Journal of Sports Sciences

Routledge
Taylor & Francis Group

ISSN: 0264-0414 (Print) 1466-447X (Online) Journal homepage: https://www.tandfonline.com/loi/rjsp20

Identifying playing talent in professional football using artificial neural networks

Donald Barron, Graham Ball, Matthew Robins & Caroline Sunderland

JOURNAL OF SPORTS SCIENCES
https://doi.org/10.1080/02640414.2019.1708036

Identifying playing talent in professional football using artificial neural networks

Donald Barron[a], Graham Ball[b], Matthew Robins[c] and Caroline Sunderland[d]

[a]School of Health and Sport Sciences, University of Suffolk, Ipswich, UK; [b]John van Geest Cancer Research Centre, School of Science and Technology, Nottingham Trent University, Nottingham, UK; [c]Institute of Sport, University of Chichester, Chichester, UK; [d]Sport, Health and Performance Enhancement Research Centre, School of Science and Technology, Nottingham Trent University, Nottingham, UK

ABSTRACT

The aim of the current study was to objectively identify position-specific key performance indicators in professional football that predict out-field players league status. The sample consisted of 966 out-field players who completed the full 90 minutes in a match during the 2008/09 or 2009/10 season in the Football League Championship. Players were assigned to one of three categories (0, 1 and 2) based on where they completed most of their match time in the following season, and then split based on five playing positions. 340 performance, biographical and esteem variables were analysed using a Stepwise Artificial Neural Network approach. The models correctly predicted between 72.7% and 100% of test cases (Mean prediction of models = 85.9%), the test error ranged from 1.0% to 9.8% (Mean test error of models = 6.3%). Variables related to passing, shooting, regaining possession and international appearances were key factors in the predictive models. This is highly significant as objective position-specific predictors of players league status have not previously been published. The method could be used to aid the identification and comparison of transfer targets as part of the due diligence process in professional football.

ARTICLE HISTORY
Accepted 7 August 2019

KEYWORDS
Soccer; talent identification; Premier League; championship; artificial intelligence

Introduction

Coaches and decision makers in professional football have traditionally used subjective observations to assess the performance of their team, to review the strengths and weaknesses of future opponents and to identify potential signings (Carling, Williams, & Reilly, 2005). Match analysis research into the individual's performance in football has focused heavily on the physical demands of the sport (Carling, 2013). Research led by sport scientists with a heavy focus upon the physical aspects of performance in football has not managed to identify key predictors of match outcome or team success (Bradley et al., 2016; Carling, 2013).

However, studies investigating physical performance during matches have also incorporated technical elements and provided some insights into the successful performance of players and teams (Bradley et al., 2016, 2013; Dellal et al., 2011; Dellal, Wong, Moalla, & Chamari, 2010). Technical factors have been identified that are prominent predictors of team success and match outcome. Shots, shots on target and ball possession are the most commonly reported predictors (Castellano, Casamichana, & Lago, 2012; Lago-Penas, Lago-Ballesteros, Dellal, & Gomez, 2010; Liu, Gomez, Lago-Penas, & Sampaio, 2015). There has been a heavy emphasis on the attacking aspects of play linked to success and more detailed analysis is required into the defensive aspects of play to gain a greater understanding of the game.

Following on from the research into team success and physical profiles, there has been an increasing interest in the technical profiles of players. Studies have found positional differences in Ligue 1 in France, the Premier League in England and in Spain's La Liga (Dellal et al., 2011, 2010). The development of advanced computer systems has supported a greater understanding of position profiles in football. However, most of the research to date has used subjective methods to select variables for analysis (Taylor, Mellalieu, & James, 2004) or they have replicated indicators used in other studies (Andrzejewski, Konefal, Chmura, Kowalczuk, & Chmura, 2016). Using subjective criteria selection rather than exploring a broad spectrum of the data points has meant that many variables have yet to be assessed. Therefore, the impact of these variables upon playing success and career progression is unknown.

A broader analysis of player performance and career progression has been provided by using artificial neural networks to assess a wide range of variables (Barron, Ball, Robins & Sunderland, 2018). Artificial neural networks have been shown to be better at identifying patterns in complex non-linear data sets than forms of regression analysis and they are capable of generalising results to solve real world problems (Basheer & Hajmeer, 2000; Lancashire, Lemetre, & Ball, 2009; Tu, 1996). In a football context, artificial neural networks have been shown to be capable of creating models that can differentiate between specific groups and identify key variables that predict career progression (Barron et al., 2018). Previous studies though have been limited by assessing players regardless of position and their accuracy could be improved by making assessments of each position and the creation of position-specific career progression models.

CONTACT Donald Barron d.barron2@uos.ac.uk School of Health and Sport Sciences, University of Suffolk, Ipswich, UK
Supplemental data for this article can be accessed here.

To the authors' knowledge there has not been an objective study carried out to develop a position-specific predictive model that could support the scouting and recruitment process in professional football. The efficient and effective identification and assessment of transfer targets is a key aspect of any professional football club and requires a thorough due diligence process. Therefore, the aim of the current study was to develop objective models to identify position-specific key performance indicators in professional football that predict out-field players league status using an artificial neural network.

Methods

Players and match data

The basis of the current study followed Barron et al.'s (2018) method but looked to build on it and focus on position-specific assessments of players. The sample consisted of 966 out-field players (mean ± SD age and height: 25 ± 4 yr, 1.81 ± 0.06 m) who had completed a full 90 minutes in the English Football League Championship during the 2008/09 and 2009/10 seasons (Table 1). Technical performance data and biographical data was collected using ProZone's MatchViewer software (ProZone Sports Ltd., Leeds, UK), the official Football League website (www.efl.com) and Scout7 Ltd's (Birmingham, UK) site. The Prozone MatchViewer system was used to collect performance data due to its accurate inter-observer agreement for the number and type of events (Bradley, O'Donaghue, Wooster, & Tordoff, 2007). The data collected from the Prozone MatchViewer software was made available by STATS LLC (Chicago, USA). Institutional ethical approval was attained from the Non-Invasive Human Ethics Committee at Nottingham Trent University.

In total, 536 variables were collected including the total number, accuracy (% success), means, medians and upper and lower quartiles of passes, tackles, possessions regained, clearances and shots. Additional data on total appearances, playing percentage, total goals and assists, international appearances and heights was also collected. The data set originally included 536 variables but low variance statistics were removed. After removing low variance data points, the data set included 340 variables for comparison. Each player's data was converted into mean 90-minute performance data before they were assigned to one of three categories (group 0, group 1 or group 2).

Table 1. Biographical data represented as means and standard deviations for player groupings.

