Science and Medicine in Football

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/rsmf20

Prediction of defensive success in elite soccer using machine learning - Tactical analysis of defensive play using tracking data and explainable AI

Leander Forcher, Tobias Beckmann, Oliver Wohak, Christian Romeike, Ferdinand Graf & Stefan Altmann

SCIENCE AND MEDICINE IN FOOTBALL
https://doi.org/10.1080/24733938.2023.2239766

Prediction of defensive success in elite soccer using machine learning - Tactical analysis of defensive play using tracking data and explainable AI

Leander Forcher[a,b], Tobias Beckmann[c], Oliver Wohak[c], Christian Romeike[c], Ferdinand Graf[c] and Stefan Altmann[a,d]

[a]Institute of Sports and Sports Science, Karlsruhe Institute of Technology, Karlsruhe, Germany; [b]Match Analysis, TSG 1899 Hoffenheim, Zuzenhausen, Germany; [c]D-Fine GmbH, Frankfurt am Main, Germany; [d]Sport Physiology, TSG ResearchLab gGmbh, Zuzenhausen, Germany

ABSTRACT

The interest in sports performance analysis is rising and tracking data holds high potential for game analysis in team sports due to its accuracy and informative content. Together with machine learning approaches one can obtain deeper and more objective insights into the performance structure. In soccer, the analysis of the defense was neglected in comparison to the offense. Therefore, the aim of this study is to predict ball gains in defense using tracking data to identify tactical variables that drive defensive success. We evaluated tracking data of 153 games of German Bundesliga season 2020/21. With it, we derived player (defensive pressure, distance to the ball, & velocity) and team metrics (inter-line distances, numerical superiority, surface area, & spread) each containing a tactical idea. Afterwards, we trained supervised machine learning classifiers (logistic regression, XGBoost, & Random Forest Classifier) to predict successful (ball gain) vs. unsuccessful defensive plays (no ball gain). The expert-reduction-model (Random Forest Classifier with 16 features) showed the best and satisfying prediction performance (F1-Score (test) = 0.57). Analyzing the most important input features of this model, we are able to identify tactical principles of defensive play that appear to be related to gaining the ball: press the ball leading player, create numerical superiority in areas close to the ball (press short pass options), compact organization of defending team. Those principles are highly interesting for practitioners to gain valuable insights in the tactical behavior of soccer players that may be related to the success of defensive play.

ARTICLE HISTORY
Accepted 18 July 2023

KEYWORDS
Football; team sports; Key performance indicators (KPI); tactics; performance analysis

Introduction

The interest in game analysis in team sports, evaluating the physical, technical, and tactical game performance of players, has increased over the last decades (Sarmento et al. 2018). Especially in soccer, the growing availability of collected data (e.g. tracking data) has initiated major developments in game performance analysis (Memmert and Raabe 2019).

Especially tracking data, based on the evolution of tracking systems (e.g. multi-camera systems) and their increasing validity and accuracy, holds high potential for game analysis (Goes et al. 2021). Tracking data comprises the exact positions of all players on the pitch and the ball. This information enables analysts to evaluate the tactical behaviour of soccer players in great detail (e.g., interaction within a team or between teams). Therefore, one can obtain deeper and more objective insights into the constraints of tactical performance compared to traditional analyzing approaches (e.g., notational analysis).

In the past, the focus of media, the public, and research in soccer was predominantly on offensive play. With it, the ball possessing team was extensively analyzed examining on-ball actions such as shooting (Rathke 2017; González-Ródenas et al. 2020) or passing (Szczepanski and McHale 2016; Forcher et al. 2021). However, soccer is played on a comparably large pitch and has the difficulty of controlling the ball with the feet, which results in relatively few goals per match. Consequently, soccer is seen as a primarily defensive team sport. Furthermore, previous studies have shown that in soccer defensive play is at least as important for the success of a team as offensive play (Lepschy et al. 2020, 2021).

In this context, tracking data enables one to analyze especially the off-ball behavior of soccer players (Herold et al. 2022) and can help to reveal information about the defending team that is not in the possession of the ball.

Alongside the increasing availability of huge data sets (e.g. full season of tracking data), there has been a similar increase in the usage of computational methods and machine learning in soccer and game analysis. Those methods enable analysts to handle and evaluate huge amounts of data. Furthermore, the use of these technologies makes it possible to identify the crucial variables related to the success of a team and can help to obtain insights into previously hidden patterns of tactical performance.

Therefore, to analyze defensive play, the combination of the use of highly informative tracking data with the use of machine learning methods can help to dive deeper into this specific part of tactical performance analysis.

Although the analysis of defensive play is a comparatively novel topic of performance analysis in soccer, there are first approaches to gain insights into the constraints of defensive performance (Forcher et al. 2022). Some studies started the analysis of defensive play evaluating tracking data by using simple variables (e.g., surface area & spread of a team) (Frencken et al. 2011; Bartlett et al. 2012; Castellano et al. 2013) or analyzing specific components of defensive

CONTACT Leander Forcher leander.forcher@kit.edu Institute of Sports and Sports Science, Karlsruhe Institute of Technology, Karlsruhe, Germany

performance (e.g. defensive pressure) (Andrienko et al. 2017). Subsequently, one approach used different basic variables to develop a defensive ability score that was shown to be higher in successful defensive plays compared to unsuccessful ones (Matsuoka et al. 2020). However, all these approaches used few predetermined variables based on the experience of the scientists or experts involved in the respective study (representing only a small part of the tactics of defensive play) (Rein et al. 2017), mostly used few games as a sample (Forcher et al. 2022), and did not take advantage of machine learning.

Besides, there are three studies that used machine learning in the context of the analysis of defensive play. Bauer and Anzer (2021) detected counter-pressing situations and identified according success drivers using a machine learning approach. Since they focussed on defensive transition (switching behavior of players after a ball loss) the results can only be partially applied to the distinctively different game phase defensive play (Hewitt et al. 2016; Navarro 2018).

Casal et al. (2016) used a logistic regression model to classify successful (ball recovery) and unsuccessful defensive plays (no ball recovery). However, they used a notational approach missing the possibilities of the utilization of tracking data. Toda et al. (2021) also predicted ball gains (as the success criterion of defensive play) using a XGBoost (eXtreme Gradient Boosting) classifier and indicated the importance of all 139 variables included in their model with SHAP values. However, most of the used variables did not represent a tactical idea (e.g., key performance indicators such as defensive pressure) that would enable the practical interpretation of the results. For example, they included variables such as the x-position of the attacker or type of event (e.g. pass or tackling), where resulting findings are not transferable to training or tactical game analysis. Due to the large number of variables and the missing use of practically applicable input variables, the practical impact of this study remains small.

Accordingly, to this date, there is no study that investigated defensive play using the possibilities of tracking data and evaluation methods of machine learning in a practically appropriate way. To be able to transfer the results of sophisticated evaluations of tracking data using machine learning into practice, the following points have to be considered: 1. The variables used in the machine learning model should contain a tactical idea to allow interpretations in a soccer game analysis way to enable coaches or practitioners to understand the results and to apply the results into training and game analysis (e.g. key performance indicators); 2. The machine learning model should be based on the tactical success of a game phase (in this case defensive play) to be able to interpret the results accordingly; 3. The contribution of all included variables to the machine learning model should be indicated to identify the potentially most important drivers for the success of a team.

By considering all these points, the gained information is thought to be highly important for practitioners and coaches. Using this information about the most important variables of defensive play, they are able to evaluate players, improve the defensive performance, or identify defensive strengths and weaknesses of players and accordingly improve their personal decisions (e.g., selection of starting formation, or squad assembling).

Accordingly, the aim of this study is to identify the tactical variables (each representing a tactical idea) that drive the success of defensive play in soccer. Therefore, we aim to predict ball gains in defense as a successful result of a defensive play using a large sample size of tracking data.

Methods

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

All developed code for the data analysis procedures presented below are available in a public repository on github which can be accessed following the provided link: https://github.com/LForcher/kit_d-fine_sports-analytics. This ensures the comprehensibility of this study and enables further projects based on this work.

Data

We used tracking data of 153 games of the second half of the German Bundesliga season 2020/21. The tracking data was measured with a semi-automatic optical tracking system (TRACAB, ChyronHego, Melville, NY, USA) that was previously proven to be a valid tool for the examination of sports performances (Linke et al. 2020).

