Modelling team performance in soccer using tactical features derived from position tracking data

Goes, F. R.; Kempe, M.; Van Norel, J.; Lemmink, K. A. P. M.

Published in:
IMA Journal of Management Mathematics

DOI:
10.1093/imaman/dpab006

IMA Journal of Management Mathematics (2021) **32**, 519–534
https://doi.org/10.1093/imaman/dpab006
Advance Access publication on 5 April 2021

Modelling team performance in soccer using tactical features derived from position tracking data

F.R. GOES†
Center for Human Movement Sciences, University of Groningen, University Medical Center Groningen (UMCG), 9713 AV Groningen, The Netherlands
†Corresponding author. Email: f.r.goes@umcg.nl
M. KEMPE
Center for Human Movement Sciences, University of Groningen, University Medical Center Groningen (UMCG), 9713 AV Groningen, The Netherlands
J. VAN NOREL
S.B.V. Vitesse, 6816 VD Arnhem, The Netherlands
K.A.P.M. LEMMINK
Center for Human Movement Sciences, University of Groningen, University Medical Center Groningen (UMCG), 9713 AV Groningen, The Netherlands

[Received on 27 November 2019; accepted on 12 February 2021]

Decision-makers in soccer routinely assess the tactical behaviour of a team and its opponents both during and after the game to optimize performance. Currently, this assessment is typically driven by notational analysis and observation. Therefore, potential high-impact decisions are often made based on limited or even biased information. With the current study, we aimed to quantitatively assess tactical performance by abstracting a set of spatiotemporal features from the general offensive principles of play in soccer using position tracking data, and to train a machine learning classifier to predict match outcome based on these features computed over the full game as well as only parts of the game. Based on the results of these analyses, we describe a proof of concept of a decision support system for coaches and managers. In an analysis of 302 professional Dutch Eredivisie matches, we were able to train a Linear Discriminant Analysis model to predict match outcome with fair to good (74.1%) accuracy with features computed over the full match, and 67.9% accuracy with features computed over only 1/4th of the match. We therefore conclude that using only position tracking data, we can provide valuable feedback to coaches about how their team is executing the various principles of play, and how these principles are contributing to overall performance.

Keywords: soccer; machine learning; position tracking data; decision support systems; tactical behaviour.

1. Introduction

Coaches routinely assess the tactical behaviour of their team and its opponents (Gudmundsson & Horton, 2017) in order to help them make strategic pre-game and in-game decisions. However, apparently coaches take these strategic in-game decisions with varying degrees of success. For example, FC Liverpool coach Jürgen Klopp was able to secure 16 points in the seven games where his team fell behind by one or more goals. Pep Guardiola, on the other hand (Manchester City), just managed to get 4 points over the course of the 2018/19 Premier League season in the six matches where his team fell behind (Transfermarkt.co.uk). This pattern was even more pronounced in last season's Champions

League competition, won by FC Liverpool despite falling behind in most of its knockout games, whereas Manchester City failed to recover in similar games. So, is Jürgen Klopp the better coach? In an attempt to answer this question and support club administrators and coaches in their decision-making, we tested a proof of concept for in-game and post-game tactical behaviour evaluation based on position tracking data. As highlighted by a recent article, decision support systems are in high demand in sport science and practice (Robertson, 2020). Such a tactical decision support system would be the first of its kind, as it analyses more than just physical performance (Robertson *et al.*, 2017). It could also aid coaches in more than just their in-game decision-making, as one could think of a wider range of applications like pre-game opponent analysis, evaluation and planning of training sessions, player and coach scouting, and staff evaluation by a club's management.

By now, tactical assessment is most frequently driven by notational assessment and observational analysis (Brink & Lemmink, 2018; Rein & Memmert, 2016). Although notational analysis is time consuming, has an ill-proven link with actual performance (Brooks *et al.*, 2016), and coaches have to make decisions under heavy time constraints, teams still heavily rely on it. However, even expert coaches have limited recall ability (Laird & Waters, 2008) and are prone to cognitive bias (Frederick, 2005). Therefore, one could argue coaches and other decision-makers in a professional soccer organization have limited tools and information available to support their decision-making, which can be assumed to affect the quality of their decisions.

Making inaccurate decisions concerning playing strategies or formations can have large consequences in professional soccer. Based on reward money alone, the difference between winning and losing one Champions League group-stage match is €2.7 million (Marca). Therefore, it is of key importance for clubs to take decisions that increase their chances of winning. Because of the importance of winning and scoring goals, research from various domains has expressed interest in modelling outcome of soccer games. For example, Cintia *et al.* (2015) have modelled match outcome (win vs. draw vs. lose) in various European leagues based on passing and shot indicators, with up to 60% accuracy. Other examples include the 2017 Soccer Prediction Challenge, which resulted in many contributions that predict match outcome in professional competitions based on variables like previous results and ranking with around 50% accuracy (Dubitzky *et al.*, 2019), the application of random forests to predict FIFA world cup outcomes, resulting in prediction accuracies around 55% using bookmaker odds and FIFA rankings (Groll *et al.*, 2019; Schauberger & Groll, 2018), or the work by Egidi *et al.* (2018) who combined pre-game data with bookmakers' information.

Although previous prediction studies provide interesting insights into how some aspects of performance are linked to match outcome, they are of little practical value to decision-makers within sport. These papers rather serve an academic purpose, in their aim for continuous improvement of state-of-the-art machine learning models, as well as practical applications off the field, primarily in the betting industry. As these studies are all based on (pre-game) aggregated statistics, they cannot be applied in real time, and more importantly, do not capture tactical behaviour. To move towards an in-game or post-game decision support system for decision-makers in soccer (coaching staff and club administrators), one would rather require a model that captures various aspects of tactical behaviour and links them to successful performance (i.e. winning).

Tactics, often referred to in research as tactical behaviour, can be defined as the management of space and time by a group of cooperating individuals, in interaction with the opponent while constantly adapting to the conditions of play, in order to achieve a common goal (Gréhaigne *et al.*, 1999; Rein & Memmert, 2016). As tactical behaviour is assumed to be dependent on a shared base of knowledge in the team (Gershgoren *et al.*, 2016), it can also be assumed to adhere to certain general principles, widely known as the general principles of play in soccer (Clemente *et al.*, 2014; da Costa *et al.*, 2009).

