Identification and Empirical Analysis of Key Influencing Factors on the Outcomes of Professional Football Matches

Jiayang Dong

Anhui Normal University, Yancheng, China
Dongjiayang2026@163.com

Abstract. In the context of strengthening youth football training nationwide, promoting open and transparent selection of national football players, fostering a healthy football competition environment, advancing the modernization of coach education, improving the construction of the national football team, and comprehensively improving the level of national football, the factors influencing the outcome of football matches have been studied. This study provides scientific data analysis for team training, prompting coaches and players to further improve their personal and team tactical strategies, increase the team's winning rate, and analyze the factors that affect the outcome of football matches. Using key match data from the English Premier League in the DCD app multiple linear regression analysis was performed via SPSS to identify the 12 most critical factors affecting match results. Based on this, practical suggestions were proposed for teams to improve their winning probability from micro to macro levels, including players themselves, coaches, and the General Administration of Sport of China.

Keywords: Football match, influencing factors, the DCD, multiple linear regression model

1. Introduction

1.1. Research background

The country is promoting systematic reforms in management systems and operational mechanisms, actively creating a clean and professional football development environment, standardizing the process of selecting players for the national team, and improving the modernization level of team selection mechanisms and coach training. In this context, the Chinese football team has introduced foreign coaches such as Branko Ivankovic to coach, continuously promoting the team's team development, ensuring player continuity, and further strengthening formation and tactical design in order to achieve better results in future competitions. The coach needs to choose a practical and efficient formation based on the characteristics of the players through data analysis. For example, the team has adjusted from a "442 diamond midfield" to a "4-2-3-1" or "three center back" system that focuses more on attacking and defending balance, in order to enhance the defensive hardness of the midfield. It is also necessary to strengthen the overall defensive discipline, emphasize the maintenance of formation and collaborative defense protection among teammates, reduce unnecessary mistakes and give opportunities to the opponent. Attacking pursues efficiency, and when facing opponents with similar or stronger strength, a strategy of steady defense and counterattack can be adopted. Don't force unnecessary ball control, but focus on the speed and passing success rate when transitioning from defense to offense, and make good use of the forward's speed and positioning. Strengthen players' skills and psychology before and after the

game, enhance their focus and stress resistance, and help them maintain high concentration during the game through psychological counseling and practical simulation. Especially when leading or falling behind, they can stabilize their mentality and perform their technical level normally. In order to improve the team's competitive performance, this study takes the first five rounds of the English Premier League season from September 2020 to September 2023 as the survey object, searches and analyzes the indicators that have a significant impact on the game results, and further analyzes the degree of influence of the indicators on the game results, providing data support for the training and matches of each team.

1.2. Research significance

With the advancement and development of technology, more and more technological means have been introduced into football matches, such as VAR video assistant referee technology. With the assistance of these technologies, the possibility of erroneous rulings is greatly reduced, and the fairness of football matches is better guaranteed, with an increasing amount of visualized data. Therefore, it is required that all teams keep up with the times and make full use of technology and visual data to make reasonable arrangements for tactical arrangements, personnel configuration adjustments, and other aspects of the team, in order to improve the team's winning rate, increase the team's and the league's visibility, and achieve greater profits. This study uses panel data composed of the first five rounds of the English Premier League matches from September 2020 to September 2023 to make the analysis of the results more reliable, and explores the underlying patterns of the influencing factors on matches in the four seasons of the English Premier League.

1.3. Definition of research scope

This study explores the relevant influencing factors on the results of football matches. By using panel data from the first five rounds of the Premier League from September 2020 to September 2023, the analysis of the results is made more reliable, and the underlying patterns of the influencing factors in the four seasons of the Premier League are explored. By analyzing the source and real-time value of team members, this study investigates the impact of the distribution rate of team value and player source on the outcome of the game.