Data processing

Prior to the calculation of key performance indicators and the modeling, there were several steps of data preprocessing. First, to reduce the amount of data we downsampled the tracking data from 25 Hz to 5 Hz in accordance with previous studies (Goes et al. 2021). Second, we solely considered possessions where a team had continuous control over the ball for a minimum of 5 sec. This procedure was chosen to exclude possessions with undeliberate short actions (e.g., just 1 pass, unforced errors, or deflections). Additionally, this investigation focuses on the tactics of defending in 11vs11 game situations. Therefore, all possessions where no regular 11vs11 game situation could be detected were excluded (e.g., red cards) as several studies have shown that player dismissals influence match performance and tactics (Lago-Peñas et al. 2016; Badiella et al. 2022).

KPIs of defensive (tactical) play

To quantify the tactical behavior of soccer players we calculated different key performance indicators (KPI) of defensive play. Each metric was critically reviewed and included in this study only if it represents a tactical idea (Claro 2010). This allows to interpret the results in a practically applicable way and accordingly transfer the results into practice. The KPIs were grouped into player metrics, evaluating the behavior of single players (individual tactics), and team metrics, comprising the behavior of a group of players or the whole team (group & team tactics). Additionally, we considered the contextual factors of

Table 1. Key performance indicators & contextual factors with associated description, calculation, and explanation of their tactical idea.

	Key Performance Indicators (KPIs)	Description	Calculation	Tactical idea
1	**Player metrics**			
1.1	defensive pressure	- On every attacker	- To quantify the defensive pressure on an attacking player an elliptical pressure area was shaped around the attacker (Andrienko et al. 2017) - This pressure area is oriented toward the opposing goal with a length of 9 m in front of the attacker and 3 m behind him - The pressure increases the closer a defender gets to the attacker inside this pressure area	- Defensive pressure represents the defensive principle of play to exert spatial and temporal pressure on a direct opposing attacker to deny his actions (e.g., passes or dribblings) and increase the chance of an opposing error to regain the ball (e.g., unsuccessful pass)
1.2	distance to the ball	- For every defender and attacker		- The distance of defenders to the ball describes a simple measure to quantify the spatial pressure on the ball carrier considering more than one defender - The distance of the closest attacker to the ball represents a measure of ball control: A large distance indicates little to no ball control (e.g., during a high ball) and a small distance indicates great control over the ball (e.g., dribbling)
1.3	velocity	- For every defender and attacker		- Different tactical ideas can be argued: - On the one hand, a higher velocity of attackers could suggest less control over the ball and thus have a positive impact on ball gains - On the other hand, a smaller velocity of attackers could suggest a calmed game situation where there is few disruption in the defensive organization with few opportunities to outplay the defense resulting in a greater chance of regaining the ball - The impact of the defenders' velocity on the probability to gain the ball is unclear and therefore no interpretation is given at this stage
2	**Team (& group) metrics**			
2.1	surface area	- Of the defending & the attacking team (each excluding the goalkeeper)	- Calculated by the smallest convex hull of all outfield players (Moura et al. 2012) (see Figure 1)	- The principle of play *defensive compact organization* of the defending team quantified with measures for compactness (surface area & spread) and measures of organization of the defending team (distances between formation lines) - This principle represents the idea that a defending team tries to increase their compactness and organization during defensive play to decrease the possibilities of opposing passes through this compact organization - This is especially true for deep defending close to the own goal - Furthermore, possible actions of attackers inside this compact organization can be denied effectively because pressure can be exerted faster due to the spatial proximity of several defenders
2.2	spread	- Of the defending team & the attacking team (each excluding the goalkeeper)	- Calculated by the sum of the squared deviation from a team's centroid of all outfield players (Bourbousson, Sève, and McGarry 2010; Bartlett et al. 2012) (see Figure 1)	
2.3	distances between the formation lines (defensive line, midfielder line, attacking line)	- Of the defending team	- We developed a formation descriptor using the mean position of players over one half - Those mean positions were clustered into three formation lines (defenders, midfielders, & attackers) utilizing a KMeans unsupervised clustering algorithm - Furthermore, we differentiated between offensive and defensive playing sequences as differences in the offensive and defensive formation were expected resulting in one offensive and one defensive formation per half and team (see Figure 1)	
2.4	numerical superiority (static pitch zones)	- In the defending half & - In the final third of the defending team	- Calculated by the difference between attacking and defending players in the respective pitch zones (see Figure 1)	- The principle of *numerical superiority* represents the idea that a defending team attempts to create numerical superiority in specific pitch areas to increase the local spatial pressure (e.g. areas close to the ball) - The defending team especially presses attacking players in those areas by closing down passing options to eventually regain the ball
2.5	numerical superiority (dynamic pitch zones)	- Around the ball (10 meters & 20 meters around the ball) & - In front of the defending line (15 meters in front of the last defender)	- Calculated by the difference between attacking and defending players in the respective pitch zones (see Figure 1)	

(*Continued*)

Table 1. (Continued).

3	**Contextual factors**			
3.1	pitch zone	- 3 vertical lanes (left, middle, right lane) - 3 horizontal thirds (first, middle, final third)	- Using the pitch position of the ball - Using the pitch position of the ball	- The pitch zone of a game situation is decisive for the tactics in soccer, as different tactical behaviors are trained and implemented in a game depending on the pitch position - In the German Bundesliga, for example, RB Leipzig is known to play high pressing while FSV Mainz 05 often plays with a deep defending playing style - A high press (in the last third in the opposing half) has differences in the tactical actions compared to a deep defending (inside the own first third) (Low, Rein, Raabe, et al. 2021) and defending tactics differ whether the opponent is playing on the outer lane or is controlling the ball in the center of the pitch
3.2	quality of teams	- 3 groups of team strength	- Dividing the final table of the considered season into 3 groups of team strength (best, middle, worst teams)	- There are differences in the defending behavior of teams with different quality - Therefore, differences in the importance of different tactical features which ultimately have an impact on the probability to gain the ball are expected

the pitch zone of the game situation and the quality of teams. All variables and their description are depicted in Table 1. Furthermore, the key performance indicators of defensive play used as variables are illustrated in Figure 1.

Modelling

To analyze the defensive success, we used supervised machine learning classifiers (logistic regression, XGBoost, & random forest classifier) to predict successful vs. unsuccessful defense (binary prediction). A successful defense was defined as game situation in which a ball gain (direct possession change) took place in the following 2 s of the game. All other game situations (where there was no direct possession change in the following 2 sec) were defined as unsuccessful defense (e.g. ball out of play). Accordingly, our classifier has to predict if the defending team is able to gain ball possession in the following 2 sec of a given game situation.

Feature engineering

We enhanced our data based on the (player & team) metrics explained above to gain extra spatial, temporal, and context information about a game situation (feature engineering). We considered the last 5 sec of a possession and raised the values at the time instances 5, 3, 1 s before the identified game situation and 0 (current game situation). Accordingly, to gain temporal information, we calculated the metrics at each time instance (*attribute*_raw_*time*), the mean values across the time intervals (e.g. average from 5 to 0 sec, *attribute*_mean_*timeinterval*), as well as the time-dependent delta values (e.g., change from 5 to 0 sec, *attribute*_delta_*timeinterval*) for all time instances for every metric.

To gain further spatial information, we generated additional features for the player metrics (see above). Next to the values about the closest player to the ball (attacker & defender), we measured the mean values of the closest 3 and 5 attackers and defenders to the ball for every player metric.

All possible time and spatial feature combinations summed up to 334 variables and are exemplarily depicted in Table 2.

Model design

We evaluated different classifiers and different models (comprising different features) using a train (60%), validation (20%), test (20%) split and used F1-Score to optimize the performance of the models. This procedure was chosen, because F1-Score is a sensitive measure for model optimization that combines precision and recall which is important in our unbalanced data set (unsuccessful 5:1 successful defensive plays).

We tested several classifiers (logistic regression, XGBoost, & random forest) to best solve our problem. All classifiers were optimized using HalvingGridSearchCV, running 200 trials for each optimization.

In the following, we focus on random forest classifiers, since they substantially outperformed logistic regression and slightly outperformed XGBoost on our data (based on F1-score). A comparison of the prediction performance between the different classifiers can be found in the appendix (see appendix 1).

