Science and Medicine in Football

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

Classification and determinants of passing difficulty in soccer: a multivariate approach

Murilo Merlin, Allan Pinto, Alexandre Gomes de Almeida, Felipe A Moura, Ricardo Da Silva Torres & Sergio Augusto Cunha

SCIENCE AND MEDICINE IN FOOTBALL
https://doi.org/10.1080/24733938.2021.1986227

Classification and determinants of passing difficulty in soccer: a multivariate approach

Murilo Merlin [a], Allan Pinto [a,b], Alexandre Gomes de Almeida [a], Felipe A Moura [c], Ricardo Da Silva Torres [d] and Sergio Augusto Cunha [a]

[a]School of Physical Education, University of Campinas, Campinas, Brazil; [b]Institute of Computing, University of Campinas, Brazil; [c]Laboratory of Applied Biomechanics, State University of Londrina, Londrina, Brazil; [d]Department of Ict and Natural Sciences, Faculty of Information Technology and Electrical Engineering, Ntnu – Norwegian University of Science and Technology, Ålesund, Norway

ABSTRACT

Introduction: Usually, the players' or teams' efficiency to perform passes is measured in terms of accuracy. The degree of difficulty of this action has been overlooked in the literature.

Objectives: The present study aimed to classify the degree of passing difficulty in soccer matches and to identify and to discuss the variables that most explain the passing difficulty using spatiotemporal data.

Results: The data used corresponds to 2,856 passes and 32 independent variables. The Fisher Discriminant Analysis presented 72.0% of the original grouped cases classified correctly. The passes analyzed were classified as low (56.5%), medium (22.6%), and high difficulty (20.9%), and we identified 16 variables that best explain the degree of passing difficulty related to the passing receiver, ball trajectory, pitch position and passing player.

Conclusions: The merit and ability of the player to perform passes with high difficulty should be valued and can be used to rank the best players and teams.In addition, the highlighted variables should be looked carefully by coaches when analyzing profiles, strengths and weaknesses of players and teams, and talent identification context.

Practical Implications: The values found for each variable can be used as a reference for planning training, such as small side games, and in future research.

ARTICLE HISTORY
Accepted 22 September 2021

KEYWORDS
Passing; passing difficulty; tactical-technical; multivariate analysis; soccer

Introduction

Tactics are the central component for success in elite soccer (Rein et al. 2017). Soccer matches have become more complex, faster, and players frequently need to work on reduced space to maintain ball possession (Wallace & Norton, 2014).

In the tactical context, the pass is the main resource used to comply with the match offensive principles, i.e., to maintain possession, to progress in the pitch and to create space and opportunity for scoring as proposed by Ouellette (2004). In addition, it has been considered one of the key performance indicators (Cintia et al. 2015; Goes et al. 2018, 2019). On average, a typical match comprises 1,000 passes (Goes et al. 2018).

In technical terms, the pass in soccer was defined as the deliberate act of touching and projecting the ball on the pitch to another teammate be able to perform a new action, maintaining the possession of the team (Cunha et al. 2011; Wallace & Norton, 2014; Horton et al. 2014). When the ball reaches its intended destination, i.e., his/her teammates, and the receiver is able to perform a new action, either by controlling the ball or performing a new pass, dribble or shot, the pass is considered successful. Usually, the players' or teams' efficiency to perform passes is measured in terms of accuracy, i.e., success rate of the passes, but the degree of difficulty of the action has been overlooked in the literature.

We consider the pass as a technical-tactical action that occurs at time and space contexts, in which the difficulty of the action depends on the interaction of several technical factors (e.g., body position and orientation, ball contact, movement speed, and pass distance) and tactical (e.g., team interaction and space occupation by individual players, group, or by the team) to the ball reaches its destination. Therefore, the passing difficulty refers to the degree of technical and tactical demands that the passing player must complete the action successfully.

The pass has been investigated since Reep & Benjamin (1968), focusing on analyses based on frequency, density, and order of events (Hughes and Franks 2005; Lago and Martín 2007; Chassy 2013; Gyarmati et al. 2014; Mitschke and Milani 2014; Peña and Navarro 2015). Spatiotemporal data provided new perspectives to analyze pass actions. The accurate position of all players on the pitch allowed the proposal of new variables (Bush et al. 2015), metrics (Gyarmati and Stanojevic 2014; Rein et al. 2017; Goes et al. 2018), indices (Cintia et al. 2015), and even predictions. Approaches based on predictive modeling, using regression or classification, has explored different concepts, such as risk and advantage of the passes (Power et al. 2017), value of the passes (Spearman et al. 2017), quality of the passes (Horton et al. 2014), players' involvement in setting up goal-scoring chances by valuing the effectiveness of their passes (Bransen and Haaren 2019).

CONTACT Murilo Merlin murilomerlin7@gmail.com Faculdade De Educação Física, Universidade Estadual De Campinas, Av. Érico Veríssimo, Cidade Universitária, Campinas, São Paulo, Brasil

Studies that aimed to predict the difficulty of the pass used as decision criterion the probability of the analyzed pass to be successful, often employing regression models. For example, the passing ability model is based on the probability that each pass is successful, given information on the environment in which the pass was made and the identity of the player making the pass (Mchale and Relton 2018). Mchale and Relton (2018) aimed to identify key players using network analysis and difficulty passes, but they defined difficulty as a synonym for importance and assumed as a criterion the probability to complete the pass. Power et al. (2017) proposed a logistic regression model to assess the risk and advantage of the pass. As a general idea, the studies start from the same principle: to assign greater weight in the efficiency in performing more difficult passes, often using regression. We found only two similar and complementary studies using classification models applied to pass analysis. Horton et al. (2014) and Chawla et al. (2017) obtained respectively 85% and 92% of accuracy when classifying passes as 'good', 'ok,' or 'bad.' Therefore, they aimed to rate the quality of the pass, not the difficulty. The proposed concept is not clear. What do 'good,' 'ok,' or 'bad' passes mean? In our view, it is essential to contextualize the phenomenon analyzed and to improve the classification model. Furthermore, existing approaches do not include unsuccessful passes, thus, limiting the quantification of the success rate in supposedly more difficult passes.

In our view, there is a need to build a specific concept for the difficulty of the pass, and not to link the difficulty to the probability of success of the action. The concept of passing difficulty can guide the eyes of experts and reduce the subjectivity when classifying pass actions at different levels of difficulty. Furthermore, there is a difference between *difficulty* and *quality or advantage* of the passes. An important pass that promotes tactical advantage, such as an assist or a key pass for example, is not necessarily a difficult task for the passing player. In addition, it is not clear in the literature what quality of the passes means. We focused on the difficulty because we sought to analyze the player's and team's ability to perform passes relativizing by the degree of difficulty. In addition, considering that the passing difficulty has a multivariate nature, it would be important to identify and discuss the variables that best explain this phenomenon, using an interpretative model. The classification model allows ranking the best passing player pondering better performance in high difficulty passes, and the highlighted variables can reveal characteristics of performing passes, identify weaknesses and strengths of players and teams, guide training processes, and contribute to the talent identification.

The present study aimed to: (i) classify the degree of passing difficulty in soccer matches; (ii) identify and discuss the variables that most explain the passing difficulty using spatiotemporal data. Our hypothesis is that the degree of passing difficulty depends on the technical and tactical variables combination associated with the passing player, receiver player, ball trajectory, and the pitch position where the action occurred.