2. Literature review

There are usually physiological, psychological, and technical factors that affect the performance of football players in a literature review. Matthew Cole [1]conducted research and evaluation on the nutritional knowledge and intake of adolescent football players, but there is still limited research on such knowledge among parents. Gygax and PascalM [2] analyzed and explored the construction of athletes' psychological representations, and found that athletes with less expertise contain more emotional factors in their psychological representations. Therefore, although there is limited research on the conditions of players themselves and the implementation of changes is difficult, the impact on the game still needs to be considered to a certain extent. In addition to factors such as the individual athlete's condition, objective factors on the day of the competition are also very important, such as home advantage, weather conditions, competition location, and local climate conditions, all of which have a critical impact on the results of the competition. Chinese scholars Liu Yang [3] , Han Zhe [4] and Yin Xiaochuan [5] have conducted research on the home and away effects of domestic and European men's professional leagues. The former investigated and analyzed their internal mechanisms and mechanisms, while the latter conducted regression analysis on the impact of the order of home and away matches on the results of qualifying matches. In addition, the research on tactical strategy management is clearly the most

extensive among various literature, with a large number of studies on specific data on the field, such as team formation, shooting frequency, foul frequency, etc. Jinmo Lee's [6] research predicts the results of the game through network graph oriented methods. In China, Ma Xiaoni [7] conducted panel data analysis on the results of Chinese Super League matches and obtained very detailed research results. However, based on existing measurement techniques, data statistics cannot achieve 100% accuracy, so the results obtained will inevitably have errors. At the same time, it is difficult to test the psychological and emotional states of players, coaches, and referees using a unified measurement scale, and the quality of survey questionnaires is relatively average. Therefore, research in this area is significantly less than that based on visual data, and there has been no breakthrough progress at present.

3. Model construction

3.1. Variable selection

3.1.1. Dependent variable

Due to the relatively high overall level of European football worldwide [8], the attention and influence of English Premier League matches in the world are far-reaching, and the data records related to the matches are relatively accurate [9], the teams participating in the first five rounds of the English Premier League from September 2020 to September 2023 were selected as the survey objects. In order to more accurately analyze the specific factors affecting football matches [10], this article selects the "win loss situation" indicator of the first five rounds of the English Premier League from September 2020 to September 2023 as the dependent variable [11], represented by Y.

3.1.2. Explanatory

Variable Ball Control Rate (X_1) Ball control rate is one of the core indicators in football game statistics, used to measure the proportion of time a team has control of the ball during the game.

The number of corner kicks (X_2) refers to the set pieces kicked by the attacking team in the corner kick area when the ball as a whole is out of bounds from the baseline and the last touch is made by the defending team. As an important offensive tool, corner kicks are often obtained in favor of the attacking team, thereby promoting goal scoring.

The number of shots (X_3) refers to the total number of times a player attempts to shoot in a game. This directly reflects the players' enthusiasm and ability to seize opportunities on the offensive end, which is crucial to the team's offensive threat.

Shooting number (X_4) refers to the number of times the trajectory of the ball crosses the goal line or hits within the range of the goal frame when shooting. The number of accurate shots can better reflect the quality of the shot, as only accurate shots have a chance of scoring.

Scoring deviation (X_5) refers to the situation where a player's shot deviates from the goal range, including hitting the door frame without being touched by the goalkeeper or defensive player.

The cross digit (X_6) refers to the number of times an attacking player is located in the opponent's half and is closer to the opponent's goal line than the ball and the second to last defensive player (usually including the goalkeeper) at the moment of a teammate's pass, and actually participates in the attack.

The number of fouls (X_7) refers to the number of times a player violates the rules during the game. Excessive fouls can bring disadvantages to the team, such as being awarded free kicks or even penalty kicks, affecting the game situation.

Yellow card count (X_8) refers to the number of times a player is shown a yellow card warning by the referee due to fouls or other reasons. Accumulating two yellow cards will result in the player being suspended for the next game.

Red card count (X_9) refers to the number of times a player is directly shown a red card and sent off by the referee due to serious fouls or other reasons. A red card will result in a team losing one player in the remaining matches, which has a significant impact on the team.

Passing frequency (X_{10}) refers to the total number of times a player passes the ball during the game. Passing the ball is an important part of the attack, constantly passing the ball to find better opportunities for attack.

Team value (X_{11}) refers to the sum of the value of a team's players in the transfer market, usually expressed in monetary terms.

