



Data-driven detection of counterpressing in professional football

A supervised machine learning task based on synchronized positional and event data with expert-based feature extraction

Pascal Bauer^{1,2} · Gabriel Anzer^{1,3}

Received: 17 August 2020 / Accepted: 4 May 2021 / Published online: 8 July 2021
© The Author(s) 2021

Abstract

Detecting counterpressing is an important task for any professional match-analyst in football (soccer), but is being done exclusively manually by observing video footage. The purpose of this paper is not only to automatically identify this strategy, but also to derive metrics that support coaches with the analysis of transition situations. Additionally, we want to infer objective influence factors for its success and assess the validity of peer-created rules of thumb established in by practitioners. Based on a combination of positional and event data we detect counterpressing situations as a supervised machine learning task. Together, with professional match-analysis experts we discussed and consolidated a consistent definition, extracted 134 features and manually labeled more than 20, 000 defensive transition situations from 97 professional football matches. The extreme gradient boosting model—with an area under the curve of 87.4% on the labeled test data—enabled us to judge how quickly teams can win the ball back with counterpressing strategies, how many shots they create or allow immediately afterwards and to determine what the most important success drivers are. We applied this automatic detection on all matches from six full seasons of the German Bundesliga and quantified the defensive and offensive consequences when applying counterpressing for each team. Automating the task saves analysts a tremen-

Responsible editor: Albrecht Zimmermann.

✉ Pascal Bauer
pascal.bauer@dfb.de

Gabriel Anzer
gabriel.anzer@sportec-solutions.de

¹ Department of Sport Psychology and Research Methods, Institute of Sports Science, University of Tübingen, Tübingen, Germany

² DFB-Akademie, Deutscher Fußball-Bund e.V. (DFB), Frankfurt, Germany

³ Sportec Solutions AG, subsidiary of the Deutsche Fußball Liga (DFL), Munich, Germany

dous amount of time, standardizes the otherwise subjective task, and allows to identify trends within larger data-sets. We present an effective way of how the detection and the lessons learned from this investigation are integrated effectively into common match-analysis processes.

Keywords Sports analytics · Football (Soccer) · Tactical performance analysis · Applied machine learning · Positional and event data

1 Introduction

Acquiring accurate and high frequency positional and event data is common in most of the world's top professional football (soccer) leagues. Manually annotated *event data* provides information about the one player carrying the ball at the time of a game relevant action only, whereas so called *positional data* can capture highly accurate positions of all 22 players up to 25 times a second.

Every professional football team spends a substantial amount of time analyzing and monitoring strategies such as *counterpressing*.—a complex team strategy for transition situations—of their own and opposing teams. Navarro and Javier (2018) defines counterpressing as simple as “[...] *pressure after losing the ball*”. Related to this, Pep Guardiola made the ‘*five second rule*’ for counterpressing famous.¹ Another coach to experience tremendous success in recent seasons is Liverpool FC manager Jürgen Klopp. He is generally accepted as the originator of the term ‘*Gegenpressing*’, which is well-known in both its German version and English translation.² It is apparent to football experts that Klopp’s counterpressing concept is closely related to Guardiola’s strategy of regaining the ball within the first five seconds.

There are significant differences in team’s defensive and offensive tactical line-ups (Bialkowski et al. 2014, 2015; Andrienko 2019; Shaw and Mark 2019). The *transition phase* describes the period following a win or loss of possession in which the team transitions between its offensive and defensive tactical line-ups and vice versa. When a team is in possession for at least a certain amount of time, we can assume that generally its tactical formation is optimized for offensive play and, consequently, sub-optimal in terms of defending its own goal (Andrienko 2019; Shaw and Mark 2019). Therefore, the first seconds after losing the ball are critical from a defensive perspective. Several studies proved that transition phases are a substantial factor for a team’s overall performance: As early as in 1968, Reep and Benjamin (1968) demonstrated in the first known football analytics study, that 30% of all regained possessions lead to shots on goal and 25% of all goals came from regained possessions in the attacking quarter. Grant et al. (1999) confirmed these findings for the 1998 World Cup. Both outcomes align perfectly with Jürgen Klopp’s statement that regaining the ball immediately after loosing it, potentially through successful counterpressing, “...is the best

¹ “[...] *after losing the ball, the team has five seconds to retrieve the ball, or, if unsuccessful, tactically foul their opponent and fall back*”, Pep Guardiola; <https://www.theblizzard.co.uk/article/peps-four-golden-rules>, accessed 06/20/2020.

² <https://www.sueddeutsche.de/sport/premier-league-bei-kloppts-liverpoolern-klemmt-das-gaspedal-1.2695408-2>, accessed 06/20/2020.

playmaker".³ Klopp hereby claims that counterpressing can also be seen as an offensive strategy. Recent studies show that regaining the ball in open play likelier leads to a goal than a save build-up from a team's own half (Vogelbein et al. 2014; Hobbs et al. 2018). Based on tracking data from the English Premier League, Hobbs et al. (2018) detected possession regains close to the opponent's goal—potential counterpressing situations—highlighting their relevance once more. Even though many coaches and clubs affected the development of this sophisticated strategy, neither an objective proof of its efficiency, nor an analysis on its usage in top leagues is presented in the literature.

Hughes and Ian (2015) point out that team sports performance analysis tends to be operationalized on the basis of *notation systems*, described as a replicable and consistent method of recording sport performance. Recent literature explained a framework, where coaches' decisions are supported by several performance analysis reports from games, teams and players (Travassos et al. 2013) and pointed out that team tactics in football refer to both a priori decisions made before the match, and also real-time adaptations during the game (Rein and Daniel 2016). In accordance to that it is described as a complex process resulting from a network of inter-dependent parameters (Kempe et al. 2014). These processes are conducted in a time-critical set-up, especially when it comes to the world's top leagues and competitions where teams need to encounter different opponents several times a week. Although many clubs extended their match-analysis departments considerably within the past years, the limited amount of time and resources during matches forced teams to seek ways to automate processes and gain insights faster in order to obtain a competitive edge.

These recent developments—the availability of accurate performance data and the need for a quick detailed tactical analysis—signifies a huge potential for the application of sophisticated machine learning techniques to football data and requires an efficient collaboration of computer-science and domain experts (Herold et al. 2019; Goes et al. 2020; Rein and Daniel 2016). Many recent scientific investigations aimed to establish new *key performance indicator (KPI)*—metrics quantifying certain aspects of the game: pass evaluation metrics were examined (Steiner et al. 2019; Goes et al. 2019), metrics to quantify controlled space were defined (Kim 2004; Fernandez and Bornn 2018; Brefeld et al. 2019) and several studies evaluated shot metrics (Lucey et al. 2014; Rathke 2017; Fairchild et al. 2018; Anzer and Bauer 2021)⁴ and goal scoring opportunities through possession values (Link et al. 2016; Spearman 2018; Fernandez and Bornn 2018; Decroos et al. 2020). Additionally, there are many approaches for measuring the defensive behavior of teams (Santos et al. 2018; Andrienko 2019; Goes et al. 2019), and even approaches aiming to quantify pressing (Bojinov and Luke 2016; Andrienko 2017; Robberechts 2019). Although pressing and counterpressing are closely related, they are two different phenomena. An interesting conference proceeding describes how specific counterpressing situations can be derived from detected general pressing scenes.⁵ Several approaches also showed, that analyzing these KPI's or even aggregating simple statistics (e.g. the pass completion rate) over one or several

³ <https://tactical-times.com/the-history-and-evolution-of-jurgen-klopp/>, accessed 06/20/2020.

⁴ Often referred to as *expected goals (xG)* values.

⁵ Will Gürpınar-Morgen (2018). "How StatsBomb data helps measure counterpressing", *Statsbomb Innovation in Football Conference 2018*; <https://statsbomb.com/2018/05/how-statsbomb-data-helps-measure-counter-pressing/>, accessed 11/11/2020.

seasons provides a helpful indication to practitioners (Power et al. 2018; Pappalardo et al. 2019). The primary goal of all these approaches is to derive new insights by processing vast amounts of information. Decroos et al. (2018) presented a first approach to detect interesting match-phases based on event data. To the best of our knowledge no peer-reviewed study focused on automating parts of the performance analysts everyday life by detecting complex tactical patterns based on positional and event data. However, a noteworthy approach aiming to detect counterattacks was presented in an established football analytics conference.⁶

With this practical need for process optimization in mind, it is the pivotal issue of this study to detect counterpressing situations without human-support and provide several ad-hoc reports for match analysts in near real-time. The outcome is optimized to fulfill their practical requirements and fit seamlessly into their tool-ecosystem. Additionally, the automated detection allows us to analyze large amounts of data that would exceed manual processing capacities. Consequently, our approach enables us to perform impartial long-term analysis of the German Bundesliga's latest seasons investigating the following research questions:

- Can we differentiate between varying regaining strategies and determine reasons for a short defensive reaction time (definition in Sect. 2.1.2), i.e. to which extent is a fast ball regain actually caused by counterpressing (*RQ1*)?
- Can we set objective benchmarks to quantify counterpressing strategies (amount and effectiveness) on a match- and season-level and point out their correlation with a team's overall success (*RQ2*)?
- Do the established rules of thumb agree with the data (i.e. counterpressing is more effective close to the sideline) (*RQ3*)?
- To what extent do team's counterpressing strategies differ in the German Bundesliga (*RQ4*)?

All together, answering these research questions helps us to define the baseline for a qualitative discussion with experts, and thus allows them to formulate requirements for the practical application (*PA*) set-up.

The remainder of this paper is structured as follows: Sect. 2 provides a detailed description of the used data, the underlying rules and definitions, the labelling process and the extracted features. The outcomes in Sect. 3 are split into three parts: First in Sect. 3.1, we describe a statistical evaluation of the detection models. Section 3.2 presents a subject-specific evaluation by interpreting our results on six seasons of German Bundesliga. Lastly, in Sect. 3.3, we demonstrate how this approach can be operationalized in the performance analysis process. This application is based on two matches of the German national teams.^{7,8} All parts of this study were developed in close cooperation with the professional match-analysts and coaches (see Acknowledgements).

⁶ Karun Singh (2020). "Learning to watch football: self-supervised representations for tracking data" In *OptaPro Analytics Forum, London*; <https://www.youtube.com/watch?v=H1iho17lnoI>, accessed 11/11/2020.

⁷ Germany against Northern Ireland; 19th of November 2019, Commerzbankarena Frankfurt.

⁸ Germany U21 against Belgium U21; 17th of November 2019, Schwarzbaldstadion Freiburg.

2 Methods

2.1 Data and definitions

2.1.1 Data collection

The present study uses *positional* and *event data* collected in more than six seasons (4118 matches) of the German Bundesliga and 2nd Bundesliga, as well as the above mentioned two matches of the German national teams. Positional data is captured by optical tracking systems⁹ and event data consists of manual annotations based on a dedicated *official match data catalogue*,¹⁰ defining around 30 events with more than 100 attributes. The event data can be seen as a log of the ball relevant actions (e.g. passes, shots, tacklings or fouls), however, it does not cover complex team-tactical behaviors such as counterpressing.

Since the two data sources are collected independently of each other, they need to be synchronized before they can be processed together. Even though several steps of quality management from independent institutions are conducted on the manually collected event data, the assigned timestamp of a given event can differ significantly to one in the positional data. The synchronization of positional and event data is conducted by dedicated rules (per event) that extracts the exact timestamp and the exact location on the pitch from tracking data given a manually tagged event. For example, when synchronizing a pass, the sudden increase in the distance between the passing player and the ball, captured by the optical tracking, can be used to align both location and timestamp of the pass. The positional data is collected at a frequency of 25 Hz and includes the longitudinal, latitudinal, and in case of the ball, also the altitudinal positions of the players, ball and referees related to the pitch markings.

The information about which team is currently in possession of the ball (hereafter referred to as *ball possession*) and whether the game is running or currently stopped (hereafter referred to as *ball status*) are crucial for our survey. Both values are collected live in the stadium for every frame of the match by a skilled operator focused exclusively on this task.

2.1.2 Definitions

Since there are conflicting definitions of ball possession in the literature (Kempe et al. 2014), we decided to adopt published definitions with expert feedback. The above mentioned operators, dedicated to acquire information about ball possession and status, are briefed to mark *ball possession for one team*, if and from that time point a player of that team touches the ball with ball control, until the ball is out of play, or an opponent player touches the ball with ball control. *Ball control* is defined in this context, as the ability to conduct a contrived action with the ball. Whenever a pass is played between two players of one team, the ball possession belongs to that team as long as no opposing player intercepted that pass or won the ball within an

⁹ <https://chyronhago.com/products/sports-tracking/tracab-optical-tracking/>, accessed 06/20/2020.

¹⁰ <https://www.sportec-solutions.de/en/index.html>, accessed 06/20/2020.

individual duel. According to the definition from Link and Hoernig (2017) we also compute ball possession on a player level (*individual ball possession*). In the case of an interception, the ball possession change is detected exactly at the time of the first ball touch of the intercepting player. We use the term *defensive reaction time*—the time it takes to regain ball possession after losing it—as defined in Vogelbein et al. (2014). All situations where either the ball is beyond the pitch markings or the play is stopped by the referee (e.g. because of a foul) are labeled as out-of-play. Hence, if the ball goes out of bounds there must typically be a change in ball possession. Situations in which the touch of the player carrying the ball outside the markings is not declared as a ball possession due to missing control (e.g. a deflected shot), or when the individual possession model disagrees with the team possession flag are excluded. For all further investigations only the *effective playing time* (also referred to as *net playing time*)—defined as all the situations while the game is running—are considered. Shots, for example, always represent the end of a ball possession phase per definition. Ball possession phases that end with the halftime-, or final-whistle or a referee ball are excluded from our analysis.

In addition to these general rules, we developed the following transition-related definitions in consultation with match-analysis experts: A *defensive transition phase* is defined as the time-window when a team loses ball possession, but is not yet into their ideal defensive formation. Within these defensive transition phases,

a team conducts *counterpressing* if at least one player exerts (spatio and/or temporal) pressure on the ball carrier, or on the opponents close to the ball.

Note that there exist many different definitions for pressing: StatsBomb¹¹ defines pressing as a defensive player being within a five-yard radius of the ball-carrying opponent.¹² Very similarly, a more granular and non-binary definition, aggregating the pressure of several defensive players, is presented by Andrienko (2017). Based on these pressing definitions, counterpressing could be defined as situations where pressing is exerted immediately after a ball possession change (Navarro and Javier 2018). Both of these rule-based definitions are used as a baseline model for our investigation.

However, according to the match-analysts involved in this project, being close to the player in ball possession is not the only way to exert pressure. Attacking or blocking the easiest pass options could, for instance, also be seen as applying pressure.

To quantify the *success of counterpressing*, we consider it as successful if the ball is regained within five seconds and shots and goals, scored or received, are accredited to the previous counterpressing phase if they occur within the following 20 seconds.