Group	Players (n)	Age (years)	Height (cm)	90 Minute Appearances	Total Minutes
Group 0 Full Back	56	24.2 ± *4.3*	180.5 ± *4.4*	10.1 ± *10.7*	1112 ± *1040*
Group 1 Full Back	125	24.9 ± *4.2*	180.2 ± *4.3*	20.0 ± *12.1*	2603 ± *1107*
Group 2 Full Back	24	25.4 ± *3.3*	179.7 ± *3.6*	18.5 ± *12.5*	1919 ± *1200*
Group 0 Centre Back	37	27.5 ± *5.1*	187.2 ± *5.1*	15.9 ± *10.9*	15901 ± *1023*
Group 1 Centre Back	131	25.6 ± *3.7*	186.7 ± *4.2*	22.5 ± *12.4*	2186 ± *1116*
Group 2 Centre Back	25	25.6 ± *3.4*	187.4 ± *3.7*	22.8 ± *12.0*	2173 ± *1141*
Group 0 Wide Midfield	42	24.4 ± *4.3*	179.1 ± *5.5*	6.6 ± *7.0*	1119 ± *858*
Group 1 Wide Midfield	103	24.6 ± *3.7*	177.2 ± *5.6*	12.6 ± *9.6*	1840 ± *1000*
Group 2 Wide Midfield	23	24.8 ± *3.7*	179.2 ± *4.8*	19.4 ± *11.5*	2425 ± *1109*
Group 0 Centre Midfield	36	25.6 ± *4.8*	179.7 ± *5.1*	12.4 ± *11.9*	1505 ± *1147*
Group 1 Centre Midfield	148	25.6 ± *3.9*	178.8 ± *5.8*	19.5 ± *11.1*	2238 ± *1006*
Group 2 Centre Midfield	21	26.3 ± *4.5*	178.5 ± *4.5*	25.6 ± *13.6*	2693 ± *1253*
Group 0 Attacker	38	26.6 ± *4.8*	182.2 ± *6.5*	6.2 ± *6.9*	1096 ± *920*
Group 1 Attacker	130	26.0 ± *3.9*	181.6 ± *5.9*	11.8 ± *9.3*	1845 ± *931*
Group 2 Attacker	27	26.2 ± *4.5*	181.7 ± *5.8*	13.2 ± *9.3*	2081 ± *930*

Player grouping

Players were allocated to one of five positions (full back, centre back, wide midfielder, central midfielder or attacker) based on where they spent most of their playing time during the season (See Table 1). They were then assigned to one of three categories (group 0, group 1 or group 2) based on where they went on to complete most of their match time during the following season. The first category (group 0) included the players who completed most of their match time in a lower league during the following season. The second group (group 1) included those players who completed most of their match time in the English Football League Championship during the following season and the final category (group 2) contained the players who progressed to complete most their match time in the English Premier League during the following season.

Sample sizes for each comparison were balanced to have an equal number of cases using a random number selector (i.e., 24 full backs were selected from group 0 to have an equal number of cases for comparisons to group 2). Players who played on loan during the 2008/09 and 2009/10 seasons were included in the study but players who moved to a club outside England were excluded due to the complications in assessing the merits of foreign competitions against those in England. The five positions for each category of playing status were subsequently analysed using a Stepwise Artificial Neural Network approach to identify the optimal collection of variables for predicting playing status.

Artificial neural network model

The artificial neural network modelling was based on the approach previously used in gene profiling with breast cancer data (Lancashire et al., 2009) and used in assessing player performances in the Football League Championship (Barron et al., 2018). It used in house code written in Microsoft Visual Basic 6 to call Statistica 10.0 (Statsoft Inc., Tulsa, USA) artificial neural network model at each loop of the stepwise procedure and output the results in a text format.

Before training the artificial neural network, the data was randomly split (60% for training purposes, 20% for validation and 20% blind test cases). A Monte-Carlo cross validation procedure was used to avoid over-fitting of the data.

The artificial neural network modelling involved a multi-layer perceptron architecture with a feed-forward back-propagation algorithm. This algorithm used a sigmoidal transfer function and weights were updated by feedback from errors. Results were provided for the average test performance and the average test error. The average test performance indicates the percentage of test cases that are correctly predicted. The average test error is the root mean square error for the test data set, this indicates the difference between the values predicted by the model and the actual values of the test data set (Salkind, 2010). Further information on the artificial neural network model can be viewed in the supplementary information.

Results

Analysis using the artificial neural network created fifteen position-specific models to predict an out-field player's league status. The models correctly predicted between 72.7% and 100% of test cases (Mean prediction of models = 85.9%), the test error ranged from 1.0% to 9.8% (Mean test error of models = 6.3%). Fourteen models correctly predicted 75% or more of the test players league status with an error of 9.6% or less (Table 2). The fifteen models, created in total, contained between five and twenty variables to predict the players league status with 134 variables in total being required to make the position models. The most prominent set of variables were those related to the players passing ability, with 48 of the 134 variables (35.8%) being passing statistics. The next most prominent type of variable was related to shooting. In total, twenty variables (14.9%) related to shooting were selected in the models. Statistics related to regaining possession accounted for eleven of the variables (8.2%) selected. Variables related to international appearances were selected nine times (6.7%). A full outline of the categories of variables selected can be viewed in full (Table 3).

Table 2. Results for all models with balanced data sets. The best average test performance = 100.0% and the best average test error = 1.0% (Using a combination of eighteen variables) – Centre Back Group 0 v 2. The worst average test performance = 72.7% and the worst average test error = 9.8% (Using a combination of five variables) – Full Back Group 0 v 1.

Position	Groups	Average Test Performance (%)	Average Test Error (%)	Number of Variables
Full Back	0 v 1	72.7	9.8	5
Full Back	0 v 2	87.5	6.5	10
Full Back	1 v 2	75	9.3	6
Centre Back	0 v 1	93.3	4.1	20
Centre Back	0 v 2	100	1.0	18
Centre Back	1 v 2	90	5.5	6
Wide Midfield	0 v 1	76.5	8	10
Wide Midfield	0 v 2	100	3.4	6
Wide Midfield	1 v 2	77.8	7.4	9
Centre Midfield	0 v 1	78.6	9.6	9
Centre Midfield	0 v 2	90.9	4.8	10
Centre Midfield	1 v 2	88.9	5.9	5
Attacker	0 v 1	80	8.7	5
Attacker	0 v 2	92.3	2.6	10
Attacker	1 v 2	81.8	7.2	6
Average	NA	85.7	6.3	9.0

Full back models

The performance of the full back models as a group were the lowest of the five positions (Average test performance = 78.4% ± *8.0%* and average test error = 8.6% ± *1.7%*) (Table 4). The group 0 v 1 comparison had the lowest average test performance and highest test error out of all the models created (Average test performance = 72.7% and average test error = 9.8%). Total appearances and mean percentage of backwards passes successful were key variables in the model (Table 5). The group 1 v 2 comparison had an average test performance of 75% and a test error of 9.3%. The percentage of sideways passes successful (upper quartile) and median total shots were the most prominent variables in the model (Table 6). The best full back model was for group 0 v 2 which had an average test performance of 87.5% and a test error of 6.6%. The mean goals scored and minimum headers were the two most prominent factors in the model (Table 7).

Table 3. Summary of the variables in all position models by grouping.

Variable Grouping	Times Selected	Selected (%)
Passing	48	35.8
Shooting	20	14.9
Regains	11	8.2
International Appearances	9	6.7
Heading	8	6.0
Fouls	5	3.7
Goals	5	3.7
Appearances	4	3.0
Entries	3	2.2
Possession Lost	4	3.0
Tackled	3	2.2
Time in Possession	3	2.2
Assists	2	1.5
Blocks	2	1.5
Clearances	2	1.5
Crossing	2	1.5
Touches	2	1.5
Balls Received	1	0.7
Possessions	1	0.7

Table 4. Comparison of overall average test performance scores from position models as means and standard deviations.