Methods

Data collection and sample

The Ethics Committee of the University of Campinas approved this research. The total samples used in this study corresponds to 2,856 passes (714.0 ± 100.3) obtained from four matches involving five teams from the first division Brazilian Football Championship 2016. Team 1 (18° ranked) played the four home matches against: team 2 (2° ranked), team 3 (17° ranked), team 4 (15° ranked) and team 5 (4° ranked) out of a total of 20 teams. Passes blocked (2.7%) and passes from corners and free kicks (8.9%) were not included in our analysis. Blocked passes limits estimating the possible pass receiver and passes from set pieces have specific characteristics which would limit the model's power to predict. Passes intercepted within a distance of less than 2 m were considered blocked. Approximately 20% of the total samples (n = 465 passes) were randomly separated for the passes labeling process.

The matches were recorded by two digital cameras Sony Handycam HDR-CX405, with HD resolution and acquisition frequency of 30 Hz. To obtain the players' 2D position data from the matches, we first sampled original data to 15 Hz using the Virtual Dub software and then we used the software DVideo, which is a semiautomatic tracking system (Pascual et al. 2002; Figueroa et al. 2006). The players of each team were labeled as p = 1, 2, . . ., 14, including starting players and substitutes. Therefore, the 2D coordinates of each player (2D matrix) were defined as Xp(t) and Yp(t), where t represents each instant of time, while the X and Y axes represent length and width of the pitch respectively.

A Butterworth third-order low-pass digital filter with a cut-off frequency of 0.4 Hz was used as an external filter according to previous study recommendations (Barros et al. 2007; Misuta 2004). DVideo software has an automatic tracking rate of 94% of the processed frames, an average error of 0.3 m for the determination of player position, and an average error of 1.4% for the distance covered (Barros et al. 2007; Misuta 2004). After the filtering step, we use the DVideo interface to record technical actions, such as pass, ball control, tackle, shot, and dribbling. For the passing action, the record was performed at the exact moment of contact with the ball (origin of the pass), and at the exact moment of the subsequent action (destination of the pass), i.e., a new pass, or ball control, dribbling, shot, and tackle.

Variables

Thirty-two predictor variables (Table 1) were proposed for this study. A part of the variables was originally proposed by the authors and soccer experts' collaboration, and the other part was based on similar previous studies about passes.

Three soccer experts were interviewed separately and answered about the following question: *'In your opinion, which information (technical and tactical actions) can we extract from the match is more relevant to determine the degree of passing difficulty in soccer?'* The soccer experts have the following profiles: Expert 1 – PhD student in sport science and assistant coach in professional soccer; Expert 2 – Master's degree student in sport science and performance analyst in professional soccer; Expert 3 –

Table 1. Tactical variables used and abbreviations, separated by groups.

Groups	Abbreviation	Variables (description)
Pitch position variables	Distance PR_{t1} to target	Distance between passing receiver and target of opponent at t1.
	Opp. btw PR_{t1} and target	Number of opponents between target and passing receiver player in relation X axis at t1.
	Distance PP_{t0} to target	Distance between passing player and target of opponent at t0.
	Distance PR_{t0} to target	Distance between passing receiver and target of opponent at t0.
Ball trajectory variables	Outplayed opp.	Number of opponents between passing player at t0 and passing receiver player at t1 in relation X axis.
	Ball progression	Variation of the ball's position in relation to the X axis between t0 and t1.
	Out ball angle	Angle (θ) between vectors $\overrightarrow{AB}$ and $\overrightarrow{AD}$. Calculation based on the angle between vectors ($\cos\theta = \overrightarrow{AB} * \overrightarrow{AD} / \lvert\overrightarrow{AB}\rvert * \lvert\overrightarrow{AD}\rvert$).
	Passing distance	Passing distance (vector modules $\overrightarrow{AB}$).
	Passing angle	Angle (θ) between vector $\overrightarrow{AB}$ and unit vector $\vec{v}$ oriented by the X axis of the pitch (θ = arctan).
	Ball velocity	Mean velocity estimated by the ratio of the passing distance to the time between t0 and t1.
Passing receiver variables	Density PR_{t0}	Number of opponents within the 1 m, 2 m, 5 m and 10 m radius in relation to the PR at t0. The distance between all opponents and the passer was calculated.
	Density PR_{t1}	Number of opponents within the 1 m, 2 m, 5 m and 10 m radius in relation to the PR at t1. The distance between all opponents and the passer was calculated.
	Nearest opp. PR_{t1}	Nearest opponent to passing receiver player at t1.
	Nearest opp. PR_{t0}	Nearest opponent to passing receiver player at t0.
	Velocity PR_{t1}	Instantaneous velocity of passing receiver player at t1.
	Displacement PR	Distance performed by passing receiver player between t0 and t1.
	Velocity PR_{t0}	Instantaneous velocity of passing receiver player at t0.
	Velocity nearest opp. PR_{t1}	Instantaneous velocity of nearest opponent to passing receiver player at t1.
Passing player variables	Nearest opp. PP_{t0}	Distance between passing player and his nearest opponent at passing moment (t0).
	Density PP_{t0}	Number of opponents within the 1 m, 2 m, 5 m and 10 m radius in relation to the PP at t0. The distance between all opponents and the passer was calculated.
	Velocity PP_{t0}	Instantaneous velocity of passing player at t0.
	Velocity nearest opp. PP_{t0}	Instantaneous velocity of nearest opponent to passing player at t0.
	Opponent angle	Angle (θ) between vectors $\overrightarrow{AB}$ and $\overrightarrow{AC}$ at t0. ($\cos\theta = \overrightarrow{AB} * \overrightarrow{AC} / \lvert\overrightarrow{AB}\rvert * \lvert\overrightarrow{AC}\rvert$).

Abbreviations: opp = opponent; PP_{t0} = passing player at the time of the pass execution; PR_{t0} = passing receiver at the time of the pass execution; PR_{t1} = passing receiver at the time of the receipt of the pass; btw = between.

Assistant coach in professional soccer. Each expert has more than 10 years of experience working with soccer. The experts' answers were compiled and analyzed by the authors of this study. Later, implemented as predictor variables from the spatiotemporal data of the two teams. The main suggestions of the experts were: ball velocity; distance and velocity of the nearest opponent to the passing player and passing receiver; number of opponents within a given radius in relation to the passing player and passing receiver; passing distance; and distance between the position of the passing player in relation to the opponent's target.

Other variables were inspired by similar studies: velocity of the player in possession and the intended receiver, nearest opponent angle to the passing line, one touch or not (Power et al. 2017), number of outplayed defenders (Rein et al. 2017), the level of pressure that the opposition team put on the passing player and passing receiver of the pass (Mchale and Relton 2018). Complementarily, some variables were proposed by the authors of the present study, such as distance performed by passing receiver, ball progress, out ball angle, and number of opponents between target and passing receiver. These variables were divided into groups and contributed as observation points for judgment (labeling process) by another group of experts. The observation points proposed were: a) pressure on the passing player; b) pressure on the passing receiver; c) ball trajectory; d) pitch position; and e) passing player techniques.