Attack frequency (X_{12}) refers to the number of times a team completes the process of organizing and passing the ball to the opponent's defensive core area (usually near the penalty area) and forming a certain threat.

3.1.3. Other factors

In addition to the factors discussed above that affect the team's outcome, the impact of invisible factors such as weather conditions [12], home and away games, historical match winning rates [13], player self-management abilities, and player mentality levels on the team's outcome [14] is represented by μ .

3.2. Data source

This study mainly estimates the parameters of the model based on the data of 200 matches in the first five rounds of the English Premier League from September 2020 to September 2023 in the DCD app, and supplemented with data from the same matches in apps such as ALLCHART, Sofascore, and LiveBar, as shown in Table 1.

Table 1. Factors influencing the winning and losing of the first five rounds of the English Premier League from September 2020 to September 2023

Team	The team's victory situation	ball control rate	corner kick count	shot count	positive shot count	off target count	The cross digit	number of fouls	yellow card count	red card count	pass count	team value	attack count
	Y	X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8	X_9	X_{10}	X_{11}	X_{12}
1	0	46	2	5	2	2	2	12	2	0	523	14795	108
2	1	54	3	13	6	4	2	12	2	0	626	65135	114
3	1	29	7	5	3	2	4	14	2	0	236	18850	91
4	0	71	3	9	5	3	0	11	1	1	580	21120	134
5	1	49	9	22	6	3	2	9	1	0	433	102000	153
6	0	51	0	6	3	1	5	6	0	0	459	12805	77
7	0	58	8	15	3	4	2	13	2	0	474	36000	112
8	1	42	7	15	2	7	4	7	2	0	342	66000	103
9	0	36	2	7	1	3	4	12	1	0	336	7615	93
10	1	64	5	13	7	3	3	9	1	0	604	43830	111
11	0	52	5	9	5	1	0	15	1	0	552	72155	126
12	1	48	3	15	4	7	1	7	0	0	529	49105	103
13	0	55	12	9	2	5	4	13	2	0	460	15415	155
14	1	45	5	11	4	5	0	7	1	0	391	33580	79
15	0	52	4	13	3	6	0	8	1	0	507	19885	106

Table 1. (continued)

16	1	48	3	10	5	2	3	13	0	0	465	93050	97
17	0	37	3	11	0	7	3	15	1	0	334	15208	103
18	1	63	4	13	7	2	2	13	2	0	566	76935	111
19	1	69	16	28	9	7	3	10	0	0	628	106000	129
20	0	31	2	11	4	4	1	9	2	1	285	32630	64

3.3. Model establishment

When studying the factors influencing the outcome of the English Premier League, there are multiple explanatory variables in the. Therefore, the following preliminary multiple linear regression function is established [15]:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + \beta_5 X_5 + \beta_6 X_6 + \beta_7 X_7 + \beta_8 X_8 + \beta_9 X_9 + \beta_{10} X_{10} + \beta_{11} X_{11} + \beta_{12} X_{12} + \mu \quad (1)$$

3.4. Parameter estimation

Linear regression analysis was performed on the data in Table 1 using SPSS, and the results are shown in Table 2. The estimated results of the model are following [16]:

$$\widehat{Y} = 0.560 + 0.023X_1 - 0.019X_2 - 0.022X_3 + 0.101X_4 + 0.024X_5 - 0.041X_6 - 0.025X_7 + 0.022X_8 - 0.113X_9 - 0.00$$

Table 2. SPSS regression results

Y	Coef.	St.Err.	t-value	p-value	Sig
β_0	0.560	0.215	2.603	0.010	
X_1	0.023	0.008	2.719	0.007	***
X_2	-0.019	0.017	-1.113	0.258	
X_3	-0.022	0.018	-1.249	0.212	
X_4	0.101	0.023	4.313	0	***
X_5	0.024	0.026	0.928	0.354	
X_6	-0.041	0.026	-1.563	0.119	
X_7	-0.025	0.012	-2.145	0.033	**
X_8	0.022	0.030	0.750	0.454	
X_9	-0.113	0.148	-0.767	0.443	
X_{10}	-0.003	0.001	-3.506	0.001	***
X_{11}	4.727E-6	0.000	2.867	0.004	***
X_{12}	0.003	0.002	1.598	0.111	