The general principles of play are irrespective of playing style or game-specific strategies, and dictate the common goals of tactical behaviour during different stages of the game (Clemente *et al.*, 2014; da Costa *et al.*, 2009). These goals are related to possession status, as teams have different tactical objectives when attacking and defending. The moment a team gains possession of the ball, their primary aim is to keep possession of the ball (*possession principle*) and to ultimately create a scoring opportunity. To achieve this, the general offensive principles of play dictate that after gaining possession, teams should increase the effective playing space (*space mobility principle)* (Fonseca *et al.*, 2012; Metulini *et al.*, 2018), create numerical superiority by positioning attacking players in key areas of the field and outplay opponents (*superiority principle)* (Rein *et al.*, 2017), disrupt the defensive organization (*disruption principle)* (Goes *et al.*, 2019) and move the ball to a position from where a scoring opportunity arises (*scoring principle*) (Clemente *et al.*, 2014; da Costa *et al.*, 2009).

The defensive objectives, on the other hand, can be summarized as preventing the opponent from achieving the offensive objectives, by moving as one defensive unit and protecting the goal, while ultimately aiming to win back the ball (Clemente *et al.*, 2014; da Costa *et al.*, 2009). According to the work by Clemente *et al.* (2014) and Costa *et al.* (2009), achieving these goals mark successful tactical behaviour and contributes to achieving the overall performance goal (i.e. scoring goals and winning the game). However, principles of play are generally defined in abstract terms, and assessing their actual relation to success therefore requires the operationalization of spatial and temporal components specific to a certain principle of play, in order to translate those components into features that can be derived from the data (Goes *et al.*, 2020b; Stein *et al.*, 2017).

The increasing availability of position tracking data has opened up the opportunity to conduct the aforementioned tactical behaviour analysis (Rein & Memmert, 2016). As position tracking data allows us to simultaneously study the actions of all 22 players on the field, it enables an increased and more accurate understanding of tactical behaviour and can therefore be used to accurately analyse adherence to the principles of play. Therefore, the overall aim of the current work is to assess the relationship between tactical behaviour and successful match performance in professional soccer (i.e. winning the game), regardless of the score line during the game or contextual factors like home advantage. The results of our study could be used to construct a support system for decision-makers in professional soccer both during and after the game. An adequate automated assessment of tactical behaviour and its relation to performance using tracking data would allow a team's manager to efficiently evaluate performance on a collective and individual level, evaluate adherence to instructions and optimize a team's strategy, as well as allowing club administrators to evaluate the performance of players and coaches more efficiently and objectively, making their decisions more robust and less prone to errors and subjective bias. We operationalize our aim by means of two research questions: first, can we adequately model the probability of winning a match, based on a set of offensive tactical features derived from position tracking data post-match, without providing information on match outcome and score line? Second, taking the time series of features computed during the match, at what timepoint in a match do predictions based on the same set of features become accurate enough to provide actionable information in order to help the coach make the decisions to change the course of game?

2. Methods

2.1 *Data*

We utilized an observational design in which we collected position tracking data and match outcome of 302 Dutch professional Eredivisie matches between 18 teams played during one full season.

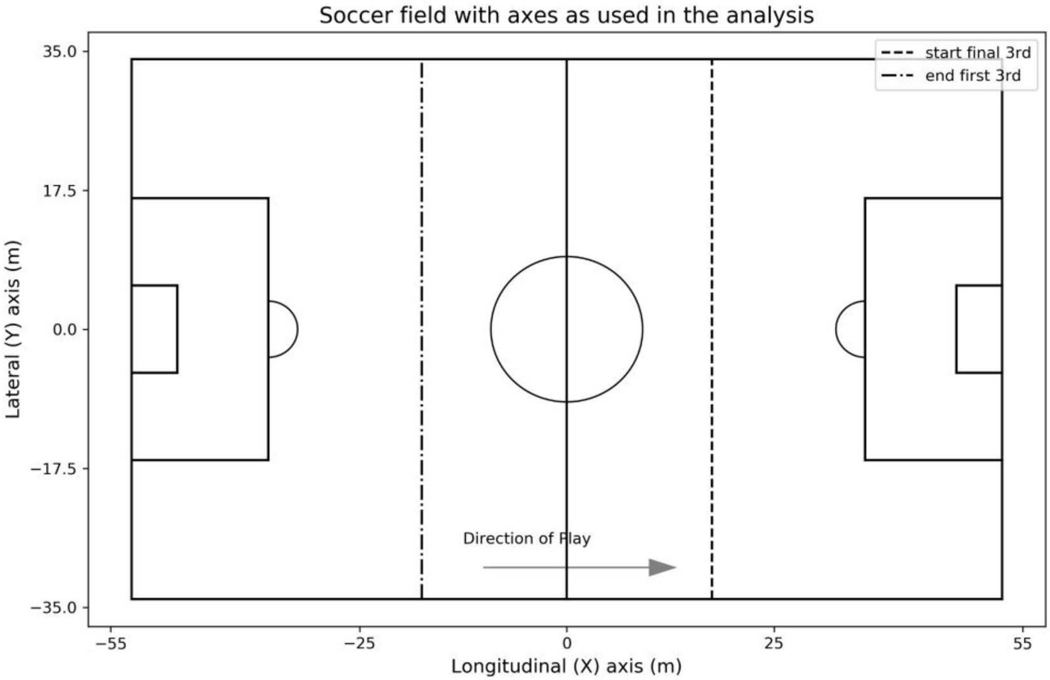

FIG. 1. Soccer field with axes and reference lines as used in our analyses.