From hereon the game is split into ball possession phases which could start and end either with an in-play ball possession change or a stoppage such as a set-piece. Note that the set-up of the in-play ball possession change, such as the defensive transition, might not be the only influence factor on the defensive reaction time—it can also occur due to short, uncontrolled ball possessions or risky passes of the opponent. Any ball possession phases that either start with a set-piece or end with a stoppage in play will not be considered further. Fig. 1 shows a heatmap displaying the occurrences of

¹¹ Statsbomb is a football event data provider based in the UK, <https://statsbomb.com/>, accessed 12/17/2020.

¹² StatsBomb event data, including the pressing tag, are can be accessed for many professional leagues <https://statsbomb.com/data/>.

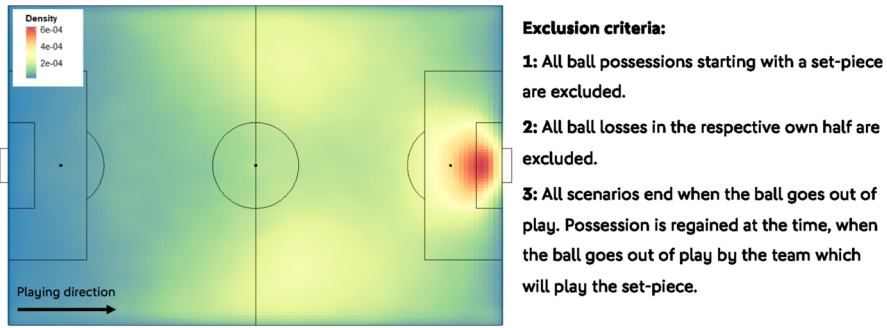


Fig. 1 Overview of where on the pitch turnovers happen most frequently (from the perspective of a team playing from left to right)

transition situations related to the pitch. It indicates, that most turnovers happen in the opposing half, especially near the sidelines. Ball possession changes due to a ball going out of bounds are added to the area touching the sideline. Easily identified is the high proportion of turnovers in the opponent's six-yard box. This is likely because both saved shots and shots missing the goal wide are counted as a change in possession as soon as the goalkeeper receives the ball.

2.2 Supervised machine learning set-up

2.2.1 Hand-crafted labeling of defensive transition situations

Since the rule-based approaches to detect counterpressing we investigated lead us to an insufficient accuracy (see Sect. 3.1), we conducted a manual tagging procedure with trained student-analysts. It was their task to label situations with a detectable counterpressing strategy. In total, out of 11, 108 relevant defensive turnovers, 3, 196 situations were labeled as counterpressing. The labeling was conducted for the first eleven Bundesliga-matchdays of the 2018/2019 season from the perspective of the home team. The percentage of counterpressings detected per transitions differs significantly per team. Borussia Mönchengladbach presented the highest value (40.07%), whereas only 21.80% of Hannover 96's transitions have been labeled as counterpressing. The aggregated outcome of the labeling process per team of the German Bundesliga is displayed in Table 7 in the Appendix A.

To quantify the inter-labeler reliability, 20 matches were labeled by three different students. To compute the pairwise accuracy for each defensive turnover, we checked if both students had identified counterpressing in the following two seconds. This yielded a pairwise accuracy of 82.01%, i.e. in 82.01% of defensive turnovers both students agreed on the nature of the actions following a turnover.

As additional information, the experts tried to detect the exact start and end-frame of the respective transition situation. The average duration of all transitions phases is 9.34 s, 9.89 s for counterpressing, whereas all non-counterpressing turnovers took in average 9.11 s.

2.2.2 Expert-based feature extraction

We defined a list of 134 features that aim to characterize the transition. The features describe the location of a ball possession change, several relevant factors describing both teams' exact positioning at the time of turnover and their movements in the first two seconds immediately after the ball loss. A time-window longer than two seconds was problematic, because it would cut off too many situations where the ball possession changed within that time.

A teams' decision to conduct counterpressing is heavily influenced by the situation of the ball possession itself. To take this into consideration, all features are also calculated at the moment of the ball possession change. According to football experts, turnovers without the chance to counterpress are often characterized by immediate clearances or aerial duels. Therefore, we included the ball position, the ball height, and the individual ball possession time (Link and Hoernig 2017)—describing the time a player of the ball possessing team was in direct control over the ball. The involved football experts suggested, that counterpressing is often characterized by achieving a local compactness close to the ball. We aimed to cover this with several metrics measuring the regaining team's positioning around the ball. For instance, we use the team's covered area, global and local stretch indices (Bourbousson and Carole Sève 2010; Santos et al. 2018) as features in our model. A team primarily aiming to defend their own goal after losing possession does this usually with high-speed towards their own goal, whereas counterpressing requires often only players close to the ball to attack their opponents with a high speed towards the ball carrier. This is addressed by calculating several speed-values and considering each team's average position, the so-called team-center (Bourbousson and Carole Sève 2010; Andrienko 2017). In contrast to a more conservative transition strategy, counterpressing's primary objective is not to place many players in a compact unit behind the ball quickly, but rather to defend more aggressively up the pitch. Therefore, we calculate both the number of players in front and behind the ball, as well as their respective compactness. Although the pressing definition from Andrienko (2017) was not sufficient as a stand-alone rule-based counterpressing detection criteria (see Sect. 3.1), it is incorporated in various features of our model.

All features were discussed, consolidated and steadily improved within workshops and based on several steps of evaluation of the detection. A detailed list and description of the features is presented in Table 1, a video describing some of the features can be accessed here.

2.3 Model training

2.3.1 Detection of counterpressing as a supervised machine learning task

We trained several classification algorithms based on the 11, 108 labeled defensive turnover situations from 97 matches fulfilling our inclusion criteria (see Fig. 1).

Table 1 The extracted features that are used for counterpressing detection. Features used in both dimensions of pitch coordinates (horizontally and vertically) and for different time points after the initial ball possession change are listed only once

Feature	Definition
Turnover position	Location (x,y-coordinate) of the ball at the timepoint of the ball possession change (BPC), hereafter 0 s.
Ball height	z-coordinate of the ball tracked at several timepoints after (BPC) (0 s, 1 s, 2 s)
Distance team	Distance closest player to the ball calculated at different timepoints after BPC (0 s, 1 s, 2 s) and for each team (opposing, regaining).
Number players close to ball	Number of players in several circles (10 m, 20 m, 30 m) around the ball counted at different timepoints after BPC (0 s, 1 s, 2 s); calculated for both teams separately.
Team center	Average position of all players of each team (goalkeeper excluded) calculated for both dimensions (x,y) and at several time points after BPC (0 s, 1 s, 2 s); calculated for both teams separately.
Covered area team	Team's covered area measured as the widest distance between two players in two dimensions (x,y) and at several time points after BPC (0 s, 1 s, 2 s); calculated for both teams separately.
Players closer to ball	Number of players from a team closer to the ball than next player from the other team calculated at different time points after BPC (0 s, 1 s, 2 s); calculated for each team separately.
Speed closest player	Speed of player closest to the ball calculated at different time points after BPC (0 s, 1 s, 2 s) and for each team separately.
Duration previous ball possession	Duration of the previous ball possession phase; sequences where the ball was out-of-play (<i>ball status</i>) have been excluded.
Speed team	Average speed of each team at different time points after BPC (0 s, 1 s, 2 s); calculated for both teams.
Individual ball possession (absolute)	Total individual ball possession time of any player from the regaining team within up to the first 3 seconds after BPC as defined in Link and Hoernig (2017).
Individual ball possession (relative)	Total individual ball possession divided by the length of the considered time window, so either divided by 3 seconds or by the duration of the ball possession change if it is shorter.
Players in front of the ball	Number of players for each team that are further away from the own goal center than the ball at different time points after BPC (0 s, 1 s, 2 s); calculated for both teams separately.
Players behind the ball	Number of players for each team that are closer to the own goal center than the ball at different time points after BPC (0 s, 1 s, 2 s); calculated for both teams separately.
Global compactness team	Normalized stretch index of each team team (excluding goalkeeper) based on definition in Bourbousson and Carole Sève (2010) calculated at several time points after BPC (0 s, 1 s, 2 s); calculated for both teams separately.

Table 1 continued

Feature	Definition
Local compactness	Normalized local stretch index of different player groups (3,4 and 5 player closest to the ball) at different time points after BPC (0 s, 1 s, 2 s) as shown in Santos et al. (2018); calculated for both teams separately.
Compactness in front of the ball	Normalized local stretch index of all players of each team that are further away from the own goal center than the ball at different time points after BPC (0 s, 1 s, 2 s); calculated for both teams separately.
Compactness behind the ball	Normalized local stretch index of all players of those players of each team that are closer to the own goal center than the ball at different time points after BPC (0 s, 1 s, 2 s); calculated for both teams separately.
Pressure regaining team on ball	Normalized pressure (definition Andrienko (2017)) the regaining team conducts on the ball at different time points after BPC (0 s, 1 s, 2 s).
Pressure on ball possessing player	Normalized pressure (definition Andrienko (2017)) the regaining team conducts on the ball possessing player at different time points after BPC (0 s, 1 s, 2 s).
Effective playing space	Defines the space covered by the convex hull of all players of a team (excluding goalkeeper) calculated at different time points after the BPC (0 s, 1 s, 2 s) as shown in Santos et al. (2018); calculated for each team separately.

We split the labeled data-set (75% training data, 25% test data) by taking randomly 25% of all transitions out of every match to avoid over-representing teams, scores, or results.

We used the above defined features (section 2.2.2 or Table 1) and evaluated the best performing models on our set of test data.

Among different basic-models, we applied extreme gradient boosting (hereafter referred to as XGBoost), a scalable tree boosting system, introduced by Chen and Carlos (2016), which outperformed traditional machine learning algorithms in numerous applications (Li et al. 2019; Liu et al. 2020; Zhang et al. 2020). For our investigation, we want to point out three major advantages of XGBoost: (a) To make use of our wide set of features without taking the risk of overfitting, an additional regularization term is added to the loss function. Additionally, (b) XGboost is a scalable machine learning model, which can be extended seamlessly with more data or more features being available. Furthermore, (c) no normalization is required.

Before training our model, a set of hyperparameters (shown in Table 2) has to be defined. As presented in Wang (2019), we applied Bayesian tree-structured Parzen Estimator hyperparameter optimization approaches to obtain the highest possible accuracy and avoid overfitting. By using tree-structured Parzen Estimators (Bergstra 2011)

Table 2 Hyperparameter-selection of the XGBoost models

	Hyperparameter	Description	Range	XGB-M1	XGB-M2
1	Learning rate	Controls the step size used per update	[0, 1]	0.045	0.05
2	Max depth	Limits the depth of the tree	[0, ∞)	5	7
3	Subsample	Controls number samples applied to the tree	(0, 1]	0.419	0.279
4	Min child weight	Controls instance weight of a node	[0, ∞)	0.855	1
5	nrounds	Limits number of iterations	1–400	400	400
6	Class balancer	Controls the balance of negative and positive weights (Number of negative cases / Number of positive cases)	(0, ∞)	1	2.5

as the surrogate model in Bayesian optimization (Dewnacker et al. 2016), we reduce the running time of hyperparameter tuning and achieve better scores on the testing set. Note that the hyperparameter *nrounds* was set to a maximum of 400 iterations.

To further guarantee the stability of our model and avoid overfitting, we applied five-fold cross-validation on the training data.

As described in Sect. 2.1.2, we also implemented two rule-based baseline models that serve as a benchmark for our detection. A naive approach defines counterpressing as follows: whenever one or more players are within a five-yard radius around the ball carrier during the first individual ball possession phase following a turnover, it is classified as a counterpress (hereafter referred to as *naive rule-based approach*).¹³ The second approach (hereafter referred to as *Andrienko-approach*) defines counterpressing as all turnovers whenever the first player in ball possession receives pressure exceeding a certain threshold according the pressure-definition in Andrienko (2017), whereby the final threshold of 0.74 was obtained by maximizing the F_1 -score on the training set.

2.3.2 Effectiveness for counterpressing and fast possession regains

In order to define some success metric of a transition phase, the low scoring nature of football causes us to examine the following actions more granularly, rather than just checking whether they are followed by a goal. For both cases—successful and unsuccessful ball recoveries through counterpressing—we extracted taken shots, expected goals¹⁴ and actual goals following a transition phase. To investigate this issue, two definitions had to be made: Which ball recovery latency of a counterpressing strategy should be considered as successful, and for how long a defensive and an offensive action would be accredited to the previous ball recovery (strategy). As a starting point for a potential threshold for successful counterpressing, a first indicator is given by Pep Guardiola's five second rule. We queried relevant video scenes with possession regains after 3, 4, 5, 6, 7 and 8 seconds and discussed them with a group of professional match-analysts. The same procedure was conducted to investigate the follow-up goal-scoring opportunities. Here scenes with shots 10, 15, 17, 20 and 25 seconds after the initial ball loss were discussed. Through this procedure, we finally agreed on the definitions described in Sect. 2.1.2.

¹³ This approach has been suggested at the *Statsbomb Innovation in Football Conference* by Will Gürpınar-Morgen (<https://statsbomb.com/2018/05/how-statsbomb-data-helps-measure-counterpressing/>), accessed 11/11/2020.

¹⁴ The "Expected Goal" (xG) value of a shot denotes the a priori probability of a shot being converted to a goal. Hence its value ranges from [0, 1]. The probability is estimated using both tracking and event data and applying a machine learning model, that was trained on more than 100, 000 shots. Details regarding the xG-model used can be found in Anzer and Bauer (2021).

Table 3 Statistical evaluation of the counterpressing outcome

	Model	Precision	Recall	F_1 -score	AUC
1	XGBoost (Model 1)	0.72	0.63	0.67	0.874
2	Logistic regression	0.69	0.51	0.59	0.841
3	Random forest	0.74	0.55	0.63	0.867
4	XGBoost with class balancer (Model 2)	0.60	0.80	0.69	0.865
5	Naive rule-based approach	0.31	0.87	0.46	0.602
6	Andrienko-approach	0.37	0.37	0.37	0.568

3 Results

3.1 Statistical evaluation

3.1.1 Detection of counterpressing

With the above described supervised machine learning set-up we are able to detect counterpressing situations with sufficient accuracy for practical applications (see also Sect. 3.3). Table 3 shows a statistical evaluation of the different models, from which XGBoost performed the best. Per team and per match, we detect around 20 to 30 counterpressing situations, out of around 90 to 200 transition situations.

