goal)	One of the main tactical objectives of a goalkeeper is to protect the own goal by saving or deflecting shots at the goal EPV: It is expected that goalkeepers who are out of position may increase the chance of scoring in the following seconds of the possession
	Off-ball behavior: Defensive Pressing	defense_press_number_defenders_10m defense_press_number_defenders_20m	Ordinal: number of players (0–11)	Number of defenders around the ball (10 & 20 [m])	One tactical objective in defense is to regain the ball Several studies have shown that pressurizing ball-near areas by placing more defenders in those areas (e.g. to cover close pass options) are highly important for the success of defensive play [13,33]
		defense_press_numerical_superiority _def_10m defense_press_numerical_superiority _def_20m	Ordinal: difference in number of defenders and attackers (−11–11)	Numerical superiority of defenders around the ball (10 & 20 [m])	xBG: It is expected that a higher number of defenders and a defensive numerical superiority in ball-near areas may increase the chance of regaining the ball in the following seconds
		defense_press_spread_5def	Numerical: [m]	Spread of the five defenders closest to the ball (sum of squared standard deviation from their average position (centroid))	Recently, a study indicated that the compact organization of the defending team is essentially shaped by the compactness of the defenders in ball-near areas [39] xBG: It is expected that higher compactness of defenders in ball-near areas may increase the chance of regaining the ball in the following seconds

Table 2. Continued.

Category	Subcategory	Feature	Exact features and expression	Calculation	Technical & Tactical Idea
		defense_press_surface_area_5def	Numerical: [m²]	Surface area of the five defenders closest to the ball (convex hull)	
	Off-ball behavior: Rest defense	defense_restdef_defenders_behind_ball	Ordinal: number of players (0–11)		Rest defense is a group tactic during defensive transition with the aim to stop opposing counter-attacks and in the best case regain the ball (definition see: [14])
		defense_restdef_number_defenders	Ordinal: number of players in rest defense area (0–11)	REST defense area is defined as 10 [m] in front of the defender closest to the own goal	Placing more defenders behind the ball [13] or in the rest defense area [14] was shown to increase the success of the defensive transition
		defense_restdef_numerical_superiority_def	Ordinal: difference in number of defenders and attackers in rest defense area (−11–11)		EPV: It is expected that placing fewer defenders behind the ball or in the rest defense area may increase the chance of scoring for the attacking team in the following seconds of the possession xBG: It is expected that placing more defenders behind the ball or in the rest defense area may increase the chance of regaining the ball in the following seconds
		defense_restdef_space_control_att_relative	Numerical: [%] (0–100)	Percentage of space control of attacking team behind the rest defense area (based on voronoi diagrams)	Recently, it was shown that the space control behind the defending line is an important feature in determining the success of the defensive transition [14]
		defense_restdef_space_control_att_total	Numerical: [m²]	Total surface area of space control of the attacking team behind the rest defense area (based on Voronoi diagrams)	By denying space control of the attacking team, the defense ensures that the possible dangerous counter-attackers are covered and the deep spaces behind the defending line are secured EPV: It is expected that a higher space control behind the opposing rest defense may increase the chance of scoring for the attacking team in the following seconds of the possession xBG: It is expected that a lower space control of the attacking team behind the rest defense may increase the chance of

Table 2. Continued.

Category	Subcategory	Feature	Exact features and expression	Calculation	Technical & Tactical Idea
					regaining the ball in the following seconds
		defense_restdef_pressure_on_attackers_mean defense_restdef_pressure_on_attackers_min	Numerical: [%] (0–100)	Defensive pressure is quantified using the defensive pressure metric that was used to quantify pass options (compare methods section, Pass Options & Appendix I)	To cover potentially dangerous counter-attackers (e.g. by directly pressurizing them) is highly important to deny opposing counterattacks and regain the ball EPV: It is expected that a lower pressure on the attackers in the rest defense area may increase the chance of scoring for the attacking team in the following seconds of the possession xBG: It is expected that a higher pressure on attackers in the rest defense area may increase the chance of regaining the ball in the following seconds

(defensive & offensive play, defensive & offensive transition, defensive & offensive set plays). First, the ball possessions, or playing sequences, of each team were identified based on a rule-based approach to calculate possession chains.[44] Afterward, the playing sequences were classified into the six specific playing phases: defensive & offensive play, defensive & offensive transition, and defensive & offensive set plays. All playing phases were defined by two professional match analysts of a Bundesliga team (LeaF & LeoF). Accordingly, offensive transitions were defined as follows: They started with a gain of possession (with a minimum of one successful on ball action, e.g. pass) of the defending team during open play and ended whenever the ball got out of play, gameplay was interrupted (e.g. foul, offside), the opposing team regained the ball, or the phase of play switched to a controlled possession (offensive play). In detail, the playing phase of offensive transition shifted to controlled possession whenever the team gaining possession executed either two consecutive backward passes or two actions while all 10 opposing players were positioned behind the ball. On the other side, defensive transitions occurred in the time interval of an opposing offensive transition.

This study focused on those (offensive & defensive) transition playing sequences.

Tactical formation

To define the relative pitch position of the ball position, the tactical formation of both the defending and the attacking team was identified. With this information, it is possible to determine the number of outplayed formation lines and, for instance, whether the ball is positioned between the midfielder- and defending-line, or behind the defending-line, etc. In detail, a k-Means clustering algorithm was used to cluster the longitudinal x-positions of all on-field players of a team into three formation lines (e.g. 5-3-2). This analysis was applied separately for each playing phase in three different time windows over the match (0–45, 46–62.5, 62.6–90 [min]). Those time windows were determined since almost all tactical formation changes occur during the second half.[40]

Deep runs

A deep run denotes an off-ball movement of an attacking player to threaten deep spaces behind the opposing defensive line with the aims (i) to create a scoring opportunity by outplaying the defensive line or (ii) to disrupt the opposing defensive structure, in consequence of defending the deep run (e.g. the defense has to fall back towards the own goal to secure the threatened spaces). This type of tactical off-ball behavior of attackers was analyzed to indicate the dynamic of an offensive transition.

While there is research on overlapping runs, which can be interpreted as deep runs at the outer lanes of the pitch,[45] to date, there is no validated definition of deep runs in soccer literature. Therefore, we defined a rule-based approach to identify deep runs based on five criteria: (i) velocity, (ii) area, (iii) direction, (iv) duration, and (v) playing phase.

During the deep run, the player has to reach (i) a minimum velocity of 21 [km/h] which is considered a high-intensity run from a physical point of view. Further, the run has to be performed (ii) in proximity to the last defending line of the opposing team. In detail, the deep running player has to have a minimum of 15 [m] vertical distance to the last defender and decrease this distance to a minimum of 4 [m] at least once during the run (i.e. come close to the defending line or cross it). Additionally, the deep run has to be performed (i) with a direction towards the opposing goal line. Accordingly, the deep run may not exceed a maximum angle of 45 degrees towards the goal line. Overall, a deep run has (iv) to last at least one second in which criteria (i-iii) have to be true. In the end, a deep run can only be performed (v) during a ball possession of the own team.

This definition was specified and validated by two professional match analysts of a Bundesliga team (LeaF & LeoF). Each analyst independently analyzed one match from the current sample. Using scouting feeds and analysis software (Hudl, Sportscode), they manually tagged instances of deep runs. During those two matches, 129 deep runs were identified by the analysts. A run was considered as detected if the manually tagged instance occurred within the timeframe defined by the rule-based identification. The rule-based approach identified 109 of them correctly (84.5%), 8 runs were identified incorrectly (false positives) and 20 runs were not identified (false negatives). This results in a Precision of 0.93 (≙proportion of true positives among all identified deep runs by the rule-based approach), Recall of 0.85 (≙proportion of true positives identified among all tagged deep runs by experts), and F1-Score of 0.89 (≙harmonic mean of Precision and Recall).

For instance, the identification of overlapping runs using a graph neural network showed a F1-Score of 0.40.[45] Concluding, the presented approach shows a powerful performance in identifying the tactical pattern of deep runs.