To optimize the selection of features for our model, automatic feature selection for instance, forward or backward selection, did not yield satisfactory results. Therefore, we deployed 6 different models. First, we implemented a baseline model implementing all calculated features ('all-features-model'). Second, the 'minimal-players-model' was realized, which solely uses the information of the closest defender and closest attacker for the player metrics (excluding player metric features of more than 1 player). Third, the 'minimal-time-model' was calculated, which solely includes information at the last time instance of a prediction (current game situation at 0, excluding all previous timepoints of 1, 3, 5 s before the game situation). Fourth, a minimal-model was computed, combining the reduction of minimal-players-model and minimal-time-model. Fifth, we created an 'expert-model' consisting of features described in the

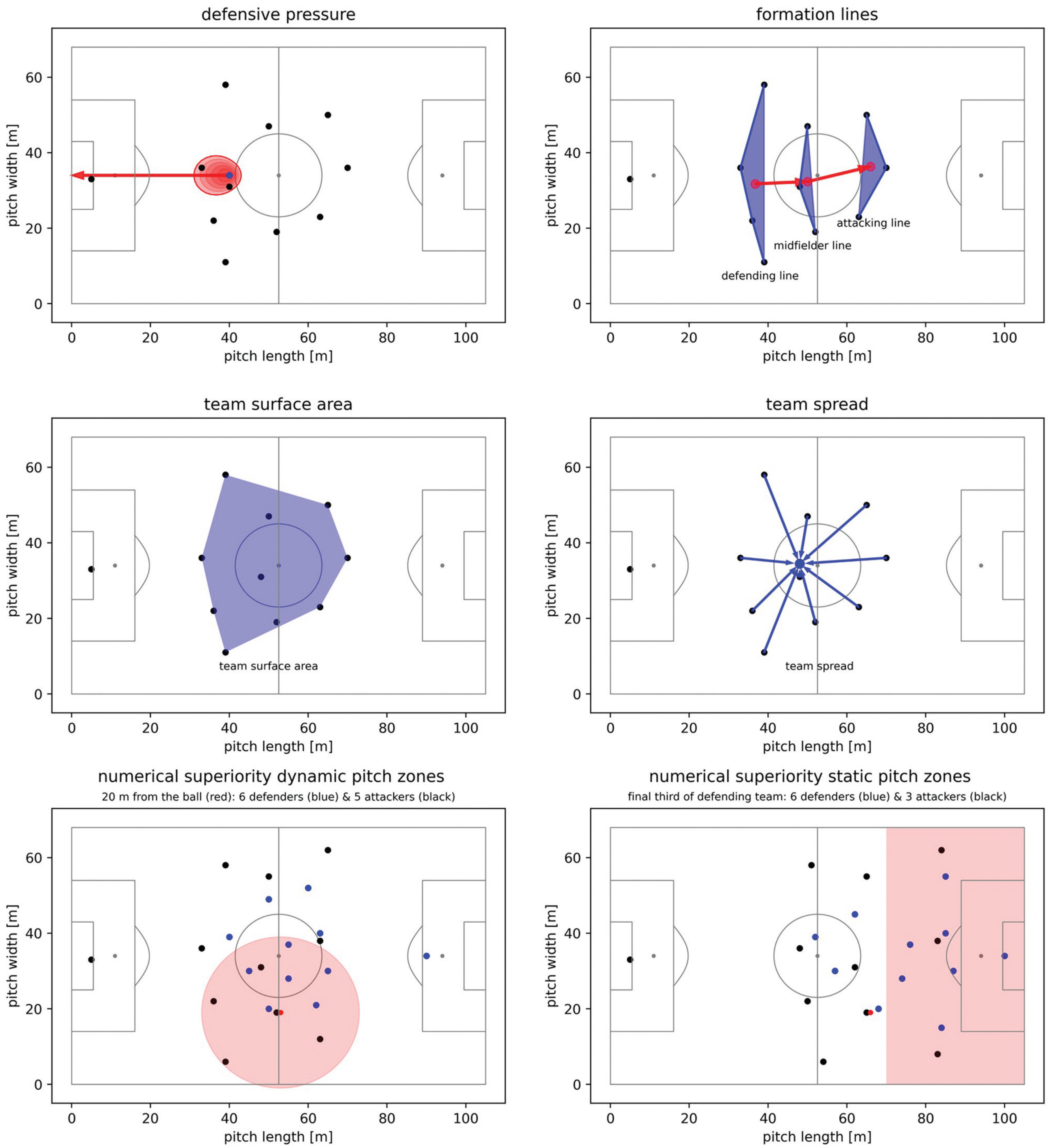

Figure 1. Visualization of specific key performance indicators of defensive play: Defensive pressure (top left), distances between the formation lines (top right), team surface area (middle left), team spread (middle right), numerical superiority in dynamic pitch zone (e.g. 20 m from the ball) (bottom left), numerical superiority in static pitch zone (e.g. final third of defending team) (bottom right).

appendix, which tries to enhance the minimal-model with as few information as necessary.

For the sixth and final model (chosen based on the F1-Score), which will be emphasized in detail in the result section, the feature importances of included features and the SHAP values were calculated.

To increase the transparency and reproducibility of our modeling approach and to identify potential data leakage which would limit the scientific claims made based on the results of the prediction model (Kapoor and Narayanan 2022; Gibney 2022), we completed the model information sheet template provided by Kapoor and Narayanan (2022). This template can be used for detecting and preventing data leakage in ML-based Science. The completed information sheet showed no evidence of possible data leakage and can be found in the appendix (see appendix 2).

The developed code was executed on a × 64 desktop PC using 24 GB RAM with a total execution time of 12 h. In detail,

Table 2. Table of all possible feature combinations for one metric as a result of the feature engineering process (combination of time information & spatial information).

Feature combinations				
		spatial information		
time instance [sec]	time information	1 player	3 players	5 players
0	current	metric raw 0 sec 1 attackers	metric raw 0 sec 3 attackers	metric raw 0 sec 5 attackers
0	current	metric raw 0 sec 1 defenders	metric raw 0 sec 3 defenders	metric raw 0 sec 5 defenders
1	current	metric raw 1 sec 1 attackers	metric raw 1 sec 3 attackers	metric raw 1 sec 5 attackers
1	current	metric raw 1 sec 1 defenders	metric raw 1 sec 3 defenders	metric raw 1 sec 5 defenders
1	mean	metric mean 1 sec 1 attackers	metric mean 1 sec 3 attackers	metric mean 1 sec 5 attackers
1	mean	metric mean 1 sec 1 defenders	metric mean 1 sec 3 defenders	metric mean 1 sec 5 defenders
1	delta	metric delta 1 sec 1 attackers	metric delta 1 sec 3 attackers	metric delta 1 sec 5 attackers
1	delta	metric delta 1 sec 1 defenders	metric delta 1 sec 3 defenders	metric delta 1 sec 5 defenders
3	current	metric raw 3 sec 1 attackers	metric raw 3 sec 3 attackers	metric raw 3 sec 5 attackers
3	current	metric raw 3 sec 1 defenders	metric raw 3 sec 3 defenders	metric raw 3 sec 5 defenders
3	mean	metric mean 3 sec 1 attackers	metric mean 3 sec 3 attackers	metric mean 3 sec 5 attackers
3	mean	metric mean 3 sec 1 defenders	metric mean 3 sec 3 defenders	metric mean 3 sec 5 defenders
3	delta	metric delta 3 sec 1 attackers	metric delta 3 sec 3 attackers	metric delta 3 sec 5 attackers
3	delta	metric delta 3 sec 1 defenders	metric delta 3 sec 3 defenders	metric delta 3 sec 5 defenders
5	current	metric raw 5 sec 1 attackers	metric raw 5 sec 3 attackers	metric raw 5 sec 5 attackers
5	current	metric raw 5 sec 1 defenders	metric raw 5 sec 3 defenders	metric raw 5 sec 5 defenders
5	mean	metric mean 5 sec 1 attackers	metric mean 5 sec 3 attackers	metric mean 5 sec 5 attackers
5	mean	metric mean 5 sec 1 defenders	metric mean 5 sec 3 defenders	metric mean 5 sec 5 defenders
5	delta	metric delta 5 sec 1 attackers	metric delta 5 sec 3 attackers	metric delta 5 sec 5 attackers
5	delta	metric delta 5 sec 1 defenders	metric delta 5 sec 3 defenders	metric delta 5 sec 5 defenders

the computation of KPIs took 2 hours for all matches, the hyperparameter tuning took 1 hour (for expert-reduction-model), and the calculation of shapely values took 8 h (for expert-reduction-model). In the practical use case of one match our approach requires less than 10 min of computation time for an individual match (calculation of metrics: < 10 minutes; prediction: < 1 second).

Results

The considered games resulted in 144,922 instances of games situations after uninterrupted possessions. Of it, 22534 (15.5%) game situations were classified as successful defense, as an active possession change took place in the following 2 sec. 122,388 (84,5%) were classified as unsuccessful.