To evaluate the passes, we considered two different moments: the origin of the pass (t_0), i.e., the exact moment of the contact with the ball by the passing player (PP); and destination of the pass (t_1), i.e., the exact moment of the contact with the ball in the subsequent action by the receiver player (RP), who may be his teammate (successful pass), or opposing team by intercepting the pass or ball out of play (unsuccessful passes). In both moments, we recorded the 2D positional information (XY) of the passing player ($PP_{(t0)}$) and the passing receiver player ($PR_{(t0)}$ and $PR_{(t1)}$), as well as all other players from both teams, team 1 ($XY_1, XY_2, \ldots, XY_{14}$) and team 2 ($XY_{15}, XY_{16}, \ldots, XY_{29}$). We consider the pass as a vector ($\overrightarrow{AB}$) originating from $PP_{(t0)}$ (A) and ending in $PR_{(t1)}$ (B), projected on the pitch (Figure 1). Another vector, $\overrightarrow{AC}$, was based on the $PP_{(t0)}$ nearest opponent, i.e., with the origin in A and the extremity in the position nearest opponent (OP) to the passing player at t_0

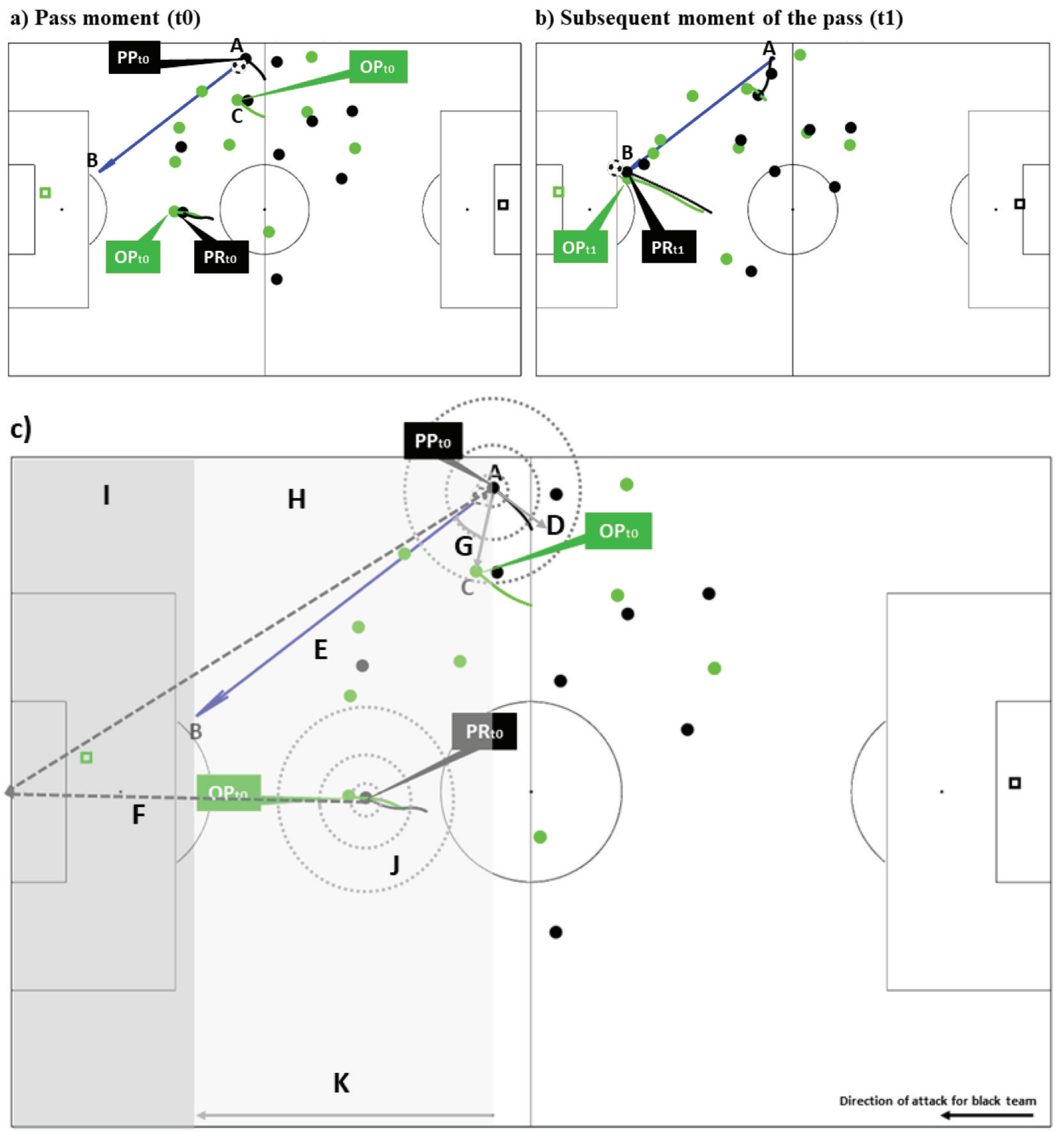

Figure 1. A) Illustration of the real pass situation, at the moment of contact with the ball (t0). PP_{t0} = passing player at the moment of the pass; PR_{t0} = receiver at the moment of the pass; OP_{t0} = nearest opponent to the passing player and receiver at the moment of the pass; A = origin of the pass; B = destination of the pass; C = OP_{t0} position. b) Illustration of the real pass situation at the moment of reception (t1). PR_{t1} = receiver at the moment of the reception of the pass. OP_{t1} = nearest opponent to the receiver when receiving the pass. c) Variables that describe the passing difficulty at the moment of the pass (t0). Abbreviations: ($\overrightarrow{AB}$) = passing distance; ($\overrightarrow{AC}$) distance between passing player and his nearest opponent at t0; ($\overrightarrow{AD}$) = fictitious vector that represents the direction PP before to perform the pass. E = distance between passing player and target of opponent at t0; F = distance between passing receiver and target of opp. at t0; G = opponent angle; H = number of outplayed opponent (into light gray shaded area); I = opponent between PRt1 and target (into dark gray shaded area); J = number of opponents within the 1 m, 2 m, 5 m and 10 m radius to passing receiver at t1; K = Ball progression. Black team attacks to the left and gray team attacks to the right.

moment, $OP_{(t0)}$ (C). The position variation of the PP also constituted an important vector, $\overrightarrow{AD}$, originating in (A), and extremity in (D), Figure 1.

In cases that the player did not perform a pass successfully (for instance, this pass was intercepted by an opponent) the position of the possible receiver of the pass (expected receiver – ER) was estimated according to the equation $ER = \frac{distance}{shortestdistance} \cdot \frac{angle}{shortestangle}$, as proposed by (Power et al. 2017). The ER position at the moment of the passing receipt, $ER_{(t1)}$, was used as $\overrightarrow{AB}$ vector extremity when passes were considered as an unsuccessful action and the calculation of other variables were based on the possible receiver position, both at t_0 and at t_1. This criterion was adopted considering that it is essential to observe characteristics of the PP intention to judge and determine its difficulty.

All variables were derived from the spatiotemporal data of all players on the pitch, at times t_0 and t_1 as explained above, and implemented using the Matlab®2018b software license number 40,604,077.