Position Comparison	Overall Average Test Performance (%)	Overall Average Test Error (%)
Full Back	78.4 ± *8.0*	8.6 ± *1.7*
Centre Back	94.4 ± *5.1*	3.5 ± *2.3*
Wide Midfield	84.8 ± *13.2*	6.3 ± *2.5*
Centre Midfield	86.1 ± *6.6*	6.8 ± *2.5*
Attacker	84.7 ± *6.6*	6.2 ± *3.2*

Table 5. Results for Group 0 v Group 1 Full Back balanced data set comparison. The best average test performance = 72.7% and the best average test error = 9.8% (Using a combination of five variables).

Rank	Variable	Average Test Performance (%)	Average Test Error (%)
1	Total Appearances	63.6	11.2
2	% Backwards Passes Successful (Mean)	72.7	10.6
3	Total Minutes	72.7	9.8
4	% Forwards Passes Successful (Mean)	72.7	9.8
5	Forwards Passes (Maximum)	72.7	9.8
6	Blocks (Mean)	70.5	9.9
7	% Unsuccessful Headers (Median)	68.2	10.0
8	Forward Passes Successful (Median)	68.2	10.0
9	% Passes Successful Own Half (Mean)	72.7	9.9
10	Passes Own Half 25% (Lower Quartile)	72.7	10.0

Table 6. Results for Group 1 v Group 2 Full Back balanced data set comparison. The best average test performance = 75.0% and the best average test error = 9.3% (Using a combination of six variables).

Rank	Variable	Average Test Performance (%)	Average Test Error (%)
1	% Sideways Passes Successful 75% (Upper Quartile)	60.0	11.3
2	Total Shots (Median)	60.0	10.9
3	International Caps	70.0	9.7
4	Tackled (Mean)	70.0	9.3
5	First Time Passes (Maximum)	70.0	9.1
6	Number of Possessions (Median)	75.0	9.3
7	Tackled (Minimum)	70.0	9.4
8	% Sideways Passes Successful 25% (Lower Quartile)	70.0	9.4
9	Total Assists	70.0	9.8
10	% First Time Passes Unsuccessful 25% (Lower Quartile)	70.0	9.8

Table 7. Results for Group 0 v Group 2 Full Back balanced data set comparison. The best average test performance = 87.5% and the best average test error = 6.6% (Using a combination of ten variables).

Rank	Variable	Average Test Performance (%)	Average Test Error (%)
1	Goals (Mean)	75.0	9.1
2	Headers (Minimum)	75.0	8.6
3	% Forward Passes Unsuccessful (Mean)	81.3	8.2
4	Shots Off Target (Exc. Blocked) (Maximum)	78.1	8.1
5	% Forward Passes Unsuccessful 75% (Upper Quartile)	75.0	8.2
6	U21 Caps	75.0	8.0
7	Shots Inside the Box (Mean)	81.3	7.7
8	Possession Lost (Mean)	81.3	7.0
9	Shots On Tgt Outside the Box (Maximum)	81.3	7.2
10	Total Assists	87.5	6.6

Centre back models

The performance of the centre back models as a group had an average test performance of 94.4% ± *5.1%* and an average test error of 3.5% ± *2.3%*. The group 0 v 1 model had an average test performance of 93.3% and an average test error of 4.1% using twenty variables. The percentage of successful passes in the opposition half (upper quartile) and shooting accuracy (upper quartile) were the most prominent variables in the model (Table 8). The group 1 v 2 model had the lowest average test performance and highest test error of the three centre back models (average test performance = 90.0% and average test error = 5.5%). Backwards passes (lower quartile) and maximum short passes were the top two factors in the model (Table 9). The group 0 v 2 model had the highest average test performance of any model and the lowest test error of any model

Table 8. Results for Group 0 v Group 1 Centre Back balanced data set comparison. The best average test performance = 93.3% and the best average test error = 4.1% (Using a combination of twenty variables).

Rank	Variable	Average Test Performance (%)	Average Test Error (%)
1	% Passes Successful Opp Half 75% (Upper Quartile)	66.7	10.9
2	Shooting Accuracy 75% (Upper Quartile)	73.3	9.3
3	% Successful Headers 75% (Upper Quartile)	80.0	7.6
4	Balls Received 75% (Upper Quartile)	80.0	7.6
5	Crosses (Median)	80.0	7.9
6	% First Time Passes Successful 25% (Lower Quartile)	80.0	6.8
7	Total Shots on Target (Mean)	86.7	6.4
8	Passes Successful Opp Half (Minimum)	86.7	6.0
9	U21 Caps	86.7	6.1
10	Shooting Accuracy 25% (Lower Quartile)	86.7	5.2
11	Medium Passes (Mean)	86.7	5.2
12	Forward Passes Successful (Minimum)	93.3	4.5
13	Total Shots on Tgt (Excluding Blocked) (Mean)	86.7	5.0
14	Goals (Mean)	86.7	4.5
15	% Unsuccessful Headers 25% (Lower Quartile)	90.0	4.7
16	Long Passes (Median)	93.3	4.5
17	% Passes Successful Opp Half (Minimum)	93.3	4.2
18	Avg Time in Possession (Mean)	86.7	4.8
19	% Forwards Passes Successful (Minimum)	86.7	4.7
20	Shooting Accuracy (Median)	93.3	4.1

Table 9. Results for Group 1 v Group 2 Centre Back balanced data set comparison. The best average test performance = 90.0% and the best average test error = 5.5% (Using a combination of six variables).

Rank	Variable	Average Test Performance (%)	Average Test Error (%)
1	Backwards Passes 25% (Lower Quartile)	70.0	10.7
2	Short Passes (Maximum)	70.0	9.4
3	Interceptions (Maximum)	80.0	8.1
4	Shots on Target Inside the Box (Mean)	80.0	6.8
5	Sideways Passes Unsuccessful (Mean)	80.0	6.6
6	Sideways Passes Successful 75% (Upper Quartile)	90.0	5.5
7	Passes Successful Own Half (Mean)	90.0	5.5
8	% Passes Successful Opp Half (Minimum)	80.0	6.3
9	% Sideways Passes Successful (Median)	90.0	6.4
10	Shots On Tgt Outside the Box (Mean)	85.0	6.6

(average test performance = 100% and test error = 1.0%). The group 0 v 2 centre back model contained eighteen variables with 0–6 assists mean (group 0 = 0.1 ± *0.1*, group 2 = 0.2 ± *0.1*), mean shots on target inside the box (group 0 = 0.2 ± *0.2*, group 2 = 0.3 ± *0.2*) and minimum penalty area entries (Group 0 = 0.2 ± *0.4*, Group 2 = 0 ± *0*) being key variables (Table 10).