Four matches of the 306 (1.3%) played during the regular season were missing from the dataset because of erroneous or missing data. Data were generated through a league-wide employed semi-automatic optical tracking system (ChyronHego, NY, USA), which captures the X and Y coordinates of all players and the ball at 25 Hz, corresponding to a measurement every 40 ms. Before the analysis, the raw position tracking files were first pre-processed on a match-by-match basis with ImoClient software (Inmotiotec Object Tracking B.V., The Netherlands). Pre-processing consisted of filtering with a weighted Gaussian algorithm (100% sensitivity), downsampling to 10 Hz, and automatic detection of possession and ball events based on synchronization of position tracking data with tagged event data. All data were mapped to the same field size where the X-axis runs longitudinally from goal to goal (−55 to +55 m), and the Y-axis runs horizontally along the midline (−35 to +35 m), including out-of-bounds regions of +2.5 m in longitudinal direction and +1 m in the lateral direction (Fig. 1).

2.2 *Quantifying offensive tactical performance*

To quantify offensive tactical behaviour, we first abstracted spatial and temporal components of the general offensive principles of play and constructed a set of spatiotemporal features specific to every principle. All features were derived from position tracking data and all the necessary events (i.e. passes, attacks) were automatically generated by synchronizing the annotated event data with the tracking data.

2.2.1 *Possession principle.* To study the *possession principle,* we computed the total number (N) of accurate passes per match, the pass accuracy (equation 1), and the average pass length (equation 2) and pass angle (equation 3) over all accurate passes per match, using the sum length of all accurate passes (LAP) and sum angle of all accurate passes (AAP).

$$\text{Pass Accuracy (\%)} = N_{\text{Accurate Passes}}/N_{\text{Attempted Passes}}{}^{*}100 \tag{1}$$

$$\text{Avg.pass length (m)} = \sum(\text{LAP})/N_{\text{Accurate Passes}} \tag{2}$$

$$\text{Avg.pass angle (m)} = \sum(\text{AAP})/N_{\text{Accurate Passes}} \tag{3}$$

2.2.2 *Offensive transition mobility principle.* To study the *offensive transition mobility principle,* we computed the mean expansion of the team surface area based on every transition from defence to offense. The team surface area (${S_A}^t$) was computed for every timepoint t in a match, as the smallest convex hull of an array (||.||) P_t containing the positions of all n outfield players, using the QHull implementation in the SciPy library (equations 4 and 5). The expansion was then computed as the difference between the minimal team surface area (m^2) and the maximal area (m^2) in window i starting after every transition and ending 8 s later, defined as the moment possession changed from one team to another in our automatically detected events. We choose the 8-s window because pilot analyses on our dataset revealed that although in general the principle that teams in possession occupy a larger surface area in comparison with teams not in possession holds, there seems to be a lag of 3–5 s right after a transition in which this expansion does not take place yet. We then aggregated this value over the match to find the mean expansion per transition.

$$\text{Pt} = \left[\left[X_i^t + Y_i^t\right], \left[X_{i+1}^t + Y_{i+1}^t\right], [\ldots, \ \ldots], \left[X_n^t, Y_n^t\right]\right] \tag{4}$$

$$S_A{}^t = \text{ConvexHull} \ \left|\left| \, \text{P}_t \right|\right| \tag{5}$$

$$\text{Transition Expansion}_i = S_{A,\max,i}^t - S_{A,\min,i}^t, \forall \ i = 1, \ldots i + 8 \text{ s} \tag{6}$$

2.2.3 *Superiority principle.* To study the *superiority principle,* we computed a set of features related to actions made with the ball like a pass or dribble (‘on-ball’) as well as a set of features related to the positioning and movement of players not in possession of the ball (‘off-ball’). To study the on-ball superiority, we computed the number of outplayed opponents and the number of outplayed defenders (last six players based on their longitudinal field position) for every accurate pass (Fig. 2). Outplaying a defender was defined based on the distance to the back line (Rein *et al.*, 2017). Both models ignored negative values (backwards passes). We then computed the total value over a full match for both features.

To study off-ball superiority, we computed the balance between attacking and defending players in both the final third section of the field and the score box, as these are considered by coaches to be key areas to gain superiority (Rein *et al.*, 2017), for every second during a possession (Fig. 2). For every second i out of a total n seconds with an attacking overload (i.e. more attacking than defending players in a certain area), the balance score was added to the match total, resulting in a total final third and total score box superiority score per possession and per match (equations 7 and 8).

$$\text{Balance Score} = \left(N_{\text{Attacking Players in Area}}\right) - \left(N_{\text{Defensive Players in Area}}\right) \tag{7}$$

$$\text{Total [Region] Superiority Score} = \sum_{i=1}^{n} \text{Balance Score} \tag{8}$$

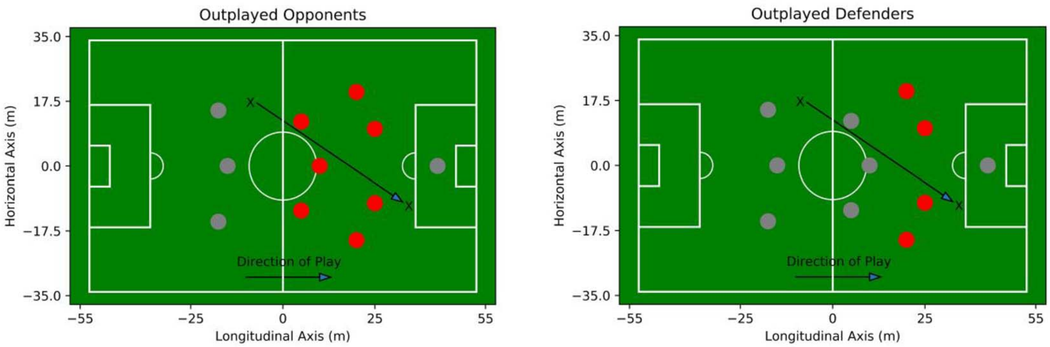

FIG. 2. Example of computation of the number of outplayed opponents (OPO—left) and the number of outplayed defenders (OPD—right). Opponents/defenders that are outplayed by a given pass are visualized in red, whereas players that were not outplayed are visualized in grey. Computed based on horizontal distance (goal-to-goal) only.