Labeling process

Two experts (researchers and coaches in soccer) performed, separately, the labeling process passes through judgment. Before judging the 465 passes, they were instructed about passing difficulty concepts, about points of observation, and were submitted to familiarization by watching examples of passes with different degrees of difficulty. For the purpose of this study, passing difficulty was defined as the degree of technical and tactical demands that the passing player must complete the action successfully. Then, they watched videos of passes and assigned a classification for each event: class 1 (low difficulty), class 2 (medium difficulty), and class 3 (high difficulty). Experts watched 10 familiarization passes events before starting the labeling process. Experts could review the passes until they have a clear judgment. When they agreed about classification of the passes, the judgments were validated. When there was disagreement, a third expert decided about the classification. We observed an inter-rater agreement between the experts of 80.2% in the labeling process, which corresponds to 373 events out of the 465 passes that comprise the data set used in this study. This result suggests a substantial agreement level (kw = 0.75) between the experts. Only the classification of the first two experts was considered for the agreement test.

The soccer experts in this step have the following profiles: Expert 1 – PhD student in sport science; Expert 2 – PhD student in sport science; Expert 3 – Master's degree student in sport science and coach in soccer. Each expert has more than 10 years of experience researching and/or working with soccer.

The labels specified by the experts comprised the dependent variables. At the end of this process, we had a data set composed by 465 events (passes), 32 indep endent variables, and three classes of dependent variables (classes): $X = \left\{ \underset{-\S_1}{x}, \underset{-\S_2}{x}, \ldots \underset{-\S}{x} n \right\}$, where $x_i \in R^m$ and $m = 32$; and $Y = \left\{ y_1, y_2, \ldots y_3 n \right\}$, where $y_i \in$ {low difficulty, medium difficulty, high difficulty}.

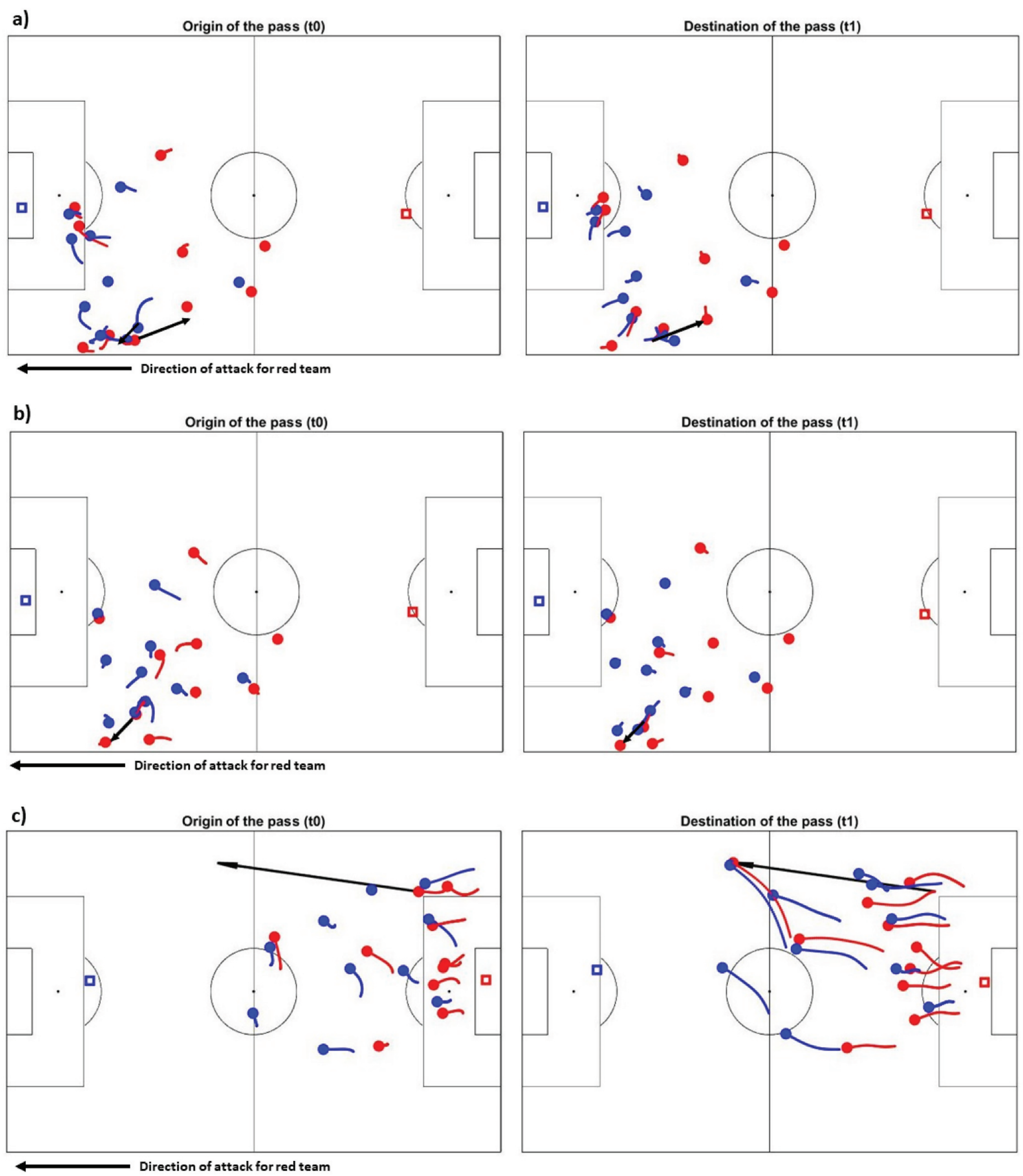

Figure 2. Illustration of real pass situation classified by model. Origin of the pass = at the moment of contact with the ball (t0); Destination of the pass = at the moment of reception (t1). a) Example of low difficulty pass. b) Example of medium difficulty pass. c) Example of high difficulty pass classified. Red team attacks to the left and blue team attacks to the right.

Statistical analysis

We adopted the use of the weighted kappa method (kw) to measure the inter-rater agreement between the experts (Cohen 1968). A fisher's discriminant analysis (FDA) was used to classify the passes into three groups and identify which variables best discriminate them. Also, we used the leave-one-out cross-validation method to validate the proposed method. The interpretation of the obtained model took into consideration the Eigenvalue and structure coefficients (greater than |0.30|) that better distinguish the groups (Pedhazur and Manning 1997).

Also, we used the One-way ANOVA method to compare sixteen variables selected into different classes (low, medium, and high difficulty pass), and Tukey's post-hoc test considering a significance level at 0.001. The statistical analyses were performed in the IBM SPSS Statistics for Windows (Armonk, NY: IBM Corp). In addition, it was observed the standardized mean differences and respective 99% confidence limits (CL), as well as magnitude of observed differences based on effect size (Cohen´s *d*), where the thresholds were <0.2, trivial; 0.6, small; 1.20, moderate; 2.0, large; and >2.0, very large (Hopkins 2009).

Results

The distributed of the 465 passes by experts into three classes considered in this study was 56.6% for the low difficulty passes (class 1), 22.6% for the medium difficulty passes (class 2), and 20.9% for the high difficulty passes (class 3). Figure 2 shows an example of a pass for each class. The FDA presented a total of 72.0% of the original grouped cases classified correctly. The percentage of successful passes within each class was 49.3% to low difficulty passes, 84.0% to medium difficulty passes, and 63.9% to high difficulty passes.