Wide midfielder models

The wide midfield models group average test performance was 84.8% ± *13.2%* with an average test error of 6.3% ± *2.5%*. The group 0 v 1 model had an average test performance of 79.4% and a test error of 8.2%. The maximum percentage of unsuccessful headers and forward passes successful (upper quartile) were the biggest predictors in the model (Table 11). The group 1 v 2 model had an average test performance of 77.8% and a test error of 7.4%. U21 international caps and median forward passes unsuccessful were the most prominent factors in the model (Table 12). The group 0 v 2 model had the second highest average test performance and third lowest test error of all the models created (average test performance = 100% and a test error of 3.4%). The group 0 v 2 wide midfielder model contained six variables including: total goals (group 0 = 1.4 ± *1.9*, group 2 = 5.5 ± 3.8), passes attempted opposition half upper quartile (group 0 = 16.2 ± *6.3*, group 2 = 21.4 ± *5.8*), fouls in the defensive third mean (group 0 = 0.2 ± *0.2*, group 2 = 0.3 ± *0.3*), total shots on target (excluding blocked) maximum (group 0 = 1.0 ± *0.8*, group 2 = 2.6 ± *1.1*), % forward passes successful mean (group 0 = 53.4% ± *14.8%*, group 2 = 55.2% ± *9.7%*) and forward passes successful median (group 0 = 5.0 ± *3.2*, group 2 = 6.1 ± *2.2*) (Table 13).

Table 10. Results for Group 0 v Group 2 Centre Back balanced data set comparison. The best average test performance = 100% and the best average test error = 1.0% (Using a combination of eighteen variables).

Rank	Variable	Average Test Performance (%)	Average Test Error (%)
1	0–6 Assists (Mean)	80.0	8.1
2	Shots on Target Inside the Box (Mean)	80.0	5.8
3	Penalty Area Entries (Minimum)	90.0	4.4
4	International Caps	90.0	3.7
5	Long Passes 25% (Lower Quartile)	90.0	3.2
6	Shots Outside the Box (Mean)	90.0	2.9
7	U21 Caps	100.0	2.4
8	Possession Gained 75% (Upper Quartile)	100.0	1.5
9	Avg Time in Possession (Median)	100.0	1.5
10	Clearances (Maximum)	100.0	1.2
11	Shots Outside the Box (Median)	100.0	1.1
12	First Time Passes (Mean)	100.0	1.3
13	Unsuccessful Passes (Minimum)	100.0	1.4
14	Interceptions 75% (Upper Quartile)	100.0	1.3
15	Possession Gained (Minimum)	100.0	1.3
16	Shots Inside the Box 25% (Lower Quartile)	100.0	1.1
17	Total Shots on Target (Mean)	100.0	1.2
18	Tackled (Minimum)	100.0	1.0
19	Final Third Entries (Mean)	100.0	1.0
20	Medium Passes 25% (Lower Quartile)	100.0	1.3

Table 11. Results for Group 0 v Group 1 Wide Midfield balanced data set comparison. The best average test performance = 79.4% and the best average test error = 8.2% (Using a combination of nine variables).

Rank	Variable	Average Test Performance (%)	Average Test Error (%)
1	% Unsuccessful Headers (Maximum)	70.6	10.8
2	Forward Passes Successful 75% (Upper Quartile)	73.5	10.0
3	Possession Won 75% (Upper Quartile)	70.6	9.8
4	Shooting Accuracy 25% (Lower Quartile)	76.5	8.9
5	% Unsuccessful Headers 75% (Upper Quartile)	79.4	8.5
6	% Successful Headers (Median)	76.5	8.4
7	Sideways Passes Successful 75% (Upper Quartile)	76.5	8.2
8	Fouls (Mean)	76.5	8.1
9	Tackled (Maximum)	79.4	8.2
10	Passes Attempted Opp Half (Mean)	76.5	8.0

Table 12. Results for Group 1 v Group 2 Wide Midfield balanced data set comparison. The best average test performance = 77.8% and the best average test error = 7.4% (Using a combination of nine variables).

Rank	Variable	Average Test Performance (%)	Average Test Error (%)
1	U21 International Caps	66.7	10.3
2	Forwards Passes Unsuccessful (Median)	77.8	9.3
3	% Sideways Passes Unsuccessful (Median)	77.8	9.1
4	Fouls (Mean)	77.8	8.9
5	Possession Won (Maximum)	77.8	8.6
6	% Unsuccessful Headers (Maximum)	77.8	8.5
7	Backwards Passes Unsuccessful (Maximum)	77.8	8.7
8	Possession Lost (Maximum)	77.8	7.9
9	Possession Won (Minimum)	77.8	7.4
10	% Unsuccessful Headers 25% (Lower Quartile)	77.8	7.6

Table 13. Results for Group 0 v Group 2 Wide Midfield balanced data set comparison. The best average test performance = 100% and the best average test error = 3.4% (Using a combination of six variables).

Rank	Variable	Average Test Performance (%)	Average Test Error (%)
1	Total Goals	84.6	7.2
2	Passes Attempted Opp Half 75% (Upper Quartile)	84.6	6.3
3	Fouls in Defensive 3rd (Mean)	84.6	6.1
4	Total Shots on Tgt (Excluding Blocked) (Maximum)	92.3	4.5
5	% Forwards Passes Successful (Mean)	92.3	3.3
6	Forward Passes Successful (Median)	100.0	3.4
7	Tackled 75% (Upper Quartile)	92.3	3.7
8	% Unsuccessful Passes 75% (Upper Quartile)	92.3	3.6
9	Backwards Passes Unsuccessful (Mean)	92.3	3.5
10	Possession Lost (Median)	92.3	3.1

Centre midfielder models

The best overall average was for the centre midfielder's models as a group (Average test performance = 86.1% ± *6.6* and average test error = 6.8% ± *2.5*). The group 0 v 1 model had the lowest average test performance of the centre midfield models and had the second highest test error across all models (Average test performance = 78.6% and average test error = 9.6%). Fouls and maximum first time passes were the most prominent variables in the model (Table 14). The group 1 v 2 model had an average test performance of 88.9% and a test error of 5.9%. Successful passes (lower quartile) and penalty area entries (lower quartile) were two key variables in the model (Table 15). The group 0 v 2 model had an average test performance of 90.9% and a test error of 4.8%. The number of starts and maximum shots on target outside the box were the highest predictors in the model (Table 16).

Attacker models

The performance of the attacker models as a group had an average test performance of 84.7% ± *6.6%* and an average test error of 6.2% ± *3.2%*. The group 0 v 1 model had an average test performance of 80% and an average test error of 8.7%. The most prominent variables in the model were international caps and the number of touches (lower quartile) (Table 17). The group 1 v 2 model had an average test performance of 81.8% and a test error of 7.2%. U21 international caps and international caps were the two most important factors in the model (Table 18). The best average test performance for an attacker model was recorded for the group 0 v 2 model and it had the lowest overall test error of all models (average test performance = 92.3% and test error = 2.6%). The group 0 v 2 attacker model contained ten variables with total goals (group 0 = 2.7 ± *3.0*, group 2 = 10.0 ± *6.2*), blocks upper quartile (group 0 = 1.0 ± *0.5*, group 2 = 1.5 ± *0.7*) and short passes minimum (group 0 = 4.9 ± *2.5*, group 2 = 4.3 ± *2.4*) being key variables (Table 19).

Table 14. Results for Group 0 v Group 1 Centre Midfield balanced data set comparison. The best average test performance = 78.6% and the best average test error = 9.6% (Using a combination of nine variables).