Expected pass reception (xPR)

To model the probability of a pass to be successfully completed we used data of 276,521 passes played during the 22/23 Bundesliga season of which 223,521 (80.83%) were successful.

The used event data provided no information on the targeted receiver for unsuccessful passes. However, this information is mandatory to analyze the coverage of the passing lane or the pressure on the receiver to effectively analyze the success probabilities of passes. To determine the targeted receiver for an unsuccessful pass, we identified the closest attacking player to the passing vector of the played pass. The passing vector was computed on the position of the ball one second after the pass and the position two seconds after the pass. This identification procedure was validated for all successful passes of a sample match which resulted in a correctly identified recipient in 93 [%] of the passes. This identification rate is similar to the 93% identification rate in.[46]

For each pass of the given sample, (i) the defensive pressure on the ball leader, (ii) the defensive pressure on the targeted receiver, (iii) the coverage of the passing lane, (iv) and the passing distance were computed. The specified feature computation can be found in the appendix (see Appendix 1).

Using those features, we trained an XGBoost classifier to predict a pass to be successfully completed (outperformed the Random Forest classifier on this problem). The prediction performance is shown in Table 3. This expected pass reception model (xPR) was used to compute the feature sub-category off-ball behavior: Pass options (see2).

Statistics

To assess the tactical objectives of both the offensive and the defensive team, two separate machine learning models were developed. For the offensive side, an EPV model that predicts the probability of the team in offensive transition to score within the next ten seconds of the possession.[29–31] Second, an expected ball gain (xBG) model that predicts the probability of the team in defensive transition to regain the ball in the next five seconds of the possession.[33]

The time windows were selected to reflect the distinct nature of offense and defense. While goals are typically scored from areas close to the goal and often result from longer sequences of attacking actions, ball regains can occur anywhere on the pitch. Therefore, in line with previous literature,[29–31] a ten second window was chosen for EPV to capture and credit key actions leading to goal-scoring opportunities, whereas a shorter five second window was used for xBG to better reflect the closer temporal link between defensive actions and their outcomes.

To solve the given binary tasks, we employed several classifiers, including Random Forest and XGBoost, as they have demonstrated robust performance on the given problems.[23,33] To effectively train the classifiers, a stratified train-test-validation split (60%, 20%, 20%) on a match-by-match basis was implemented. This approach prevented match situations from the same possession from appearing in different data sets, which could lead to overfitting. The test set was held out to assess the model's performance on unseen data. The validation set was used for hyperparameter tuning, and the models were optimized on a composite score of Brier score, expected calibration error

Table 3. Prediction performance of machine learning classifiers of xPR (top), EPV (middle), & xBG (bottom). The selected models for further analysis are highlighted in green.

Expected Pass Reception (xPR)

Description	Number of features	Algorithm	Optimization	Training Set (60%)								Test Set (20%)								
				Accuracy	Precision	Recall	F1-Score	log-loss	Brier Score	RPS	ECE	Accuracy	Precision	Recall	F1-Score	log-loss	Brier Score	RPS	ECE	AUC
XGBoost	4	XGBoost	ECE, Brier score & Recall	0.840	0.904	0.907	0.905	0.335	0.106	0.106	0.055	0.847	0.913	0.909	0.911	0.325	0.102	0.102	0.067	0.842

Expected Possession Value (EPV)

Description	Number of features	Algorithm	Optimization	Training Set (60%)								Test Set (20%)								
				Accuracy	Precision	Recall	F1-Score	log-loss	Brier Score	RPS	ECE	Accuracy	Precision	Recall	F1-Score	log-loss	Brier Score	RPS	ECE	AUC
Rule-based: Distance to goal	1											0.626	0.023	0.864	0.044	0.650	0.228	0.370	0.427	0.858
Rule-based: Chance-evaluation (∅ conversion rate, sitter vs. no sitter)	1											0.983	0.278	0.451	0.344	0.090	0.013	0.870	0.061	0.720
Logistic Regression	30	Logistic Regression	elastic net	0.576	0.281	0.586	0.380	0.675	0.241	0.241	0.269	0.572	0.279	0.583	0.378	0.677	0.242	0.242	0.268	0.605
RandomForest (all features)	30	RandomForest	ECE, Brier score & Recall	0.975	0.278	0.967	0.432	0.109	0.025	0.025	0.084	0.966	0.128	0.388	0.193	0.125	0.030	0.030	0.084	0.878
XGBoost (all features)	30	XGBoost	ECE, Brier score & Recall	0.991	0.740	0.176	0.284	0.039	0.008	0.008	0.006	0.990	0.479	0.103	0.170	0.045	0.010	0.010	0.006	0.883
XGBoost (best 15 features)	15	XGBoost	ECE, Brier score & Recall	0.991	0.706	0.166	0.269	0.039	0.008	0.008	0.006	0.990	0.568	0.106	0.178	0.045	0.010	0.010	0.005	0.878

Expected Ball Gain (xBG)

Description	Number of features	Algorithm	Optimization	Training Set (60%)								Test Set (20%)								
				Accuracy	Precision	Recall	F1-Score	log-loss	Brier Score	RPS	ECE	Accuracy	Precision	Recall	F1-Score	log-loss	Brier Score	RPS	ECE	AUC
Rule-based: Defensive pressure	1											0.532	0.219	0.432	0.291	1.550	0.292	0.420	0.266	0.506
Logistic Regression	30	Logistic Regression	elastic net	0.576	0.281	0.586	0.380	0.675	0.241	0.241	0.269	0.572	0.279	0.583	0.378	0.677	0.242	0.242	0.268	0.605
RandomForest (all features)	30	RandomForest	ECE, Brier score & Recall	0.669	0.373	0.726	0.493	0.604	0.209	0.209	0.242	0.609	0.305	0.589	0.401	0.639	0.225	0.225	0.241	0.646
XGBoost (all features)	30	XGBoost	ECE, Brier score & Recall	0.782	0.601	0.045	0.083	0.493	0.160	0.160	0.025	0.776	0.474	0.034	0.063	0.509	0.166	0.166	0.024	0.642
XGBoost (best 15 features)	15	XGBoost	ECE, Brier score & Recall	0.782	0.550	0.085	0.147	0.495	0.161	0.048	0.689	0.773	0.442	0.067	0.116	0.517	0.170	0.170	0.048	0.635
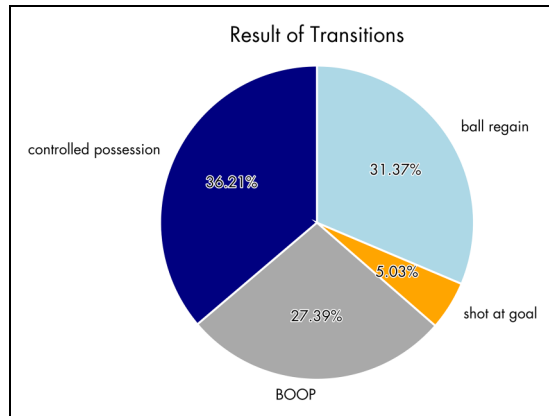
Figure 1. Result of transitions. Distribution of the ending of transitions in the Bundesliga seasons 2022/23 and 2023/24 including transitions that shifted to controlled possession, ending in the ball going out of play or a stoppage of play (BOOP), ending in a shot at goal of the attacking team, or ending in a ball regain of the defending team.

(ECE), and Recall. The weights were set to $\alpha = 0.025$, $\beta = 0.95$, and $\gamma = 0.025$, thereby emphasizing calibration while still accounting for probabilistic accuracy and sensitivity. Several different weighting configurations were tested, and the reported combination yielded the most robust balance between calibration and discrimination for the present study. This transparent, rule-based weighting scheme was deliberately chosen to enhance the interpretability and reproducibility of our approach, explicitly prioritizing calibration and discrimination. The composite score was defined as:

$$\text{Composite Score} = \alpha \cdot (-\text{Brier Score}) + \beta \cdot (-\text{ECE}) + \gamma \cdot \text{Recall}.$$

We optimized on Recall to account for adequate discrimination in the highly unbalanced data set ($\approx 1\%$ successful outcomes for EPV) while also minimizing the Brier score and ECE to maintain well-calibrated probability estimates (calibration). In detail, in the case of EPV, it is the main aim to correctly predict goals (Recall) while minimizing the number of false positives and preventing overestimation of scoring probabilities.