The results of all computed models are depicted in Table 3. All models showed satisfactory results in the prediction of defensive success (0.56 ≤ F1-Score (test) ≤ 0.58).

The exclusion of earlier time instances in the minimal-time-model (5, 3, 1 sec before the observation) in the prediction of defensive success showed a small decrease in the model performance (F1-Score (test) = 0.56) compared to the all-features-model (F1-Score (test) = 0.58). The exclusion of additional spatial information (for the player metrics beyond the information about the closest player to the ball (3 and 5 defenders closest to the ball)) in the minimal-player-model showed similar performance (F1-Score (test) = 0.58) to the all-features-model. The minimal-model, combining the reduced information of the minimal-time-model and the minimal-player-model showed a slight decrease in model performance (F1-Score (test) = 0.56) (detailed results of all models see Table 3).

According to the gathered information stated above, the expert-model was derived based on the minimal-model with the addition of one time information feature (the average of the last 3 sec for the player metric distance to the ball). This expert-model showed a higher prediction performance (F1-Score (test) = 0.57) compared to the minimal-model. The expert-reduction-model was derived by using the 16 most important uncorrelated features of the expert-model (based on the feature importance). As this model uses the fewest features with a consistent prediction performance (F1-Score (test) = 0.57) this model was selected for further consideration.

The expert-reduction-model comprises a total of 16 features, the feature importances and according SHAP-values are depicted in Figure 2.

While the expert-reduction-model did consider the horizontal pitch zones (lanes) as an important feature (see Figure 2), the context information about the individual team strength and the vertical pitch zones (thirds) were not included as they did not improve the prediction.

Discussion

The aim of this study was to identify tactical variables that drive the success of defensive play in soccer. We predicted ball gains in defense based on a large sample of tracking data of elite soccer (German Bundesliga) returning satisfactory performance results of the trained models (Accuracy ≥0.82, Precision ≥ 0.46, Recall ≥ 0.65, F1-Score ≥0.56). With it, we defined tactical principles of defensive play (e.g. press the ball leading player, create numerical superiority in ball near areas), which seem to be closely connected to defensive success and, therefore, could

Table 3. Results of all models.

Model		Information		Training Set (60%)				Test Set (20%)			
Name	Description	Number of features	Algorithm	Accuracy	Precision	Recall	F1-Score	Accuracy	Precision	Recall	F1-Score
All-features-model	including all features	334	Random Forest	0.96	0.81	0.99	0.89	0.85	0.52	0.65	0.58
Minimal-time-model	including solely features at time instance 0	46	Random Forest	0.92	0.67	0.99	0.8	0.82	0.47	0.71	0.56
Minimal-player-model	including solely features of the closest defender to the ball for all player metrics	234	Random Forest	0.96	0.79	0.99	0.88	0.85	0.52	0.65	0.58
Minimal-model	Combining the reduction of minimal-time- & minimal-player-model	36	Random Forest	0.91	0.64	0.99	0.78	0.82	0.46	0.71	0.56
Expert-model	Minimal-model with addition of one time feature (average 3 seconds of distance to the ball metric)	38	Random Forest	0.92	0.66	0.99	0.79	0.83	0.48	0.71	0.57
Expert-reduction-model	Expert-model including the 16 most important uncorrelated features	16	Random Forest	0.91	0.64	0.99	0.78	0.82	0.47	0.70	0.57

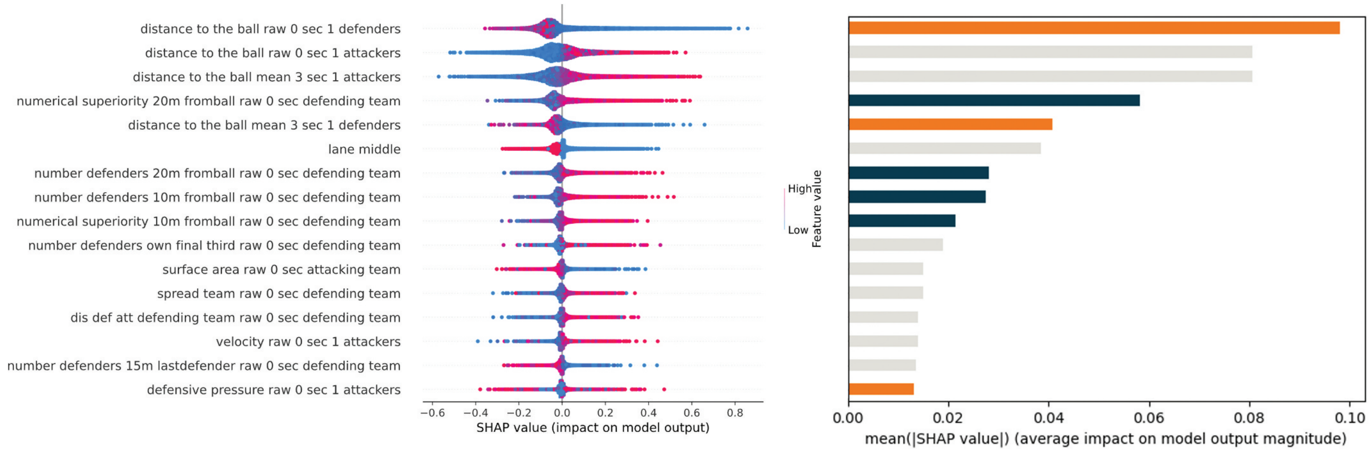

Figure 2. SHAP-values and feature importance of expert-reduction-model (sorted according to their importance for the model). Feature importances of the most important principles of defensive play are highlighted (on the right) with 'press the ball leading player' in orange and 'create numerical superiority in ball near areas' in dark blue. (Metrics that were not included in the expert-reduction-model: 'dis def mid defending team', 'dis mid att defending team', 'number defenders 15 m lastdefender', 'numerical superiority 15 m lastdefender', 'numerical superiority own half', 'numerical superiority own final third', 'number defenders 15 m lastdefender', 'surface area defending team', 'velocity attackers', 'velocity defenders', 'Team strength', 'attacking third', 'spread attacking team')

be applied in practice to possibly enhance defensive success. Furthermore, we identified contextual variables that are associated to defensive play (e.g. the pitch position where the defensive play took place) and some that appear not to have a particular influence (e.g. strength of teams). Overall, the results could help practitioners to objectively measure players performance in the analysis of defensive play. Since this is one of the first studies to analyse defensive success in its entirety, it solely provides first results on defensive play that will need to be confirmed in the future.

Comparison between models

It has been shown that reducing spatial information in the player metrics does not affect the prediction performance (see minimal-player-model). This is primarily due to other team metrics which are still included in the prediction and contain similar information. For instance, the team metric number of *defenders 20 from the ball* covers similar information as the player metric *distance to the ball of the closest 5 defenders*, indicated by the high correlation of 0.92 (Spearman) between those features. Even if the correlation is not that high in all cases, this expansion of player metrics (average values of *3 & 5 defenders closest to the ball*) does not provide additional information and can be excluded in our case. However, in cases where similar team metrics are not captured those metrics might still have to be considered.

In comparison, the exclusion of further time information about the past of the initial observation (1, 3, 5 sec before the observation) slightly decreases the prediction performance (see minimal-time-model). Either way, the correlations between the time instances are high (e.g., the *number defenders 20 m from ball* at 0 sis highly correlated with the value at 1 second (0.81 Spearman) and the value at 3 seconds (0.45 Spearman)). One could conclude that the moment of an analyzed game situation is most important for the tactical analysis of defensive play, no matter how this certain game situation was created in the past. Still, the use of further time information, about the closest defender and attacker to the ball in our case, leads to a more robust model (see expert-model). This emphasizes the significance of the playing situation in near ball areas for defensive success. The additional information about the closest defender and attacker is highly practicable and important for the prediction of defensive success. This importance of areas close to the ball for the success of defensive play is explained in more detail in the identified tactical principles of defensive play below.

Summing up all modelling evidence of this study, we selected the expert-reduction-model for further analysis as it is the model with the fewest variables and satisfactory prediction performance. Accordingly, with the detailed analysis of the expert-reduction-model (e.g., using SHAP values) we are able to define tactical principles of defensive play based on their connection to defensive success.

Principles of defensive play

The identified principles of play are ranked according to the importance of features related to each principle.