Rank	Variable	Average Test Performance (%)	Average Test Error (%)
1	Fouls	57.1	11.5
2	First Time Passes (Maximum)	64.3	10.9
3	Backwards Passes 75% (Upper Quartile)	64.3	10.6
4	Number of Touches (Median)	64.3	10.6
5	Fouls (Maximum)	64.3	10.5
6	Total Minutes	71.4	9.9
7	% Forward Passes Unsuccessful 25% (Lower Quartile)	71.4	9.6
8	Sideways Passes (Median)	71.4	9.6
9	Passes Attempted Opp Half (Minimum)	78.6	9.6
10	Height	71.4	9.7

Table 15. Results for Group 1 v Group 2 Centre Midfield balanced data set comparison. The best average test performance = 88.9% and the best average test error = 5.9% (Using a combination of five variables).

Rank	Variable	Average Test Performance (%)	Average Test Error (%)
1	Successful Passes 25% (Lower Quartile)	66.7	10.2
2	Penalty Area Entries 25% (Lower Quartile)	66.7	9.6
3	Goals (Mean)	77.8	8.4
4	Backwards Passes Unsuccessful (Mean)	88.9	6.2
5	First Time Passes Successful (Maximum)	88.9	5.9
6	Backwards Passes (Median)	88.9	6.2
7	% Sideways Passes Successful 25% (Lower Quartile)	88.9	6.4
8	Total Shots 25% (Lower Quartile)	88.9	6.4
9	Passes Own Half (Mean)	88.9	6.9
10	Dribbles 75% (Upper Quartile)	83.3	7.2

Table 16. Results for Group 0 v Group 2 Centre Midfield balanced data set comparison. The best average test performance = 90.9% and the best average test error = 4.8% (Using a combination of ten variables).

Rank	Variable	Average Test Performance (%)	Average Test Error (%)
1	No. Of Starts	72.7	9.6
2	Shots On Tgt Outside the Box (Maximum)	81.8	8.6
3	Possession Lost (Maximum)	77.3	8.0
4	Forwards Passes (Mean)	81.8	7.2
5	Possession Won (Median)	81.8	6.0
6	Clearances 25% (Lower Quartile)	81.8	5.5
7	Total Shots on Target (Mean)	90.9	5.2
8	Total Blocked Shots (Maximum)	90.9	5.2
9	Forwards Passes (Median)	90.9	4.9
10	% Passes Successful Opp Half 75% (Upper Quartile)	90.9	4.8

Table 17. Results for Group 0 v Group 1 Attacker balanced data set comparison. The best average test performance = 80.0% and the best average test error = 8.7% (Using a combination of five variables).

Rank	Variable	Average Test Performance (%)	Average Test Error (%)
1	International Caps	73.3	10.4
2	Number of Touches 25% (Lower Quartile)	73.3	9.2
3	First Time Passes (Maximum)	73.3	9.1
4	Blocks (Maximum)	73.3	8.9
5	Final Third Entries (Mean)	80.0	8.7
6	Passes Successful Own Half (Median)	73.3	8.9
7	% Successful Passes (Maximum)	73.3	9.2
8	Tackled 25% (Lower Quartile)	73.3	9.0
9	% Forwards Passes Successful (Minimum)	73.3	9.1
10	% Passes Successful Opp Half (Minimum)	73.3	9.1

Table 18. Results for Group 1 v Group 2 Attacker balanced data set comparison. The best average test performance = 81.8% and the best average test error = 7.2% (Using a combination of six variables).

Rank	Variable	Average Test Performance (%)	Average Test Error (%)
1	U21 International Caps	63.6	11.0
2	International Caps	72.7	9.9
3	Unsuccessful Passes (Maximum)	72.7	9.6
4	Interceptions (Maximum)	72.7	8.7
5	Possession Won (Median)	81.8	7.2
6	% Unsuccessful Passes 75% (Upper Quartile)	81.8	7.2
7	Final Third Entries 25% (Lower Quartile)	81.8	7.8
8	Tackles (Maximum)	81.8	7.4
9	% Unsuccessful Passes (Minimum)	81.8	7.5
10	Penalty Area Entries (Minimum)	81.8	7.3

Table 19. Results for Group 0 v Group 2 Attacker balanced data set comparison. The best average test performance = 92.3% and the best average test error = 2.6% (Using a combination of ten variables).

Rank	Variable	Average Test Performance (%)	Average Test Error (%)
1	Total Goals	76.9	7.6
2	Blocks 75% (Upper Quartile)	84.6	5.6
3	Short Passes (Minimum)	92.3	5.0
4	Passes Own Half 25% (Lower Quartile)	92.3	4.4
5	% Unsuccessful Headers (Maximum)	92.3	4.0
6	Crosses (Mean)	92.3	3.0
7	Avg Time in Possession 75% (Upper Quartile)	92.3	2.9
8	Interceptions (Median)	92.3	3.0
9	Passes Successful Opp Half 75% (Upper Quartile)	92.3	3.0
10	Backwards Passes 25% (Lower Quartile)	92.3	2.6

Table 20. Comparison of overall average test performance scores from position models as means and standard deviations.

Group Comparison	Overall Average Test Performance (%)	Overall Average Test Error (%)
Group 0 v 1 Comparisons	80.8 ± *7.6*	8.1 ± *2.3*
Group 1 v 2 Comparisons	82.7 ± *6.6*	7.1 ± *1.5*
Group 0 v 2 Comparisons	94.1 ± *5.6*	3.7 ± *2.1*

Model comparisons

The models produced comparing positions for group 0 v 1 had the lowest overall average test performance and highest test error (mean test performance = 80.8% ± *7.6%* and average test error = 8.1% ± *2.3%*). The overall average test performance across all five positions for group 1 v 2 comparisons was 82.7% ± *6.6%* and the average test error was 7.1% ± *1.5*. The highest overall average test performance across the five positions was for group 0 v 2 (mean test performance = 94.1% ± *5.6%* and average test error = 3.7% ± *2.1%*) (Table 20). The top three models produced by the neural network were for 0 v 2 centre back (average test performance 100% and 1.0% test error), group 0 v 2 wide midfielder (average test performance 100% and 3.4% test error) and group 0 v 1 centre back (average test performance 93.3% and 4.1% test error). The means and standard deviations for key variables for the top three models can be reviewed in full (Tables 21–23).

Discussion

The aim of the current study was to develop objective models to identify position-specific key performance indicators in professional football that predict out-field players league status using an artificial neural network. The artificial neural network created fifteen position-specific models to predict out-field players league status. The artificial neural network's ability to correctly classify more than 75% of the players league status for fourteen different position comparisons is a key result. This surpasses the previous prediction rates reported using artificial neural networks in other team sports, such as those undertaken in cricket

Table 21. Group 0 v 2 Centre Back model variables represented as means and standard deviations for all player groupings.