2.2.4 *Disruption principle.* To study the *disruption principle*, we utilized two features that were introduced in previous work (Goes *et al*., 2019). The defensive disruption feature (Def-D) quantifies the disruption of the defensive organization induced by a pass, whereas the induced movement (I-Mov) feature quantifies the amount of individual movement of defensive players in response to a pass. To construct the Def-D feature, we first computed the displacement of the line centroids (average positions of the defensive, midfield and attacking lines) in horizontal (y) and longitudinal (x) planes, team centroids (x and y), as well as the change in team surface area (excluding the goalkeeper) and the team spread, between the moment of passing and the moment of receiving, for every successful pass. We subsequently scaled all 10 variables using Z-scores and reduced them to three principal components that represent longitudinal disruption (PC1), lateral disruption (PC2) and shape disruption (PC3), and in their absolute form constitute the feature as a whole (equation 9). These three principal components where chosen based on their eigenvalues, and together explain 83.3% of the variance, as discussed in our previous work (Goes *et al*., 2019). In contrast with our previous publication, we utilized a K-Means clustering model (mean silhouette score $= 0.63 \pm 0.07$) for dynamic line definition instead of defining them manually, and an improved temporal aggregation method that standardizes the feature in disruption per second (equation 10). For further details, we refer to our previous work (Goes *et al*., 2019, 2020a).

$$\text{Def-D} = |\text{PC1}| + |\text{PC2}| + |\text{PC3}| \tag{9}$$

$$\text{Std.Def-D} = \text{Def-D}/\text{Pass Duration (ms)}\,{}^{*}1000 \tag{10}$$

The second feature quantifies the amount of individual movement per player in the defensive team for n players as the sum of displacement on the X-axis and Y-axis in response to a pass. For the current work, this feature was also updated with an improved temporal aggregation method that standardizes the feature as the amount of movement per second (equation 11). We finally computed the mean values per pass and the total value over a full match for both features.

$$\text{I-Mov} = \left(\left|\text{Disp.}X_1\right| + \left|\text{Disp.}Y_1\right| + \cdots + \left|\text{Disp.}X_n\right| + \left|\text{Disp.}Y_n\right|\right)/n \tag{11}$$

2.2.5 *Scoring principle.* To study the *scoring principle*, we computed the potential for every pass to result in a scoring opportunity (POT). This feature is partly derived from the dangerousity model previously published by Link *et al.* (2016), where the 'danger' of an individual ball possession at a point in time is computed based on its field location relative to the goal, and the pressure and organization of the defensive players relative to the ball and the goal. Every individual ball possession is assigned a value Z (0–1) based on its field position relative to the goal using a 2×2 m grid (Fig. 3), which is subsequently penalized using features related to the defensive organization and degree of ball control. Values close to 1 can be thought of a highly dangerous moments where the potential to create a scoring opportunity is large, whereas values close to 0 represent the opposite. For the current paper, we defined our POT feature using the same 2×2 m grid to assign a field value Z and computed the penalty (PR) using an adaptation of the pressure model published by Andrienko *et al.* (2017), in which pressure on the ball possessing player is derived from a model accounting for inter-player distances, relative position of defensive players respective to the threat direction (the goal) while controlling for direction and velocity of movement of all involved players. Both Z and PR range from 0 to 1 and are used to construct POT based on equation 12. For the current paper, we computed the POT for every individual reception i and aggregated this into a total POT per match, as well as a mean POT per reception.

$$Z_i\left(1\text{–}\mathrm{PR}_i\right), \quad \mathrm{POT}_{\mathrm{match}} = \sum\nolimits_i \mathrm{POT}_i \tag{12}$$

2.3 *Analysis*

Every match resulted in two observations (both teams), for which features were computed. All computed features were aggregated into mean and/or total values per team per match, resulting in a total of 15 features (Table 1). As one could express performance in soccer both in absolute terms (scoring N goals) as well as in relative terms (scoring N goals more than your opponent), we created two instances of the same feature set: a set of *absolute* features and a set of *relative* features. For example, if Team A achieved an *absolute* Total Potential of 20, and Team B achieved an *absolute* Total Potential of 15, Team A has a *relative* Total Potential of 133%, whereas Team B scores 75%.

To achieve our aim of creating a tool that could support coaches in their tactical decision-making, we then defined the following problem: deriving a win probability for a team based on the classification of ordinal match outcome in terms of winning or losing. To solve this classification problem, we first split the dataset into a training set containing 75% of the data and a test set that contained 25% of the data. We then trained a set of classifiers on the training set and validated and compared their performance on the test set.

As we wanted to compute the win probability for a team to allow giving feedback to decision-makers, and because they are typically noisy without clear patterns, we first omitted draws ($N = 74$) from the training and validation process, as we were mainly interested in predicting winners and losers, and draws—which can be regarded a research line of their own—would be a source of added noise (Hvattum, 2017). Furthermore, we needed classifiers that allow interpretation on the level of individual features and perform relatively well on a small dataset. Therefore, we trained a Decision Tree (DT) classifier, a Gradient Boosting classifier, a Linear Discriminant Analysis (LDA) classifier and a Quadratic Discriminant Analysis classifier. As we have a relatively small dataset, we first used the DT classifier for a two-step feature selection process, as decision trees can be efficiently used for feature selection (Sugumaran *et al.*, 2007). By training the classifier on both the set of all absolute as well as

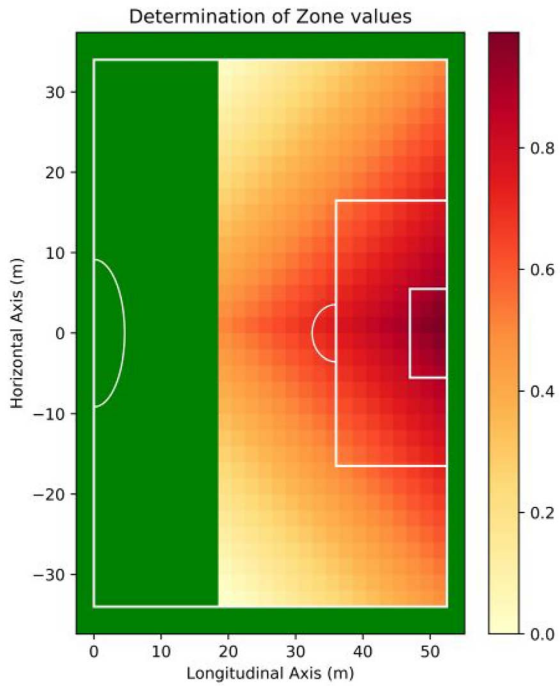

FIG. 3. Illustrative example of the determination of 'field-values' in the computation of the potential of creating a scoring opportunity. Colour bar values represent the values assigned to the 'zone' parameter in equation 12.