Table 2. (continued)

R-squared	0.149	Adjusted R-squared	0.120
F-test	4.994	P-value	0.000

***p<0.01,**p<0.05,*p<0.1

4. Interpretation of results

Based on the data of the first five rounds of Premier League matches from September 2020 to September 2023, we conducted linear regression analysis using SPSS. According to the regression analysis results:

(1) we observed significant explanatory variables and found that, while controlling for other independent variables, X_1 changed by an average of 1 unit, Y changed by an average of 0.023 units, X_4 changed by an average of 1 unit, and Y changed by an average of 0.101 units, The average change of X_7 is 1 unit, and the average change of Y is 0.025 units.

(2) The p-value of the overall F-statistic of the model is less than 0.01, indicating that the model is significant, that is, the explanatory variable set has a certain degree of credibility in influencing the competition results.

(3) The F-test value is not negative or too small, indicating normal data quality and good model significance.

(4) The R^2 of the model and the adjusted R^2 values are 0.149 and 0.120, respectively, indicating that the selected explanatory variables have a high degree of fit of the selected explanatory variables and the results are relatively reliable, and can avoid problems such as false correlations caused by increased variables

5. Model testing and correction

Due to the variety of professional indicators that may affect the results of football matches, a total of 16 variables were selected before the initial regression analysis, including 12 variables that were ultimately recorded as explanatory variable [17], as well as the final number of blocked, dangerous attacks, free kicks, and lineup completeness. In the testing process, all 16 explanatory variables were first included in the linear regression analysis and tested [18]. A preliminary model with a P-value of 0.01 was obtained, indicating that the overall model and data quality were good, but some individual data had poor significance. Therefore, the 16 explanatory variables involved in the regression analysis were screened. The first revision removed the completeness of the lineup, and with almost no change in the value of R^2, the adjusted R^2 increased, enhancing the significance of the model; The second correction removed the number of free kicks, further increasing the adjusted R^2 without any change in its value; The third correction removed the number of dangerous attacks and blockages, resulting in a slight increase in adjusted R^2 with almost no change in R and R^2, accurately reducing explanatory variables that have very little impact on the model.

6. Conclusion

6.1. Suggestions and prospects

6.1.1. Suggestions for team training

During the training process, the first step should be to strengthen the practical ability of the attacking end [19]. Coaches should focus on improving the accuracy of players' shooting. It is recommended to

conduct training with defensive personnel to simulate practical situations and improve the quality of shooting. At the same time, attention should be paid to the training of players' stealing ability, and the enthusiasm of players to attack and steal should be mobilized [20]. Divide the field vertically into three zones, and after a successful steal, a shot or effective cross must be completed within 10 seconds. Set a reward for "second attack": If the first shot is blocked, a successful supplementary shot within 3 seconds will earn double points and cultivate a sense of continuous attack [21]. In terms of penalty kicks, the first step is to enhance players' awareness of making penalty kicks and corner kicks, and reduce the situation of giving opponents corner kicks due to mistakes. The coaching team needs to refer to various historical game data before the game, analyze the characteristics of the opponent's goalkeeper's saves, and conduct targeted penalty, corner, and free kick training during the training process, striving to improve the scoring rate of penalty kicks and minimize the success rate of the opponent's penalty kicks. In other words, defensive players should be more standardized and cautious in their defensive actions in the penalty area. Secondly, in terms of foul management, encourage midfielders and defenders to decisively disrupt the opponent's attack. Train players to use reasonable technical fouls outside the penalty area at critical moments to disrupt the opponent's attacking opportunities and reduce unnecessary fouls [22]. For example, in a halftime 8v8 confrontation, the coach suddenly gave a "foul order", requiring the player to choose "destroy the ball path" or "delay the pursuit" within 0.5 seconds outside the penalty area, and immediately raise their hand to signal to the referee, cultivating a habit of taking responsibility proactively. Especially beware of the situation where two yellow cards turn into one red card. Therefore, before the game, the coach should pay attention to the emotional management of players who already have a yellow card. During the game, the coach must also pay attention to the replacement of players who already have a yellow card. While increasing the number of attacks, continuously improve the level of tacit understanding among team members, reduce attack failures caused by offside, and avoid the occurrence of red cards.