With the highest overall area under the curve (AUC) the above presented optimization (Table 3, row 1) is best suited for the long term analysis of several seasons with the goal to identify trends and underpin practitioner rules (*RQ1-RQ4*, Sect. 3.2). The XGBoost model with a class balancer (Table 3, row 4) has a higher recall of 80% with still an acceptable false positive rate. Thus, it can be applied for specific performance analysis of either the own match or several matches of the next opponent (PA, Sect. 3.3)—where match-analysts spend a lot of time analyzing video footage either way. The optimal hyperparameters used for both models can be found in Table 2. When examining the results of the two baseline approaches, they exhibit a very low overall accuracy. The naive rule-based approach (Table 3, row 5) classifies 72.41% of all turnovers as counterpressing, which leads to a high recall but also a large number of false positives. For the Andrienko-approach, selecting the threshold by optimizing the F_1 -score lead to a more realistic percentage of predicted counterpressing situations in the test set (25.65%), but are, nevertheless, significantly outperformed by either machine learning approach.

Another advantage of the XGboost approach is that the individual influences of our rich feature set can be somewhat quantified and interpreted by analyzing the respective SHAP-values.¹⁵ The naming was both coined by their originator Lord Shapley, who introduced them in the context of cooperative game theory (Roth and Thomson 1988), but also by Lundberg and Su (2017), who used the concept to interpret the features for machine learning models. In comparison to traditional feature importance models (e.g. gain or Saabas method), SHAP-values present a consistent and locally accurate

¹⁵ The abbreviation **SHAP** stands for **SH**apley **A**dditive **eX**planation.

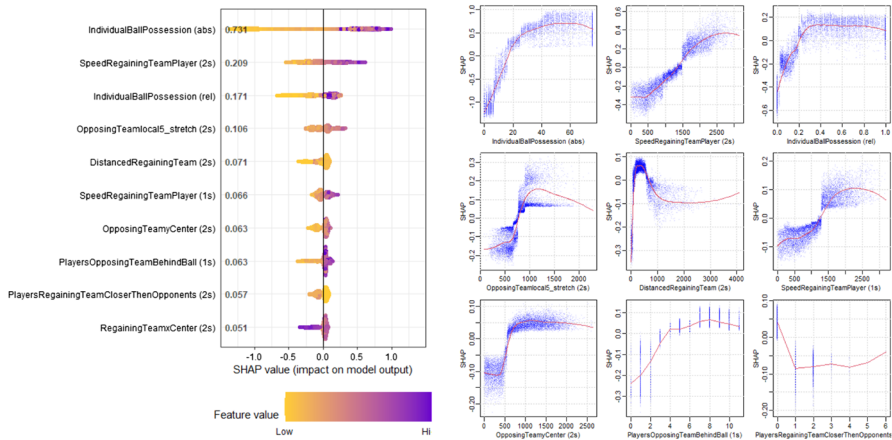


Fig. 2 Feature influence to the counterpressing prediction based on SHAP-values

method to identify the individualized feature contribution to machine learning models. This method has been effectively used in different applications (Antipov et al. 2020; Meng et al. 2020; Ibrahim et al. 2020; Anzer and Bauer 2021).

Figure 2 displays the most influential features according to the SHAP values in two different representations. In the left Fig., each dot represents the contribution of the feature to the model, whereas the color-coding describes the value of that feature. Both, the absolute individual ball possession time (*IndividualBallPossession (abs)*) and the speed of the regaining team two seconds after the change in ball possession (*SpeedRegainingTeamPlayer (2s)*), have a very strong and linear impact on the predictions. Besides the fact, that both features have the highest overall influence on the prediction (widest dispersion of the dots on the left part of Fig. 2, the interpretation of the SHAP-values can be expressed as follows: the higher the absolute amount of individual ball possession time within the first three seconds and the higher the speed of the regaining team two seconds after the turnover, the more likely a defensive turnover is classified as counterpressing. The number of opposing players behind the ball after one second (*PlayersOpposingTeamBehindBall (1s)*) influences the prediction in a different way: A high number of players behind the ball increases the chances for a classification of counterpressing, but the relation is clearly non-linear. To get a better idea of the influence, we will have a look at the right part of Fig. 2. The value per feature is now displayed by the x-axis, whereas the model influence is shown on the y-axis. If less than four players are behind the ball, this feature on average decreases the chance of a counterpressing classification. The number of four defenders—almost half of the team—seems to be a decisive threshold. If four or more players are behind the ball, this feature has a positive contribution to the prediction. This not only aligns with the expectation of the practitioners, but also led to a very valuable discussion among the professional analysts. A more complex relation is shown by the local stretch index of the five closest player to the ball of the opposing team two seconds after the ball possession change (*OpposingTeamlocal5 stretch (2s)*). After a steep increase starting at 600 cm, the influence of that feature reaches its maximum at roughly 1, 000 cm (see

Table 4 For all 4118 considered matches of the German Bundesliga and 2nd Bundesliga, this table shows the probability of shots and goals per team following counterpressing situations

Strategy	Goals scored %	Goals conceded %	Shots for %	Shots against %
Counterpressing (all)	0.35	0.78	3.22	5.15
Unsuccessfull counterpressing	0.16	1.02	1.83	6.7
Successfull counterpressing	0.75	0.25	6.27	1.76

right plot in Fig. 2) but decreases afterwards. This indicates that a higher stretch index (lower compactness) of the opposing team after two seconds increases the chances for counterpressing.

Excluding features with little to no influence according to the SHAP-values, did not improve neither the F_1 -score nor the AUC of our prediction on the test data-set.

3.1.2 Effects of counterpressing

Table 4 shows the outcome of counterpressing regarding goals and shots scored or conceded within 20 seconds. If one is successful, i.e. wins back the ball within five seconds the chance of scoring increases tremendously, but if unsuccessful one is far likelier to concede a goal.

It is no surprise, that the chances to either shoot or score are significantly higher when counterpressing was applied successfully, since it implies the crucial attacking advantage of having gained the possession of the ball.

While using goals scored versus conceded would theoretically be ideal to measure success, the low scoring nature often prevents us from doing so. Therefore, we use shots to compare teams and coaches. Nevertheless, according to the experts, looking at both shot- and goal- (or even expected goal-)balance is a very valuable key-performance-indicator for counterpressing.

3.2 Subject-specific evaluation of six seasons of German Bundesliga data

In the following section, we use our quantitative results as a baseline for a qualitative, subject-specific evaluation and interpretation with the involved experts.

3.2.1 Lessons learned about defensive transitions (RQ1)

A common procedure when analyzing a team's transition strategy is looking at the easily acquirable defensive reaction time (Vogelbein et al. 2014). This, however, comes with the drawback that it is not able to distinguish between situations with intentional counterpressing behavior and noise. Note that in non-trivial defensive turnover situations typically a team can choose between falling back or conduct counterpressing. But for the purpose of this study, the distinction between fallback and other non-counterpressing situations was removed for the sake of simplicity. An additional analysis of the expert-based labeling (see Appendix A) showed, that around 62.5%

of all ball losses fulfilling the inclusion criteria cannot be assigned to either defensive strategy (counterpressing or fallback). The sheer number of these situations—with a very short defensive reaction time (on average 7.83 s) and without any defensive tactical choices being detectable—significantly influences the defensive reaction time when applied to all transitions. Further analysis (see Appendix A) shows that specifically turnovers with very short individual ball possession times fall in this category. Their exclusion presents a crucial step for a better understanding of transition situations.

In general, counterpressing is not always advantageous and needs to be executed well. Although, the expectation of some practitioners (invigorated by Jürgen Klopp's statement "*counterpressing is the best playmaker*") may be different, it is still intuitive that the team in possession of the ball has a higher chance to perform a successful offensive action (e.g. a shot or goal; see Table 4, row 1). However, this is an average over all teams independent of their skill level and as we will later see, there are some teams/coaches that were able to apply counterpressing so successfully that they even ended up with a positive shot balance by creating more shots after counterpressing than conceding. In order to properly assess the risk versus reward nature of counterpressing, one would ideally compare it to its strategic counterpart falling back. But even then, one would need to carefully address potentially confounding variables describing the original situations, since, as it seems from the feature importance discussion, the situation typically dictates the strategic response. This goes, however, beyond the scope of this study, but could be the ground for interesting future work. Additionally, since all non-counterpressing situations consist of myriad of different circumstances, they do not serve as reasonable baseline to effectiveness of counterpressing.

3.2.2 Define and statistically underpin objective Benchmarks (RQ2)

Based on the above explained definitions and trained prediction models, several quotients and ratios were discussed with the experts. Aggregated on a season level, we analyzed the correlation with a team's final ranking. Our detection provides several different metrics with a significant correlation to success. These performance indicators can be calculated per match, per match phase (e.g. one half-time) or even per turnover, which allows practitioners to objectively compare their teams performance with pre-defined benchmarks. With a negative Pearson correlation, the ratio of successful counterpressing-situations to the total number of transitions predicts a team's final ranking the best ($r = -0.44$). Another metric, correlating with the final ranking that was very valuable to the experts due to its direct interpretability, was shot and goal balance ($r = -0.36$ shots, $r = -0.42$ goals)—describing, whether more shots or goals are taken after successful counterpressing, than are conceded after failed attempts. However, these metrics should be used carefully since they are based on small sample size and could contain confounding effects with the overall offensive or defensive qualities of a team. Loosing the ball, increases the probability of an opponent conducting an offensive action. Thus we present an effective strategy to monitor the outcome of counterpressing strategies, such as several performance indicators that enables coaches to objectively benchmark a team's defensive transition behavior.

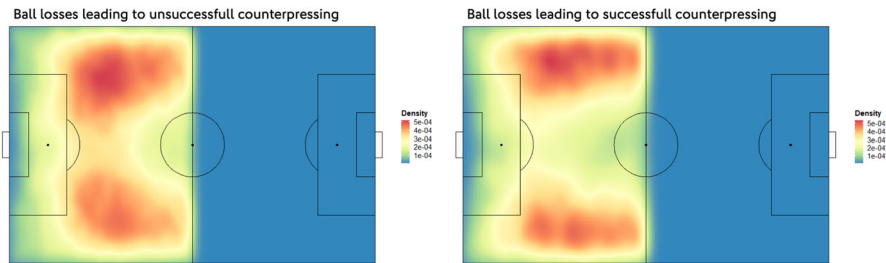


Fig. 3 Heatmap of ball possession change locations before unsuccessful counterpressing situations (left) and successful situations (right)

3.2.3 Approve established rules of thumb (RQ3)

A widely spread rule of thumb is that counterpressing is ideal after ball losses close to the sideline or close to the corners in the opponent's half. Fig. 3 presents two heatmaps that underpin this statement. Secondly, we want to examine, whether a numerical superiority of players close to the ball increases the chance of a successful counterpress as assumed by many experts. For that we examine the 109, 852 detected counterpressing situations satisfying the inclusion criteria. Whenever the team out of possession has a numerical superiority within a 10m radius around the ball, at the time of the turnover, they regain ball possession within 5 seconds 36.2% of the time, compared to only 30.2%, when the other team has more players in that area. This indicates that the rule of thumb has some truth to it, but is far from the only influencing factor, deciding whether a counterpress will be successful.

3.2.4 Compare Team's and Coach's Counterpressing in German Bundesliga (RQ4)

Further investigations highlight to which extent teams use completely different defensive transition strategies. We investigated the coaches that were expected to have a pronounced counterpressing-behavior by the experts. The respective ranking among all coaches finishing a full season can be found in the Appendix C.

Table 5 also gives a first indication of which further aspects could be considered. Teams ending up in the top 5 of a season perform above average in almost all defined metrics. No consistent tendency over the considered seasons is detectable: games do not get more intensive in terms of total in-play transitions per match, nor are there significant changes in teams average counterpressing behavior. FC Bayern München has the best goal balance after counterpressing, which might be heavily influenced by their offensive efficiency. Jürgen Klopp and Ralf Ragnick performed extraordinary well in terms of ending up with more created than received shots after attempting counterpressing. Given that Jürgen Klopp ended only seventh place in one of his two considered seasons, this should be seen as an outstanding performance. Over the course of a whole season only nine coaches achieved a non-negative shot balance within 20 s after their own counterpressing—the average final ranking of the respective teams was four. Note that teams playing at home tend to conduct counterpressing slightly more often than away teams. Considering only home-teams, our model classifies 27.24%

Table 5 Comparison of counterpressing-related performance indicators in Bundesliga. The colored arrows display the rankings within the respective groups compared to the average (teams by ranking; all teams by seasons; teams and coaches). For the columns presenting the shot and goal balance we used colored dots for the teams and coaches, showing whether the total outcome per match is positive, neutral or negative

	All Transitions						Off. / Def. Balance		Successful Counterpressing			Unsuccessful Counterpressing		
	Team	Coach/Match	Matches	per Match	%CP	%Successful CP	Goals	Shots	All per CP	%Shots	%Goals	All per CP	%Shots	%Goals
	All Teams	All Coaches	4118	115.60	23.08%	7.25%	-0.11	-1.15	31.42%	2.65%	1.10%	68.58%	7.51%	1.13%
	Top 5 Teams	All Coaches/Season	1142	114.99	23.74%	7.80%	-0.04	-0.28	32.85%	10.94%	1.42%	67.15%	8.85%	0.94%
	Bottom 5 Teams	All Coaches/Season	1145	116.62	22.42%	7.03%	-0.16	-0.65	31.35%	9.78%	0.95%	68.65%	8.10%	1.32%
	All Teams	Season 2013/2014	610	98.72	24.50%	7.47%	-0.12	-0.47	30.50%	10.94%	1.07%	69.50%	7.59%	1.16%
	All Teams	Season 2014/2015	612	126.33	21.80%	7.19%	-0.07	-0.28	32.99%	10.36%	1.10%	67.01%	6.61%	0.90%
	All Teams	Season 2015/2016	612	121.58	22.56%	7.39%	-0.09	-0.50	32.49%	8.97%	1.10%	67.51%	7.01%	1.00%
	All Teams	Season 2016/2017	612	121.79	22.28%	7.39%	-0.08	-0.49	33.16%	8.75%	1.09%	66.84%	7.02%	0.99%
	All Teams	Season 2017/2018	612	116.32	23.46%	7.16%	-0.09	-0.59	30.51%	10.56%	1.22%	69.49%	7.74%	1.03%
	All Teams	Season 2018/2019	612	112.42	23.97%	7.15%	-0.18	-0.64	29.81%	11.68%	1.04%	70.19%	8.32%	1.41%
	All Teams	Season 2019/2020	448	110.65	23.53%	7.04%	-0.20	-0.69	29.91%	11.17%	1.09%	70.09%	8.53%	1.55%
	FC Bayern München	All coaches	229	112.21	24.63%	8.02%	0.06	-0.24	32.56%	11.98%	2.04%	67.44%	7.10%	0.66%
	FC Bayern München	Pep Guardiola	102	109.47	25.24%	8.61%	0.02	-0.21	34.10%	11.34%	1.46%	65.90%	7.00%	0.65%
	FC Bayern München	Jupp Heynckes	26	109.73	22.43%	6.98%	-0.08	-0.31	31.09%	12.56%	1.51%	68.91%	7.48%	1.13%
	FC Bayern München	Carlo Ancelotti	40	118.80	22.84%	7.60%	-0.13	-0.68	31.86%	7.48%	2.22%	68.14%	6.99%	0.99%
	FC Bayern München	Hansi Flick	15	121.87	25.33%	8.56%	0.07	0.00	32.55%	11.61%	2.52%	67.75%	8.38%	0.80%
	Borussia Dortmund	All Coaches	229	119.57	24.87%	8.51%	-0.04	-0.16	34.20%	10.56%	1.29%	65.80%	6.29%	0.87%
	Borussia Dortmund	Jürgen Klopp	68	129.00	23.51%	8.54%	-0.04	0.26	36.32%	11.62%	0.80%	63.68%	5.26%	0.69%
	Borussia Dortmund	Thomas Tuchel	68	121.59	24.73%	8.84%	0.00	-0.16	35.75%	9.58%	1.64%	64.25%	6.16%	0.91%
	Borussia Dortmund	Lucien Favre	58	105.07	25.98%	8.34%	-0.10	-0.34	30.90%	10.24%	1.38%	69.10%	6.34%	1.14%
	Borussia Dortmund	Thomas Tuchel	34	95.82	24.31%	7.98%	-0.18	-0.50	32.83%	10.15%	1.15%	67.17%	8.65%	1.69%
	Mainz 05	Thomas Tuchel	102	113.00	24.61%	8.60%	-0.06	-0.27	34.93%	9.99%	1.51%	65.07%	6.88%	1.14%
	RB Leipzig	All Coaches	127	126.69	22.13%	7.84%	-0.06	-0.02	35.41%	10.71%	1.43%	64.59%	6.00%	1.13%
	RB Leipzig	Julian Nagelsmann	25	110.60	22.21%	7.88%	-0.12	-0.20	35.50%	11.47%	0.92%	64.50%	7.58%	1.26%
	RB Leipzig	Ralf Rangnick	34	131.47	21.17%	7.00%	-0.03	0.56	32.33%	14.38%	1.18%	67.77%	3.95%	0.76%
	TSG Hoffenheim	Julian Nagelsmann	116	111.01	23.98%	7.46%	-0.03	-0.53	31.09%	13.44%	1.77%	68.91%	8.98%	0.99%
	All Teams	Julian Nagelsmann	141	110.94	23.67%	7.53%	-0.05	-0.48	31.82%	13.07%	1.61%	68.18%	8.76%	1.03%