TABLE 1. *Overview of included features and units of measurement (in absolute units)*

Feature	Approximate range (units)	Principle of play
Number of passes	0–1000 (passes)	Possession
Pass accuracy	0–100 (%)	Possession
Average pass length	0–50 (m)	Possession
Average pass angle	−90–90 (°)	Possession
Average transition expansion	−1000–1000 (m^2)	Space mobility
Total outplayed opponents	0–750 (players)	Superiority
Total outplayed defenders	0–400 (players)	Superiority
Total score box superiority	0–1000 (seconds)	Superiority
Total final third superiority	0–1000 (seconds)	Superiority
Average pass Def-D	0–1 (unitless)	Disruption
Average pass I-MOV	0–3 (m/player)	Disruption
Total Def-D	0–400 (unitless)	Disruption
Total I-MOV	0–2000 (m/player)	Disruption
Average pass potential	0–1 (unitless)	Scoring
Total potential	0–50 (unitless)	Scoring

the set of all relative features as our baseline models, we identified which set of features resulted in the best model performance. Subsequently, we selected a subset of features from the best performing dataset based on collinearity and feature importance in the DT model, selecting the most important feature in all instances where multiple features had a high collinearity. Classifier performance was evaluated based on mean accuracy score, recall and precision in predicted wins (given that wins are our main focus), and mean Brier score over a 5-fold cross-validation. We first optimized all models by hyperparameter tuning every model using a cross-validated grid search over the hyperparameter space of every classifier, and finally selected the best performing model based on classifier performance evaluated using mean accuracy score, recall and precision in predicted wins (given that wins are our main focus), and mean Brier score over a 5-fold cross-validation.

After selecting the best performing classifier, we repeated the training process of both predictors (the predictor trained with absolute features and the one trained with relative features), this time using features only computed over the first 25, 50 or 75% of the match time to train our model. We then compared the performance of the classifier based on partial information to performance based on the full match, using similar evaluation metrics and 5-fold cross-validation as discussed above.

All processing and analyses were conducted using custom routines programmed in Python 3.7, using sklearn (Pedregosa *et al*., 2011) for all machine learning routines. The study has been approved by the local ethical committee of the university and adheres to the declaration of Helsinki.

3. Results

Our dataset contained 302 professional matches in which 975 goals were scored and 209,562 accurate passes were played. The raw position tracking data of all these matches combined resulted in a dataset of approximately 989,805,000 datapoints. Out of the 302 matches, 228 matches had a winner and 74 matches resulted in a draw. On average, winning teams scored 2.9 ± 1.4 goals vs. 0.8 ± 0.9 goals for losing teams, which means teams in your sample win with a mean 2.1 goal difference.

3.1 *Modelling match outcome*

After fitting our baseline DT models to the set of absolute and the set of relative features and testing its performance on test set, the set of relative features was picked as the best performing baseline model based on clear difference in accuracy scores (Table 2). As *total pass potential* and *average pass potential*, as well as *total pass I-MOV*, *total pass Def-D*, *average pass Def-D* and *average pass I-MOV* had collinearity issues ($r > 0.5$), we selected *the average pass potential* and *average pass I-MOV* for subsequent training of the other models, as these had the highest feature importance in the DT compared with their correlating counterparts. Furthermore, we omitted the *pass accuracy* feature because of a comparably poorer feature importance as well (Fig. 4). As a result, we selected 10 of the 15 original features from the set off all relative features for subsequent model training and hyperparameter tuning. The results (Table 2) show that overall the LDA classifiers have the best performance.

3.2 *In-game modelling*

To study the potential of in-game performance prediction and the value of in-game feedback to support strategic decision-making by the coaching staff, we retrained the LDA model using partial information from the first 25, 50 and 75% of the match. We then compared the model performance metrics to the model trained on the full match and found that using partial information results in only a limited decline in model performance (Table 3).

TABLE 2. *Overview of model performance in predicting binary match outcome in the test set*

	Accuracy score (%)	Brier score	Precision (wins)	Recall (wins)
DT (absolute)	59.2	—	—	—
DT (relative)	67.1	0.36	0.64	0.52
GBoost	71.2	0.25	0.72	0.73
LDA	**74.1**	**0.19**	**0.71**	**0.71**
QDA	64.5	0.29	0.73	0.48

GBoost, gradient boosting; QDA, quadratic discriminant analysis.