For feature selection, the Pearson correlation between the features was analyzed and for all feature pairs with a high correlation (Pearson's $r > 0.75$) one feature was excluded (see Appendix 2). Further, to reduce the complexity of the models and enable more comprehensible interpretability, we additionally computed reduced feature models with the best 15 features for both given problems. Finally, for the selected models SHAP values and feature importances were analyzed. SHAP values, grounded in cooperative game theory, calculate the impact of individual features on the model's output.[47] Thereby, they reveal the overall influence of each feature on the prediction (feature importance) as well as the effect of specific feature values on a

given prediction. This enables a detailed interpretability of a machine learning approach.

Baseline approaches

To further evaluate the classifiers and judge their performance on the given problems, we established several naïve baseline appoaches to model the binary tasks (i.e. EPV to predict the probability of scoring in the next ten seconds of the attacking possession & xBG to estimate the probability of a ball regain in the next five seconds of the opposing possession).

For the EPV task, we implemented two rule-based baselines:

(i) A distance to the goal approach, which linearly scales the ball's current distance to the goal from 1 directly on the goal line to 0 at the farthest possible position at the opposing corner (110.37 [m] away from goal). In contrast to other international competitions, all pitches in the German Bundesliga have standardized dimensions and a uniform coordinate system (68 [m]×105 [m]). Accordingly, the definition of the maximum possible distance to goal was consistent across all matches.

(ii) A chance-evaluation approach similar to Anzer & Bauer,[23] that assigns each match situation in the possession leading to a big scoring chance (hand-labelled as a sitter in the notated event data) its mean conversion rate in the current data set (mean conversion rate sitter = 0.52) and assigns all other situations the mean conversion rate of other scoring chances not labeled as sitter (mean conversion rate no sitter = 0.06). In the given event data sitters are defined as a shooting opportunity in a highly promising position, typically close to the goal with only the goalkeeper to beat, where the ball-leading player faces little defensive pressure, has a clear lane and sufficient angle to the goal, and can control the ball without external factors (e.g. poorly played pass).[42]

For the xBG task, we implemented one naïve rule-based baseline:

(i) A defensive pressure approach, which used the defensive pressure on the ball-leading player in every match situation (between 0–100%) as estimate to determine xBG.

Results

All 612 matches of two Bundesliga seasons resulted in a total of 220,226 match situations (on-ball actions of the attacking team) during 58,868 playing sequences of transition. Accordingly, on average 96 transition playing

sequences took place during each Bundesliga match consisting of about 3,74 actions per transition.

Across the considered Bundesliga seasons 2022/23 and 2023/24, a total of 1956 goals were scored, with 480 of them (24.54%) scored during the playing phase of offensive transitions. The other transition playing sequences shifted to controlled possession according to the derived definition 21,142 times (35.91%), while 15,992 playing sequences resulted in the ball going out of play or a stoppage of play (27.17%). Additionally, 2938 transitions (4.99%) concluded with a shot at the opposing goal. Meanwhile, defending teams actively regained possession in 18,316 playing sequences of transition (31.11%) (see Figure 1).

In 2344 (1.06%) of all match situations, the attacking team scored within the next 10 s of their possession (EPV target variable). In 51,448 (23.36%) of all match situations the defending team regained the ball within the next 5 s of the opposing possession (xBG target variable).

All Classifiers showed a satisfactory prediction performance (see Table 3). For both EPV and xBG the reduced feature models of XGBoost were selected for further analysis since they showed similar prediction performance to the all feature models. SHAP values and feature importances of those models are depicted in Figure 2 and their calibration as well as their probability distribution on the test set are illustrated in Figure 3.

Discussion

The aim of this study was to develop and deploy two separate machine learning models (EPV & expected ball gain (xBG)) to analyze the tactical behavior of players during the playing phase of defensive and offensive transition in the context of the objectives of those mutually opposing playing phases (objectives offensive transition: control the ball & score, objectives defensive transition: save the own goal & regain the ball). The combination of those models enables a *risk-reward* assessment of the tactical match performance. In detail, the defending team's *risk* of conceding a goal in the following seconds of the opposing possession can be determined by the magnitude of opposing EPV, while their potential *reward* to regain the ball can simultaneously be measured by the size of xBG. Conversely, for the attacking team, the *reward* of scoring in the next few seconds of the possession can be analyzed by their EPV, and the *risk* of losing the ball can be evaluated by the size of the opposing xBG. With it, the achievement of all main objectives of the transition playing phases can be quantified which enables a detailed analysis of tactical match performance of teams and players. By applying those models in the described way, the *risk* and *reward* of every match situation in transition can be quantified and the quality of every player decision can be evaluated in detail.

The presented EPV model in transition showed robust prediction performance (Accuracy: 0.99, Recall: 0.11, F1-Score: 0.18, AUC: 0.88) and outperformed the baseline approaches as well as standard statistical models (e.g. logistic regression, see Table 3). Furthermore, the selected model indicated higher Accuracy and AUC compared to a comparable EPV model in the playing phase of offensive play (Accuracy: 0.97, Recall: 0.32, F1-Score: 0.17, AUC: 0.80). Accordingly, the presented model successfully identifies 11% of all successful match situations (a goal is scored in the next 10 s of the possession) which is comparable to benchmark xG models (Recall: 0.20[23]; Recall: 0.30[48]) in identifying goals scored. This worse performance in identifying true positives may be traced back to the higher imbalance of the dataset in EPV (1.06% of actions entail a goal in the next ten seconds) compared to xG (13.08% of shots result in goals, across the 2022/23 and 2023/24 Bundesliga seasons: 14,956 shots, 1956 goals). Overall, this robust performance of the presented EPV model in discriminating between the classes may be attributed to the large dataset of over 600 matches analyzed in the current study,[49] and the focus on Recall in model optimization.

Next to the discrimination, the calibration of the presented models is also of interest, as the predicted probabilities should reflect the true likelihood of an action resulting in a goal (in the case of EPV) or a ball recovery (in the case of xBG) to be interpretable.[50] Although the predicted probability distribution for the selected EPV model (Figure 3) is concentrated toward values near zero, the selected EPV model maintained robust calibration even in higher probability bins (Brier score: 0.01, RPS: 0.01, log-loss: 0.05). However, the model slightly overestimates the true probabilities, which should be noted when interpreting the estimates.

While the presented EPV model's most important features belong to the match situational (three features included), it included more offensive features (six features included) and defensive features (six features). This underlines the importance of taking both the offensive and the defensive team into account when analyzing offensive match performance.[7,51]

The most important variable in the EPV model was distance to the goal (see Figure 2). This finding is in line with previous studies on EPV or xG.[23,48] The information on the distance towards the opposing goal is complemented by the relative pitch position of horizontal zones (3rd most important variable in EPV). Accordingly, a match situation is becoming more likely to result in a goal if more opposing formation lines are outplayed and/or the distance to the goal is decreased.

The second most important feature in the current EPV prediction in transitions is the box occupation of the attacking team. Besides, this feature is more relevant in transitions compared to offensive play. Accordingly, the attacking team should occupy deep spaces especially occupy the opposing box to effectively threaten the opposing goal.