6.1.2. Suggestions for the sports bureau

Firstly, the General Administration should organize experts and excellent coaches, establish unified tactical concepts and strategies based on the characteristics of Chinese players, promote the materialization of professional leagues, and enable professional leagues truly composed of clubs to operate, manage, and develop commercially. For example, the establishment of the "National Football Strategy Committee", consisting of technical directors, data scientists, youth coaches, and sports psychology experts, will formulate and release the "Chinese Player DNA Tactical White Paper" within six months, and enforce the gradual implementation of the Chinese Super League, China League One, and U-series competitions. The General Administration of Sport should give more play to its regulatory and service functions, respect market laws and the development of football, and promote the healthy development of the league. Secondly, it is necessary to vigorously promote football into schools at all levels as a basis for popularizing sports and selecting talents, expand the youth football population, enable more children to play football and love it, and actively build a multi-level and large-scale football competition system, creating a good atmosphere for the whole society to pay attention to and participate in football. Actively organize various levels of football events and establish tradition and scale, promoting public attention to football events. At the same time, it is necessary to strengthen the cultivation and improvement of coach skills, select young coaches with potential to study in developed football countries, and introduce relevant experience into training practice. Hiring world-class coaches to come to China to provide strategic and tactical guidance to players and domestic coaches, updating knowledge and awareness. Finally, it is necessary to ensure logistics and intelligence work, and conduct reliable investigations into the competition venue, climate, time difference, etc. in advance. Establish a powerful intelligence system that not only collects videos of opposing teams, but also gathers in-depth

information on their player characteristics, tactical details, and provides comprehensive support for the coach's decision-making.

6.1.3. Suggestions for individual players

Players should identify their weaknesses [23] and receive special training based on their weaknesses [24], such as wingers' crosses and forwards' kicks. At the same time, players need to enhance their self-discipline and focus, and engage in strict self-management. For example, they can collaborate with nutritionists to create a "three color plate" - green leaf fiber, red and white protein, and orange carbohydrates. Each meal will be photographed and uploaded for AI recognition, and the next day they violate the rules, they will run an additional 800 meters. Reduce situations that are detrimental to the game, such as injuries and poor performance due to lack of self-discipline. Consciously engage in self psychological suggestion during training to enhance self-confidence, stress resistance, and competition focus [25]. Manage emotions well during training, reduce the number of fouls caused by emotional reactions, and enhance team awareness and dedication. Sacrifice too much personal heroism for more favorable attacking situations.

6.2. Shortcomings

Firstly, the data in this study is mainly quantitative, with less discussion on qualitative factors such as the head coach's on-site command ability, player mentality, team unity, etc. These factors also have a significant impact on the game results. Secondly, the data only comes from the competition data of Premier League matches from September 2020 to September 2023, and the universality of the analysis results in different countries, types, and levels of competitions still needs further verification.

References

[1] Matthew Cole; Jennie L Carter; Frances Brooks; Charlie J Roberts(2025).Caregivers nutrition knowledge and perspectives on the enablers and barriers to nutrition provision for male academy football players.Journal of the International Society of Sports Nutrition, 22.

[2] Gygax, Pascal M; Wagner-Egger, Pascal; Parris, Ben; Seiler, Roland; Hauert, Claude-Alain(2008).A psycholinguistic investigation of football players' mental representations of game situations: Does expertise count? European Journal of Psychology Open, 67.

[3] Liu Yang.A cognitive study on the home and away effects of Chinese men's professional football league players [D]. Shanxi University, 2024.

[4] Han Zhe.Research on home advantage in the second leg of the UEFA Champions League and Europa League [D]. Southwest University of Finance and Economics, 2023.

[5] Yin Xiaochuan.Research on the influencing factors of home advantage in China's Jia A Football League [C]. Chinese Psychological Association Selected Abstracts from the 8th National Conference on Psychology Beijing Institute of Physical Education; , 1997: 233.

[6] Jinmo Lee; Eunil Park; Angel P. del Pobil(2025)We know who wins: graph-oriented approaches of passing networks for predictive football match outcomes.Journal of Big Data, 12.