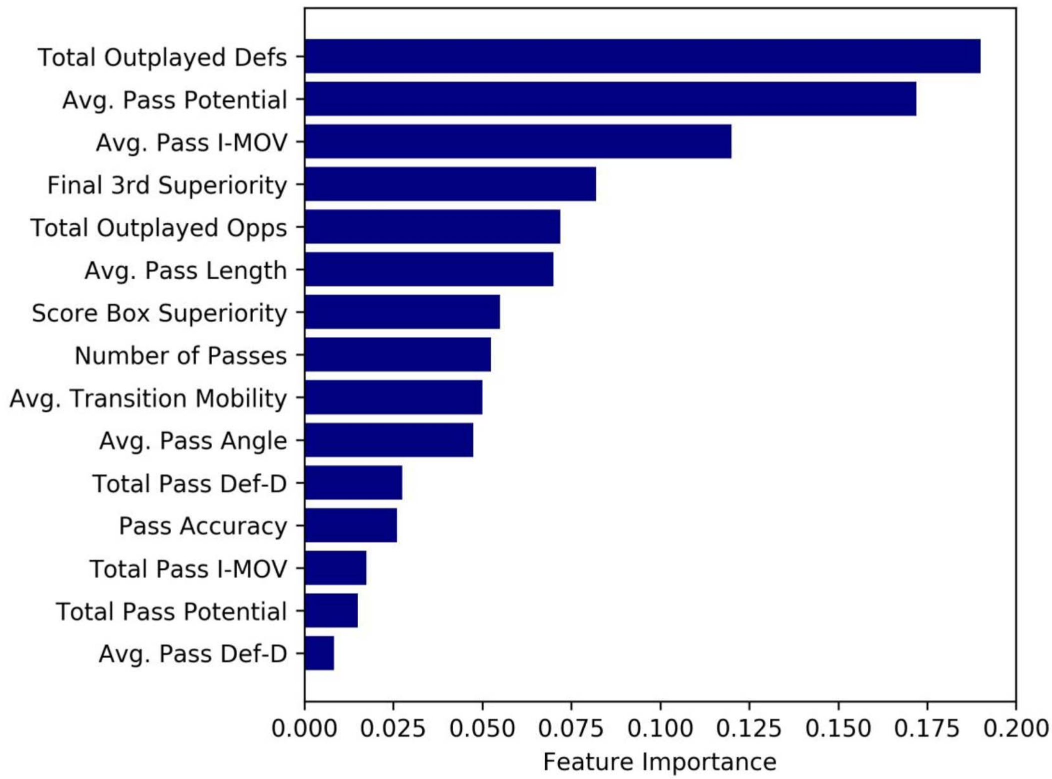

FIG. 4. Feature importance in predicting binary match outcome derived from the baseline decision tree classification model.

3.3 *Proof of concept real-time tactical feedback*

As our results indicated that—assuming one has access to reliable real-time position tracking data—match outcome can be predicted with fair accuracy after only 22.5 min in the game, we subsequently constructed a proof of concept for real-time outcome predictions. We scaled the in-game outcome predictor to provide predictions every 5 min by extrapolating the performance of team to that over a full match based on the features computed up to a given timepoint. We subsequently computed the win probability per team using our trained model, resulting in a probability to win the game if performance levels stay similar to that of performance up to that timepoint. The result provides a near continuous

TABLE 3. *Overview of LDA model performance in predicting binary match outcome using partial information derived from the first 25, 50 or 75% of the match*

	Accuracy score (%)	Brier score	Precision (wins)	Recall (wins)
100% (full match)	**74.1**	**0.19**	**0.71**	**0.71**
75% (67.5 min)	69.6	0.23	0.63	0.68
50% (45.0 min)	67.1	0.21	0.66	0.62
25% (22.5 min)	67.9	0.31	0.53	0.88

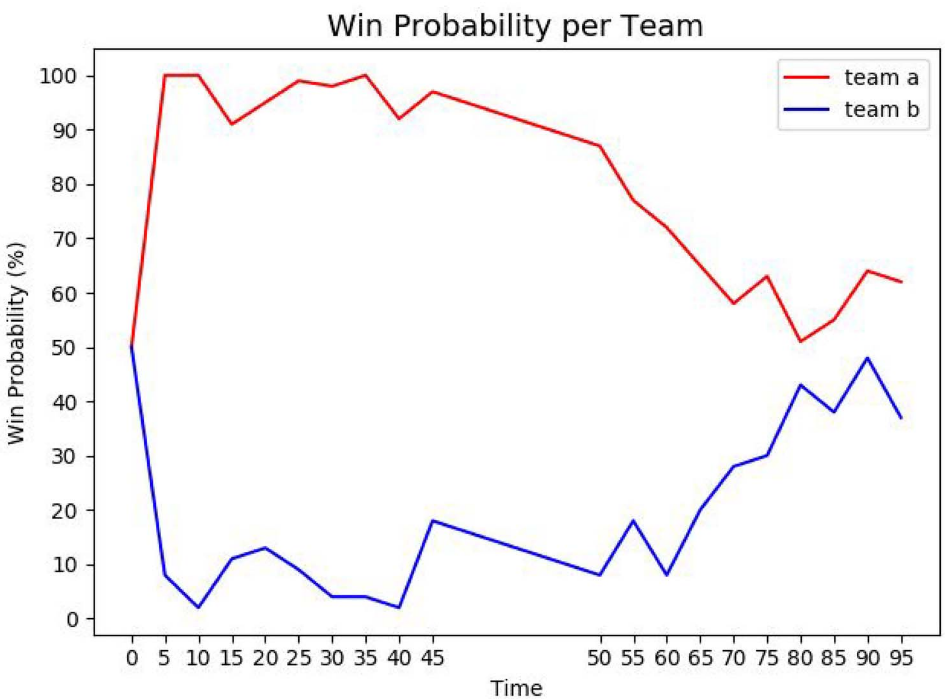

FIG. 5. Example of the flow of a match based on predictions of match outcome every 5 min using real data. The game was won by the red team who scored one goal near the end of the first half. Note that both teams changed their formations because of substitutions between the 60th and the 75th min (team names are made anonymous for privacy reasons). Please note that as the first predictions are made after 5 min, the win probability at minute 0 is equal for both teams, which would theoretically result in a draw.

representation of the flow of the game (Fig. 5). The figure shows that the first predictions occur after 5 min into the game and will gravitate to a more reliable state further in the game. As our predictions are independent of the score line, we found in this specific example that the win probability of the red team dropped in the second half, despite scoring a goal to put them in the lead at the end of first half, indicating that although they scored, their tactical performance declined relative to that of the blue team.

4. Discussion

The current study aims to assess the relationship between tactical behaviour and match performance in professional soccer. The purpose of this analysis is to construct a support system for decision-makers in professional soccer both during and after the game, using an automated approach based on position tracking data. To achieve this, we first trained a model to classify match outcome in terms of a probability of winning based on tactical features computed over the full match, and then trained a model to classify that same match outcome based on partial information of the match. Our findings indicate that we can model match outcome with fair to good accuracy both in-game as well as post-game based only on tactical features derived from position data. These findings have led to a proof of concept for a decision support system that could aid decision-makers in their performance evaluation.