of all included defensive transitions as counterpressing, which is roughly in line with the labeled training data, that was conducted only on home teams (28.77%).

Figure 4 shows a shortlist of teams' counterpressing outcomes. Note that, since on average more shots/goals are conceded than created when counterpressing, all four axes are scaled differently. For both sub-figures, teams on the upper left, including, for example, 1. FC Nürnberg¹⁶, Hannover 96¹⁷ and SV Darmstadt 98¹⁸ in the left figure, perform worse. Teams on the bottom right tend to generate more shots/goals and allow fewer while counterpressing. Teams with high values in the top right quadrant like Borussia Mönchengladbach¹⁹, FC Augsburg²⁰ or TSG Hoffenheim²¹ seem to employ risky defensive transition strategy, by both creating and also receiving many shots after their own counterpressing.

As a general recap of the Bundesliga analysis, we would like to point out that teams use significantly different transition strategies (RQ4). The experts' expectations of which coaches use counterpressing more often and/or more efficiently were underpinned by the results.

3.3 Proof of concept

The central objective of this study is to automate the detection of counterpressing situations. This helps match-analysts in their daily processes by saving them time, but also by providing objective and comparable benchmarks. First, we describe the general

¹⁶ Highest in the left figure.
¹⁷ Green circle with the inscription 96.
¹⁸ Blue circle around a white lily.
¹⁹ Black and white hatched diamond logo roughly in the center of both plots.
²⁰ Fifth highest in the left figure.
²¹ Second from the right in the right figure.

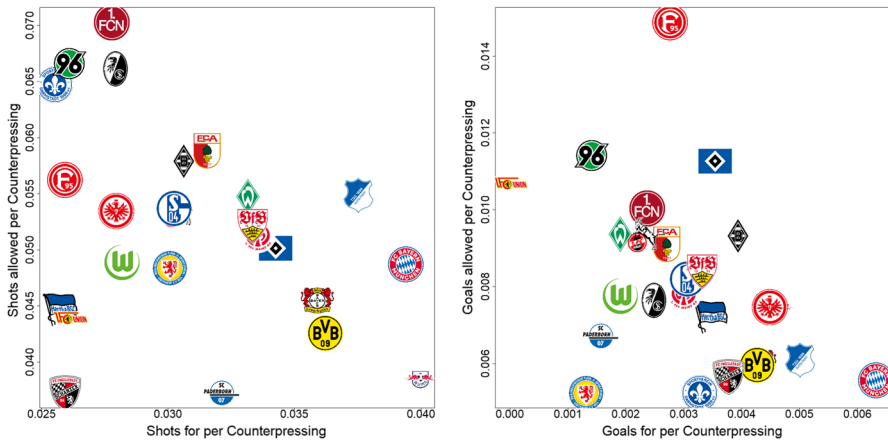


Fig. 4 For both shots (left figure) and goals (right figure) teams are displayed depending on their percentage of counterpressings leading to either shots or goals for (x-axis) or against (y-axis). Team values are aggregated across all considered seasons

set-up, whereas a prototypical application for two exemplary matches of German national teams is conducted.^{22,23} Based on the results described above, we are now able to provide match-analysts with two files fully automatically in virtually real-time: First, they receive a list of all detected counterpressing situations. To integrate this efficiently into their ecosystem, the files are produced in different file-formats, which can be imported into their video-analysis tool of choice (e.g. Hudl Sportscode,²⁴ Stats Edge Viewer²⁵). Such tools basically help to handle tags or labels in combination with the video footage. Figure 5 shows how this eliminates the usual process of the match-analysts labeling the videos manually in an exemplary tool (here Hudl's Sportscode). Usually, match-analysts use these tools to tag important situations live during the match but also in detail post-match for opponent analysis. Once a match or parts of it are tagged, the tool allows the analyst to output the tags either as a video-playlist or as an xml-file, containing the category and the time-frame of each tag. Depending on the coaches needs, the outcome is either presented as a video-playlist or a quantitative report giving an aggregated overview—which also is typically produced manually. An automatically generated counterpressing-playlist for the U21 match based on our prediction can be viewed here.

Second, we automatically provide coaches and analysts a counterpressing match-report with visualization after the respective match or entire season reports. For the U21 match an excerpt of the automated match report is presented in Fig. 6. Only two shots occurred within 20 seconds after either team counterpressed. However, these two shots after unsuccessful counterpressing attempts by the German team lead to two goals, which were decisive for the total match-outcome (Germany vs. Belgium, final score

²² Germany against Northern Ireland; 19th of November 2019, Commerzbankarena Frankfurt.

²³ Germany U21 against Belgium U21; 17th of November 2019, Schwarzbaldstadion Freiburg.

²⁴ <https://www.hudl.com/products/sportscode>, accessed 06/20/2020.

²⁵ <https://www.statsperform.com/team-performance/football/stats-edge/>, accessed 06/20/2020.

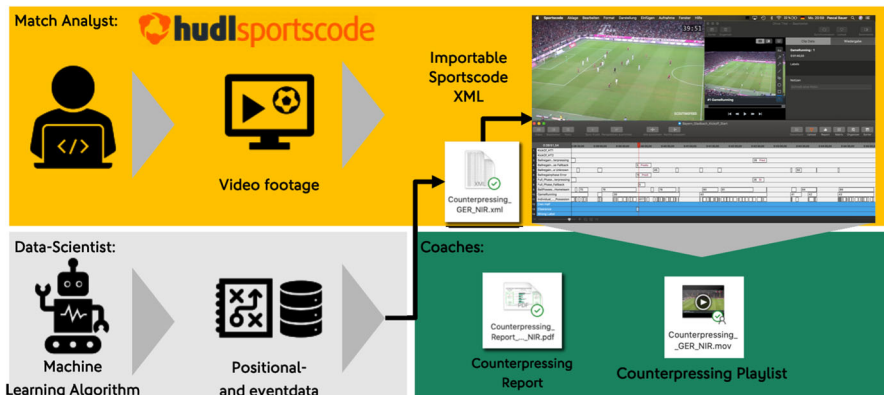


Fig. 5 Integration of the counterpressing-analysis into the match-analysts and coaches daily business. The yellow colored part shows the traditional process, the grey part displays how the automated analysis assimilates seamlessly

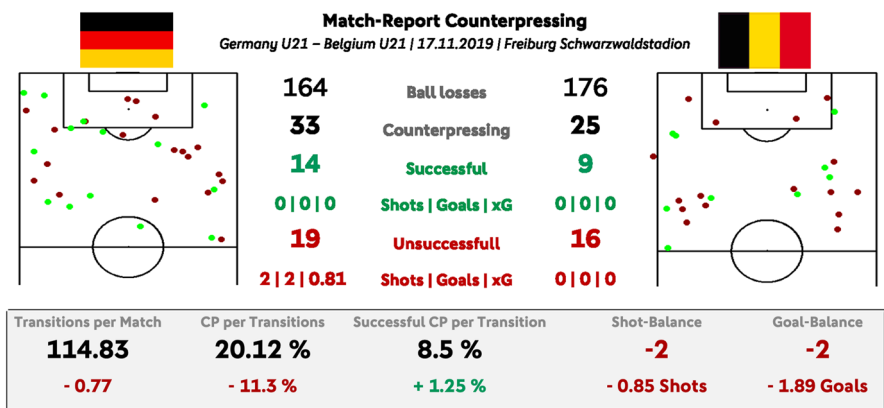


Fig. 6 Automated match report for the match Germany U21 against Belgium U21 in Freiburg. The pitch visualizations shows the positions of ball losses leading to a successful (green) or (unsuccessful) counterpressing. Whereas the absolute values are presented in the middle block, a benchmark of the central performance indicators against Bundesliga average is shown in the lower grey box

2–3). Whereas situations leading to goals are analyzed on a highly detailed level by coaches and analysts either way, they arrived at the same conclusion. Nonetheless, this report helps coaching staffs to evaluate whether the German team had a general bias in their defensive transition behavior or whether the two goals happened as a consequence of extraordinary opposing actions or defensive errors from the German team. In this case, bad counterpressing behavior had a significant stake in the origin of both goals. Another outcome was that many successful counterpressing situations (1.23% above Bundesliga average) did not end up in a single shot.

The practical implementation of this study was prototyped by the example of two recent matches of the German national teams. After each match, both the Sportscode-xml file and the automated match-report sheets were produced and shared with the experts for their post-match-analysis process. Additionally, to validate the report, the

Table 6 Statistical expert-evaluation of the counterpressing outcome. For the defensive transition column, the number in brackets displays the number of scenes excluded by our criteria (see Fig. 1)

Match	Team	Defensive transitions	Counterpressing detected	Manually excluded	Additionally detected
GER-NIR	Germany	153 (85)	25	6	0
GER-NIR	Northern Ireland	155 (95)	24	3	2
GER-BEL	Germany	164 (123)	33	5	0
GER-BEL	Belgium	176 (87)	25	2	1

matches were manually analyzed and provided a ground truth to compare our results with. The overall results are shown in Table 6.

For the first game 153 relevant defensive transition situations of the German national team were queried by the above defined rules. Out of these, 25 scenes were detected as counterpressing from which 6 were manually excluded from the final counterpressing playlist. For 164 defensive transitions from the second match 33 situations were detected as counterpressing and analysts ruled out 5 manually. The manually excluded scenes were discussed with all experts and it turned out that different definitions and interpretations lead to different labels. In this case study, ten of the eleven manually excluded situations consisted of only one player exerting pressure. Although this fulfills our definition, some of the experts would classify situations with only one player defending actively towards the ball as fallback (definition in Appendix A). All scenes that were additionally labeled by the respective match-analysts (in total five) contained a clearance closely after the initial ball loss and where thus not clearly related to the counterpressing strategy.

Both the Sportscodex-xml and the automated match report turned out to be valuable for the coaching staff. Due to the interpretability of the inaccuracies, the experts trust the outcome for further applications. They deemed the results to be sufficient in terms of a practical usage and noted that it saved vast amounts of resources in the pre- and post-match-analysis. The automated match-report (see Fig. 6) allows us to provide an objective comparison, with a flexible benchmark (here Bundesliga average) and therefore provide a new way to approach complex tactical strategies. The shot- and goal-balance are very intuitive and present a direct monitoring of the efficiency of conducted strategies.

4 Discussion

This paper shows that complex tactical strategies, such as counterpressing, can be detected automatically based on synchronized positional and event data. Comparing team's counterpressing strategies objectively and on longer periods of time creates insights that could not have been achieved with traditional methods.

The interdisciplinary cooperation turned out to be a very beneficial factor for this study. In our opinion, such a set-up of competencies is necessary to obtain relevant results. Machine learning techniques are required to detect complex strategies from spatio-temporal data, but also tactical football expertise are inevitable to determine

definitions, extract features and evaluate and interpret the resulting outcomes. A key lesson we learned through this study is that both definitions of complex strategies and their reading vary between football experts—this became apparent during an intensive process of expert-supported evaluations. One of our most meaningful key-performance-indicator is the shot-balance after counterpressing. Here, shots are used as a proxy for a successful attack. This is a common procedure in football analytics, however, the approach could be extended by using expected goals (e.g. Anzer and Bauer (2021)) or expected possession values (e.g. Spearman (2018)).

Since there does not exist a comparable approach for detecting counterpressing in the literature, we implemented two naive baseline approaches to benchmark our model against (see Table 3). While the approach based on Andrienko (2017) was originally not designed to quantify counterpressing but rather pure pressure, we build this rule based approach on Fernandez and Bornn (2018), who defined counterpressing as immediate pressure after losing the ball. Hence, it may not be the ideal approach, but due to complete the lack of alternatives in the literature, we use it as a benchmark model. Even though our model outperforms different rule-based baseline models (see Table 3) and the prediction accuracy is sufficient for practical application (see section 3.3), the basic limitation to achieve further accuracy is the inter-labeler reliability of 82.01%. After discussing the definitions, no further steps of consolidation between the labelers were conducted—but we would highly recommend such a step including the strict monitoring of the inter-labeler reliability for similar investigations. However, data labeling is a time-consuming process which cannot be conducted for each occurring philosophy and definition. Furthermore, methodologies that reduce labeling efforts, such as weak supervision, should be implemented on top of general detections as the one presented here to adjust definitions to the specific needs and to improve both the accuracy and the degree of individualization (Ratner et al. 2016, 2017). With an even larger amount of labeled data, one could consider using continuous features, or even the raw positional data of all players instead of features at discrete time points. The application of labeling-support methods could lead to more individualized and accordant labels and thus to a better prediction.

With an even more accurate model using one of the above described approaches, an improved and team-individual success prediction model for counterpressing could support the reflection of teams' decision making processes significantly. Also, the adaption of the counterpressing detection itself to team-specific definitions, provide a huge potential for further investigations.