In contrast, the sophisticated feature deep runs was solely included as the 14th most important feature in the EPV model. Still the presented rule-based approach to identify deep runs showed exceptional performance (Accuracy:
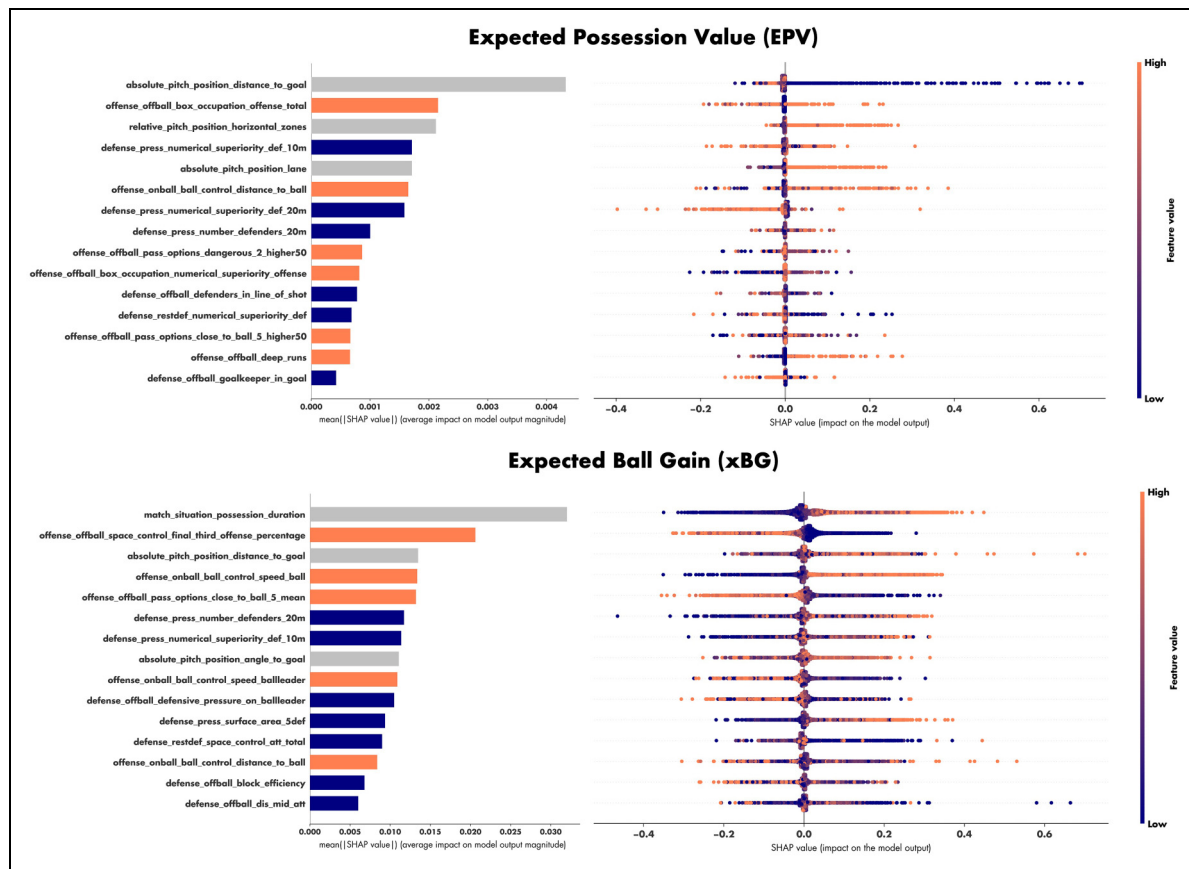
Figure 2. SHAP values of EPV (top) and xBG (bottom). Feature importances on the left are illustrated according to their category (defense: blue, offense: orange, match situation: gray). The feature expressions are depicted on the right.

0.85, F1-Score: 0.92) in comparison to other studies that focused on the identification of related tactical patterns such as overlapping runs (F1-Score: 0.40)[45] (see methods section). This feature was designed to reflect the dynamic of possession and increase the chance of scoring by threatening deep spaces and disrupting the opposing defensive organization. One possible argument why deep runs were not highly important for EPV may be that most transitions are highly dynamic due to the nature of the match situations and this feature therefore holds few information for the predictions. In this context, further analysis showed that over all analyzed match situations of offensive transition, 18.52% had a minimum of one deep run (40,781). Of all match situations that entailed a scored goal (successful EPV), 50.77% of match situations included a deep run (1190). While this exemplary result shows a clear trend of more deep runs in successful offensive match situations, it did not improve the machine learning model. Further, deep runs may hold more crucial information for match situations that are not highly dynamic, such as controlled possessions, as they have a greater impact on differentiating dynamic and static match situations.

In comparison to the presented EPV model, the presented xBG model that predicts the probability for the defending team to regain the ball within the next five seconds of the possession showed a decreased but satisfactory prediction performance (Accuracy: 0.77, Recall: 0.07, F1-Score: 0.12, AUC: 0.64). Relative to the EPV model, the xBG model yields a broader spread of predicted probabilities (see Figure 3), reflecting the more balanced dataset (23.36% successful outcomes) compared to EPV (1.06% successful outcomes). While the model's calibration indicates robust prediction results (Brier score: 0.17, RPS: 0.17, log-loss: 0.52), it also slightly overestimates the true probabilities.

In this context, it could be argued that objective of regaining the ball in defense is more difficult to identify compared to goals in offense (predicted in EPV). In detail, goals are predominantly scored after a sum of positive tactical and technical behavior of the attackers. In contrast, regaining possession for the defending team frequently appears to be a consequence of opponent misconduct (e.g. technical errors in passing) and thereby is not solely dependent on positive tactical behavior of the defense. This makes the prediction of positive tactical situations in

defense (predicted in xBG) more difficult. In this context, recently, an approach of xBG in controlled possession showed higher prediction performance (Accuracy: 0.85, Recall: 0.65, F1-Score: 0.58).[33] However, in this case, a positive defending outcome was determined with a ball gain within two seconds after the match situation (compared to five seconds in the present study). With it, it is important to note that prediction models with other basic principles for the division of classes (successful vs. unsuccessful) are hardly comparable. Still, it can be generally assumed that outcomes that are further in the future (three seconds more in the presented model) are more difficult to predict.

Overall, the presented xBG model included seven defensive features and three features of match situation context. While there were only five offensive features included, they showed high feature importances (three features under the five most important features for the xBG model, see Figure 2). This also includes the second most important feature of the model, the space control of the attacking team in the final third. The SHAP values for this feature indicate that a ball regain becomes more likely when the defense has more spatial control in the final third (see Figure 2). Accordingly and in line with previous research on defensive transition, the defending team should therefore safeguard deep spaces behind their defending line to be successful.[13,14] With this tactical behavior, the defending team is also able to decrease the chance of successful opposing counterattacks (see SHAP values of EPV in Figure 2). The safeguarding of deep spaces can be achieved by falling back to cover those pitch areas and denying pass receptions behind the defending line in rest defense. However, to achieve a high *reward* and simultaneously minimize the *risk* in defensive transition the results also suggest to pressurize the ball near areas in counter-pressing (e.g. by covering close pass options, 4th most important feature for xBG). Consequently, the ball-leader may not be able to play a controlled deep pass. This defending principle of covering close pass options to regain the ball is supported by a number of investigations on defensive performance.[9,13,33,34]

Next to safeguarding deep spaces in rest defense and pressurizing ball-near areas in counter-pressing, the possession duration of the transition is of interest. This feature is the most important in the xBG model. Several studies on defensive transition also identified the "defensive reaction time" (time to regain the ball after a ball loss) to be an important KPI in transitions.[36,37] However, the SHAP values of the presented xBG model suggest that the probability to regain the ball is increased with increasing duration of opposing possession. In this context, the presented EPV model indicates that the longer the attacking transition lasts the fewer the chance is to score (see Figure 2). Overall, one may therefore argue that fast transitions are most dangerous as they take advantage of an unorganized defense after the ball loss, and long transitions increase the chance of the possession to end in

another game state rather than conceding. Those results suggest that the defending team needs to deny fast counter-attacks as the chance to score for the attacking team is reduced the longer the possession takes. Accordingly, the defending team in defensive transition should focus on preventing fast-counter attacks by rest defending or in the best case by directly counter-pressing ball near areas. Still, this comes with high *risks* in the case the counter-pressing is overplayed leaving large spaces for the attacking team.