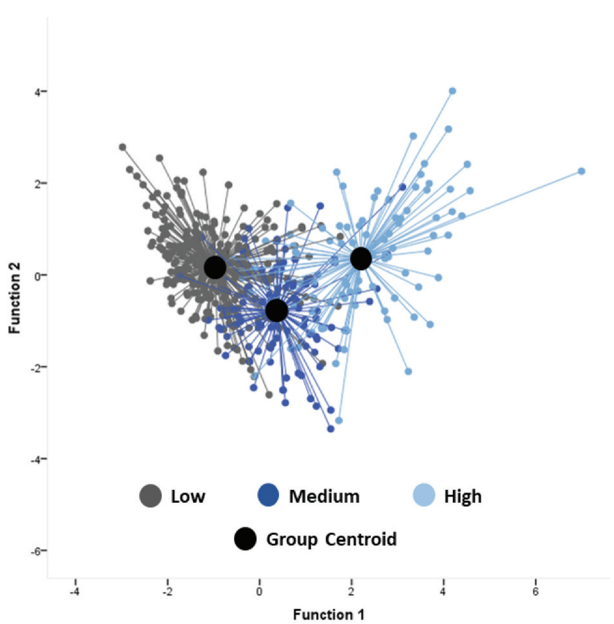

Figure 3. Territorial maps of the group centroid and their respective passes groups (low = low difficulty; medium = medium difficulty; long = long difficulty) based on two canonical discriminant functions.

Subsequently, the FDA was used to identify which variables most explain the passes classification in low, medium, and high difficulty. The model consisted of two discriminant functions, with function 1 representing 89.6% of the total variance and function 2 representing 10.4% (Figure 3). The canonical correlations of functions 1 and 2 were, respectively, 0.78 and 0.39, with both functions being statistically significant (p <0.0001), (Wilks' Lambda = 0.32 and 0.84 for functions 1 and 2, respectively). The discriminant scores of the variables for each function are shown in Table 2.

The variables highlighted in function 1 in order of relevance based on structure coefficient (SC) were: Opponents between PR_{t1} and target, Density (5 m) PRt0, Outplayed opponents, Density (5 m) PR_{t1}, Nearest opponent PR_{t1}, Nearest opponent PR_{t0}, Ball progress, Density (2 m) PR_{t1}, Density (10 m) PRt1, Velocity PR_{t1}, Density (10 m) PR_{t0}, Displacement PR, Distance PR_{t1} to target. For function 2, the variables highlighted were: Nearest opponent PP, Density (10 m) PP, Density (5 m) PP. Table 2 presents the descriptive and inferential analysis for each variable, for the three classes, as well as the structure coefficients (SC) and discriminant function coefficients (FC) for each function. Figure 4 shows the comparison between three classes for each of sixteen variables highlighted by FDA.

The FDA revealed through function 1 that the most important variables to determine the passing difficulty in soccer matches are related to the passing receiver, ball trajectory, and pitch position. In relation to the passing receiver, pressure variables at moment of the pass Density (5 m and 10 m) PRt0 and Nearest opponent PRt0 and at moment of the receipt, Density (2 m, 5 m, and 10 m) PRt1 were highlighted. In addition, kinematic variables related to the displacement of the receiver, Displacement PR, and Velocity PRt1 were also highlighted. For the ball trajectory, function 1 highlighted variables that quantify the number of opponents beat with the pass (outplayed opponents) and the progression of the ball in relation to the depth of the pitch (Ball progress). Besides that, two other highlighted variables, Opponents between PRt1 and target and Distance PRt1 to target represent, respectively, how many players there are between the receiver and the opposing target, and the position of the receiver when receiving the pass. Function 2, which explained only 10.4% of the variance, highlighted variables related to the pressure on the passing player at the time of the pass, Nearest opponent PP, and Density (5 m and 10 m) PP.

Discussion

The present study aimed to classify the degree of passing difficulty in soccer matches and identify and discuss the variables that most explain the passing difficulty using spatiotemporal data. In the first step, the FDA presented 72.0% of accuracy when classifying the degree of passing difficulty into three classes. The function coefficient for each highlighted variable is shared in Table 2 and can be used to classify future datasets. In the second step we identified 16 variables that best explain the degree of passing difficulty in soccer. Besides contributing to the predictive ability of the model, the present study discussed the variables highlighted under the perspective of practical implications for the match.

Table 2. Descriptive and inferential statistics of three different classes (low, medium and high difficulty) of the passes.

Variables	Low (Mean ± SD)	Medium (Mean ± SD)	High (Mean ± SD)	Low vs Med	Low vs High	Med vs High	F1 (SC) 89.6%	F2 (SC) 10.4%	F1 (FC) 89.6%	F2 (FC) 10.4%
Opp. btw PRt1 and target	$8.84^{ab} \pm 2.20$	$7.00^{c} \pm 2.32$	4.90 ± 2.25	−1.84 ± 0.66 −0.83 (Moderate)	−3.94 ± 0.68 −1.78 (Large)	−2.1 ± 0.83 −0.93 (Moderate)	−0.562*	0.062	−0.227	0.340
Distance PRt1 to target	$56.14^{b} \pm 16.79$	$51.64^{c} \pm 15.62$	37.84 ± 19.75	−4.50 ± −4.92 −0.27 (Small)	−18.30 ± 5.42 −1.03 (Moderate)	−13.80 ± 6.49 −0.78 (Moderate)	−0.324*	−0.190	−1.196	−2.696
Outplayed opponents	$0.54^{ab} \pm 1.04$	$1.28^{c} \pm 1.69$	2.82 ± 2.68	0.74 ± 0.37 0.58 (Small)	2.29 ± 0.51 1.38 (Large)	1.55 ± 0.81 0.69 (Moderate)	0.426*	0.143	0.180	0.534
Ball progress	$0.02^{b} \pm 8.71$	$4.35^{c} \pm 11.43$	12.82 ± 15.76	4.33 ± 2.85 0.45 (Small)	12.8 ± 3.40 1.16 (Moderate)	8.47 ± 5.01 0.62 (Moderate)	0.356*	0.102	−0.568	−0.641
Density PRt0 (5 m)	$0.18^{ab} \pm 0.44$	$0.46^{c} \pm 0.57$	1.08 ± 0.85	0.27 ± 0.14 0.57 (Small)	0.90 ± 0.18 1.55 (Large)	0.63 ± 0.26 0.87 (Moderate)	0.480*	0.188	0.316	0.245
Density PRt1 (5 m)	$0.40^{ab} \pm 0.66$	$0.74^{c} \pm 0.69$	1.35 ± 0.85	0.34 ± 0.20 0.51 (Small)	0.95 ± 0.22 1.32 (Large)	0.61 ± 0.28 0.78 (Moderate)	0.415*	0.089	0.105	0.239
Nearest opponent PRt1	$8.09^{ab} \pm 4.60$	$4.82^{c} \pm 3.23$	3.16 ± 2.72	−3.27 ± 1.27 −0.77 (Moderate)	−4.93 ± −1.28 −1.18 (Moderate)	−1.66 ± 1.09 −0.55 (Small)	−0.406*	0.278	−0.026	0.277
Nearest opponent PRt0	$10.11^{ab} \pm 5.43$	$6.74^{c} \pm 4.39$	4.06 ± 3.36	−3.38 ± 1.54 −0.65 (Moderate)	−6.05 ± 1.52 −1.22 (Large)	−2.68 ± 1.44 −0.68 (Moderate)	−0.403*	0.153	−0.226	−0.171
Density PRt1 (2 m)	$0.04^{ab} \pm 0.20$	$0.19^{c} \pm 0.39$	0.42 ± 0.52	0.15 ± 0.08 0.55 (Small)	0.38 ± 0.09 1.20 (Moderate)	0.23 ± 0.16 0.51 (Small)	0.354*	0.044	0.094	0.060
Density PRt1 (10 m)	$1.36^{ab} \pm 1.23$	2.11 ± 1.15	2.73 ± 1.42	0.75 ± 0.36 0.62 (Moderate)	1.37 ± 0.39 1.07 (Moderate)	0.62 ± 0.47 0.48 (Small)	0.353*	0.125	0.181	−0.338
Velocity PRt1	$7.34^{ab} \pm 4.97$	$11.04^{c} \pm 6.15$	13.63 ± 7.30	3.70 ± 1.59 0.69 (Moderate)	6.29 ± 1.75 1.10 (Moderate)	2.59 ± 2.46 0.38 (Small)	0.352*	−0.165	0.251	−0.248
Density PRt0 (10 m)	$1.09^{b} \pm 1.24$	$1.59^{c} \pm 1.16$	2.48 ± 1.58	0.50 ± 0.36 0.41 (Small)	1.39 ± 0.41 1.04 (Moderate)	0.89 ± 0.50 0.65 (Moderate)	0.334*	0.075	−0.087	−0.378
Displacement PR	$3.55^{ab} \pm 3.01$	$5.91^{c} \pm 5.69$	8.48 ± 6.96	2.36 ± 1.18 0.59 (Small)	4.92 ± 1.36 1.11 (Moderate)	2.56 ± 2.32 0.40 (Small)	0.330*	−0.048	−0.188	−0.320
Nearest opp. PP	$6.02^{ab} \pm 4.23$	3.19 ± 1.92	3.53 ± 2.56	−2.83 ± 1.11 −0.76 (Moderate)	−2.49 ± 1.18 −0.64 (Moderate)	0.34 ± 0.82 0.15 (Trivial)	−0.251	0.482*	−0.020	0.369
Density PP (10 m)	$1.62^{ab} \pm 1.20$	2.50 ± 1.17	2.30 ± 1.28	0.88 ± 0.35 0.73 (Moderate)	0.68 ± 0.37 0.55 (Small)	−0.20 ± 0.44 −0.16 (Trivial)	0.204	−0.463*	0.198	−0.301
Density PP (5 m)	$0.67^{ab} \pm 0.75$	1.18 ± 0.81	1.05 ± 0.74	0.52 ± 0.22 0.67 (Moderate)	0.39 ± 0.23 0.52 (Small)	−013 ± 0.28 −0.16 (Trivial)	0.186	−0.443*	0.087	−0.129