Vogelbein et al. (2014) evaluated 306 matches of the 2010/2011 Bundesliga season and showed the time it takes to regain the ball also depends of the score at the time. They pointed out that teams with a lead tend to regain the ball slower than the ones that are trailing, and that teams finishing their season in the top third of the table regain ball possession significantly faster than the other teams—especially in drawn and losing match states. We found that the defensive reaction time, which serves as a baseline for our success definition, typically includes many noisy situations, where no clear strategy is observable. Further lessons learned regarding defensive transitions are described in section 3.2.1 and extended in Appendix A. The high influence of individual ball possession times on the predictions (see Fig. 2) can be attributed to uncontrolled situations without the possibility for either defensive strategy (counterpressing, fallback).

This also shows a limitation of our work, that some of the model's most important features focus on the situation itself and fewer on the strategy conducted thereafter. A possible explanation for this is that besides filtering out noisy situations, the model found that most often the situation dictates the defensive response.

Nevertheless, a tendency of the opponent to play counterattacks, or especially risky passes could lead to many fast ball recoveries independent from the defensive transition strategy. This issue could be considered by either combining this approach with an equivalent offensive transition strategy detection as shown for example in Hobbs et al. (2018) or by including more features of the opposing team or even the raw data of all 22 players and the ball. Not only the question in which situations counterpressing induces a high chance for a possession regain, but furthermore—given a situation where counterpressing is conducted—how likely is it to take/receive a shot when conducting that strategy are of high interest for practitioners. Regaining the ball fast might not be the only objective of counterpressing. Consequently, future investigations should also consider quantifying alternative success definitions, e.g. slowing down the opposing attack, forcing back-passes etc.

Another missing piece which should be investigated further is an accurate selection of fallback situations. Comparing their risk-reward structure to counterpressing situations, could lead to crucial insights by evaluating a teams' decision to counterpress versus falling back objectively. Since different teams may have their own club-specific definitions, our experimental set-up could be applied to arbitrary counterpressing-definitions or even other tactical patterns, as long as they gather sufficient labeled data. An interesting follow-up study, could investigate how many labeled matches would be necessary to achieve a sufficient accuracy depending on the definition. In our case we found 100 labeled matches to be sufficient, but also stress that a high inter-labeler reliability is necessary.

Note that counterpressing is only one example of a complex tactical pattern, that is of interest to match-analysts, but not covered in typical event level data. The needs of match-analysis departments combined with the growing availability and accuracy of positional and event data present a huge potential for task automating approaches.

5 Conclusions

Based on expert-evaluated definitions and hand-crafted labels, we are able to detect counterpressing strategies automatically with a sufficient accuracy in a supervised machine learning set-up. By producing both an understandable match-report and tagging-files suitable for conventional video-analysis software, the integration of the process into a match-analyst's daily business saves a significant amount of time. The outcome helps to analyze the own team's performance and provides helpful information about the next opponent's defensive transition behavior (*PA*).

We can differentiate between intended counterpressing strategies and the many uncontrolled transition situations with short defensive reaction times. This provides not only a better understanding of transitions but also several more granular performance indicators describing defensive transitions (*RQI*). The respective performance indicators, consolidated by statistical influences and expert opinions, derived inter-

pretable and intuitive metrics (*RQ2*), such as the goal- or shot balance presenting an effective efficiency quantification for the counterpressing strategy—that were not used before but seem to have a huge potential according to the experts. Two of the proven rules of thumb are that counterpressing is more likely to succeed closer to the sidelines and a numerical superiority close to the ball increases the chance of winning it back (*RQ3*). Through analyzing different facets over several seasons we are also able to quantify trends over a large period of time: teams within the German Bundesliga follow appreciably different transition strategies (*RQ4*). Furthermore, successful teams—measured against their final ranking—tend to use the counterpressing strategy more efficiently, giving credence to the notion of declaring it as an offensive strategy (*RQ2*).

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1007/s10618-021-00763-7>.

Acknowledgements This work would not have been possible without the perspective of professional match-analysts from world class teams who helped us to define relevant features and spend much time evaluating (intermediate) results. We would cordially like to thank Dr. Stephan Nopp and Christofer Clemens (head match-analysts of the German mens National team), Jannis Scheibe (head match-analyst of the German U21 mens national team) as well as Sebastian Geißler (former match-analyst of Borussia Mönchengladbach). Additionally, the authors would like to thank Dr. Hendrik Weber and Deutsche Fußball Liga (DFL) / Sportec Solutions AG for providing the positional and event data.

Funding Open Access funding enabled and organized by Projekt DEAL.

Declarations

Ethical Approval By informing all participating players, all tracking is compliant to the general data protection regulation (GDPR) <https://gdpr-info.eu/>, accessed 07/20/20. An ethics approval for wider research program using the respective data is authorized by the ethics committee of the Faculty of Economics and Social Sciences at the University of Tübingen. The data are property of the DFL e.V. / DFB e.V. and cannot be shared public. However, interested researchers can request samples of data under non-disclosure agreement constraints at the respective institutions. With the description of the respective tracking vendors and systems, peers working in the football industry can reproduce the results by using any kind of professional football data.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

Appendix

Appendix A gives additional information regarding the labeling conducted for this study including further tactical explanation of defensive transitions. Appendix B, C, and D show the outcome of the defined counterpressing performance indicator per team, per coach, and per season on the full Bundesliga data-set.

For all columns in Appendix B, C and D, T stands for turnovers, M for matches, CP for counterpressing, S for Shots and G for Goals. Whereas + indicates successful or positive offensive actions and – the other way around, +/– points out the offensive/defensive balance as described above. The first table in Appendix B compares all teams playing in Bundesliga between the 2013/2014 and 2019/2020 (until match-day 26) seasons. The teams are ordered by their average final ranking, the number of matches considered are shown in brackets. In a second table in Appendix C all coaches playing in Bundesliga between the 2013/2014 and 2018/2019 seasons are presented, whereas a third table shows all teams per season ordered first by their final ranking and secondly by the column CP+/T. Only full seasons with 34 matches one and the same coach where considered. The succession is made based on the highest correlating feature with teams final ranking, number of successful counterpressings per transition. Appendix D compares teams on a season-level and is sorted by the respective final ranking which is shown in brackets.

Appendix A

As discussed in Sect. 3.2, the actual labeling was simplified for our analysis. During the expert labeling process (see Sect. 2.2.1), not only counterpressing, but also *fallback* was labeled by the experts. This strategy is said to be the alternative to counterpressing and was defined as a defensive transition phase, where all players' intention is to either react inactively, or move backwards to their defensive line-up without exerting pressure on the ball. The terms *forward-* and *backward-defending* are often used in this context interchangeably to counterpressing and fallback. Figure 7 shows two exemplary situations for both strategies and lists some typical observations for each strategy.

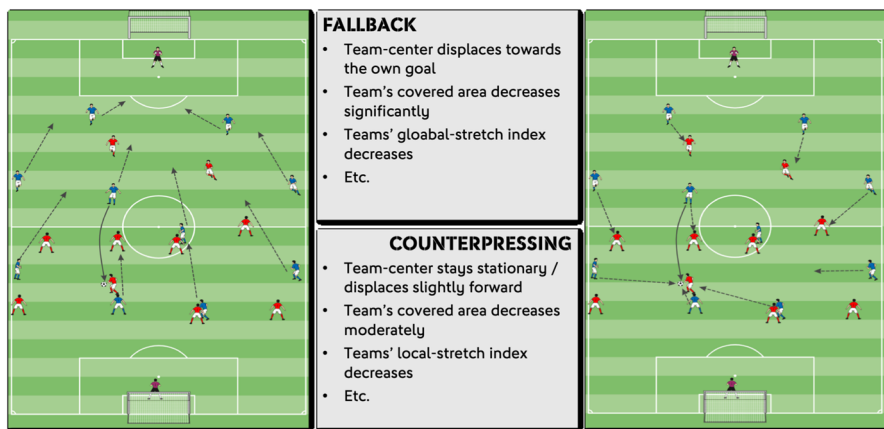


Fig. 7 This figure shows an exemplary ball loss of the blue team, who conducts counterpressing in the right figure and a fallback in the left picture. The pass leading to the turnover is displayed as a solid arrow and the dotted arrows show the expected player movements in the seconds after the ball loss. The metrics *stretch-index* and *team-center* are defined in Table 1

Table 7 Outcome of the manual expert labeling. The first number per entry shows the number of scenes that fulfill the criteria defined in Sect. 2.1.2 and are used for the model building. The total number of tagged scenes is displayed in brackets

Team (Matches)	Defensive Turnovers	Counterpressing	Fallback	Undefined	% CP
All (97)	11,108 (20,928)	3,196 (4,523)	970 (1,268)	6,942 (15,137)	28.77
1. FC Nürnberg (6)	666 (1,167)	178 (228)	94 (111)	394 (828)	26.72
1. FSV Mainz 05 (6)	689 (1,253)	214 (276)	76 (88)	399 (889)	31.06
Bayer 04 Leverkusen (5)	622 (1,092)	175 (262)	57 (63)	390 (767)	28.14
Borussia Dortmund (6)	656 (1,227)	210 (340)	38 (53)	408 (834)	32.01
Borussia Mönchengladbach (5)	534 (963)	214 (289)	46 (50)	274 (624)	40.07
Eintracht Frankfurt (5)	656 (1,092)	206 (273)	61 (75)	389 (744)	31.40
FC Augsburg (5)	583 (1,075)	130 (188)	49 (59)	404 (828)	22.30
FC Bayern München (4)	443 (684)	133 (165)	46 (53)	264 (466)	30.02
FC Schalke 04 (5)	623 (1,054)	149 (190)	47 (56)	427 (808)	23.91
Fortuna Düsseldorf (6)	660 (1,233)	186 (233)	74 (96)	400 (904)	28.18
Hannover 96 (5)	610 (1,079)	133 (156)	44 (55)	433 (868)	21.80
Hertha BSC (5)	548 (1,010)	178 (258)	35 (41)	335 (711)	32.48
RB Leipzig (6)	752 (1,298)	196 (246)	43 (47)	513 (1005)	26.06
Sport-Club Freiburg (5)	556 (1,044)	128 (167)	34 (45)	394 (832)	23.02
SV Werder Bremen (6)	667 (1,176)	224 (304)	60 (66)	383 (806)	33.59
TSG Hoffenheim (6)	677 (1,187)	226 (313)	52 (59)	399 (815)	33.38
VfB Stuttgart (5)	488 (946)	143 (192)	50 (63)	295 (691)	29.30
VfL Wolfsburg (6)	678 (1,181)	173 (215)	64 (77)	441 (889)	25.52

Turnovers in which neither of the two strategies can be identified are labeled as *undefined* or *uncontrolled transition situations*. As pointed out in this study, this usually occurs due to short or very uncontrolled ball possession phases (e.g. headers after a corner situation). These situations typically exhibit a very low (relative) individual ball possession time in the three seconds after the turnover. Therefore, it has a high impact on the resulting XGBoost model (see Fig. 2). In total, 20,928 defensive transition situations from the first eleven matchdays of the 2018/2019 Bundesliga season were labeled based on the definitions formulated above (and in Sect. 2.1.2). The task was to label situations with a detectable strategy with either counterpressing or fallback and not to label any defensive transitions where no strategy was noticeable. Through this procedure the students with a background in football tactics excluded on average 62.50% of all turnover situations by implicitly labeling them as undefined. In total, out of 11,108 relevant defensive turnovers (after the inclusion criteria), 3,196 situations were labeled as counterpressing, 970 as fallback, and 6,942 were explicitly dropped as uncontrolled. Table 7 shows the outcome of the full labeling.

Further evaluation of the labeling outcome showed that more than 95% of all as fallback detected situations start with a goalkeeper catching the ball, what can also be queried solely rule-based on the event data. Due to this fact and because it was our focus to investigate counterpressing situations we decided to exclude the distinction

between fallbacks and undefined ball possession changes for all further investigations. Note that fallback situations last on average 18.30 s, whereas undefined situations were the shortest with 7.83 s on average, which again highlights the tremendous influence on the defensive reaction time (counterpressing: 9.89 s; all turnovers average: 9.34 s).

Appendix B

Table 8 Counterpressing per team for Bundesliga seasons 2013/2014 to 2019/2020

Team (Games)	T/M	CP+/T (%)	%CP	S+/–	G+/-	%CP+	%CP–	%S+	%G+	%S–	%G–
FC Bayern München (229)	112.21	8.02	24.63	– 0.24	0.03	32.56	67.44	3.90	0.66	7.10	0.80
Borussia Dortmund (229)	119.57	8.51	24.87	– 0.16	– 0.04	34.20	65.80	3.61	0.44	6.29	0.87
RB Leipzig (127)	126.69	7.84	22.13	– 0.02	– 0.06	35.41	64.59	3.79	0.51	6.00	1.13
FC Bayer 04 Leverkusen (229)	125.59	8.13	22.91	– 0.24	– 0.04	35.48	64.52	3.66	0.46	6.96	0.92
Borussia Mönchengladbach (229)	104.44	7.27	24.60	– 0.69	– 0.12	29.54	70.4	3.18	0.41	8.32	1.23
Schalke 04 (229)	113.48	7.06	23.27	– 0.59	– 0.11	30.33	69.67	3.11	0.31	7.64	1.07
TSG 1899 Hoffenheim (228)	115.36	7.48	23.27	– 0.43	– 0.04	32.16	67.84	3.82	0.47	8.02	0.94
VfL Wolfsburg (229)	111.14	6.96	23.88	– 0.49	– 0.15	29.14	70.86	2.93	0.20	6.76	1.07
AVERAGE (4118)	115.60	7.25	23.08	– 0.51	– 0.11	31.42%	68.58%	3.22%	0.35%	7.51%	1.13%
Hertha BSC (229)	113.14	7.36	23.67	– 0.50	– 0.10	31.09	68.91	2.58	0.31	6.46	1.02%
1. FC Köln (161)	115.12	6.46	22.45	– 0.54	– 0.16	28.79	71.21	2.91	0.24	7.02%	1.18%
1. FC Augsburg (229)	111.49	6.75	22.53	– 0.62	– 0.14	29.95	70.05	3.25	0.28%	8.14%	1.19%
Eintracht Frankfurt (228)	119.78	7.44	23.50	– 0.68	– 0.10	31.66	68.34	2.77	0.37	7.61	1.07%
SV Werder Bremen (228)	115.01	6.87	23.06	– 0.58	– 0.19	29.79	70.21	3.31	0.20	7.84	1.32%
1. FSV Mainz 05 (229)	120.28	7.48	22.95	– 0.45	– 0.13	32.58	67.42	3.40	0.32	7.46	1.15%
Sport-Club Freiburg (195)	111.33	6.14	20.87	– 0.83	– 0.11	29.40	70.60	2.85	0.26	9.07	1.06%
Fortuna Düsseldorf (59)	103.44	6.83	24.56	– 0.78	– 0.37	27.82	72.18	2.94	0.33	8.32	2.50%
1. FC Union Berlin (25)	114.96	5.50	18.65	– 0.48	– 0.24	29.48	70.52	2.99	0.00	7.41	1.59%
VfB Stuttgart (170)	113.80	7.44	23.41	– 0.50	– 0.14	31.80	68.20	3.33	0.33	7.64	1.23%
ING (68)	147.91	7.55	21.10	– 0.35	– 0.06	35.77	64.23	2.59	0.38	5.80	0.88
Hannover 96 (170)	114.62	6.43	22.05	– 1.02	– 0.25	29.14	70.86	2.61	0.14	9.40	1.61%
Hamburger SV (170)	120.76	7.10	22.05	– 0.42	– 0.21	32.21	67.79	3.42	0.35	7.40	1.66%
SV Darmstadt 98 (68)	118.96	6.21	18.90	– 0.88	– 0.04	32.83	67.17	2.55	0.33	9.64	0.78%
1. FC Nürnberg (68)	102.16	6.36	24.37	– 1.06	– 0.19	26.11	73.89	2.78	0.24	9.51	1.36%
Eintracht Braunschweig (33)	104.42	6.99	22.20	– 0.42	– 0.09	31.50	68.50	3.01	0.13	7.06	0.76%
SC Paderborn 07 (59)	115.51	7.31	23.15	– 0.12	– 0.12	31.56	68.44	3.36	0.19	5.56	0.93%