Press the ball leading player

The most important feature in the expert-reduction-model is the *distance to the ball* of the closest defender at the instance of the observation (see Figure 2 in orange). Furthermore, the average of 3 seconds of the distance feature (5^{th} most important feature) and *defensive pressure* on the ball-leading player (15^{th} most important feature) did also contribute to the prediction. In this context, a smaller distance of the closest defender to ball (at time instances 0 and average 3 sec) may increase the chance of gaining the ball (see Figure 2). As the feature defensive pressure did not reveal clear results in the SHAP values, one can argue that the simple measure of defensive pressure (solely the distance of the closest defender to the ball) provides clearer results for the model compared to the sophisticated defensive pressure model (based on Andrienko et al. (2017)). Overall, it can be concluded that for defensive success, it may be most

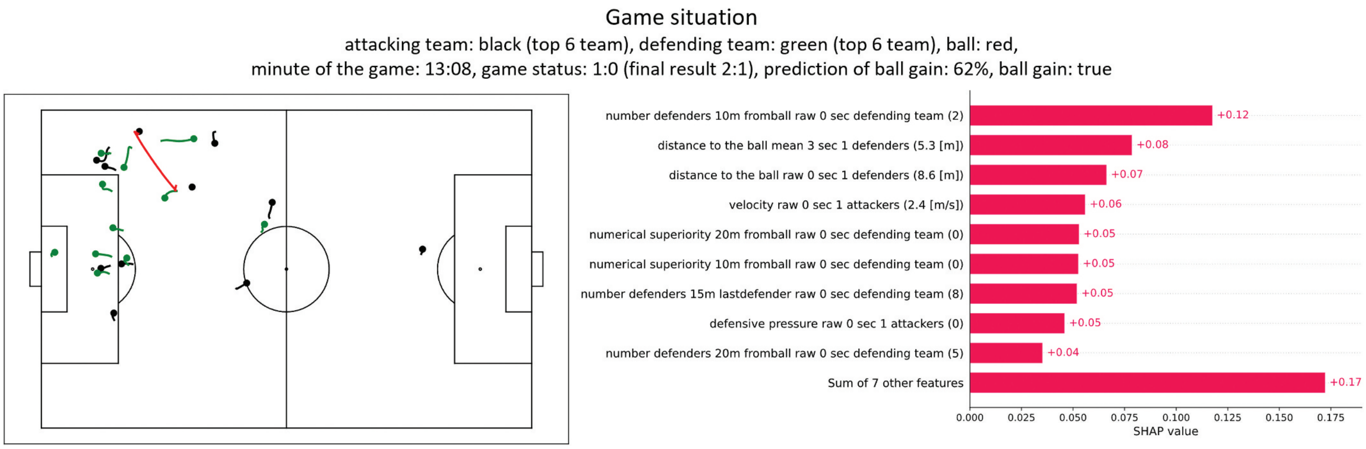

Figure 3. Showcase how this model can be applied in practice and used by game analysts to enable more in-depth game analysis in professional soccer. The ball gain prediction of an exemplary game situation (and the following two seconds = lines) based on the expert-reduction-model is presented with the individual variable characteristics in this game situation, their effect on the prediction, and the resulting probability of gaining the ball.

important to 'pressurize the ball-leading player' to increase the chance of a ball gain.

Create numerical superiority in ball near areas (press close pass options)

With 4 features (*numerical superiority around the ball 20 m* & *10 m*, *number of defenders around the ball 20 m* & *10 m*) ranked in the top 9 most important features of the expert-reduction-model of defensive success (see Table 3 in dark blue) the principle of defensive play 'create numerical superiority around the ball' is the second most important principle of play identified. According to the SHAP values (see Figure 2), a greater numerical superiority of defenders and a greater number of defenders around the ball may increase the chance of a successful defensive play. Those findings are in accordance with previous work of Grehaigne et al. (2002) who analyzed defending behavior that led to ball gains. They found that it seems important for defensive success to be in numerical superiority.

In particular, the findings indicate that numerical superiority of defenders around the ball may be more crucial to gain the ball compared to the sole number of defenders in those areas (numerical superiority features are ranked higher compared to the number of defenders features in the feature importances, see Table 3). Furthermore, it seems to be more important to increase this superiority 20 m around the ball compared to closer areas (10 m).

Overall, those results suggest that the defending team may press areas around the ball to cover possible short pass options for the ball-leading player to possibly increase the chance of a ball gain.

Compact organization of defending team

The *spread* (13th most important feature) and the *distance between the defending- and attacking-line* (12th most important feature) of the defending team can be condensed to the principle of defensive play 'compact organization'. Thereby, a more compact organization of the defending team (smaller spread, smaller distance between defending- & attacking-line) appears to be related to defensive success (see Figure 2). However, the results indicate that in comparison to the compactness of near ball areas (see 'numerical superiority around the ball') this principle of play seems less important. This confirms previous findings that more meaningful information can be gained by the analysis of subgroups compared to team-level analysis (Bartlett et al. 2012; Low et al. 2021). Additionally, the information about the distances between the formation lines (defending-, midfielder-, attacking-line) can be summed up by one feature without losing significant information in our study: the distance between the attacking- and the defending-line excluding the information about the placement of the midfielder-line. Overall, a compact organization seems to enhance the chance of regaining the ball.

Defend the own goal

The last principle of defensive play that can be identified by the analysis of our expert-reduction-model is 'defending the own goal'. The features number of defenders in the own final third (10th most important feature) and number of defenders 15 m in front of the last defender (16th most important feature) are less important to the prediction compared to the features of the other principle of plays. The main goal of these features is to save deep spaces behind or in front of the defensive-line by placing more defenders inside those areas to prevent the opposing team from scoring. Therefore, as the principle of defending the own goal represents the second goal of defending next to the regain of the ball, its importance for the prediction of winning the ball is limited.

Practical application

There are different ways to apply the findings of this study in practice. For instance, the principles of play could be used to objectively measure tactical match performance in defense. Furthermore, soccer game analysts or coaches can use the prediction of defensive success in real match situations to gain deeper insights in the defending tactics of their team to improve their match analyses (e.g., data analyses) or player evaluations. An example of the practical application of this approach is illustrated in Figure 3.

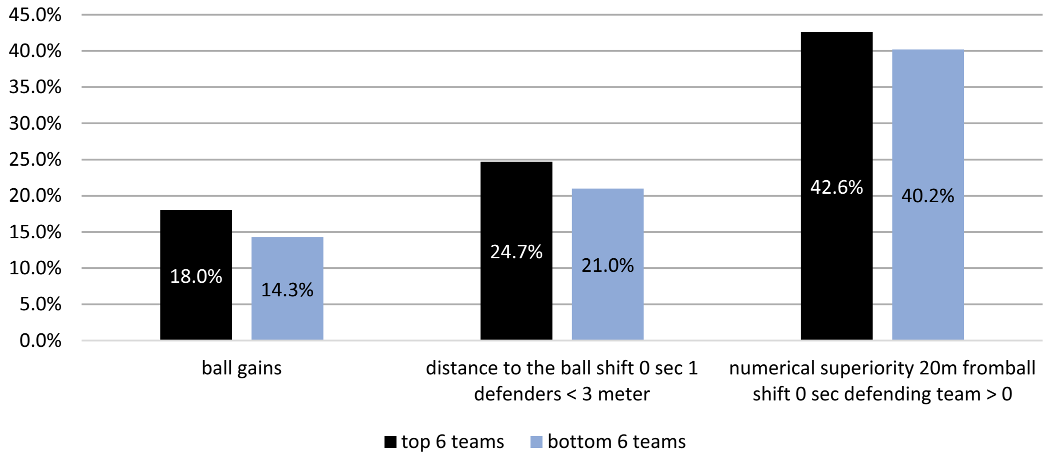

Figure 4. Proportion of ball gains, proportion of the distance of the closest defender to the ball at the last moment of the observation smaller than 3 meters, & proportion of positive defensive numerical superiority 20 meters around the ball to all observations in the data set compared between teams of different quality.

Contextual information

Gaining the ball on the side differs from gaining the ball in the middle

One notable result is that there is solely one contextual variable, which is *middle lane*, that was important enough for the prediction to be included in the final model. There, the information on whether the ball was located on the middle lane or not, had a remarkably high feature importance of 3.7% (ranked 6th most important feature, see Figure 2). This encouraged us to investigate the influence of the lane position of the ball in more detail. We trained two different Random Forest classifiers, one solely based on the data of the middle lane and one solely on the data of the outer lanes. Indeed, the classifier trained on the middle lane proved to be much less efficient on the data of outer lanes and vice versa based on F1-Score (see Table 4) – indicating that there is a significant difference in the manner of gaining the ball on outer vs. middle lane. The information on whether the ball is on the left or right lane seems, however, not important, which was expected due to the fact that plays on both sides of the pitch can usually be mirrored.