Mean ± standard deviation (SD), mean difference and respective 99% confidence limit (CL), effect size based on Cohen´s *d*, structure coefficient (SC), function coefficient (FC) of 16 variables selected by the FDA model. *Variable better explained by function 1 or 2. One-way ANOVA and the Bonferroni post hoc to differentiate between groups (a = difference between Low and Medium; b = difference between Low and High; c = difference between Medium and High; $p < 0.001$). Abbreviations: Opp = opponent.; F1 = Function 1; F2 = Function 2; Med = Medium.

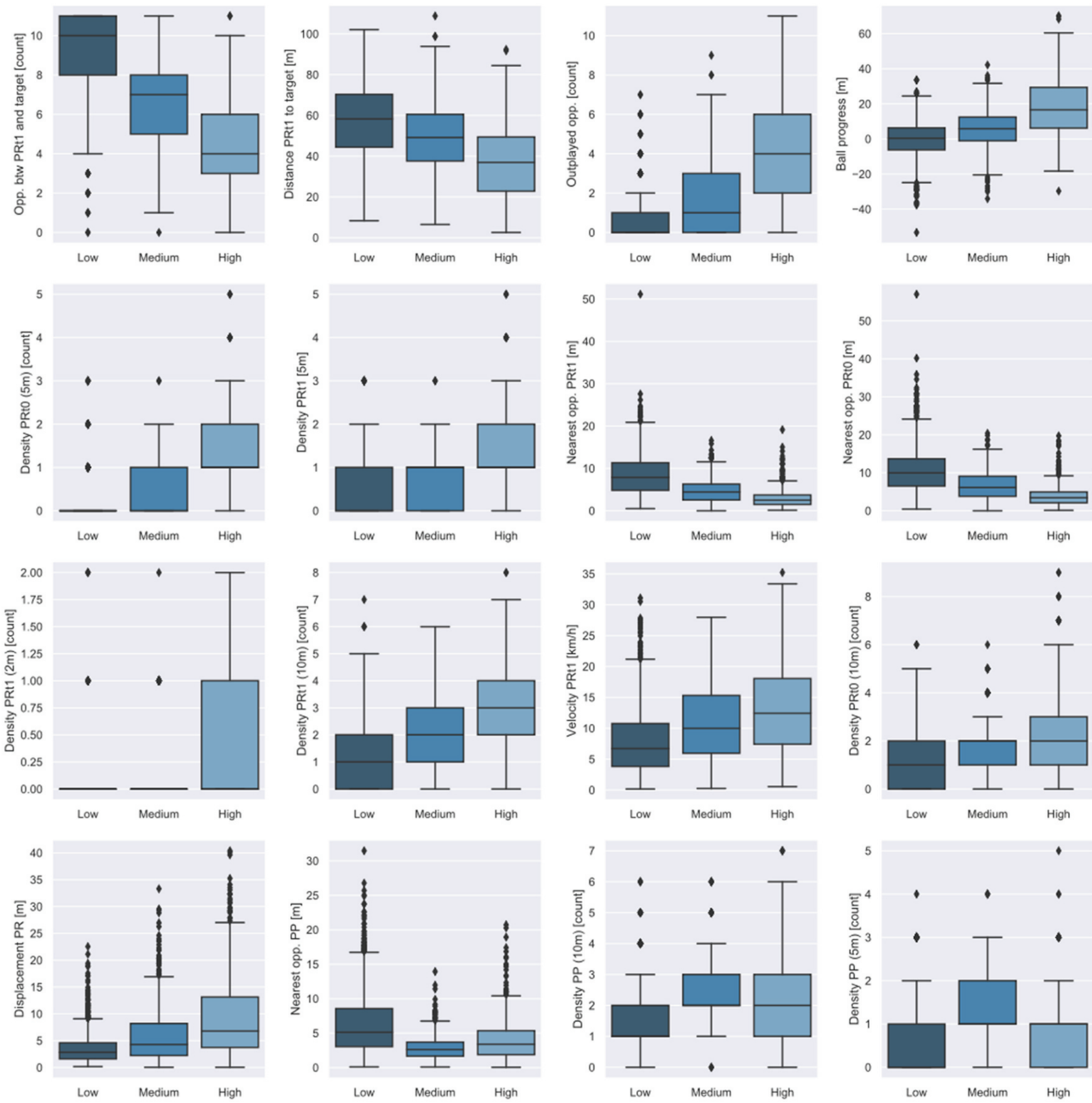

Figure 4. Comparison between three classes (low, medium, and high difficulty) of the passes for each of sixteen variables highlighted by FDA.