Our results show that by using spatiotemporal features, one can train a model that makes accurate classifications and which output could potentially support decision-makers. The classification model we trained was able to classify nearly three out of four match outcomes correctly in terms of winning or losing (74.1%), thereby outperforming existing models (Cintia *et al.*, 2015; Dubitzky *et al.*, 2019). However, one has to note that a direct comparison with these studies is not justified, as our model and its underlying features represent entirely different constructs compared with those presented in previous work, as well as the exclusion of draws. This result in significant methodological differences from the work mentioned above. Still, the use of position tracking data opens up a range of possibilities for analysing performance in much more detail, as it allows features to be exponentially more complex than those derived from annotated event data, and it provides an analysis that allows understanding what behaviour drives successful performance, and in extension can be used as a starting point to optimize performance. Furthermore, it could be that our dataset presented a near-optimal case for the classification problem as the margin between winning and losing teams was relatively large (1.8 goals). Therefore, in future work, our model should also be tested on different datasets of different leagues. Nevertheless, as an answer to our first research question, we conclude that we are able to adequately model wins and losses solely based on a set of offensive tactical features.

Our secondary objective was to study the potential for near-time in-game predictions and to construct a proof of concept of such a system. Our results have shown that in-game predictions are only 5–10% less accurate when compared with classifications made based on data from the full game. To find at what time timepoint predictions reach an accuracy that has practical relevance, we used data from the first 25, 50 and 75% of the match to train a model similar to the one we used for classification based on the full match. Our results show that there is only a limited drop in model performance when the model is trained with partial information, and that even after playing only the first quarter of the match, relatively accurate predictions about match outcome can be made. As even predictions modelled on information from the first quarter of the match seemed to be relatively accurate, one might conclude that our system could provide adequate support for coaches. However, in the current project, all predictions were made using post-processed data to improve reliability. Using this system in-game would require data that is reliable in real-time without the need for post-processing. As current evidence suggests that semi-automatic optical tracking data can be collected with good accuracy in real time (Taberner *et al.*, 2020; Takahashi *et al.*, 2018), this might actually be feasible in the near future.

The in-game and post-game tactical performance evaluation system that we proposed has multiple practical implications for the coaching staff. First, it allows the coach to quickly scan how well his team is performing at any given timepoint during a game: if the team has a low win probability for a prolonged period of time, or if the win probability suddenly starts to drop during the game, the coach might decide to intervene by making strategical changes or substitutions. This is not dissimilar to the current practice, with the main difference being that the coach currently has to make his decisions based on only his personal recall of events and that of his staff, something that is a known source of bias given the limited recall ability of even the most expert coaches (Laird & Waters, 2008). Of course, our model does not provide a specific solution, nor does it tell the coach what do. However, by assessing the different features related to various principles of play, a coach would be able to get a clear picture on why his team is underperforming, with the actual decision still depending on the coach's expertise. Second, one can think of similar implications for the post-game tactical evaluation, as well as the pre-game opponent analysis. Both are currently predominantly based on recollection of events, notational statistics (event data) and—most importantly—video analysis. Given the limited recall ability mentioned before (Laird & Waters, 2008), the poor relation between notational statistics and performance (Brooks *et al.*, 2016; Cintia *et al.*, 2015) and the amount of time required for video analysis, this can be considered sub-optimal at best. Implementing our system would allow for the analysis of much more games of an upcoming opponents, thereby providing a more robust assessment, as well as more objective analysis of tactical performance of both the own team as well as that of opponents. Finally, one can extend these implications to club administrators and media. Teams can play well and still lose, due to individual mistakes, or even due to uncontrollable circumstances like misjudged calls by the referee. Although this is widely acknowledged, teams, coaches and player are still mainly evaluated based on match outcome and interpretation of what the observers have seen. Both means of evaluations are known to often not reflect the actual performance of a team, and can therefore lead to bad decisions. Integrating our system in the process of decision-making could prevent coaches from being hired or fired for the wrong reasons.

The current study is the first large-scale study using position tracking data to support the analysis of tactical behaviour and subsequent decision-making in relation to strategy in professional soccer. Over the last years, the role of data in the support of decision-making in professional soccer has already been gaining interest (Rein & Memmert, 2016; Stein *et al.*, 2017). The results of the current work are also promising, and our proof of concept could already aid decision-makers, we see several opportunities to advance our work in future research. Future work could unravel the complex game of soccer further by looking into very specific and rare match events, controlling for team-specific strategic determinants, and predicting performance on an individual level. However, doing so would ideally require data over multiple seasons. Furthermore, it would be interesting to combine the prediction models based on historical and metadata with our current model, as they measure completely different constructs. Doing so would not necessarily lead to more insights into tactical behaviour, but it would almost certainly result in even more adequate predictions. That, in turn, would allow for our model to be applied for other purposes like gambling and media. Finally, future work could resolve some of the limitations we currently encountered. For example, given their added noise (Hvattum, 2017), we deliberately omitted draws from the current study, but it would definitely be interesting to study what differentiates draws from matches with a clear winner and loser.

With the current study, we have been able to illustrate the potential of position tracking data in the support of strategic decisions by coaches and other staff in professional soccer. By abstracting a set of spatiotemporal features from the general principles of play of soccer, we have been able to identify several offensive tactical performance indicators that can be used in post-game and in-game feedback systems. We have showed that even with a limited amount of playing time, one can already make predictions about the result of the game that are more accurate than previous models, and that positional data can indeed revolutionize the way we are looking at and talking about tactical behaviour in soccer.

4. Funding

Netherlands Organization for Scientific Research (project title: 'The Secret of Playing Football: Brazil vs. The Netherlands') grant (629.004.012-SIA).

References

Andrienko, G., Andrienko, N., Budziak, G., Dykes, J., Fuchs, G., von Landesberger, T. & Weber, H. (2017) Visual analysis of pressure in football. *Data Min. Knowl. Discov.*, **31**, 1793–1839.

Brink, M. S. & Lemmink, K. A. P. M. (2018) Performance analysis in elite football: all in the game? *Sci. Med. Football*, **2**, 253–254.

Brooks, J., Kerr, M. & Guttag, J. (2016) Using machine learning to draw inferences from pass location data in soccer. *Stat Anal Data Min*, **9**, 338–349.

Cintia, P., Giannotti, F., Pappalardo, L., Pedreschi, D. & Malvaldi, M. (2015). The harsh rule of the goals: data-driven performance indicators for football teams. *2015 IEEE International Conference on Data Science and Advanced Analytics (DSAA)*, Paris, France, pp. 1–10.