The interpretation is somewhat difficult though. Comparing feature importances (see Table 4), we see that on the outer lanes, it seems much more important having more defenders in the own half. Looking at the data, we can indeed see, that on the outer lanes, 30.6% of all situations are located in the own half, whereas in the middle lane, 35.5% of all situations are located in the own half. On the other hand, the model did not choose the variables attacking, middle, or defending third – indicating that distance to the goals is not associated to the manner of ball gains.

Successful defending does not differ between teams of different quality

Another interesting result is that the context information in which third of the final table a team finished did not influence the model. Therefore, the tactics of successful defending appear to not differ between teams of different quality. Nevertheless, we could show that teams finishing in the top third of the final table have significantly more ball gains in all considered situations (see Figure 4). Analyzing the most important features of the prediction (distance to the ball of the closest defender) in detail, the top teams more often created defensive situations where one defender is closer than 3 m to the ball compared to teams with less quality. The same holds for the second most important feature (numerical superiority in the area of 20 m around the ball) as the top teams were able to create numerical superiority around the ball more frequently. In conclusion, while the tactics of successful defending do not seem to differ between teams, top teams appear to score better in the identified success variables of our model by placing their defenders more effectively and thus being more successful in gaining the ball.

Limitations and future directions

A main limitation of our investigation is that we solely analyzed the primary goal of defensive play (gaining the ball) as success criterion without analyzing the second goal of defensive play: defending the own goal. Therefore, the results cannot be generalized to the second success criterion of defensive play. However, this success criterion for defending an own goal is thought to be distinctively different in the resulting tactical behavior and could be a target for future research in defensive play in soccer.

Furthermore, all analyses of this study were done across all teams without regard to defending tactics of specific individual teams. This allows one to make general statements about the defensive play. However, since different teams may have different defensive strategies or patterns, there may be differences in tactical principles that are important to defense in individual cases which should be examined in detail in the future. Additionally, due to this used procedure, we cannot assess the risk and reward of tactical defensive behavior. We can solely analyze tactical behavior according to its value to gain the ball. However, in the future, it could be interesting to assess whether specific tactical behavior that increases the chance of gaining the

Table 4. Prediction Performance of Random Forest Classifiers for middle lane and outer lanes at the top. This features the importance of both models at the bottom. Both models include the same features as the expert-reduction-model excluding the feature of vertical lane position (≙ 15 features).

	middle-lane-model				outer-lanes-model			
	Accuracy	Precision	Recall	F1-Score	Accuracy	Precision	Recall	F1-Score
Model performance								
data middle lane	0.84	0.43	0.68	0.53	0.85	0.42	0.55	0.48
data outer lanes	0.79	0.46	0.41	0.43	0.8	0.5	0.74	0.59

Feature	middle-lane-model	outer-lanes-model	difference in importance
Feature importance			
number defenders own final third raw 0 sec defending team	0.02	0.08	−0.06
numerical superiority 20m fromball raw 0 sec defending team	0.02	0.08	−0.06
numerical superiority 10m fromball raw 0 sec defending team	0.01	0.04	−0.03
distance to the ball raw 0 sec 1 attackers	0.09	0.07	0.02
distance to the ball mean 3 sec 1 attackers	0.08	0.06	0.02
distance to the ball raw 0 sec 1 defenders	0.09	0.09	−0.01
number defenders 10m fromball raw 0 sec defending team	0.03	0.02	0.01
distance to the ball mean 3 sec 1 defenders	0.03	0.04	−0.01
defensive pressure raw 0 sec 1 attackers	0.03	0.01	0.01
spread team raw 0 sec defending team	0.03	0.01	0.01
dis def att defending team raw 0 sec defending team	0.02	0.01	0.01
velocity raw 0 sec 1 attackers	0.01	0.02	−0.01
number defenders 20m fromball raw 0 sec defending team	0.03	0.03	0.00
number defenders 15m lastdefender raw 0 sec defending team	0.02	0.02	0.00
surface area raw 0 sec attacking team	0.02	0.02	0.00

ball is also increasing the counter-risk of allowing easier opposing scoring opportunities.

Finally, it has to be acknowledged that this is an exploratory study using machine learning. While, this study provides an insightful model interpretation approach to increase the explainability of the model, the causality of the findings cannot be established (Holzinger et al. 2019; Molnar et al. 2020). Accordingly, future studies need to confirm the replicability of the observed associations and the causal hypotheses proposed in the current study (e.g., principles of defensive play).

Conclusion

Overall, using the current study design based on highly informative tracking data with a sophisticated machine learning approach, we are able to effectively analyze the tactics of successful defending in soccer. The deployed models achieved a satisfactory prediction performance of ball gains in defense. By analyzing the input variables, we identified several tactical principles of defensive play that appear to be related to gaining the ball, such as press the ball-leading player, create numerical superiority in ball near areas (press close pass options), and defensive compact organization. Furthermore, our results indicated that the defending behavior seems distinctively different depending on the vertical pitch position of the ball (lanes) and that, even though the quality of a defending team did not influence the prediction, better teams created situations with an increased chance of gaining the ball more often.

Due to the connection to success and the identification of the most important variables using explainable AI, those results can change the way how soccer is analyzed.

Acknowledgements

The authors thank the German Football League (Deutsche Fußball Liga, DFL) for providing the match data used in this study.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

The author(s) reported that there is no funding associated with the work featured in this article.

ORCID

Leander Forcher http://orcid.org/0000-0002-6428-8643

Data availability statement

The used data is a 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 data of other soccer leagues or data providers). To process the tracking data in this project, the floodlight package was used (a high-level data-driven sports analytics framework) (Raabe et al. 2022). This python package can be used to process similar match data to make use of the developed code of this investigation, which is publicly available on github: https://github.com/LForcher/kit_d-fine_sports-analytics.

Open scholarship

This article has earned the Center for Open Science badge for Open Materials. The materials are openly accessible at https://github.com/LForcher/kit_d-fine_sports-analytics

References

Andrienko G, Andrienko N, Budziak G, Dykes J, Fuchs G, von Landesberger T, Weber H. 2017. Visual analysis of pressure in Football. Data Min Knowl Discov. 31(6):1793–1839. doi: 10.1007/s10618-017-0513-2.

Badiella L, Puig P, Lago-Peñas C, Casals M. 2022May. Influence of red and yellow cards on team performance in elite soccer. Ann Oper Res. 325 (1):149–165. 10.1007/s10479-022-04733-0

Bartlett R, Button C, Robins M, Dutt-Mazumder A, Kennedy G. 2012. Analysing team coordination patterns from player movement trajectories in soccer: Methodological considerations. Int J Perform Anal Sport. 12(2):398–424. doi: 10.1080/24748668.2012.11868607.

Bauer P, Anzer G. 2021. Data-Driven detection of counterpressing in professional football. Data Min Knowl Discov. 35(5):2009–2049. doi: 10.1007/s10618-021-00763-7.

Bourbousson J, Sève C, McGarry T. 2010. Space–time coordination dynamics in basketball: Part 2. The interaction between the two teams. J Sports Sci. 28(3). Article 3. doi: 10.1080/02640410903503640.

Casal CA, Ángel Andujar M, Luis Losada J, Ardá T, Maneiro R. 2016. Identification of defensive performance factors in the 2010 FIFA World Cup South Africa. Sports (Basel, Switzerland). 4(4):54. doi: 10.3390/sports4040054.

Castellano J, Álvarez D, Figueira B, Coutinho D, Sampaio J. 2013. Identifying the effects from the quality of opposition in a football team positioning strategy. Int J Perform Anal Sport. 13(3):822–832. doi: 10.1080/24748668.2013.11868691.

Claro R. 2010. Tactical principles of soccer: Concepts and application. 15.

Forcher L, Altmann S, Forcher L, Jekauc D, Kempe M. 2022. The use of player tracking data to analyze defensive play in professional soccer - a Scoping review. Int J Sports Sci Coach. 17(6):17479541221075734. doi: 10.1177/17479541221075734.

Forcher L, Kempe M, Altmann S, Forcher L, Woll A. 2021. The 'Hockey' assist makes the difference - validation of a defensive disruptiveness model to evaluate passing sequences in elite soccer. Entropy. 23(12):1607. doi: 10.3390/e23121607.