Appendix C

Table 9 Counterpressing per coach for Bundesliga seasons 2013/2014 to 2019/2020. Only full seasons with in total 34 matches per coach are taken into consideration

Coach (Games)	T/M	CP+/T (%)	%CP	S+/-	G+/-	% CP+	%CP-	% S+	% G+	% S-	% G-
Pep Guardiola (102)	109.47	8.61	25.24	- 0.21	0.02	34.10	65.90	3.87	0.50	7.00	0.65
Thomas Tuchel (102)	113.00	8.60	24.61	- 0.27	- 0.06	34.93	65.07	3.49	0.53	6.88	1.14
Jürgen Klopp (68)	129.00	8.54	23.51	0.26	- 0.04	36.32	63.68	4.22	0.29	5.26	0.69
Roger Schmidt (34)	156.29	8.49	18.84	0.91	0.09	45.05	54.95	5.29	0.80	4.00	0.91
Jos Luhukay (34)	104.97	8.27	24.99	- 0.12	- 0.09	33.07	66.93	2.80	0.11	4.86	0.67
Adi Hütter (34)	128.94	8.05	23.27	- 0.59	- 0.09	34.61	65.3	3.63	0.39	8.55	1.05
Markus Gisdol (34)	141.03	8.03	22.34	0.24	0.12	35.95	64.05	4.30	0.56	5.54	0.29
Ralph Hasenhtütl (102)	136.35	8.00	21.99	- 0.27	- 0.03	36.38	63.62	2.91	0.56	6.01	1.03
Sandro Schwarz (68)	120.76	7.93	24.05	- 0.26	- 0.18	32.96	67.0	3.70	0.25	6.87	1.28
Carlo Ancelotti (34)	120.03	7.69	23.52	- 0.56	0.15	32.71	67.29	2.60	0.83	6.81	0.46
Heiko Herrlich (34)	115.71	7.68	23.64	- 0.44	- 0.03	32.47	67.53	3.76	0.54	7.96	0.96
Martin Schmidt (68)	128.29	7.62	21.92	- 0.46	- 0.06	34.78	65.22	2.98	0.37	7.06	0.88
Domenico Tedesco (34)	113.94	7.56	23.88	- 0.12	- 0.09	31.68	68.32	3.35	0.22	5.54	0.79
Niko Kovac (102)	118.18	7.46	23.92	- 0.28	- 0.03	31.18	68.82	3.09	0.55	5.95	0.96
AVERAGE (2176)	116.22	7.40	23.06	- 0.39	- 0.08	32.07	67.93	3.28%	0.38%	6.99%	1.01%
Julian Nagelsmann (102)	110.10	7.36	23.70	- 0.38	0.03	31.03	68.97	4.43	0.64	8.55	0.76
Thomas Schaaf (34)	127.44	7.29	22.62	- 0.24	- 0.12	32.24	67.76	2.96	0.31	5.57	1.05

Table 9 continued

Coach (Games)	T/M	CP+/T (%)	%CP	S+/-	G+/-	% CP+	%CP-	% S+	% G+	% S-	% G-
Dieter Hecking (170)	106.00	7.27	24.13	- 0.41	- 0.14	30.13	69.87	3.40	0.39	7.18	1.32
Florian Kohfeldt (34)	111.24	7.22	23.43	- 0.71	- 0.26	30.81	69.19	3.61	0.11	9.14	1.63
Markus Weinzierl (136)	109.07	7.21	23.53	- 0.46	- 0.10	30.66	69.34	3.30	0.32	7.31	0.99
Pal Dardai (136)	113.18	7.15	23.36	- 0.69	- 0.08	30.62	69.38	2.50	0.33	7.37	0.92
Bruno Labbadia (68)	114.91	7.06	24.75	- 0.63	- 0.19	28.54	71.46	2.95	0.21	7.24	1.23
Robin Dutt (34)	106.94	7.04	23.93	- 0.56	- 0.24	29.43	70.57	3.45	0.23	7.98	1.63
Ralf Rangnick (34)	131.47	7.00	21.72	0.56	- 0.03	32.23	67.77	4.63	0.41	3.95	0.76
Lucien Favre (68)	94.10	6.95	25.38	- 0.87	- 0.03	27.40	72.60	2.65	0.43	8.65	0.76
Armin Veh (34)	93.38	6.93	25.98	- 0.79	- 0.24	26.67	73.33	2.91	0.00	8.43	1.32
Andr Breitenreiter (102)	117.86	6.82	22.20	- 0.67	- 0.18	30.72	69.28	3.07	0.26	8.11	1.35
Viktor Skripnik (34)	125.35	6.69	21.73	- 0.32	- 0.18	30.78	69.22	3.02	0.32	6.08	1.40
Friedhelm Funkel (34)	100.44	6.62	24.77	- 0.59	- 0.47	26.71	73.29	3.43	0.35	7.90	3.06
Peter Stöger (102)	116.09	6.60	21.81	- 0.35	- 0.12	30.25	69.75	2.48	0.19	5.55	0.94
Manuel Baum (34)	121.62	6.41	21.45	- 0.65	0.00	29.88	70.12	2.71	0.34	7.40	0.48
Christian Streich (136)	114.29	6.26	20.34	- 0.68	- 0.07	30.77	69.23	2.66	0.25	8.04	0.78
Dirk Schuster (34)	125.82	5.82	16.88	- 0.65	- 0.03	34.49	65.51	2.35	0.28	8.25	0.63

Appendix D

Table 10 Counterpressing of Bundesliga teams on a season level (1/2). The number in brackets shows the respective position in the final ranking

Team (Ranking)	T/M	CP+/T	%CP	S+/-	G+/-	%CP+	%CP-	%S+	%G+	%S-	%G-
FCB - 2013/2014 (1)	97.29	8.98%	26.18%	-0.71	0.03	34.30%	65.70%	3.58%	0.46%	9.67%	0.53%
FCB - 2014/2015 (1)	116.65	8.50%	24.79%	0.00	0.06	34.28%	65.72%	3.87%	0.51%	5.88%	0.46%
FCB - 2015/2016 (1)	114.47	8.40%	24.90%	0.09	-0.03	33.75%	66.25%	4.13%	0.52%	5.76%	0.93%
FCB - 2019/2020 (1)	116.68	8.40%	26.02%	-0.32	-0.04	32.28%	67.72%	4.87%	0.66%	8.75%	1.17%
FCB - 2016/2017 (1)	120.03	7.69%	23.52%	-0.56	0.15	32.71%	67.29%	2.60%	0.83%	6.81%	0.46%
FCB - 2018/2019 (1)	111.15	7.33%	24.21%	0.32	0.12	30.27%	69.73%	5.14%	1.20%	5.64%	1.10%
FCB - 2017/2018 (1)	110.41	7.03%	23.39%	-0.56	-0.06	30.07%	69.93%	3.30%	0.46%	7.82%	0.98%
BVB - 2015/2016 (2)	120.79	9.40%	24.71%	-0.09	0.03	38.03%	61.97%	3.15%	0.59%	5.56%	0.79%
BVB - 2013/2014 (2)	113.12	8.71%	24.60%	-0.12	-0.09	35.41%	64.59%	3.59%	0.21%	6.22%	0.82%
BVB - 2018/2019 (2)	105.82	8.59%	26.57%	-0.32	-0.12	32.32%	67.68%	2.41%	0.21%	5.26%	0.93%
RBL - 2016/2017 (2)	134.97	8.35%	21.16%	-0.24	-0.06	39.44%	60.56%	2.68%	0.72%	5.78%	1.53%
BVB - 2019/2020 (2)	104.00	8.19%	27.65%	-0.32	-0.08	29.62%	70.38%	4.17%	0.70%	7.51%	1.38%
S04 - 2017/2018 (2)	113.94	7.56%	23.88%	-0.12	-0.09	31.68%	68.32%	3.35%	0.22%	5.54%	0.79%
WOB - 2014/2015 (2)	118.44	7.18%	22.00%	0.15	0.00	32.62%	67.38%	4.63%	0.45%	6.03%	0.67%
LEV - 2015/2016 (3)	143.85	8.79%	22.33%	-0.24	0.15	39.38%	60.62%	2.66%	0.55%	5.59%	0.15%
BVB - 2016/2017 (3)	122.38	8.29%	24.75%	-0.24	-0.03	33.50%	66.50%	3.69%	0.58%	6.72%	1.02%

Table 10 continued

Team (Ranking)	T/M	CP+T	%CP	S+/-	G+/-	%CP+	%CP-	%S+	%G+	%S-	%G-
RBL - 2019/2020 (3)	110.60	7.88%	22.21%	- 0.20	- 0.12	35.50%	64.50%	4.07%	0.33%	7.58%	1.26%
HOF - 2017/2018 (3)	109.15	7.49%	23.31%	- 0.18	0.12	32.14%	67.86%	4.74%	0.81%	8.01%	0.51%
RBL - 2018/2019 (3)	131.47	7.00%	21.72%	0.56	- 0.03	32.23%	67.77%	4.63%	0.41%	3.95%	0.76%
S04 - 2013/2014 (3)	96.74	6.87%	23.44%	- 0.03	0.03	29.31%	70.69%	4.02%	0.65%	5.87%	0.73%
BMG - 2014/2015 (3)	103.12	6.85%	24.02%	- 0.62	- 0.03	28.50%	71.50%	2.85%	0.36%	7.48%	0.66%
LEV - 2014/2015 (4)	156.29	8.49%	18.84%	0.91	0.09	45.05%	54.95%	5.29%	0.80%	4.00%	0.91%
LEV - 2019/2020 (4)	114.76	8.16%	23.60%	- 0.64	- 0.12	34.56%	65.44%	3.10%	0.00%	8.35%	0.68%
BVB - 2017/2018 (4)	121.85	7.89%	24.79%	- 0.71	0.00	31.84%	68.16%	3.51%	0.49%	8.57%	0.71%
BMG - 2015/2016 (4)	117.32	7.60%	24.34%	- 0.47	- 0.15	31.20%	68.80%	3.30%	0.41%	7.19%	1.35%
HOF - 2016/2017 (4)	107.94	7.52%	23.98%	- 0.35	- 0.06	31.36%	68.64%	3.64%	0.34%	7.28%	0.83%
LEV - 2013/2014 (4)	101.26	7.44%	26.08%	- 0.38	- 0.06	28.51%	71.49%	3.01%	0.56%	6.23%	1.09%
LEV - 2018/2019 (4)	114.68	7.28%	24.26%	- 0.29	- 0.15	30.02%	69.98%	4.65%	0.21%	8.16%	1.06%
LEV - 2017/2018 (5)	115.71	7.68%	23.64%	- 0.44	- 0.03	32.47%	67.53%	3.76%	0.54%	7.96%	0.96%
AUG - 2014/2015 (5)	121.35	7.54%	22.95%	- 0.85	- 0.21	32.84%	67.16%	2.75%	0.21%	8.65%	1.42%
WOB - 2013/2014 (5)	94.82	7.44%	24.88%	- 0.41	- 0.12	29.93%	70.07%	2.24%	0.12%	5.69%	0.89%
BMG - 2018/2019 (5)	99.29	7.35%	26.04%	- 0.85	- 0.29	28.21%	71.79%	3.98%	0.34%	10.14%	2.06%
KOE - 2016/2017 (5)	113.41	6.95%	22.15%	- 0.26	- 0.12	31.38%	68.62%	2.34%	0.00%	4.95%	0.68%
S04 - 2015/2016 (5)	115.44	6.80%	22.39%	- 0.76	- 0.12	30.38%	69.62%	3.30%	0.34%	8.99%	1.14%

Table 10 continued

Team (Ranking)	T/M	CP+/T	%CP	S+/-	G+/-	%CP+	%CP-	%S+	%G+	%S-	%G-
BMG - 2019/2020 (5)	110.48	6.59%	22.95%	- 0.56	- 0.24	28.71%	71.29%	2.84%	0.32%	7.08%	1.77%
RBL - 2017/2018 (6)	125.44	8.14%	23.56%	- 0.26	- 0.06	34.53%	65.47%	3.88%	0.50%	7.29%	1.06%
M05 - 2015/2016 (6)	123.71	7.75%	22.92%	- 0.50	0.03	33.82%	66.18%	3.01%	0.31%	7.21%	0.31%
HBER - 2016/2017 (6)	115.29	7.73%	22.65%	- 0.47	- 0.03	34.12%	65.88%	2.48%	0.34%	6.50%	0.68%
WOB - 2018/2019 (6)	110.38	7.49%	27.15%	- 0.79	- 0.21	27.58%	72.42%	2.85%	0.29%	7.59%	1.36%
BMG - 2013/2014 (6)	85.09	7.09%	27.03%	- 1.12	- 0.03	26.21%	73.79%	2.43%	0.51%	9.88%	0.87%
WOB - 2019/2020 (6)	123.36	6.74%	23.09%	- 0.40	- 0.12	29.21%	70.79%	3.23%	0.14%	6.55%	0.79%
S04 - 2014/2015 (6)	123.56	6.19%	20.97%	- 1.03	0.00	29.51%	70.49%	2.50%	0.34%	9.18%	0.48%
BVB - 2014/2015 (7)	144.88	8.40%	22.66%	0.65	0.00	37.10%	62.90%	4.75%	0.36%	4.42%	0.57%
FRA - 2018/2019 (7)	128.94	8.05%	23.27%	- 0.59	- 0.09	34.61%	65.39%	3.63%	0.39%	8.55%	1.05%
M05 - 2013/2014 (7)	95.82	7.98%	24.31%	- 0.50	- 0.18	32.83%	67.17%	3.66%	0.38%	8.65%	1.69%
STU - 2017/2018 (7)	114.26	7.54%	24.53%	- 0.71	- 0.06	30.75%	69.25%	2.31%	0.31%	6.97%	0.76%
HBER - 2015/2016 (7)	111.53	7.30%	22.97%	- 0.56	- 0.09	31.80%	68.20%	2.18%	0.34%	6.40%	1.01%
SCF - 2016/2017 (7)	122.44	5.79%	19.10%	- 0.76	- 0.03	30.31%	69.69%	2.26%	0.25%	7.94%	0.54%
SCF - 2019/2020 (7)	106.48	5.48%	22.46%	- 0.84	- 0.44	24.41%	75.59%	3.18%	0.17%	14.38%	2.65%
HOF - 2014/2015 (8)	141.03	8.03%	22.34%	0.24	0.12	35.95%	64.05%	4.30%	0.56%	5.54%	0.29%
BRE - 2018/2019 (8)	111.24	7.22%	23.43%	- 0.71	- 0.26	30.81%	69.19%	3.61%	0.11%	9.14%	1.63%
S04 - 2019/2020 (8)	113.68	7.14%	24.77%	- 0.88	- 0.24	28.84%	71.16%	2.84%	0.14%	8.38%	1.40%