The availability of close pass options was the 4th most important variable in the presented xBG model, indicating that more available close pass options increase the probability of the attacking team to remain ball control. This feature is calculated using the advanced expected pass reception model (xPR), which incorporates the average xPR of close passing options (xPR, see methods section). The presented xPR model (Accuracy: 0.85, F1-Score: 0.91, AUC: 0.84) showed comparable prediction performance to related studies which concentrated on predicting passing probabilities. For instance, Goes et al. (Accuracy: 0.85)[11] and Spearman et al. (Accuracy: 0.82, AUC: 0.84)[52] demonstrated similar results in predicting the success of passing attempts. Contrary, Bauer and Anzer's[46] pass model showed an increased prediction performance (F1-Score: 0.95, AUC: 0.96), however, they excluded passes that were blocked immediately which has a direct effect on the prediction. Those results indicate that our model effectively predicts the success probability of an attempted pass by solely including four powerful features. Thereby, it provides important information on the match situation.

Overall, the combination of both presented models hold high potential in performance analysis in various use cases which are outlined below.

Practical risk-reward assessment

In practice-oriented soccer data analysis, the true value of any approach or model is reflected in its practical application. Showcasing cutting-edge models through clear and compelling use cases demonstrates their usability and defines their significance in real-world scenarios. Furthermore, the gap between complex data points and actionable insights is bridged, making them accessible, hands-on, and interpretable. Thereby, the ultimate goal is to translate these advanced approaches into tangible impacts on the pitch, supporting teams, coaches, and analysts to enable more informed decisions and gain a competitive advantage. Therefore, the possibilities of the presented EPV and xBG models to analyze tactical match performance are illustrated.

Next to the independent analysis of EPV and xBG, the most interesting insights can be gained by analyzing the *risk* and *reward* of match situations by combining both models. For instance, in defensive transition, the *risk* to concede can be measured by the magnitude in opposing EPV (minus EPV) while simultaneously the *reward* of regaining
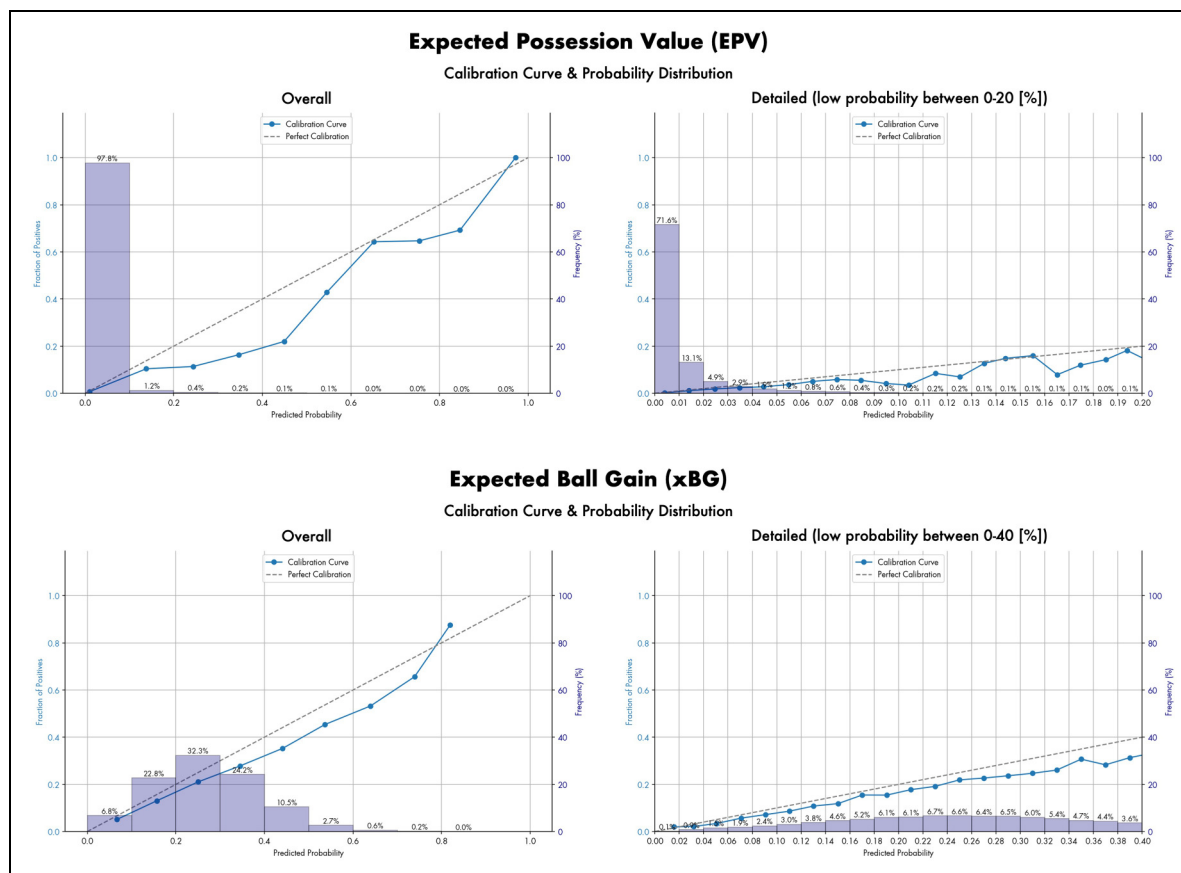
Figure 3. Calibration curve and the distribution of predicted probabilities of the selected models for EPV (top) and xBG (bottom) illustrating overall performance (left) and detailed analysis of the low-probability range (right). The calibration curves (dark blue) compare observed target variable rates (e.g. EPV: Goal in the next ten seconds of the possession) to the model's predicted probabilities across each probability bin. The probability distributions (light blue) indicate the relative frequency distribution of all predictions across each bin.

the ball can be quantified by xBG (plus xBG) which results in an overall measure of: $[-EPV + xBG]$ with a range from $[-1$ to $1]$. On the other side, in offensive transition, this evaluation of *reward* (to score) and *risk* (to loose possession) can be evaluated by the composite measure of: $[+EPV - xBG]$ with the same range $[-1$ to $1]$. For both composite measures, positive values indicate a favorable *risk-reward* scenario of a match situation or action, while negative values represent an unfavorable game state.

Those measures can be applied in both post-match and pre-match analyses. For instance, in post-match analysis, the possessions of a whole match can be analyzed focusing on EPV-xBG values to efficiently filter decisive match situations in seconds. Examples of those approaches to analyze a whole match or individual match situations using the presented models can be found in Forcher et al.[33]

Next to the analysis of whole teams, an individual player analysis can be applied to gain insights about individual performances. For soccer players the main focus is to make the right tactical decisions[11] and with the presented

metrics this decision-making process becomes measurable based on the achievement of the tactical objectives. This type of analysis is exemplarily performed on the passing performance of players as passing is the most frequent way of player interactions in offense.[12] Identifying which players create the most value for the team's performance in achieving the tactical objectives of both offense and defense can be a competitive edge in scouting or pre-match analysis.

In this case, passes with a high EPV-xBG gain can be interpreted as passes that successfully put a teammate into a position where the chance to score is increased while simultaneously the *risk* of losing ball possession is minimized. On the other hand, low EPV-xBG gain values (or even decreasing values) may indicate passes with a high *risk* of losing possession (e.g. due to a highly pressurized receiver and limited nearby passing options to maintain ball control), even if the EPV *reward* for scoring may be maximized.