Frencken W, Lemmink K, Delleman N, Visscher C. 2011. Oscillations of centroid position and surface area of soccer teams in small-sided games. Eur J Sport Sci. 11(4):215–223. doi: 10.1080/17461391.2010.499967.

Gibney E. 2022. Could machine learning fuel a reproducibility crisis in science? Nat. 608(7922):250–251. doi: 10.1038/d41586-022-02035-w.

Goes F, Kempe M, van Norel J, Lemmink KAPM. 2021. Modelling team performance in soccer using tactical features derived from position tracking data. IMA J Manag Math. 32(4):519–533. doi: 10.1093/imaman/dpab006.

Goes F, Meerhoff LA, Bueno MJO, Rodrigues DM, Moura FA, Brink MS, Elferink-Gemser MT, Knobbe AJ, Cunha SA, Torres RS, et al. 2021. Unlocking the potential of big data to support tactical performance analysis in professional soccer: A systematic review. Eur J Sport Sci. 21 (4):481–496. doi:10.1080/17461391.2020.1747552.

González-Ródenas JOAQUIN, López-Bondia IGNACIO, Aranda-Malavés RODRIGO, Desantes AT, Sanz-Ramírez ENRIQUE, Malaves RA. 2020. Technical, tactical and spatial indicators related to goal scoring in european elite soccer. J Hum Sport Exerc. 15(1):186–201. doi: 10.14198/jhse.2020.151.17.

Grehaigne J-F, Marchal D, Duprat E. 2002. Regaining possession of the ball in the defensive area in soccer. In: Murphy A, Reilly T, and Spinks W, editors. Science and Football IV. London, UK: Routledge; p. 108–117.

Herold M, Hecksteden A, Radke D, Goes F, Nopp S, Meyer T, Kempe M. 2022. Off-Ball behavior in association football: A data-driven model to measure changes in defensive pressure. J Sport Sci. 40(12):1412–1425. *Accepted*. doi:10.1080/02640414.2022.2081405.

Hewitt A, Greenham G, Norton K. 2016. Game style in soccer: What is It and can we quantify It? Int J Perform Anal Sport. 16(1):355–372. doi: 10.1080/24748668.2016.11868892.

Holzinger A, Langs G, Denk H, Zatloukal K, Müller H. 2019. Causability and explainability of artificial intelligence in medicine. Data Min Knowledge Discovery. 9(4):e1312. doi: 10.1002/widm.1312.

Kapoor S, Narayanan A. 2022. Leakage and the reproducibility crisis in ML-Based science. doi:10.48550/ARXIV.2207.07048.

Lago-Peñas C, Gómez-Ruano M-Á, Lee Owen A, Sampaio J. 2016. The effects of a player dismissal on competitive technical match performance. Int J Perform Anal Sport. 16(3):792–800. doi: 10.1080/24748668.2016.11868928.

Lepschy H, Waesche H, Woll A. 2020. Success factors in football: An analysis of the German Bundesliga. Int J Perform Anal Sport. 20(2):150–164. doi: 10.1080/24748668.2020.1726157.

Lepschy H, Woll A, Waesche H. 2021. Success factors in the FIFA 2018 World Cup in Russia and FIFA 2014 World Cup in Brazil. Front Psychol. 12:1–9. doi:10.3389/fpsyg.2021.638690.

Linke D, Link D, Lames M, Kerhervé HA. 2020. Football-specific validity of TRACAB's Optical video tracking systems. PLoS One. 15(3):1–17. doi: 10.1371/journal.pone.0230179.

Low B, Rein R, Raabe D, Schwab S, Memmert D. 2021. The porous high-press? An experimental approach investigating tactical behaviours from two pressing strategies in football. J Sports Sci. 39(19):1–12. doi: 10.1080/02640414.2021.1925424

Low B, Rein R, Schwab S, Memmert D. 2021. Defending in 4-4-2 or 5-3-2 Formation? Small differences in footballers' collective tactical behaviours. J Sports Sci. 40(3):1–13. doi: 10.1080/02640414.2021.1993655.

Matsuoka H, Tahara Y, Ando K, Nishijima T. 2020. Development of criterion-referenced measurement items for soccer defensive tactical play from tracking data. Football Sci. 17:29–40.

Memmert D, Raabe D. 2019. Revolution im Profifußball: Mit Big Data zur Spielanalyse 4.0.2. 2019 Edition. Berlin: Aktualisierte und erweiterte AuflSpringer. doi:10.1007/978-3-662-59218-2.

Michalis, et al. 2019. The creation of goal scoring opportunities in professional soccer. Tactical differences between Spanish La Liga, English Premier League. German Bundesliga and Italian Serie A. International Journal of Performance Analysis in Sport. 19(3):452–465.

Molnar C, König G, Herbinger J, Freiesleben T, Dandl S, Scholbeck CA, Casalicchio G, Grosse-Wentrup M, Bischl B. 2020. General pitfalls of model-agnostic interpretation methods for machine learning models. In International workshop on extending explainable ai beyond deep models and classifiers. Springer; pp. 39–68. doi: 10.1007/978-3-031-04083-2_4.

Moura FA, Martins LEB, Anido RDO, de Barros RML, Cunha SA. 2012. Quantitative analysis of Brazilian football players' organisation on the pitch. Sport Biomech. 11(1):85–96. doi: 10.1080/14763141.2011.637123

Navarro FJF 2018. Analysis of Styles of Play in Soccer and Their Effectiveness. [PhD Thesis]. Universidad de Granada.

Raabe D, Biermann H, Bassek M, Wohlan M, Komitova R, Rein R, Kuppens Groot T, Memmert D. 2022. Floodlight - a high-level, data-driven sports analyticsframework. J Open Source Softw. 7(76):4588. *ArXiv Preprint ArXiv:2206.02562*. doi:10.21105/joss.04588.

Rathke A. 2017 January. An examination of expected goals and shot efficiency in soccer. J Hum Sport Exerc. 12(2). 10.14198/jhse.2017.12.Proc2.05.

Rein R, Raabe D, Memmert D. 2017. 'Which pass is better?' Novel approaches to assess passing effectiveness in elite soccer. Hum Mov Sci. 55(October):172–181. doi: 10.1016/j.humov.2017.07.010.

Sarmento H, Clemente FM, Araújo D, Davids K, McRobert A, Figueiredo A. 2018. What performance analysts need to know about research trends in association football (2012–2016): A systematic review. Sports Med. 48(4):799–836. doi: 10.1007/s40279-017-0836-6.

Szczepanski L, McHale I. 2016. Beyond completion rate: Evaluating the passing ability of footballers. J Royal Stat Soc Ser A-Stat In Soc. 179(2):513–533. doi: 10.1111/rssa.12115.

Toda K, Teranishi M, Kushiro K, Fujii K. 2021 March. Evaluation of soccer team defense based on prediction models of ball recovery and being attacked. Plos One. 17(1):1–12. doi:10.13711/journal.pone.0263051

Appendix 1

Table A1. Results of prediction performance of different classifiers.

Model		Information		Training Set (60%)				Validation Set (20%)			
Name	Description	Number of features	Algorithm	Accuracy	Precision	Recall	F1-Score	Accuracy	Precision	Recall	F1-Score
All-features-model	including all features	334	Random Forest	0.95	0.81	0.99	0.89	0.85	0.51	0.65	0.58
All-features-model	including all features	334	XGBoost	0.86	0.58	0.33	0.42	0.86	0.59	0.32	0.41
All-features-model	including all features	334	Logistic Regression	1.00	1.00	1.00	1.00	0.87	0.65	0.41	0.51
Expert-reduction-model	Expert-model including the 16 most important uncorrelated features	16	Random Forest	0.91	0.64	0.99	0.78	0.85	0.50	0.65	0.57
Expert-reduction-model	Expert-model including the 16 most important uncorrelated features	16	XGBoost	0.90	0.79	0.52	0.63	0.87	0.61	0.38	0.47
Expert-reduction-model	Expert-model including the 16 most important uncorrelated features	16	Logistic Regression	0.85	0.52	0.16	0.25	0.85	0.51	0.15	0.25

Appendix 2

Model Info Sheet

Section 1: **Information about paper**

1) Author(s): Names of the authors of the paper
   Leander Forcher, Tobias Beckmann, Oliver Wohak, Christian Romeike, Ferdinand Graf, Stefan Altmann
2) Title of the paper or report which introduces the model
   Prediction of defensive success in elite soccer using machine learning – Tactical analysis of defensive play using tracking data and explainable AI.
3) DOI or permanent link to the paper or report (for example, link to arxiv.org webpage)
   DOI not available at the time of completion of the info sheet.
4) License: Under which license(s) are the data and/or model shared?
   Not applicable.
5) Email address of the corresponding author
   leander.forcher@kit.edu

Section 2: **Scientific claim(s) of interest**

6) Does your paper make a generalizable claim based on the ML model? If yes, what is the scientific claim?