Recently, some studies have been proposing metrics, indices or predictions to improve the level of pass information and surpass the traditional information about the success rate of passes. The main proposals aimed to attribute merit to the pass, i.e., the advantage that the pass provides for the match (Gyarmati and Stanojevic 2014; Goes et al. 2018; Bransen and Haaren 2019), or to predict the difficulty of the pass (Power et al. 2017; Mchale and Relton 2018, 2018). The studies that aimed to predict the difficulty of the pass used regression-based models, where the classifiers are trained to produce continuous output, between 0 and 1.

The fundamental difference between the studies cited and the present study is that we have proposed a model of difficulty of the pass centered on an original concept and that represents the phenomenon analyzed from experts' perspective. In addition, we focused on the difficulty because we wanted to analyze the player's ability to perform passes relativizing by the degree of difficulty, i.e., what is the success rate of players and teams in performing difficult passes? In our view, players and teams with a higher success rate on difficult passes should be valued. In our sample, 87.5% of passes were classified as successful. When we analyzed the percentage of successful passes in each class, we observed that high difficulty passes had a success rate of 52.1% only, followed by 91.5% for medium difficulty passes and 98.9% for low difficulty passes. These numbers justify the importance of analyzing successful and unsuccessful passes relativizing by the difficulty of the action. These success rates in different classes allow ranking the best passing player pondering better performance in high difficulty passes. Thus, the merit and ability of the player to perform passes with high difficulty are contemplated. The only two studies with similar design explored prediction by passes classification. Horton et al. (2014) obtained 85% accuracy and Chawla et al. (2017) obtained 85% accuracy when classifying passes as 'good', 'ok', or 'bad'. In both studies, the authors designed a model that computes a vector of predictor variables for each pass made and uses machine learning techniques to determine a classification function that can accurately rate passes. The limitations of their study in relation to ours are the absence of the concept of pass quality of the pass, and the fact that their work did not include unsuccessful passes, limiting the analysis of the ability of players and teams to perform supposedly more difficult passes.

Another novelty of this study was the identification and discussion of the variables that best explain the difficulty in performing passes and bring this information to a more applied context. Studies usually test variables to improve the accuracy of the prediction, but do not necessarily discuss the impact of each variable in the context of the match. In this study step, we identified 16 between 32 variables that best explain the degree of passing difficulty in soccer. These variables made it possible to quantitatively describe low, medium, and high difficulty passes and allow to classify further datasets with the discriminant function coefficients presented.

The most determining variable in function 1 was Opponents between PRt1 and target, originally proposed by this study. High difficulty passes have approximately five opponents between the receiver player and the target. This variable joint with the variable Distance PRt1 to target compose the group of variables that represent the position in the pitch. The results showed that the position of the passing receiver is more important than the position of the passing player in determining the difficulty of the pass, and put the forward in a more promising position to perform the shot is more difficult task for the passing player, and therefore, must be valued.

Another important attention point was the variables related to the ball trajectory. It has been common to use angle and distance information from the pass to improve the level of information about this action (Bush et al. 2015; Goes et al. 2018). In a similar predictive study, the authors highlighted the variable passing distance as important for predicting successful passes (Mchale and Relton 2018). In the present study, angle and distance demonstrated not to have a relevant influence on the passing difficulty. An offensive, but short pass probably does not require difficulty for the passing player, as well as a long and defensive pass. On the other hand, a pass that progresses towards the target and that beats opponents is more challenging. Therefore, the variables Ball progress and mainly Outplayed opponents were more determinant for the model. The variable Outplayed opponents was also an object of investigation in other studies. Rein et al. (2017) observed that passes that won more opponents are more effective and are related to the success in matches. In addition, this variable represents the relationship of interaction between teams, which emphasize the importance of using a tracking system able to obtain data from both teams, such as multicamera systems.

Other variables highlighted by the FDA are related to the passing player and passing receiver, mainly the pressure variables. Pressure variables have been widely used in the literature, especially on-the-ball player in possession (Link et al. 2016, 2017). In similar studies, the authors highlighted the importance of pressure variables on passing and receiver player in predicting the difficulty (Mchale and Relton 2018; Power et al. 2017) or quality of the pass (Chawla et al. 2017). In the present study, the pressure variables on the passing receiver were highlighted in function 1, which explains 89.6% of the total variance, and therefore, they are more determinant than the pressure variables on the passing player, highlighted in function 2. In addition, the Nearest opponent PRt0 variable showed a large difference when comparing low and high difficulty passes. In a practical context, we can suggest the importance of the passing receiver moving farther away from the opponents and facilitate the passing action of his teammates. Also, two other highlighted variables were, Velocity PRt1 and Displacement PR. Both variables were originally proposed in the present study and explains a higher degree of requirement for the passing player when the pass receiver is in greater and faster displacement.

Another novelty of this study is that we showed and compared the values of the variables in the three classes of passing difficulty which can be used as a reference in similar studies and practical context. In general, high difficulty passes can be characterized as high pressure on the receiver player at the passing moment (4.06 ± 3.36 m), as well as at the receipt moment (3.16 ± 2.72 m), greater displacement (8.48 ± 6.96 m), and speed (13.63 ± 7.30 km/h) of the receiver between t0 and t1, greater progression of the ball (12.82 ± 15.76 m) and rupture of opponents on the pitch (2.82 ± 2.68), greater proximity to the opponent's goal (37.84 ± 19.75 m), and fewer opponents between the receiver and the opponent's target (4.90 ± 2.25). With less relevance, greater pressure on the passing player at the passing moment (3.53 ± 2.56 m) may be considered.

As practical implications, we highlight three main reasons for using the highlighted variables within the context of match analysis in soccer. First, the highlighted variables can reveal characteristics of performing passes by players and teams. For example, it is possible to identify passing players that win more opponents and/or put their teammates in a better condition to shot, with fewer opponents and closer to target. In addition, it is possible from the highlighted variables to identify weaknesses of players and teams, i.e., which variables best explain unsuccessful passes. These first two practical implications could compose individual and collective performance indicators for match and season reports, or even talent identification implications. The third practical implication concerns the training process. From the previous information, it is possible to guide training processes in order to reduce weaknesses and enhance detected strengths for effective offensive and defensive actions. In addition, the values of the variables can be used as a reference for specific pass training, providing tasks with different levels of difficulty.

The main limitation of this study was the number of events analyzed. Although it was sufficient to support the proposed model, a larger sample would be needed to compare players and teams, and to explore some potential practical implications such as those described. In addition, the model could be applied in other leagues and different contexts such female soccer and young soccer to generalize the results. Another limitation of this study can be attributed to the DVideo software. Although it has been widely used in research on soccer and other sports, it still lacks validity to measure displacements at high speed and intra- and inter-evaluator reproducibility, considering that it is a semi-automatic instrument. Therefore, the results obtained must be analyzed with caution.

We confirmed our hypothesis, where the technical and tactical variables combination associated with the passing player, receiver player, ball trajectory, and the pitch position were determinant to classify degree of passing difficulty in soccer matches.

Conclusions

The present study contributed to a more accurate analysis of an extremely frequent and determinant action in soccer matches. Passes in soccer matches can be classified not only for their success rate, but also based on their difficulty degree. This allows determining the ability of players and teams to successfully perform low, medium, and high difficulty passes.