An exemplary passing performance analysis using the described metrics is illustrated in Figure 4. This analysis
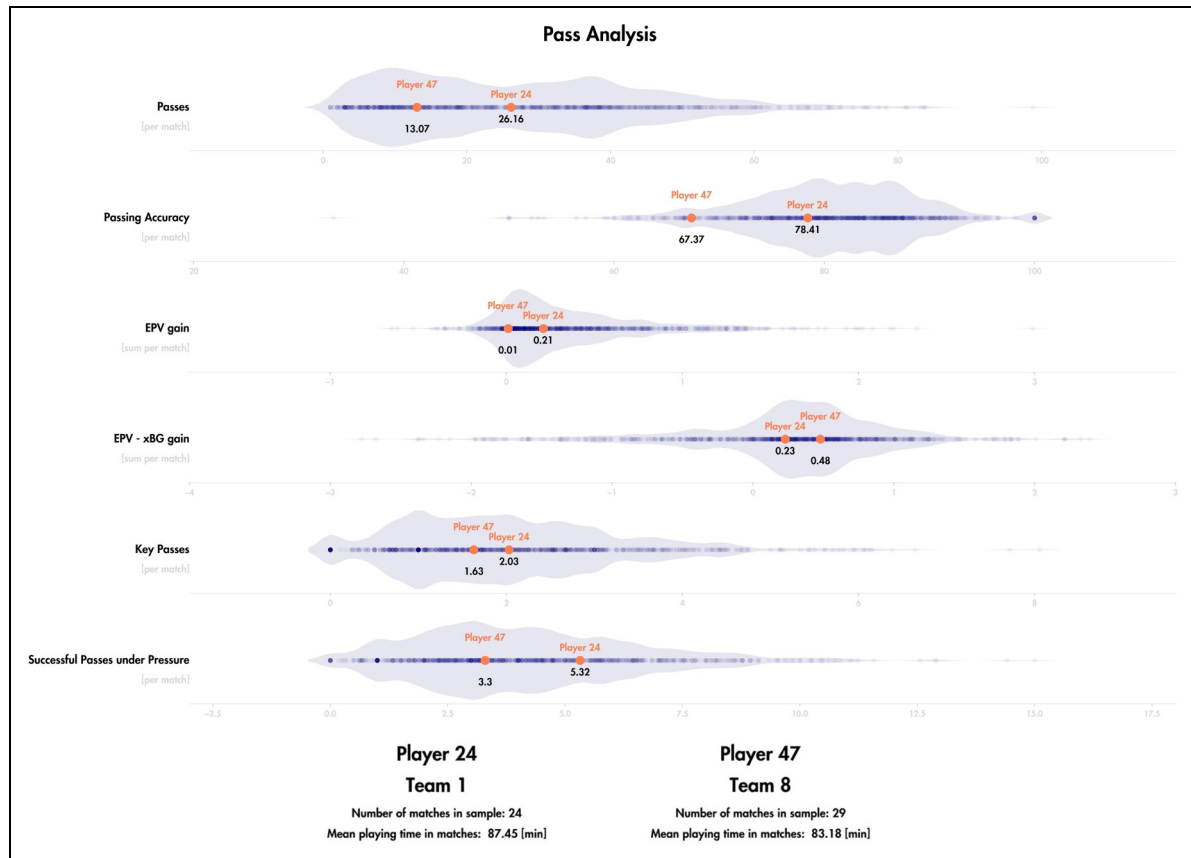
Figure 4. Individual player analysis based on the passing performance and the presented xBG and EPV models in Bundesliga season 23/24. (EPV gain ≙ sum of EPV gained by successful passes of the individual player, EPV − xBG gain ≙ sum of EPV-xBG gained by successful passes of the individual player, Key Passes ≙ number of successful passes of the individual player with an EPV-xBG gain of over 0.2, Successful Passes under Pressure ≙ number of successful passes of the individual player exposed with a defensive pressure over 60%).

allows for comparing two players based on their risk-reward decision-making in passing, contextualized by their playing time over the course of a season. It highlights their respective strengths and weaknesses in managing complex match situations, where both the structure of the attacking team (e.g. nearby passing options) and the defensive structure of the opposition (e.g. probability to regain the ball) are considered using the two presented models (EPV & xBG). This enables deeper insights into players' tactical performance (e.g. success rate under pressure or chance creation with passing) and can thereby support various processes in elite soccer (e.g. scouting).

Next to the analysis of passing performance this type of analysis can be applied to other player actions such as dribblings (players that increase the probability of a successful possession by dribbling) or pass receptions (players that create valuable target options).

Limitations and future research

When interpreting the results of this study, it is important to note its limitations.

One limitation is that the optimization and evaluation of machine learning models developed in this study also focus on discrimination of classes (i.e. Recall) rather than focusing solely on calibration, which accounts for the accuracy of predicted probabilities.[50] While the resulting models indicate robust discrimination and calibration results, the models slightly overestimate the true probabilities (see Figure 3). Therefore, the probability estimates need to be interpreted with caution. Therefore, future studies should also focus on calibrating the models presented in this study (EPV & xBG) to enable a more accurate and nuanced interpretation of probability outputs.

Further, this study was conducted exclusively using data from the highest performance level of the first-division German Bundesliga. Therefore, transfers of the presented findings to other leagues or performance levels should be made with caution.

Additionally, while some information on the match situational context was included in the presented analysis (e.g. duration of the possession) other contextual information was not considered (e.g. match venue, score-line). This procedure was chosen to focus on tactical match performance

and present findings that are valuable independent of the general match context. Future studies should therefore analyze the differences in EPV and xBG depending on different match contexts. For instance, for the match context of score-line future studies could analyze whether leading teams achieve an increased EPV-xBG in offensive transition due to a high *risk* opposing team that needs to score an equalizer which may result in an unstable defense after a ball loss.

Furthermore, the presented models solely included features on tactical and technical match performance. Further information about the analyzed players (e.g. players' anthropometry such as height or weight) and their physical and psychological match performance may improve the accuracy of applied models. Even though the analysis of psychological match performance (such as anticipation) is challenging and the availability of anthropometric data in the current data formats is scarce, future studies should aim to analyze the performance of soccer players as comprehensively as possible (e.g. considering all main facets of match performance: tactical, technical, physical & psychological) in order to gain deeper insights into overall soccer performance.

Moreover, the weighting scheme used for the composite optimization metric (α, β, γ; see Statistics in Methods section) was selected to best align with our study objectives, without conducting a systematic sensitivity analysis. Future methodological work could therefore explore alternative weighting configurations.

Finally, the features derived using rule-based approaches (e.g. deep runs) or machine learning models (e.g. expected pass reception in feature close pass options) used in this study contain a certain degree of error (e.g. accuracy of deep run identification of 84.5%). This may result in the possibility of incorrect labels within the input data which has direct effects on the performance and inner mechanisms of the developed models. This limitation needs to be noted when interpreting the results of this study.

Conclusion

This study presented a comprehensive big data analysis of over 600 elite-level soccer matches of two seasons of German Bundesliga, leveraging machine learning models (xBG & EPV) built upon match analysis expert knowledge and hand-crafted feature engineering. With the analysis of xBG and EPV during transitions valuable and practice-oriented insights into the tactical match performance of teams and players were gained. By simultaneously quantifying both the *risk* and *reward* of every match situation an in-depth assessment of player and team performance for the offensive and defensive team is enabled. As a result, these findings assist in measuring tactical strategy trade-offs, for instance in counter-pressing in defensive transition (vs. falling back) or counter-attacking in offensive transition (vs. saving the ball). Taking this analysis further to practical use cases, such as the individual player analysis presented, showcases the practical usability of the presented approach for advancing in the analysis of tactical decision-making in soccer.

ORCID iDs

Leander Forcher (iD) https://orcid.org/0000-0002-6428-8643
Leon Forcher (iD) https://orcid.org/0000-0003-4291-8787
Stefan Altmann (iD) https://orcid.org/0000-0001-5009-9545

Ethical considerations

This study was conducted according to the guidelines of the Declaration of Helsinki and approved by the local ethics committee (Human and Business Sciences Institute, Saarland University, Germany, identification number: 22–02, 10 January 2022).

Declaration of conflicting interests

The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Data availability statement

The used data is property of the German Football League (Deutsche Fußball Liga, DFL) and is not publicly available. The authors do not have permission to share the data publicly. This work can be reproduced using similar data from professional soccer (e.g. tracking and event data of other soccer leagues).

References

1. Hackinger J. Ignoring millions of euros: transfer fees and sunk costs in professional football. *J Econ Psychol* 2019; 75: 1–19.
2. Morgulev E and Lebed F. Beyond key performance indicators – theoretical-methodological discussion of performance analysis (sports analytics) research. *Ger J Exerc Sport Res* 2024; 54: 335–340.
3. Forcher L. *The beautiful game unveiled – the influence of tactical factors on soccer match performance.* PhD Thesis, Karlsruher Institute of Technology (KIT), 2024. Epub ahead of print 2024. DOI: 10.5445/IR/1000169718.
4. Forcher L, Forcher L and Altmann S. How soccer coaches can use data to better develop their players and be more successful. In: P Düking and B Sperlich (eds) *Individualizing*

training procedures with wearable technology. Cham: Springer International Publishing, 2024, pp.99–123.