Table 10 continued

Team (Ranking)	T/M	CP+T	%CP	S+/-	G+/-	%CP+	%CP-	%S+	%G+	%S-	%G-
AUG - 2013/2014 (8)	93.50	7.14%	25.01%	- 0.12	0.00	28.55%	71.45%	3.65%	0.63%	5.81%	0.88%
FRA - 2017/2018 (8)	120.18	7.12%	25.01%	- 0.97	- 0.12	28.47%	71.53%	2.25%	0.39%	7.66%	1.09%
BRE - 2016/2017 (8)	110.15	6.92%	23.68%	- 0.76	- 0.24	29.20%	70.80%	3.16%	0.34%	8.60%	1.75%
WOB - 2015/2016 (8)	109.85	6.53%	23.51%	- 0.35	- 0.26	27.79%	72.21%	2.62%	0.34%	5.52%	1.89%
BMG - 2017/2018 (9)	107.59	7.90%	24.69%	- 0.59	0.00	32.00%	68.00%	3.43%	0.66%	8.31%	0.98%
HOF - 2013/2014 (9)	108.24	7.61%	23.91%	- 0.33	- 0.18	31.85%	68.15%	3.28%	0.23%	6.70%	1.37%
FRA - 2014/2015 (9)	127.44	7.29%	22.62%	- 0.24	- 0.12	32.24%	67.76%	2.96%	0.31%	5.57%	1.05%
BMG - 2016/2017 (9)	109.79	7.26%	23.36%	- 0.59	- 0.12	31.08%	68.92%	3.21%	0.23%	7.99%	1.00%
AVERAGE (9)	115.60	7.25%	23.08%	- 0.51	- 0.11	31.42%	68.58%	3.22%	0.35%	7.51%	1.13%
HOF - 2018/2019 (9)	113.21	7.07%	23.82%	- 0.62	0.03	29.66%	70.34%	4.91%	0.76%	10.23%	0.93%
KOE - 2015/2016 (9)	115.47	6.11%	21.88%	- 0.59	- 0.12	27.94%	72.06%	2.10%	0.12%	6.14%	0.81%
HOF - 2019/2020 (9)	102.44	5.66%	22.73%	- 1.28	- 0.12	24.91%	75.09%	2.75%	0.69%	10.98%	1.60%
S04 - 2016/2017 (10)	114.38	7.17%	23.68%	- 0.47	- 0.09	30.29%	69.71%	3.04%	0.43%	6.85%	1.09%
BRE - 2014/2015 (10)	127.21	6.68%	21.39%	- 0.44	- 0.12	31.24%	68.76%	3.57%	0.00%	7.55%	0.63%
HSV - 2015/2016 (10)	119.44	6.67%	22.53%	- 0.47	- 0.18	29.62%	70.38%	3.06%	0.11%	6.83%	1.09%
DUE - 2018/2019 (10)	100.44	6.62%	24.77%	- 0.59	- 0.47	26.71%	73.29%	3.43%	0.35%	7.90%	3.06%
H96 - 2013/2014 (10)	98.32	6.61%	23.24%	- 0.38	- 0.15	28.44%	71.56%	4.12%	0.26%	8.09%	1.26%

Table 10 continued

Team (Ranking)	T/M	CP+/T	%CP	S+/-	G+/-	%CP+	%CP-	%S+	%G+	%S-	%G-
KOE - 2019/2020 (10)	106.52	6.57%	23.73%	- 0.96	- 0.28	27.69%	72.31%	3.16%	0.00%	9.63%	1.53%
HBer - 2017/2018 (10)	112.94	6.25%	23.41%	- 0.62	- 0.09	26.70%	73.30%	2.89%	0.33%	7.13%	0.91%
HBer - 2013/2014 (11)	104.97	8.27%	24.99%	- 0.12	- 0.09	33.07%	66.93%	2.80%	0.11%	4.86%	0.67%
FRA - 2016/2017 (11)	123.21	7.90%	22.58%	- 0.21	- 0.09	34.99%	65.01%	2.01%	0.11%	4.23%	0.65%
ING - 2015/2016 (11)	148.65	7.58%	21.43%	- 0.32	0.03	35.36%	64.64%	2.22%	0.46%	5.00%	0.57%
HBer - 2019/2020 (11)	104.04	7.38%	25.03%	- 0.52	- 0.24	29.49%	70.51%	2.76%	0.31%	6.75%	1.74%
HBer - 2018/2019 (11)	112.94	7.32%	24.43%	- 1.12	- 0.12	29.96%	70.04%	2.45%	0.32%	9.28%	1.07%
BRE - 2017/2018 (11)	116.06	6.79%	24.00%	- 0.68	- 0.12	28.30%	71.70%	2.75%	0.11%	7.22%	0.74%
M05 - 2014/2015 (11)	129.59	6.13%	21.56%	- 0.59	- 0.09	28.42%	71.58%	3.47%	0.21%	7.79%	0.74%

Table 11 Counterpressing of Bundesliga teams on a season level (2/2). The number in brackets shows the respective position in the final ranking

Team (Ranking)	T/M	CP+/T	%CP	S+/-	G+/-	%CP+	%CP-	%S+	%G+	%S-	%G-
LEV - 2016/2017 (12)	129.71	8.64%	23.70%	- 0.71	- 0.18	36.46%	63.54%	3.06%	0.38%	8.43%	1.51%
M05 - 2018/2019 (12)	121.79	7.49%	24.70%	- 0.62	- 0.26	30.30%	69.70%	3.42%	0.10%	7.85%	1.40%
BRE - 2013/2014 (12)	106.94	7.04%	23.93%	- 0.56	- 0.24	29.43%	70.57%	3.45%	0.23%	7.98%	1.63%
AUG - 2015/2016 (12)	107.03	6.95%	22.73%	- 0.38	- 0.09	30.59%	69.41%	3.87%	0.00%	7.84%	0.52%
AUG - 2019/2020 (12)	104.40	6.82%	22.68%	- 0.52	- 0.36	30.07%	69.93%	3.55%	0.68%	8.21%	3.14%
KOE - 2014/2015 (12)	119.38	6.73%	21.41%	- 0.21	- 0.12	31.42%	68.58%	2.99%	0.46%	5.54%	1.34%
AUG - 2017/2018 (12)	121.62	6.41%	21.45%	- 0.65	0.00	29.88%	70.12%	2.71%	0.34%	7.40%	0.48%
FRA - 2013/2014 (13)	93.38	6.93%	25.98%	- 0.79	- 0.24	26.67%	73.33%	2.91%	0.00%	8.43%	1.32%
BRE - 2015/2016 (13)	125.35	6.69%	21.73%	- 0.32	- 0.18	30.78%	69.22%	3.02%	0.32%	6.08%	1.40%
H96 - 2017/2018 (13)	115.74	6.45%	22.21%	- 1.06	- 0.35	29.06%	70.94%	2.52%	0.11%	9.35%	2.10%
AUG - 2016/2017 (13)	118.97	6.18%	21.63%	- 0.71	- 0.09	28.57%	71.43%	3.20%	0.23%	8.32%	0.80%
H96 - 2014/2015 (13)	130.62	6.10%	19.88%	- 1.06	- 0.21	30.69%	69.31%	1.81%	0.00%	8.50%	1.14%
SCF - 2018/2019 (13)	103.03	6.08%	22.01%	- 0.68	- 0.06	27.63%	72.37%	3.37%	0.39%	8.78%	0.90%
UBER - 2019/2020 (13)	114.96	5.50%	18.65%	- 0.48	- 0.24	29.48%	70.52%	2.99%	0.00%	7.41%	1.59%
M05 - 2017/2018 (14)	119.74	8.38%	23.38%	0.09	- 0.09	35.82%	64.18%	3.99%	0.42%	5.73%	1.15%
S04 - 2018/2019 (14)	116.68	7.71%	24.33%	- 0.88	- 0.32	31.71%	68.29%	2.80%	0.10%	8.65%	1.82%
HSV - 2016/2017 (14)	139.88	7.59%	20.90%	0.24	0.03	36.32%	63.68%	3.52%	0.50%	4.27%	0.63%
STU - 2014/2015 (14)	120.24	7.46%	22.06%	- 0.65	- 0.26	33.81%	66.19%	2.88%	0.22%	8.04%	1.84%
FRA - 2019/2020 (14)	123.00	7.35%	21.71%	- 1.12	- 0.12	33.85%	66.15%	3.28%	0.62%	11.32%	1.65%
SCF - 2013/2014 (14)	96.24	6.57%	22.31%	- 0.79	- 0.12	29.45%	70.55%	2.33%	0.14%	8.54%	0.97%

Table 11 continued

Team (Ranking)	T/M	CP+/T	%CP	S+/-	G+/-	%CP+	%CP-	%S+	%G+	%S-	%G-
DAR - 2015/2016 (14)	125.82	5.82%	16.88%	- 0.65	- 0.03	34.49%	65.51%	2.35%	0.28%	8.25%	0.63%
STU - 2013/2014 (15)	92.38	8.28%	25.66%	0.03	- 0.03	32.26%	67.74%	4.22%	0.50%	6.04%	0.92%
HOF - 2015/2016 (15)	121.85	8.21%	22.95%	- 0.74	- 0.24	35.75%	64.25%	2.73%	0.00%	8.35%	1.31%
M05 - 2016/2017 (15)	132.88	7.50%	20.98%	- 0.41	- 0.15	35.76%	64.24%	2.95%	0.42%	6.90%	1.48%
HBER - 2014/2015 (15)	127.85	7.34%	22.87%	- 0.12	- 0.12	32.09%	67.91%	2.52%	0.40%	4.30%	1.19%
M05 - 2019/2020 (15)	117.72	7.24%	23.48%	- 0.68	- 0.16	30.82%	69.18%	3.33%	0.43%	8.37%	1.46%
SCF - 2017/2018 (15)	114.32	6.51%	20.63%	- 0.68	- 0.18	31.55%	68.45%	2.87%	0.12%	8.38%	1.28%
AUG - 2018/2019 (15)	111.68	6.29%	21.83%	- 1.06	- 0.29	28.83%	71.17%	3.26%	0.00%	10.68%	1.69%
HSV - 2013/2014 (16)	95.32	7.44%	22.96%	- 0.59	- 0.44	32.39%	67.61%	4.17%	0.40%	10.14%	3.58%
FRA - 2015/2016 (16)	123.26	7.25%	23.50%	- 1.00	0.06	30.86%	69.14%	2.54%	0.81%	8.66%	0.88%
DUE - 2019/2020 (16)	107.52	7.11%	24.29%	- 1.04	- 0.24	29.25%	70.75%	2.30%	0.31%	8.87%	1.73%
WOB - 2016/2017 (16)	117.50	6.93%	22.38%	- 0.74	- 0.18	30.98%	69.02%	2.46%	0.00%	7.62%	0.97%
STU - 2018/2019 (16)	116.62	6.86%	22.65%	- 0.59	- 0.18	30.29%	69.71%	3.79%	0.11%	8.63%	1.12%
HSV - 2014/2015 (16)	119.79	6.58%	21.46%	- 0.50	- 0.21	30.66%	69.34%	3.20%	0.46%	7.43%	1.82%
WOB - 2017/2018 (16)	106.88	6.38%	24.38%	- 0.88	- 0.15	26.19%	73.81%	2.48%	0.00%	7.95%	0.76%
ING - 2016/2017 (17)	147.18	7.51%	20.76%	- 0.38	- 0.15	36.19%	63.81%	2.98%	0.29%	6.64%	1.21%
STU - 2015/2016 (17)	125.50	7.27%	22.73%	- 0.59	- 0.15	31.96%	68.04%	3.61%	0.52%	8.33%	1.52%
HSV - 2017/2018 (17)	129.35	7.21%	22.71%	- 0.79	- 0.24	31.73%	68.27%	3.30%	0.30%	8.80%	1.61%

Table 11 continued

Team (Ranking)	T/M	CP+/T	%CP	S+/-	G+/-	%CP+	%CP-	%S+	%G+	%S-	%G-
NUR - 2013/2014 (17)	99.47	6.86%	24.78%	- 1.09	- 0.12	27.68%	72.32%	3.58%	0.36%	11.06%	1.16%
BRE - 2019/2020 (17)	105.25	6.81%	24.03%	- 0.62	- 0.21	28.34%	71.66%	3.79%	0.33%	8.74%	1.61%
H96 - 2018/2019 (17)	109.35	6.70%	23.02%	- 1.59	- 0.32	29.09%	70.91%	2.22%	0.12%	12.03%	1.98%
SCF - 2014/2015 (17)	124.18	6.25%	19.78%	- 0.47	0.06	31.62%	68.38%	3.11%	0.48%	7.36%	0.35%
PAD - 2019/2020 (18)	106.12	7.50%	24.95%	- 0.04	- 0.20	30.06%	69.94%	3.32%	0.00%	4.97%	1.08%
PAD - 2014/2015 (18)	122.41	7.18%	22.01%	- 0.18	- 0.06	32.64%	67.36%	3.38%	0.33%	6.00%	0.81%
BRA - 2013/2014 (18)	104.42	6.99%	22.20%	- 0.42	- 0.09	31.50%	68.50%	3.01%	0.13%	7.06%	0.76%
DAR - 2016/2017 (18)	112.09	6.64%	21.18%	- 1.12	- 0.06	31.35%	68.65%	2.73%	0.37%	10.83%	0.90%
H96 - 2015/2016 (18)	119.06	6.35%	22.38%	- 1.03	- 0.24	28.37%	71.63%	2.54%	0.22%	8.94%	1.54%
KOE - 2017/2018 (18)	118.53	6.00%	23.50%	- 0.79	- 0.18	25.55%	74.45%	3.91%	0.53%	9.08%	1.56%
NUR - 2018/2019 (18)	104.85	5.89%	23.98%	- 1.03	- 0.26	24.56%	75.44%	1.99%	0.12%	8.06%	1.55%