Variables	Group 0 Centre Back	Group 2 Centre Back
0–6 Assists (Mean)	0.1 ± *0.1*	0.2 ± *0.1*
Shots on Target Inside the Box (Mean)	0.2 ± *0.2*	0.3 ± *0.2*
Penalty Area Entries (Minimum)	0.2 ± *0.4*	0.0 ± *0.0*
International Caps	4.8 ± *18.3*	9.2 ± *14.6*
Long Passes 25% (Lower Quartile)	4.3 ± *2.2*	4.9 ± *2.0*
Shots Outside the Box (Mean)	0.1 ± *0.2*	0.1 ± *0.1*
U21 Caps	0.3 ± *0.9*	3.5 ± *6.6*
Possession Gained 75% (Upper Quartile)	34.2 ± *5.5*	36.7 ± *5.7*
Avg Time in Possession (Median)	2.4 ± *2.2*	2.6 ± *0.3*
Clearances (Maximum)	10.9 ± *3.2*	11.4 ± *3.2*
Shots Outside the Box (Median)	0.0 ± *0.2*	0.0 ± *0.0*
First Time Passes (Mean)	6.5 ± *1.9*	7.0 ± *1.2*
Unsuccessful Passes (Minimum)	1.4 ± *1.8*	1.0 ± *1.2*
Interceptions 75% (Upper Quartile)	29.9 ± *4.2*	31.1 ± *5.3*
Possession Gained (Minimum)	21.1 ± *4.9*	18.5 ± *6.3*
Shots Inside the Box 25% (Lower Quartile)	0.1 ± *0.4*	0.0 ± *0.1*
Total Shots on Target (Mean)	0.2 ± *0.2*	0.3 ± *0.2*
Tackled (Minimum)	0.2 ± *0.7*	0.0 ± *0.2*

Table 22. Group 0 v 2 Wide Midfield model variables represented as means and standard deviations for all player groupings.

Variables	Group 0 Wide Midfield	Group 2 Wide Midfield
Total Goals	1.4 ± 1.9	5.5 ± 3.8
Passes Attempted Opp Half 75% (Upper Quartile)	16.2 ± 6.3	21.4 ± 5.8
Fouls in Defensive 3rd (Mean)	0.2 ± 0.2	0.3 ± 0.3
Total Shots on Tgt (Excluding Blocked) (Maximum)	1.0 ± 0.8	2.6 ± 1.1
% Forwards Passes Successful (Mean)	53.4 ± 14.8	55.2 ± 9.7
Forward Passes Successful (Median)	5.0 ± 3.2	6.1 ± 2.2
Total Goals	1.4 ± 1.9	5.5 ± 3.8
Passes Attempted Opp Half 75% (Upper Quartile)	16.2 ± 6.3	21.4 ± 5.8
Fouls in Defensive 3rd (Mean)	0.2 ± 0.2	0.3 ± 0.3
Total Shots on Tgt (Excluding Blocked) (Maximum)	1.0 ± 0.8	2.6 ± 1.1

Table 23. Group 0 v 1 Centre Back model variables represented as means and standard deviations for all player groupings.

Variables	Group 0 Centre Back	Group 1 Centre Back
% Passes Successful Opp Half 75% (Upper Quartile)	81.2 ± *22.3*	92.4 ± *13.5*
Shooting Accuracy 75% (Upper Quartile)	23.5 ± *35.6*	20.1 ± *33.8*
% Successful Headers 75% (Upper Quartile)	51.0 ± *8.7*	52.7 ± *6.6*
Balls Received 75% (Upper Quartile)	16.9 ± *5.8*	20.6 ± *8.9*
Crosses (Median)	0.1 ± *0.3*	0.1 ± *0.3*
% First Time Passes Successful 25% (Lower Quartile)	59.3 ± *13.0*	59.9 ± *12.7*
Total Shots on Target (Mean)	0.2 ± *0.2*	0.3 ± *0.3*
Passes Successful Opp Half (Minimum)	0.2 ± *0.5*	0.3 ± *1.0*
U21 Caps	0.3 ± *0.9*	1.3 ± *3.2*
Shooting Accuracy 25% (Lower Quartile)	0.0 ± *0.0*	1.9 ± *11.4*
Medium Passes (Mean)	7.9 ± *2.9*	9.6 ± *5.1*
Forward Passes Successful (Minimum)	1.5 ± *1.3*	1.6 ± *2.5*
Total Shots on Tgt (Excluding Blocked) (Mean)	0.1 ± *0.1*	0.2 ± *0.2*
Goals (Mean)	0.0 ± *0.1*	0.1 ± *0.1*
% Unsuccessful Headers 25% (Lower Quartile)	49.0 ± *8.7*	47.2 ± *6.7*
Long Passes (Median)	5.5 ± *2.1*	6.3 ± *2.5*

(Iyer & Sharda, 2009; Saikia, Bhattacharjee, & Lemmer, 2012). Their studies could predict classification of batsmen and bowlers with accuracy levels ranging from 49% to 77%. In individual sports, artificial neural networks have been able to predict 80.2% of gymnast's future classifications based on a multidimensional testing process (Pion, Hohmann, Liu, Lenoir, & Segers, 2017). Therefore, the current artificial neural network prediction rates are among the highest reported to date in an athlete classification study.

Passing variables

The most prominent set of variables were those related to the players passing ability, with 48 of the 134 total variables included in models (35.8%) being passing statistics. Many passing variables have been highlighted previously as key indicators when differentiating between players of various playing levels and linked to team success (Bradley et al., 2013; Rampinini, Impellizzeri, Castagna, Coutts, & Wisloff, 2009). Comparisons between players within the English football pyramid showed that players in the Premier League performed a greater number of total passes, successful passes and forward passes (Bradley et al., 2013). Out of the 48 passing variables identified in the models, 29 were related to the success of the passing variables. The passing variables related to their success were a mixture of 27 different statistics accounting for the direction (forwards, sideways and backwards) of the pass, the origin of the pass (own half or opposition half) and the mean, median, minimum, maximum and upper and lower quartile figure for different variables.

In further agreement with Bradley et al.'s (2013) findings, thirteen of the passing variables were related to forward passing. Forward passes have been shown to have the lowest chance of success when compared to sideways or backwards passes (Szczepański & Mchale, 2015). Yet, to create scoring opportunities and in turn score goals players are required to progress the play with forward passing. Variables relating to forward passes appeared in models for full backs (group 0 v 1 and group 0 v 2), centre backs (group 0 v 1), wide midfield (group 0 v 1, group 1 v 2 and group 0 v 2), centre midfield (group 0 v 1 and group 0 v 2) but did not feature prominently in any models for attackers. This would appear logical as attackers play in more advanced areas and have fewer opportunities to perform forward passes. The prevalence of forward passing variables for a number of positions and different comparisons highlights its importance in playing success.

The current study also highlighted two variables related to short passing with the maximum and minimum variables being selected in two models (group 1 v 2 centre back and group 0 v 2 attacker). Research into factors that distinguish between top four and bottom four English Premier League teams highlighted short passes as a key variable (Adams, Morgans, Sacramento, Morgan, & Williams, 2013). Specifically, the mean frequency of successful short passes played by centre backs and full backs was the biggest factor differentiating between the two groups.

Using the artificial neural network methodology has highlighted some overlap between factors previously identified by research articles. The current study has also identified novel findings for variables that have not previously been analysed or identified as key variables. Eight passing variables were related to those in the opposition half and they appeared in six different position models (group 0 v 1 centre back, group 0 v 1 and 0 v 2 centre midfield, group 0 v 1 and 0 v 2 wide midfield and 0 v 2 attacker models). Six of the variables were also related to first time passes played and they appeared in the group 0 v 1 and 0 v 2 centre back, group 1 v 2 full back, group 0 v 1 and 1 v 2 centre midfield and group 0 v 1 attacker models. Passes in the opposition half indicate possession taking place in more offensive pitch locations and could indicate the involvement of players in attacking moves. The ability to pass the ball accurately over a range of distances and directions is a key factor in performance and for differentiating between players of varying ability. This is accepted knowledge amongst coaches but the models have accurately identified specific key variables and provided an objective assessment of their impact on league status.