5. Goes F, Brink MS, Elferink-Gemser MT, et al. The tactics of successful attacks in professional association football: large-scale spatiotemporal analysis of dynamic subgroups using position tracking data. *J Sports Sci* 2021; 39: 523–532.

6. Fernandez-Navarro J. *Analysis of styles of play in soccer and their effectiveness*. PhD Thesis, Universidad de Granada, 2018. https://digibug.ugr.es/bitstream/handle/10481/54554/61550.pdf?sequence=4.

7. Forcher L. *Success factors in soccer defense – match analysis in soccer based on positional tracking data*. PhD Thesis, Karlsruher Institute of Technology (KIT), 2024. Epub ahead of print 2024. DOI: 10.5445/IR/1000173445.

8. Memmert D, Lemmink KAPM and Sampaio J. Current approaches to tactical performance analyses in soccer using position data. *Sports Med* 2017; 47: 1–10.

9. Andrienko G, Andrienko N, Budziak G, et al. Visual analysis of pressure in football. *Data Min Knowl Discov* 2017; 31: 1793–1839.

10. Neuberg L and Thiem S. Risk-taking in contests with heterogeneous players and intermediate information—evidence from handball. *J Sports Econ* 2022; 23: 851–880.

11. Goes F, Schwarz E, Elferink-Gemser M, et al. A risk-reward assessment of passing decisions: comparison between positional roles using tracking data from professional men's soccer. *Sci Med Footb* 2021; 3: 1–9.

12. Power P, Ruiz H, Wei X, et al. Not all passes are created equal: objectively measuring the risk and reward of passes in soccer from tracking data. In: Proceedings of the 23rd ACM SIGKDD international conference on knowledge discovery and data mining, New York, NY, USA, pp.1605–1613: Association for Computing Machinery.

13. Bauer P and Anzer G. Data-driven detection of counterpressing in professional football. *Data Min Knowl Discov* 2021; 35: 2009–2049.

14. Forcher L, Forcher L, Altmann S, et al. The success factors of rest defense in soccer – a mixed-methods approach of expert interviews, tracking data, and machine learning. *J Sports Sci Med* 2023; 22: 707–725.

15. Tenga A, Holme I, Ronglan LT, et al. Effect of playing tactics on goal scoring in Norwegian professional soccer. *J Sports Sci* 2010; 28: 237–244.

16. Anzer G. *Large scale analysis of offensive performance in football - using synchronized positional and event data to quantify offensive actions, tactics, and strategies*. PhD Thesis, Eberhard Karls Universität Tübingen, 2022. https://publikationen.uni-tuebingen.de/xmlui/handle/10900/124678.

17. Bauer P. *Automated detection of complex tactical patterns in football—using machine learning techniques to identify tactical behavior*. PhD Thesis, Eberhard Karls Universität Tübingen, 2022, https://publikationen.uni-tuebingen.de/xmlui/handle/10900/124679.

18. Goes F. *Tactical behaviour in professional soccer - the secret of successful attacks*. PhD Thesis, University of Groningen, 2023.

19. Goes F, Meerhoff LA, Bueno MJO, et al. Unlocking the potential of big data to support tactical performance analysis in professional soccer: a systematic review. *Eur J Sport Sci* 2021; 21: 481–496.

20. Memmert D, Raabe D, Knyazev A, et al. Big Data im Profi-Fußball. *Leistungssport*. 2016: 1–13.

21. Coutts AJ. Evolution of football match analysis research. *J Sports Sci* 2014; 32: 1829–1830.

22. Rein R and Memmert D. Big data and tactical analysis in elite soccer: future challenges and opportunities for sports science. *SpringerPlus* 2016; 5: 1–13.

23. Anzer G and 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: 1–15.

24. Mead J, O'Hare A and McMenemy P. Expected goals in football: improving model performance and demonstrating value. *PLoS One* 2023; 18: 1–29.

25. Singh K. Introducing expected threat (xT), https://karun.in/blog/expected-threat.html (2018).

26. Decroos T, Bransen L, Van Haaren J, et al. Actions speak louder than goals: valuing player actions in soccer. In: Proceedings of the 25th ACM SIGKDD international conference on knowledge discovery & data mining, Anchorage, AK, USA, pp.1851–1861: ACM.

27. Decroos T and Davis J. Valuing on-the-ball actions in soccer: a critical comparison of XT and VAEP. In: Proceedings of the AAAI-20 workshop on artificial intelligence in team sports, pp.1–8: AI in Team Sports Organising Committee.

28. Olthof S and Davis J. Perspectives on data analytics for gaining a competitive advantage in football: computational approaches to tactics. *Sci Med Footb* 2025: 1–13.

29. Fernández J, Bornn L and Cervone D. Decomposing the immeasurable sport: a deep learning expected possession value framework for soccer. In: 13th MIT Sloan sports analytics conference, 2019. http://www.lukebornn.com/papers/fernandez_sloan_2019.pdf.

30. Fernández J, Bornn L and Cervone D. A framework for the fine-grained evaluation of the instantaneous expected value of soccer possessions. *Mach Learn* 2021; 110: 1389–1427.

31. Rahimian P, Van Haaren J and Toka L. Towards maximizing expected possession outcome in soccer. *Int J Sports Sci Coach* 2024; 19: 230–244.

32. Toda K, Teranishi M, Kushiro K, et al. Evaluation of soccer team defense based on prediction models of ball recovery and being attacked. *PLoS One* 2021; 17: 1–12.

33. Forcher L, Beckmann T, Wohak O, et al. Prediction of defensive success in elite soccer using machine learning – tactical analysis of defensive play using tracking data and explainable AI. *Sci Med Footb* 2024; 8: 317–332.

34. Forcher L, Forcher L, Altmann S, et al. The keys of pressing to gain the ball – characteristics of defensive pressure in elite soccer using tracking data. *Sci Med Footb* 2024; 8: 161–169.

35. Lee M, Jo G, Hong M, et al. *Express: contextual valuation of individual players within pressing situations in soccer*. Boston: United States of America, 2025, pp.1–16.

36. Casal CA, Andujar MÁ, Losada JL, et al. Identification of defensive performance factors in the 2010 FIFA World Cup South Africa. *Sports MDPI* 2016; 4: 54–65.

37. Vogelbein M, Nopp S and Hökelmann A. Defensive transition in soccer – are prompt possession regains a measure of success? A quantitative analysis of German Fußball-Bundesliga 2010/2011. *J Sports Sci* 2014; 32: 1076–1083.

38. Sahasrabudhe A and Bekkers J. *A Graph Neural Network deep-dive into successful counterattacks*. Boston: United States of America, 2024, pp.1–15.

39. Forcher L, Forcher L, Altmann S, et al. Is a compact organization important for defensive success in elite soccer? – Analysis based on player tracking data. *Int J Sports Sci Coach* 2024; 19: 757–768.

40. Forcher L, Preine L, Forcher L, et al. Shedding some light on in-game formation changes in the German Bundesliga: frequency, contextual factors, and differences between offensive and defensive formations. *Int J Sports Sci Coach* 2023; 18: 2051–2060.

41. Linke D, Link D and Lames M. Football-specific validity of TRACAB's optical video tracking systems. *PLoS One* 2020; 15: 1–17.

42. DFL. Definitionskatalog Offizielle Spieldaten (Definitions for Official Gama Data). *DFL: Definitionskatalog Offizielle Spieldaten*, https://www.bundesliga.com/de/bundesliga/news/dfl-definitionskatalog-offizielle-spieldaten_0000265687.jsp (2014, accessed 5 December 2022).

43. Costa IT, da Silva JMG, Greco PJ, et al. Tactical principles of soccer: concepts and application. *Rev Mot* 2009; 15: 657–668.