[7] Ma Xiaoni.Analysis of Influencing Factors on the Results of Chinese Super League Matches in the 2009-2018 Season [D]. Capital Institute of Physical Education, 2020.

[8] Saragih Joe Pratama; Widodo Achmad; Bawono Mokhamad Nur; Firmansyah Awang(2025).Analysis of The Playing Technique Skills of The English Manchester City Football Club in Winning The UEFA Champions League Title for The 2022–2023 Season.Proceedings of the International Seminar of Sport and Exercise Science, 14-19.

[9] Fernando Leiva Bertrán(2025), Ranking in Incomplete Tournaments: The Generalized Win Percentage Method, Efficiency, and NCAA Football.Journal of Sports EconomicsVolume 26, 3-34.

[10] Hill S. E.(2022).In-game win probability models for Canadian football.Journal of Business Analytics, 5, 164-178

[11] Gamble Declan; McCarren Andrew; Bradley Jonathan; Moyna Niall(2020), Novel component variables can be used to distinguish between winning and losing across match periodsin elite Gaelic football.International Journal of Performance Analysis in Sport, 20, 264-279.

[12] Lionel Frost; Luc Borrowman; Vinod Mishra; Abdel K. Halabi(2019).Stadium construction and the distribution of winning percentages in the English and Scottish football leagues, 1888-89 to 1938-9.Historical Perspectives on Sports Economics: Lessons from the Field, 37-51.

[13] Gamble; Bradley; McCarren; Moyna(2019).Team performance indicators which differentiate between winning and losing in elite Gaelic football.International Journal of Performance Analysis in Sport, 19, 478-490.

[14] Silvestri R.; D'isanto T.; Izzo R.; Di Domenico F.(2019).The winning shares in the major football leagues.Journal of Human Sport and Exercise, 14, S631-S637.

[15] Cheng Kang.Research on the Model Construction of Winning Factors in the Chinese Football Super League [D]. Shaanxi Normal University, 2021.

[16] Li Chunhua.Analysis of Winning Factors in the 2019 Chinese Football Association Super League [J]. Sports Excellence, 2022, 41 (7): 93-95

[17] Hu Guangming.Technical continuity and relative advantages of teams that have won consecutive championships in the five major European football leagues [D]. Fujian Normal University, 2021.

[18] Par Prateek; Kumar Gupt Ankit; Singh Samarth; Khare Neelu; Bhattachrya Sweta(2017).Football league win prediction based on online and league table data.IOP Conference Series: Materials Science and Engineering, 263.

[19] Sigrid Olthof; Jesse Davis(2025).Perspectives on data analytics for gaining a competitive advantage in football: computational approaches to tactics.Science and Medicine in Football.

[20] Xu Lulu.Quantitative Study on Factors Influencing the Performance of Chinese Super League Teams [D]. Shanghai Normal University, 2020.

[21] Tahir Sheikh Badar Ud Din(2025).A Comparative and Statistical Analysis of xG to Predict Football Shot Outcomes in Event Data.Statistics for Innovation I: SIS 2025, Short Papers, Plenary, Specialized, and Solicited Sessions, 408-414.

[22] Mohamed Elsharkawi; Raja Hashim Ali; Talha Ali Khan(2025).Crafting a Player Impact Metric through analysis of football match event data.Journal of Computational Mathematics and Data Science, 15.

[23] Victor Chang; Sreeram Sajeev; Qianwen Ariel Xu; Mengmeng Tan; Hai Wang(2024).Football Analytics: Assessing the Correlation between Workload, Injury and Performance of Football Players in the English Premier League.Applied Sciences, 14.

[24] Brager Allison J; Mistovich R Justin(2017).Game Times and Higher Winning Percentages of West Coast Teams of the National Football League Correspond With Reduced Prevalence of Regular Season Injury.Journal of Strength and Conditioning Research, 31, 462-467.

[25] Brigitte, Nana; André, Kaboré; Gnynayen Xavier, Ziba(2024).Emotional Analysis of First Division Football Referees in Burkina Faso during Matches.Journal of Psychology & Behavior Research, 6.