The merit and ability of the player to perform passes with high difficulty should be valued, and can be used to rank and discriminate the best players and teams when performing passes. In addition, the highlighted variables should be looked at more carefully by coaches when analyzing profiles, strengths and weaknesses of players and teams, and talent identification context. The values found for each variable can be used as a reference for planning training, such as small side games, and in future research.

Future research could focus on increasing the number of events, based on other competitive leagues, levels, age groups. In addition, the highlighted variables can help as a basis for other predictive models aiming at improving the accuracy in the classification of the passing difficulty in soccer matches.

Disclosure statement

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

Funding

This work was supported by the Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP) under Grants [#2016/50250-1, #2017/20945-0, #2018/19007-9 and #2019/16253-1]; Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brazil (CAPES), Finance Code 001".

ORCID

Murilo Merlin http://orcid.org/0000-0002-6156-4980
Allan Pinto http://orcid.org/0000-0003-3765-8300
Alexandre Gomes de Almeida http://orcid.org/0000-0002-0374-2718
Felipe A Moura http://orcid.org/0000-0002-0108-7246
Sergio Augusto Cunha http://orcid.org/0000-0003-1927-0142

References

Barros R, Misuta M, Menezes R, Figueroa P, Moura F, Cunha S, ... Leite N. 2007. Analysis of the distances covered by first division Brazilian soccer players obtained with an automatic tracking method. J Sports Sci Med. 6:233–242.

Bransen, Lotte, Van Haaren, Jan and van de Velden, Michel. "Measuring soccer players' contributions to chance creation by valuing their passes" Journal of Quantitative Analysis in Sports, vol. 15, no. 2, 2019, pp. 97–116. https://doi.org/10.1515/jqas–2018–0020

Bush M, Barnes C, Archer DT, Hogg B, Bradley PS. 2015. Evolution of match performance parameters for various playing positions in the English Premier League. Hum Mov Sci. 39:1–11. doi:10.1016/j.humov.2014.10.003

Chassy, P. (2013). Team Play in Football: How Science Supports F. C. Barcelona's Training Strategy. Psychology, 4, 7–12. doi:10.4236/psych.2013.49A2002

Chawla S, Estephan J, Gudmundsson J, Horton M. 2017. Classification of passes in football matches using spatiotemporal data. ACM Trans Spatial Algorithms Syst. 3(2):2. doi:10.1145/3105576.

Cintia P, Giannotti F, Pappalardo L, Pedreschi D, Malvaldi M 2015. The harsh rule of the goals: data-driven performance indicators for football teams. Proceedings of the 2015 IEEE International Conference on Data Science and Advanced Analytics, DSAA 2015; Paris, France. 10.1109/DSAA.2015.7344823

Cohen, J. (1968). Weighted kappa: Nominal scale agreement provision for scaled disagreement or partial credit. Psychological Bulletin, 70(4), 213–220. https://doi.org/10.1037/h0026256

Cunha SA, Moura FA, Santiago PRP, et al. Futebol : aspectos multidisciplinares para o ensino e treinamento. Edição 1. Rio de Janeiro: : Editora Guanabara Koogan Ltda 2011

Figueroa PJ, Leite NJ, Barros RML. 2006. Background recovering in outdoor image sequences: an example of soccer players Background recovering in outdoor image sequences: an example of soccer players segmentation. Image Vis Comput. 24(4):363–374. doi:10.1016/j.imavis.2005.12.012.

Goes F, Kempe M, Lemmink K, Goes F, Kempe M, Lemmink K 2019. Predicting match outcome in professional Dutch football using tactical performance metrics computed from position tracking data Predicting match outcome in professional Dutch soccer using tactical performance metrics computed from position tracking data.

Goes F, Kempe M, Meerhoff M, Lemmink K. 2018. Not every pass can be an assist: a data-driven model to measure pass effectiveness in professional soccer matches. Big Data. 6(4):1–28. doi:10.1089/big.2018.0067.

Gyarmati L, Kwak H, Rodriguez P. 2014. Searching for a Unique Style in Soccer. arXiv. 5–8. Recuperado de http://arxiv.org/abs/1409.0308

Hopkins WG, Marshall SW, Batterham AM, Hanin J. 2009. Progressive statistics for studies in sports medicine and exercise science. Med Sci Sports Exerc. 41(1):3–12. doi:10.1249/MSS.0b013e31818cb278.

Horton M, Gudmundsson J, Chawla S, Estephan J. 2014. Classification of passes in football matches using spatiotemporal data. ACM Trans Spatial Algorithms Syst. 3:2. doi:10.1145/3105576

Hughes M, Franks I. 2005. Analysis of passing sequences, shots and goals in soccer. J Sports Sci. 23(5):509–514. doi:10.1080/02640410410001716779.

Lago C, Martín R. 2007. Determinants of possession of the ball in soccer. J Sports Sci. 25(9):969–974. doi:10.1080/02640410600944626.

Link D, Hoernig M, Nassis G, Laughlin M, Witt JD. 2017. Individual ball possession in soccer. Plos One. 12(7):e0179953. doi:10.1371/journal.pone.0179953.

Link D, Lang S, Seidenschwarz P. 2016. Real time quantification of dangerousity in football using spatiotemporal tracking data. PLoS ONE. 11 (12):1–16. doi:10.1371/journal.pone.0168768.

Mchale I, Relton S. 2018. Identifying key players in soccer teams using network analysis and pass difficulty. Eur J Oper Res. 268(1):339–347. doi:10.1016/j.ejor.2018.01.018.

Misuta, Milton Shoiti. Rastreamento automatico de trajetorias de jogadores de futebol por videogrametria: validação do metodo e analise dos resultados. 2004. 75f. Dissertação (mestrado) - Universidade Estadual de Campinas, Faculdade de Educação Fisica, Campinas, SP. Disponível em: . Acesso em: 6 ago. 2018

Ouellette J. 2004. Principles of Play for Soccer. Strategies. 17 (October2014):3. doi:10.1080/08924562.2004.10591082.

Pascual F, Leite N, Barros R. 2002. A flexible software for tracking of markers used in human motion analysis. Comput Methods Programs Biomed. 72:155–165.

Pedhazur EJ, Manning S. 1997. *Multiple regression in behavioral research* (Trird). Florida: Christopher P. Klein.

Pena JL, Navarro RS. Who can replace Xavi? A passing motif analysis of football players. arXiv. 2015;arXiv:1506.07768

Power P, Ruiz H, Wei X, Lucey P 2017. Not all passes are created equal: objectively measuring the risk and reward of passes in soccer from tracking data. Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining; 1605–1613. https://doi.org/10.1145/3097983.3098051

Reep, C, e B Benajmin. 1968. “Skill and Chance in Association Football”. Journal of the Royal Statistical Society 131(4): 581–85. http://www.jstor.org/stable/2343726

Rein R, Raabe D, Memmert D. 2017. Human movement science “ which pass is better? ” Novel approaches to assess passing effectiveness in elite soccer. Hum Mov Sci. 55(August):172–181. doi:10.1016/j.humov.2017.07.010.

Spearman, William et al. 2017. “Physics-Based Modeling of Pass Probabilities in Soccer”. In *Sports Analytics Conference*, 1–14.

Wallace, Jarryd Luke, e Kevin Ian Norton. 2014. “Evolution of World Cup soccer final games 1966–2010: Game structure, speed and play patterns”. Journal of Science and Medicine in Sport 17(2): 223–28. http://dx.doi.org/10.1016/j.jsams.2013.03.016