44. Sumpter D and Andrzejewski A. Soccermatics, https://soccermatics.readthedocs.io/en/latest/ (2022).

45. Anzer G, Bauer P, Brefeld U, et al. Detection of tactical patterns using semi-supervised graph neural networks. In: MIT Sloan sports analytics conference, Boston, pp.1–15: United States of America.

46. Anzer G and Bauer P. Expected passes: determining the difficulty of a pass in football (soccer) using spatio-temporal data. *Data Min Knowl Discov* 2022; 36: 295–317.

47. Marcílio WE and Eler DM. From explanations to feature selection: assessing SHAP values as feature selection mechanism. In: 33rd conference on graphics, patterns and images (SIBGRAPI), 2020, pp.340–347: IEEE.

48. Cavus M and Biecek P. Explainable expected goal models for performance analysis in football analytics. In: 2022 IEEE 9th international conference on data science and advanced analytics (DSAA), pp.1–9: IEEE.

49. Robberechts P, Davis J. How data availability affects the ability to learn good xG models. In: U Brefeld, J Davis, J Van Haaren, et al. (eds) *Machine learning and data mining for sports analytics*. Cham: Springer International Publishing, 2020, pp.17–27.

50. Klemp M and Rein R. Comment on: "a machine learning approach to concussion risk estimation among players exhibiting visible signs in professional hockey". *Sports Med* 2025; 55: 2039–2040.

51. Lepschy H, Woll A and Waesche H. Success factors in the FIFA 2018 World Cup in Russia and FIFA 2014 World Cup in Brazil. *Front Psychol* 2021; 12: 1–9.

52. Spearman W, Basye A, Dick G, et al. Physics-based modeling of pass probabilities in soccer. In: MIT Sloan sports analytics conference, Boston, 2017: United States of America.

53. Herold M, Hecksteden A, Radke D, et al. Off-ball behavior in association football: a data-driven model to measure changes in defensive pressure. *J Sport Sci* 2022; 40: 1412–1425.

Appendix I

Feature computation for expected pass reception (xPR, see methods)

(i) & (ii) Defensive pressure. Defensive pressure was quantified based on the approaches to quantify spatial pressure of.[9,53] Thereby, an oval pressure area is formed around the attacker which is oriented towards the center of the opposing goal. Based on the distance and the orientation of a defender towards the pressurized attacker the amount of pressure that is exerted on an attacker is quantified. The pressure increases with a decreased distance and a positioning between the attacker and the defendants' goal.

(1) $L = D_{back} + (D_{front} - D_{back})(z^3 + 0.3z)\,/1.3$
(2) $z = (\cos \alpha + 1)\,/2$
(3) $D_{front} = (GoalDis * 0.02) + 6.9$
(4) $D_{back} = D_{front}\,/3$
(5) $q = 1.75$
(6) $Pr = (1 - d/L)^q \times 100\%$

$L \triangleq$ Shape of the pressure area
$\alpha \triangleq$ Angle between defender and attacker
$D_{front} \triangleq$ Length of pressure area in front of attacker (in threat direction)
$D_{back} \triangleq$ Length of pressure area behind attacker (in opposite direction to threat direction)
$GoalDis \triangleq$ Distance of the player in ball possession to the goal, if attacker enters penalty box (GoalDis = GoalDis/2) to further decrease the pressure area)
$q \triangleq$ Weight of pressure increase from 0 to 100 [%] inside the pressure area
$d \triangleq$ Distance from defender to attacker
$Pr \triangleq$ Pressure value in [%]

(iii) Coverage of passing lane. The coverage of a passing lane was quantified using a triangular pass coverage area along the passing lane of a given pass. The coverage is increased the closer a defender is positioned towards the passing lane.

(1) $s = 0.5*(a + b + c)$
(2) $h = 2/s * \sqrt{s*(s-a)*(s-b)*(s-c)}$
(3) $d = \sqrt{a^2 - h^2}$
(4) $h\_max = d*0.5$

(5) q = 1.75
(6) CPL = (1 − h/h_max)^q *100

h $\triangleq$ Shortest distance of defender to passing lane (interception point)
   d $\triangleq$ Distance of interception point to passer
   h_max $\triangleq$ Size of CPL area

q $\triangleq$ Weight of CPL increase from 0 to 100 [%] inside the CPL area
   CPL $\triangleq$ CPL value in [%]

(iv) pass distance. Pass distance was computed based on the x and y positions of the passer and the receiver in the moment of a match situation.

Appendix 2
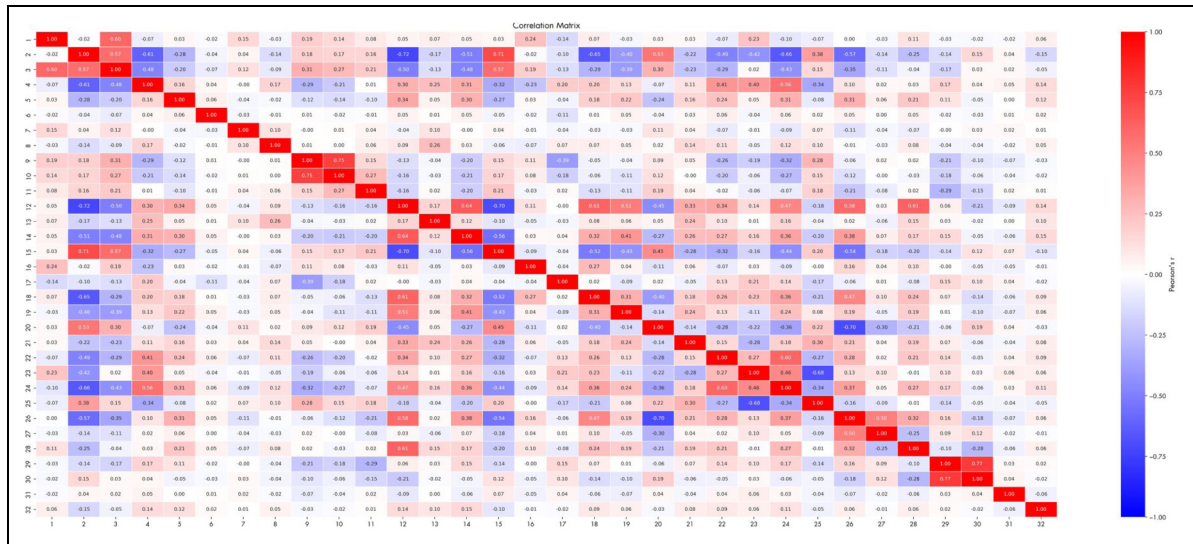
Figure A2. Correlation matrix of all features and the success criteria of both machine learning models (xBG & EPV).

Table A2. Decoding of feature names in Figure A2.

1	absolute_pitch_position_lane
2	absolute_pitch_position_distance_to_goal
3	absolute_pitch_position_angle_to_goal
4	relative_pitch_position_horizontal_zones
5	match_situation_possession_duration
6	offense_onball_ball_control_distance_to_ball
7	offense_onball_ball_control_speed_ball
8	offense_onball_ball_control_speed_ballleader
9	offense_offball_pass_options_close_to_ball_5_mean
10	offense_offball_pass_options_close_to_ball_5_higher50
11	offense_offball_pass_options_dangerous_2_higher50
12	offense_offball_space_control_final_third_offense_percentage
13	offense_offball_deep_runs
14	offense_offball_box_occupation_offense_total
15	offense_offball_box_occupation_numerical_superiority_offense
16	defense_offball_defenders_in_line_of_shot
17	defense_offball_defensive_pressure_on_ballleader
18	defense_offball_goalkeeper_in_goal
19	defense_offball_block_efficiency
20	defense_offball_dis_def_mid
21	defense_offball_dis_mid_att
22	defense_press_numerical_superiority_def_10m
23	defense_press_number_defenders_20m
24	defense_press_numerical_superiority_def_20m
25	defense_press_surface_area_5def
26	defense_restdef_number_defenders
27	defense_restdef_numerical_superiority_def
28	defense_restdef_space_control_att_total
29	defense_restdef_pressure_on_attackers_mean
30	defense_restdef_pressure_on_attackers_min
31	xbg_success
32	epv_success