Clemente, F. M., Martins, F. M. L., Mendes, R. S. & Figueiredo, A. J. (2014) A systemic overview of football game: the principles behind the game. *J. Hum. Sport Exerc.*, **9**, 656–667.

da Costa, I. T., da Silva, J. M. G., Greco, P. J. & Mesquita, I. (2009) Tactical principles of soccer: concepts and application. *Motriz*, **15**, 657–668.

Dubitzky, W., Lopes, P., Davis, J. & Berrar, D. (2019) The Open International Soccer Database for machine learning. *Mach. Learn.*, **108**, 9–28.

Egidi, L., Pauli, F. & Torelli, N. (2018) Combining historical data and bookmakers' odds in modelling football scores. *Stati. Model.*, **18**, 436–459.

Fonseca, S., Milho, J., Travassos, B. & Araújo, D. (2012) Spatial dynamics of team sports exposed by Voronoi diagrams. *Hum. Mov. Sci.*, **31**, 1652–1659.

Frederick, S. (2005) Cognitive reflection and decision making. *J. Econ. Perspect.*, **19**, 25–42.

Gershgoren, L., Basevitch, I., Filho, E., Gershgoren, A., Brill, Y. S., Schinke, R. J. & Tenenbaum, G. (2016) Expertise in soccer teams: a thematic inquiry into the role of shared mental models within team chemistry. *Psychol. Sport Exerc.*, **24**, 128–139.

Goes, F. R., Brink, M. S., Elferink-Gemser, M. T., Kempe, M. & Lemmink, K. A. P. M. (2020a) The tactics of successful attacks in professional soccer: large-scale spatiotemporal analysis of dynamic subgroups using position tracking data. *J. Sports Sci.*, 1–10.

Goes, F. R., Kempe, M., Meerhoff, L. A. & Lemmink, K. A. P. M. (2019) Not every pass can be an assist: a data-driven model to measure pass effectiveness in professional soccer matches. *Big Data*, **7**, 57–70.

Goes, F. R., Meerhoff, L. A., Bueno, M. J. O., Rodrigues, D. M., Moura, F. A., Brink, M. S., Elferink-Gemser, M. T., Knobbe, A. J., Cunha, S. A., Torres, R. S. & Lemmink, K. A. P. M. (2020b) Unlocking the potential of big data to support tactical performance analysis in professional soccer: a systematic review. *Eur. J. Sport Sci.*, 1–16.

Gréhaigne, J.-F., Godbout, P. & Bouthier, D. (1999) The foundations of tactics and strategy in team sports. *J. Teach. Phys Educ.*, **18**, 159–174.

Groll, A., Ley, C., Schauberger, G. & Van Eetvelde, H. (2019) A hybrid random forest to predict soccer matches in international tournaments. *J. Quant. Anal. Sports*, **15**, 271–287.

Gudmundsson, J. & Horton, M. (2017) Spatio-temporal analysis of team sports. *ACM Comput. Surv.*, **50**, 1–34.

Hvattum, L. M. (2017) Ordinal versus nominal regression models and the problem of correctly predicting draws in soccer. *Int. J. Comput. Sci. Sport*, **16**, 50–64.

Laird, P. & Waters, L. (2008) Eyewitness recollection of sport coaches. *Int. J. Perform. Anal. Sport*, **8**, 76–84.

Link, D., Lang, S. & Seidenschwarz, P. (2016) Real time quantification of dangerousity in football using spatiotemporal tracking data. *PLoS One*, **11**, e0168768.

Marca. *Winners of 2019/20 Champions League could receive over 80 millions in prize money*. Retrieved 29 October 2019 from https://www.marca.com/en/football/international-football/2019/07/11/5d27616646163f138c8b45aa.html.

Metulini, R., Manisera, M. & Zuccolotto, P. (2018) Modelling the dynamic pattern of surface area in basketball and its effects on team performance. *J. Quant. Anal. Sports*, **14**, 117–130.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., et al. (2011) Scikit-learn: machine learning in python. *J. Mach. Learn. Res.*, **12**, 2825–2830.

Rein, R. & Memmert, D. (2016) Big data and tactical analysis in elite soccer: future challenges and opportunities for sports science. *SpringerPlus*, **5**, 1410.

Rein, R., Raabe, D. & Memmert, D. (2017) “Which pass is better?” Novel approaches to assess passing effectiveness in elite soccer. *Hum. Mov. Sci.*, **55**, 172–181.

Robertson, S. (2020) Man & machine: adaptive tools for the contemporary performance analyst. *J. Sports Sci.*, **38**, 2118–2126.

Robertson, S., Bartlett, J. D. & Gastin, P. B. (2017) Red, amber, or green? Athlete monitoring in team sport: the need for decision-support systems. *Int. J. Sports Physiol. Perform.*, **12**, S273–S279.

Schauberger, G. & Groll, A. (2018) Predicting matches in international football tournaments with random forests. *Stat. Model.*, **18**, 460–482.

Stein, M., Janetzko, H., Lamprecht, A., Breitkreutz, T., Zimmermann, P., Goldlucke, B., et al. (2017) Bring it to the pitch: combining video and movement data to enhance team sport analysis. *IEEE Trans. Vis. Comput. Graph.*, **24**, 13–22.

Sugumaran, V., Muralidharan, V. & Ramachandran, K. I. (2007) Feature selection using decision tree and classification through proximal support vector machine for fault diagnostics of roller bearing. *Mech. Syst. Signal Process.*, **21**, 930–942.

Taberner, M., Keefe, J. O., Flower, D., Phillips, J., Close, G., Cohen, D. D., Richter, C. & Carling, C. (2020) Interchangeability of position tracking technologies; can we merge the data? *Sci. Med. Football*, **4**, 76–81.

Takahashi, M., Yokozawa, S., Mitsumine, H. & Mishina, T. (2018) Real-time ball-position measurement using multi-view cameras for live football broadcast. *Multimed. Tools Appl.*, **77**, 23729–23750.

Transfermarkt.co.uk. *Points after falling behind*. Retrieved 28 October 2019 from https://www.transfermarkt.co.uk/premier-league/punktenachrueckstand/wettbewerb/GB1/plus/?saison_id=2018&spiele=alle.