   (1) Our model can be used in real game situations in soccer to gain deeper insights in the defending tactics of their team to improve their game analyses (e.g. data analyses) or player evaluations.
   (2) Interpreting the results of our model we identified tactical principles of defensive play in soccer that are related to gaining the ball in defense.

7) Is the scientific claim made about a distribution or population from which you can sample?
   Yes.

   (a) Population: Professional soccer matches/teams/players.
   (b) Sample: Professional soccer matches of German Bundesliga season 2020/21.

8) Does the scientific claim only apply to certain subsets of the distribution mentioned in Q6?
   No.

Section 3: **Train-test split is maintained across all steps in creating the model**

9) Train-test split type: How was the dataset split into train and test sets? (For example, cross-validation; separate train and test sets).
   We used a train (60%), validation (20%), test (20%) split. This split was completed prior to the model training.
10) Are there duplicates in the dataset? If yes, explain how duplicates are handled to ensure the train-test split.
   There were no duplicates in the dataset.
11) In case the dataset has dependencies (e.g., multiple rows of data from the same patient), describe how the dependencies were addressed (for example, using block-cross validation).
   One could argue that teams have several possessions against the same defending teams which would allow the suggestion of dependent data. However, according to previous research in soccer performance (see below), we assumed that each possession is not related to any other possession, since no possession can be compared to a previous one (due to the complexity of team sport soccer). We therefore, assumed that each data point is independent. (e.g. Moura et al. (2012). Quantitative analysis of Brazilian football players' organisation on the pitch. Sports Biomech, 11, 85–96.; Bartlett et al. (2012). Analysing Team Coordination Patterns from Player Movement Trajectories in Soccer: Methodological Considerations. Int J Perform Anal Sport, 12, 398–424.)
   However, despite the assumed independence of the data, all game situations of a single match were assigned to one of the sets when divided into validation, train, test sets. Accordingly, no game situations of the same match were assigned to the training or the test set at the same time.
12) List all the pre-processing steps used in creating your model. For example, imputing missing data, normalizing feature values, selecting a subset of rows from the dataset for building the model.
   All preprocessing steps are included in the methods section of this paper.
13) How was the train-test split observed during each pre-processing step? If applicable, use a separate line for each step mentioned in Q12.
   All preprocessing steps described in the method section (e.g. selection of game situations) were done independently of each other data point. The train-validation-test split was completed after all preprocessing steps. Since the data points were assumed to be independent, this has no impact on possible data leakage.
14) List all the modeling steps used in creating your model. For example, feature selection, parameter tuning, model selection.
   All modeling steps are included in the method section of this paper.
15) How was the train-test split observed during each modeling step? If applicable, use a separate line for each step mentioned in Q14.
   The train-validation-test split was maintained through all modeling steps.
16) List all the evaluation steps used in evaluating model performance. For example, cross-validation, out-of-sample testing.
   We used the test data set to evaluate the model performance. We focused on F1 score, since our data set was unbalanced.
17) How was the train-test split observed during each evaluation step? If applicable, use a separate line for each step mentioned in Q16.
   The train-validation-test split was maintained through all evaluation steps. We solely used the test-set to evaluate the modeling performance.

Section 4: **Test set is drawn from the distribution of scientific interest.**

18) Why is your test set representative of the population or distribution about which you are making your scientific claims?

The test set is representative for the population because it consists of defensive plays of professional soccer. Several studies showed that data of different leagues are comparable (e.g. Forcher et al. (2021). The "Hockey" Assist Makes the Difference-Validation of a Defensive Disruptiveness Model to Evaluate Passing Sequences in Elite Soccer. Entropy, 23(12), 1607.; Michalis et al. (2019). The creation of goal scoring opportunities in professional soccer. Tactical differences between Spanish La Liga, English Premier League, German Bundesliga and Italian Serie A. International Journal of Performance Analysis in Sport, 19:3, 452-465.)

However, to this date, there is no investigation that compared the defensive performance between different leagues of professional soccer.

19) Explain the process for selecting the test set and why this does not introduce selection bias in the learning process.

The test set was randomly choosen from all defensive plays identified in a large and representative sample of 153 matches of German Bundesliga. Therefore, the test set does not introduce a bias in the learning process.

Furthermore, all game situations of a single match were assigned to one of the sets when divided into validation, train, test sets to account for possible dependencies in the data, which were not assumed (compare section 2, 9)).

20) In case your model is used to predict a future outcome of interest using past data, detail how data in the training set is always from a date earlier than the data in the test set.

The model is solely partly used to make predictions about the future. It is mainly used to analyze defensive plays post event using the identified principles of defensive play which are stated and reasoned in the discussion section of this paper.

Section 5: **Each feature used in the model is legitimate for the task**

21) List the features used in the model, alongside an argument for their legitimacy. A legitimate feature is one that would be available when the model is used in the real world and is not a proxy of the outcome being predicted. You can also include this list in an appendix and reference the relevant section of your Appendix here.

All features included in this paper and in the prediction of defensive success were only used if they represented a tactical idea. This allows to interpret the results in a practical applicable way and accordingly transfer the results into practice.

All metrics included (explained in detail in the methods section of this paper (please see Table A1)):

- defensive pressure
- distance to the ball
- velocity
- surface area
- spread
- distances between formation lines
- numerical superiority
- pitch zone
- quality of teams

For each metric the following features were extracted:

Feature combinations

time instance [sec]	time information	spatial information		
		1 player	3 players	5 players
0	current	metric_raw_0_sec_1_attackers	metric_raw_0_sec_3_attackers	metric_raw_0_sec_5_attackers
0	current	metric_raw_0_sec_1_defenders	metric_raw_0_sec_3_defenders	metric_raw_0_sec_5_defenders
1	current	metric_raw_1_sec_1_attackers	metric_raw_1_sec_3_attackers	metric_raw_1_sec_5_attackers
1	current	metric_raw_1_sec_1_defenders	metric_raw_1_sec_3_defenders	metric_raw_1_sec_5_defenders
1	mean	metric_mean_1_sec_1_attackers	metric_mean_1_sec_3_attackers	metric_mean_1_sec_5_attackers
1	mean	metric_mean_1_sec_1_defenders	metric_mean_1_sec_3_defenders	metric_mean_1_sec_5_defenders
1	delta	metric_delta_1_sec_1_attackers	metric_delta_1_sec_3_attackers	metric_delta_1_sec_5_attackers
1	delta	metric_delta_1_sec_1_defenders	metric_delta_1_sec_3_defenders	metric_delta_1_sec_5_defenders
3	current	metric_raw_3_sec_1_attackers	metric_raw_3_sec_3_attackers	metric_raw_3_sec_5_attackers
3	current	metric_raw_3_sec_1_defenders	metric_raw_3_sec_3_defenders	metric_raw_3_sec_5_defenders
3	mean	metric_mean_3_sec_1_attackers	metric_mean_3_sec_3_attackers	metric_mean_3_sec_5_attackers
3	mean	metric_mean_3_sec_1_defenders	metric_mean_3_sec_3_defenders	metric_mean_3_sec_5_defenders
3	delta	metric_delta_3_sec_1_attackers	metric_delta_3_sec_3_attackers	metric_delta_3_sec_5_attackers
3	delta	metric_delta_3_sec_1_defenders	metric_delta_3_sec_3_defenders	metric_delta_3_sec_5_defenders
5	current	metric_raw_5_sec_1_attackers	metric_raw_5_sec_3_attackers	metric_raw_5_sec_5_attackers
5	current	metric_raw_5_sec_1_defenders	metric_raw_5_sec_3_defenders	metric_raw_5_sec_5_defenders
5	mean	metric_mean_5_sec_1_attackers	metric_mean_5_sec_3_attackers	metric_mean_5_sec_5_attackers
5	mean	metric_mean_5_sec_1_defenders	metric_mean_5_sec_3_defenders	metric_mean_5_sec_5_defenders
5	delta	metric_delta_5_sec_1_attackers	metric_delta_5_sec_3_attackers	metric_delta_5_sec_5_attackers
5	delta	metric_delta_5_sec_1_defenders	metric_delta_5_sec_3_defenders	metric_delta_5_sec_5_defenders