Shooting variables

The next most prominent type of variable was related to players shooting ability. In total, twenty variables (14.9%) related to shooting were selected in the models. This agrees with previous research into team success in football, with total shots and shooting accuracy being the most commonly reported predictors in matches (Castellano et al., 2012; Lago-Penas et al., 2010; Liu et al., 2015). Surprisingly, all positions except attacker included shooting variables in the models created in the current study. However, one of the attacker models (group 0 v 2) did include total goals as a key variable. Many teams now prefer to play with one lone attacker in their line-up that spreads the need for scoring goals throughout the team and the requirements of the centre forward position could be changing as a result (Adams et al., 2013).

Attacking entries

Other attacking variables selected as part of the models were related to crossing and entries into the final third and penalty area. Final third and penalty area entries were selected three times and in three different models. Crosses are a factor that have been repeatedly identified as being key to differentiating between successful and unsuccessful teams (Lago-Penas et al., 2010; Lagos-Penas, Lago-Ballesteros, & Rey, 2011). They have not been identified as key when differentiating between players of different performance levels previously, they were only selected twice in the current study meaning they did not play a prominent role in the position models. The mean number of crosses were selected in the group 0 v 2 attacker model (crosses mean group 0 1.0 ± *0.8*, group *2* 1.75 ± 1.23). The inclusion of the number of crosses in the attacker model and the higher values reported for group 2 may offer more evidence for the evolving role of the attacker.

As well as crosses, final third and penalty area entries were selected three times and in three different models. Previous research has indicated that penalty area entries differentiate between winning and losing teams (Ruiz-Ruiz, Fradua, Fernandez-Garcia, & Zubillaga, 2013). However, in the current study they were selected in one model for centre backs (group

0 v 2), the centre backs from players dropping down to a lower playing level reported higher values (minimum penalty area entries group 0 0.2 ± 0.4, group 2 0.0 ± 0.0). The identification of minimum penalty area entries in the centre back model and group 0 having a higher value is a novel finding. It may appear counter intuitive but centre backs who drop down to a lower level may play in teams who use a more direct style of play and play longer passes from their centre backs as opposed to building the play with shorter passing combinations.

Defensive variables

The models also highlighted several defensive variables as key predictors of league status. Statistics related to regaining possession accounted for eleven of the variables (8.2%). Previous research into match outcomes and players technical and tactical ability has heavily focused on the attacking aspects of play (Mackenzie & Cushion, 2013), passing (Adams et al., 2013; Szczepanski & McHale, 2016) and possession (Castellano et al., 2012; Collett, 2013; Lago-Penas et al., 2010; Liu et al., 2015). A limited number of defensive variables have been researched or identified that are linked to success. A balanced defensive shape (Tenga, Holme, Ronglan, & Bahr, 2010), defensive reaction after losing possession (Vogelbein, Nopp, & Hokelmann, 2014) and regaining possession in the final third have been identified previously (Almeida, Ferreira, & Volossovitch, 2014).

The current study highlighted possession won based on the minimum, median, maximum and upper quartile variables as being key predictors of league status. Possession gained upper quartile and interceptions median and maximum were also selected as key variables in models. The defensive variables were not selected as part of any of the full back models. They were commonly selected as part of the wide midfield (group 0 v 1 and group 1 v 2) and attacker models (group 1 v 2 and group 0 v 2). This may appear counter intuitive and these factors would not normally be assessed when profiling more attacking positions in research studies. Modern playing philosophies valuing high pressing tactics from forward players to regain possession in more advanced areas of the pitch may explain the importance of these factors in wide midfield and attacker models within the current study (Perarnau, 2014).

International recognition

Other key variables selected throughout several models relate to international appearances, international caps and U21 international caps were selected nine times (6.7%) in total. This is a novel finding as previous assessments of players' performances have limited themselves to match performance and season totals of performance data. Previous research into international recognition and team or playing success has not been undertaken to the authors' knowledge. However, international recognition has been found to be linked with player salary allocation, particularly at the higher levels of the game (Frick, 2011).

Position-specific models

The current study created a number of strong predictive models for players league status, there were also some key findings relating to the prediction rates of specific positions. Three of the five positions had very similar levels of classification accuracy (centre midfield 86.1%, wide midfield 85.7% and attacker 84.7%) but the full back position's overall accuracy was only 78.4% and the centre back position's overall accuracy was 94.4%. The full back results are still an important finding but below the levels reported for other positions. The group 0 v 1 full back model had the lowest classification accuracy of all the models and the group 1 v 2 full back model had the second lowest classification accuracy. The full back position is one that requires a complex set of technical and tactical skills as it requires a wide array of attacking and defensive qualities (Bush, Archer, Hogg, & Bradley, 2015).

Recent evaluations of the changes within performance data for playing positions has shown extensive changes over time in the Premier League (Bush et al., 2015). Pronounced increases were found for the levels of high-speed running and the distances covered while sprinting, with full backs showing the largest increases between 2006–07 and 2012–13 (Bush et al., 2015). Therefore, the full back position may be influenced more by the physical aspects of performance. This could explain the lower prediction rates for full backs due to the lack of physical tracking data being available.

Study limitations

Strong models were identified for fourteen out of the fifteen position comparisons assessed but there are some limitations to the present study that should be addressed in future research. The match running performance data for players was not available for the current study. There is an acceptance amongst the sports science community that running performance is not a predictor of team success or match outcome (Bradley et al., 2016; Carling, 2013). However, including match running performance data could provide a higher level of classification accuracy for some of the positions assessed. Another limitation of the study is the lack of contextual data available and the inability of the data to provide a detailed assessment for off the ball parameters. The final limitation of the study relates to the sample size for players progressing to play in the Premier League. The samples for the players progressing from the five positions to play in the Premier League were the smallest of all the groupings. Statistical power tests on similar sample sizes have reached the required levels (Lancashire et al., 2009). However, future studies should look to increase the sample available to increase confidence that the results are repeatable to new cases.

Conclusions

The current study has shown that artificial neural networks are a valid and highly effective tool to classify and predict players league status. Fourteen models across all five positions were created that provided strong prediction accuracy levels for players league status. This is an important result as it outlines an objective methodology that can aid the scouting and recruitment process in professional football. The process of identifying and recruiting players in professional football has largely been a subjective process in the past. Further research should look to combine assessments of physical and technical performance data to provide a more accurate prediction of league status. Studies should also look to create models to predict the career progression of players from multiple leagues to provide a better practical tool for scouting and recruitment purposes. The combination of subjective assessments and more objective tools could lead to a more effective overall process in the highly competitive football transfer market.

Acknowledgments

The authors would like to thank STATS for providing access to the performance data that is used in this study. We would also like to thank Scout7 for providing access to their system to include biographical and international appearance data in the current study.

Disclosure statement

No potential conflict of interest was reported by the authors.