References

- Andrienko G et al (2017) Visual analysis of pressure in football. *Data Mining Knowl Discov* 31(6):1793–1839. <https://doi.org/10.1007/s10618-017-0513-2>
- Andrienko G et al (2019) Constructing Spaces and Times for Tactical Analysis in Football. *IEEE Trans Vis Comput Graph*. <https://doi.org/10.1109/tvcg.2019.2952129>
- Antipov EA, Pokryshevskaya EB (2020) Interpretable machine learning for demand modeling with high-dimensional data using Gradient Boosting Machines and Shapley values. *J Revenue Pricing Manag* 19(5):355–364. <https://doi.org/10.1057/s41272-020-00236-4>
- Anzer G, Bauer P (2021) “A Goal Scoring Probability Model based on Synchronized Positional and Event Data”. *Frontiers in Sports and Active Learning* (in print) Using Artificial Intelligence to Enhance Sport Performance, pp. 1–18. <https://doi.org/10.3389/fspor.2021.624475>.
- Bergstra J et al (2011) “Algorithms for hyper-parameter optimization”. In: *Advances in Neural Information Processing Systems 24: 25th Annual Conference on Neural Information Processing Systems 2011, NIPS 2011*, pp. 1–9
- Bialkowski A et al (2014) “Large-Scale Analysis of Soccer Matches Using Spatiotemporal Tracking Data”. In: *Proceedings - IEEE International Conference on Data Mining, ICDM 2015-January*, pp. 725–730. issn:15504786. <https://doi.org/10.1109/ICDM.2014.133>
- Bialkowski A et al (2015) “Identifying team style in soccer using formations learned from spatiotemporal tracking data”. In: *IEEE International Conference on Data Mining Workshops, ICDMW 2015-January*, pp. 9–14. issn: 23759259. <https://doi.org/10.1109/ICDMW.2014.167.29> <https://gdpr-info.eu/>, accessed 07/20/20. Data-Driven Detection of Counterpressing in Professional Football 15
- Bojinov I, Bornn L (2016) “The Pressing Game: Optimal Defensive Disruption in Soccer”. In: *MIT Sloan Sports Analytics Conference*, pp. 1–8
- Bourbousson Jérôme, Carole Sève Tim McGarry (2010) “Space-time coordination dynamics in basketball: Part 2. the interaction between the two teams”. In: *Journal of Sports Sciences* 28.3, pp. 349–358. issn: 02640414. <https://doi.org/10.1080/02640410903503640>
- Brefeld U, Lasek J, Mair S (2019) Probabilistic movement models and zones of control. *Mach Learn* 108(1):127–147. <https://doi.org/10.1007/s10994-018-5725-1>
- Chen T, Guestrin C (2016) “XGBoost: A scalable tree boosting system”. In: *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining 13-17-Aug.*, pp. 785–794. <https://doi.org/10.1145/2939672.2939785>.
- Decroos T, Van Haaren J, Davis J (2018) “Automatic discovery of tactics in spatio-temporal soccer match data”. In: *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. isbn: 9781450355520. <https://doi.org/10.1145/3219819.3219832>. url: <https://people.cs.kuleuven.be/~jesse.davis/decroos-kdd18.pdf>
- Decroos T et al (2020) “VAEP: An Objective Approach to Valuing On-the-Ball Actions in Soccer (Extended Abstract)”. In: *Proceedings of the Twenty-Ninth International Joint Conference on Artificial Intelligence (IJCAI-20)*, pp. 4696–4700. issn: 10450823. <https://doi.org/10.24963/ijcai.2020/648>.
- Dewnacker Ian, Michael McCourt, Scott Clark (2016) “Bayesian Optimization for Machine Learning. A Practical Guidebook”. [arXiv:1612.04858](https://arxiv.org/abs/1612.04858)
- Fairchild A, Pelechrinis K, Kokkosis M (2018) “Spatial analysis of shots in MLS: A model for expected goals and fractal dimensionality”. In: *Journal of Sports Analytics* 4.3, pp. 165–174. issn: 2215020X. <https://doi.org/10.3233/jsa-170207>.
- Navarro Fernández, J (2018) *Analysis of Styles of Play in Soccer and Their Effectiveness*. isbn: 9788413060576
- Fernandez J, Bornn L (2018) “Wide Open Spaces : A statistical technique for measuring space creation in professional soccer”. In: *MIT Sloan Sports Analytics Conference*, pp. 1–19
- Goes F, Kempe M, Koen L (2019). “Predicting match outcome in professional Dutch football using tactical performance metrics computed from position tracking data”. In: *MathSport International Conference* June, pp. 4–5
- Goes F et al (2020) Interpretable machine learning for demand modeling with high-dimensional data using Gradient Boosting Machines and Shapley values. *Euro J Sport Sci*. <https://doi.org/10.1080/17461391.2020.1747552>
- Grant AG et al (1999) Analysis of the goals scored in the 1998 World Cup. *J Sports Sci* 17(10):826–827

- Herold M et al (2019) Machine learning in men's professional football: current applications and future directions for improving attacking play. *Int J Sports Sci Coaching*. <https://doi.org/10.1177/1747954119879350>
- Hobbs J et al. (2018). "Quantifying the Value of Transitions in Soccer via Spatiotemporal Trajectory Clustering". In: MIT Sloan Sports Analytics Conference, pp. 1–11
- Hughes M, Ian M Franks (2015) Essentials of performance analysis in sport. Vol. 53. 04, pp. 53–1831. <https://doi.org/10.5860/choice.193440>. url:https://books.google.de/books?hl=de&lr=&id=KorCDwAAQB&oi=fnd&pg=PT14&dq=the+essentials+of+performance+analysis&ots=ZhJd6413Fq&sig=F_MnUcL21bpzACnWTzNhxESEgUY4#v=onepage&q=theessentials+of+performance+analysis&f=false
- Ibrahim L et al (2020) Explainable prediction of acute myocardial infarction using machine learning and shapley values. *IEEE Access*. <https://doi.org/10.1109/access.2020.3040166>
- Kempe M et al (2014) Possession vs. direct play: evaluating tactical behavior in elite soccer. *Int J Sports Sci* 4(6A):35–41. <https://doi.org/10.5923/s.sports.201401.05>
- Kim S (2004) Voronoi Analysis of a Soccer Game. *Nonlinear Anal Model Control* 9(3):233–240. <https://doi.org/10.15388/na.2004.9.3.15154>
- Li TR et al (2019). Sentiment-based prediction of alternative cryptocurrency price fluctuations using gradient boosting tree model. <https://doi.org/10.3389/fphy.2019.00098>
- Link D, Lang S, Seidenschwarz P (2016) Real time quantification of dangerousity in football using spatiotemporal tracking data. *PLoS ONE*. <https://doi.org/10.1371/journal.pone.0168768>
- Link D, Hoernig M (2017) Individual ball possession in soccer. *PLoS ONE* 12(7):1–15. <https://doi.org/10.1371/journal.pone.0179953>
- Liu K, Chen W, Lin H (2020) XG-PseU: an eXtreme Gradient Boosting based method for identifying pseudouridine sites. *Mole Genet Genom* 295(1):13–21. <https://doi.org/10.1007/s00438-019-01600-9>
- Patrick L et al (2014) "Quality vs Quantity": Improved Shot Prediction in Soccer using Strategic Features from Spatiotemporal Data". In: MIT Sloan Sports Analytics Conference, pp. 1–9. url: <http://www.sloansportsconference.com/?p=15790>
- Lundberg SM, SI Lee (2017) "Consistent feature attribution for tree ensembles". In: Proceedings of the 34th International Conference on Machine Learning, pp. 1–9
- Yuan M et al (2020) What Makes an Online Review More Helpful: An Interpretation Framework Using XGBoost and SHAP Values. *J Theor Appl Electron Commerce Res* 16(3):466–490. <https://doi.org/10.3390/jtaer16030029>
- Pappalardo L et al (2019) A public data set of spatio-temporal match events in soccer competitions. *Scientific Data* 6(1):236. <https://doi.org/10.1038/s41597-019-0247-7>
- Paul P et al (2018) "Mythbusting Set-Pieces in Soccer". In: MIT Sloan Sports Analytics Conference, pp. 1–12
- Rathke A (2017) "An examination of expected goals and shot efficiency in soccer". In: *Journal of Human Sport and Exercise* 12.Proc2. issn: 1988-5202. <https://doi.org/10.14198/jhse.2017.12.proc2.05>. url: <http://www.redalyc.org/articulo.oa?id=301052437005>
- Ratner Alexander J et al (2017) "Learning to compose domain-specific transformations for data augmentation". In: *Advances in Neural Information Processing Systems Nips*, pp. 3237–3247. issn: 10495258
- Ratner A et al (2016) "Data programming: Creating large training sets, quickly". In: *Advances in Neural Information Processing Systems Nips*, pp. 3574–3582. issn: 10495258
- Reep C, Benjamin B (1968) Skill and Chance in Association Football. *J Royal Stat Soc Series A (General)*. <https://doi.org/10.2307/2343726>
- Robert R, Memmert D (2016) "Big data and tactical analysis in elite soccer: future challenges and opportunities for sports science". In: *SpringerPlus* 5.1. issn: 21931801. <https://doi.org/10.1186/s40064-016-3108-2>.
- Robberechts P (2019) "Valuing the Art of Pressing". In: *StatsBomb Innovation in Football Conference* 2019, p. 11. url: <http://statsbomb.com/wp-content/uploads/2019/10/Pieter-Robberechts-Valuing-the-Art-of-Pressing.pdf>
- Roth Alvin E, Thomson W (1988) *The Shapley Value: Essays in Honor of Lloyd S. Shapley*. isbn: 052136177X. <https://doi.org/10.2307/2554979>.
- Santos A, Benito al (2018) "Data-driven visual performance analysis in soccer: An exploratory prototype". In: *Frontiers in Psychology* 9. issn: 16641078. <https://doi.org/10.3389/fpsyg.2018.02416>.

- Shaw L, Glickman M (2019) “Dynamic analysis of team strategy in professional football”. In: Barça sports analytics summit. Retrieved from https://static.capabliaser.com/frontend/clients/barca/wp_prod/wp-content/uploads/2020/01/56ce723e-barca-conferencepaper-laurie-shaw.pdf
- Spearman W (2018) “Beyond Expected Goals”. In: MIT Sloan Sports Analytics Conference, pp. 1–17
- Steiner S et al (2019) Outplaying opponents-a differential perspective on passes using position data. *German J Exerc Sport Res*. <https://doi.org/10.1007/s12662-019-00579-0>
- Travassos B et al (2013) Performance analysis in team sports: advances from an ecological dynamics approach. *Int J Perform Anal Sport* 13(1):83–95. <https://doi.org/10.1080/24748668.2013.11868633>
- Vogelbein M, Nopp S, Hökelmann A (2014) “Defensive transition in soccer - are prompt possession regains a measure of success? A quantitative analysis of German Fußball-Bundesliga 2010/2011”. In: *Journal of Sports Sciences* 32.11, pp. 1076–1083. issn: 1466447X. <https://doi.org/10.1080/02640414.2013.879671>.url: <http://dx.doi.org/10.1080/02640414.2013.879671>
- Wang Y (2019) “A Xgboost Risk Model Via Feature Selection and Bayesian Hyper-Parameter Optimization”. [arXiv:1901.08433](https://arxiv.org/abs/1901.08433)
- Zhang W et al (2020) “Prediction of undrained shear strength using extreme gradient boosting and random forest based on Bayesian optimization”. *Geoscience Frontiers* 12(1): 469–477. <https://doi.org/10.1016/j.gsf.2020.03.007>

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Terms and Conditions

Springer Nature journal content, brought to you courtesy of Springer Nature Customer Service Center GmbH (“Springer Nature”).

Springer Nature supports a reasonable amount of sharing of research papers by authors, subscribers and authorised users (“Users”), for small-scale personal, non-commercial use provided that all copyright, trade and service marks and other proprietary notices are maintained. By accessing, sharing, receiving or otherwise using the Springer Nature journal content you agree to these terms of use (“Terms”). For these purposes, Springer Nature considers academic use (by researchers and students) to be non-commercial.

These Terms are supplementary and will apply in addition to any applicable website terms and conditions, a relevant site licence or a personal subscription. These Terms will prevail over any conflict or ambiguity with regards to the relevant terms, a site licence or a personal subscription (to the extent of the conflict or ambiguity only). For Creative Commons-licensed articles, the terms of the Creative Commons license used will apply.

We collect and use personal data to provide access to the Springer Nature journal content. We may also use these personal data internally within ResearchGate and Springer Nature and as agreed share it, in an anonymised way, for purposes of tracking, analysis and reporting. We will not otherwise disclose your personal data outside the ResearchGate or the Springer Nature group of companies unless we have your permission as detailed in the Privacy Policy.

While Users may use the Springer Nature journal content for small scale, personal non-commercial use, it is important to note that Users may not:

1. use such content for the purpose of providing other users with access on a regular or large scale basis or as a means to circumvent access control;
2. use such content where to do so would be considered a criminal or statutory offence in any jurisdiction, or gives rise to civil liability, or is otherwise unlawful;
3. falsely or misleadingly imply or suggest endorsement, approval, sponsorship, or association unless explicitly agreed to by Springer Nature in writing;
4. use bots or other automated methods to access the content or redirect messages
5. override any security feature or exclusionary protocol; or
6. share the content in order to create substitute for Springer Nature products or services or a systematic database of Springer Nature journal content.

In line with the restriction against commercial use, Springer Nature does not permit the creation of a product or service that creates revenue, royalties, rent or income from our content or its inclusion as part of a paid for service or for other commercial gain. Springer Nature journal content cannot be used for inter-library loans and librarians may not upload Springer Nature journal content on a large scale into their, or any other, institutional repository.

These terms of use are reviewed regularly and may be amended at any time. Springer Nature is not obligated to publish any information or content on this website and may remove it or features or functionality at our sole discretion, at any time with or without notice. Springer Nature may revoke this licence to you at any time and remove access to any copies of the Springer Nature journal content which have been saved.

To the fullest extent permitted by law, Springer Nature makes no warranties, representations or guarantees to Users, either express or implied with respect to the Springer nature journal content and all parties disclaim and waive any implied warranties or warranties imposed by law, including merchantability or fitness for any particular purpose.

Please note that these rights do not automatically extend to content, data or other material published by Springer Nature that may be licensed from third parties.

If you would like to use or distribute our Springer Nature journal content to a wider audience or on a regular basis or in any other manner not expressly permitted by these Terms, please contact Springer Nature at

onlineservice@springernature.com