ORCID

Graham Ball http://orcid.org/0000-0001-5828-7129
Caroline Sunderland http://orcid.org/0000-0001-7484-1345

References

Adams, D., Morgans, R., Sacramento, J., Morgan, S., & Williams, M. D. (2013). Successful short passing frequency of defenders differentiates between top and bottom four English Premier League teams. *International Journal of Performance Analysis in Sport*, *13*(3), 653–668.

Almeida, C. H., Ferreira, A. P., & Volossovitch, A. (2014). Effects of match location, match status and quality of opposition on regaining possession in UEFA Champions League. *Journal of Human Kinetics*, *41*(2014), 2013–2214.

Andrzejewski, M., Konefal, M., Chmura, P., Kowalczuk, E., & Chmura, J. (2016). Match outcome and distances covered at various speeds in match play by elite German soccer players. *International Journal of Performance Analysis in Sport*, *16*(3), 817–828.

Barron, D., Ball, G., Robins, M., & Sunderland, C. (2018). Artificial neural networks and player recruitment in professional soccer. *PLoSONE*, *13* (10). doi: 10.1371/journal.pone.0205818

Basheer, I. A., & Hajmeer, M. (2000). Artificial neural networks: Fundamentals, computing, design and application. *Journal of Microbiological Methods*, *43*(2000), 3–31.

Bradley, P. S., Archer, D., Hogg, B., Schuth, G., Bush, M., Carling, C., & Barnes, C. (2016). Tier-specific evolution of match performance characteristics in the English Premier League: It's getting tougher at the top. *Journal of Sports Sciences*, *34*(10), 980–987.

Bradley, P. S., Carling, C., Diaz, A. G., Hood, P., Barnes, C., Ade, J., … Mohr, M. (2013). Match performance and physical capacity of players in the top three competitive standards of English professional soccer. *Human Movement Science*, *32*(2013), 808–821.

Bradley, P. S., O'Donaghue, P., Wooster, B., & Tordoff, P. (2007). The reliability of ProZone MatchViewer: A video-based technical performance analysis system. *International Journal of Performance Analysis in Sport*, *7*(3), 117–129.

Bush, M. D., Archer, D. T., Hogg, R., & Bradley, P. S. (2015). Factors influencing physical and technical variability in the English Premier League. *International Journal of Sports Physiology and Performance*, *10*, 865–872.

Carling, C. (2013). Interpreting physical performance in professional soccer match-play: Should we be more pragmatic in our approach. *Sports Medicine*, *43*(2013), 655–663.

Carling, C., Williams, A. M., & Reilly, T. R. (2005). *Handbook of soccer match analysis: A systematic approach to improving performance*. London: Routledge.

Castellano, J., Casamichana, D., & Lago, C. (2012). The use of match statistics that discriminate between successful and unsuccessful soccer teams. *Journal of Human Kinetics*, *31*(2012), 139–147.

Collett, C. (2013). The possession game? A comparative analysis of ball retention and team success in European and international football, 2007–2010. *Journal of Sports Sciences*, *31*(2), 123–136.

Dellal, A., Chamari, K., Wong, D. P., Ahmaidi, S., Keller, D., Barros, R., … Carling, C. (2011). Comparison of physical and technical performance in European soccer match-play: FA Premier League and La Liga. *European Journal of Sport Science*, *11*(1), 51–59.

Dellal, A., Wong, D. P., Moalla, W., & Chamari, K. (2010). Physical and technical activity of soccer players in the French First League – With special reference to their playing position. *International SportMed Journal*, *11*(2), 278–290.

Frick, B. (2011). Performance, salaries, and contract length: Empirical evidence from German soccer. *International Journal of Sport Finance*, *6*, 87–118.

Iyer, S. R., & Sharda, R. (2009). Prediction of athletes performance using neural networks: An application in cricket team selection. *Expert Systems with Applications*, *36*(2009), 5510–5522.

Lago-Penas, C., Lago-Ballesteros, J., Dellal, A., & Gomez, M. (2010). Game-related statistics that discriminated winning, drawing and losing teams from the Spanish soccer league. *Journal of Sports Science and Medicine*, *9*, 288–293.

Lago-Penas, C., Lago-Ballesteros, J., & Rey, E., (2011). Differences in performance indicators between winning and losing teams in the UEFA Champions League. *Journal of Human Kinetics*, *27*(1),135-146.

Lancashire, L. J., Lemetre, C., & Ball, G. R. (2009). An introduction to artificial neural networks in bioinformatics – Application to complex microarray and mass spectrometry datasets in cancer studies. *Briefings in Bioinformatics*, *10*(3), 315–329.

Liu, H., Gomez, M. A., Lago-Penas, C., & Sampaio, J. (2015). Match statistics related to winning in the group stage of 2014 Brazil FIFA World Cup. *Journal of Sports Sciences*, *33*(12), 1205–1213.

Mackenzie, R., & Cushion, C. (2013). Performance analysis in football: A critical review and implications for future research. *Journal of Sports Sciences*, *31*(6), 639–676.

Perarnau, M. (2014). *Pep confidential: The inside story of Pep Guardiola's first season at Bayern Munich*. Edinburgh: Arena Sport.

Pion, J., Hohmann, A., Liu, T., Lenoir, M., & Segers, V. (2017). Predictive models reduce talent development costs in female gymnastics. *Journal of Sports Sciences*, *35*(8), 806–811.

Rampinini, E., Impellizzeri, F. M., Castagna, C., Coutts, A. J., & Wisloff, U. (2009). Technical performance during soccer matches of the Italian Serie A league: Effect of fatigue and competitive level. *Journal of Science and Medicine in Sport*, *12*, 227–233.

Ruiz-Ruiz, C., Fradua, L., Fernandez-Garcia, A., & Zubillaga, A. (2013). Analysis of entries into the penalty area as a performance indicator in soccer. *European Journal of Sport Science*, *13*(3), 241–248.

Saikia, H., Bhattacharjee, D., & Lemmer, H. H. (2012). Predicting the performance of bowlers in IPL: An application of artificial neural network. *International Journal of Performance Analysis in Sport*, *12*(1), 75–89.

Salkind, N. J. (2010). *Encyclopedia of research design*. California: Sage.

Szczepański, L., & Mchale, I.G. (2015). Beyond completion rate: evaluating the passing ability of footballers. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*. 179(2).

Taylor, J. B., Mellalieu, S., & James, N. (2004). Behavioural comparisons of positional demands in professional soccer. *International Journal of Performance Analysis in Sport*, *4*(1), 81–97.

Tenga, A., Holme, I., Ronglan, L. T., & Bahr, R. (2010). Effects of playing tactics on goal scoring in Norwegian professional soccer. *Journal of Sports Sciences*, *28*(3), 237–244.

Tu, J. V. (1996). Advantages and disadvantages of using artificial neural networks versus logistic regression for predicting medical outcomes. *Journal of Clinical Epidemiology*, *49*(11), 1225–1231.

Vogelbein, M., Nopp, S., & Hokelmann, A. (2014). Defensive transition in soccer – Are prompt possession regains a measure of success? A quantitative analysis of German Fusball-Bundesliga 2010/11. *Journal of Sports Sciences*, *32*(11), 1076–1083.