



# Predicting the Outcomes of Football Matches Using Machine Learning Approach

Usman Haruna<sup>1,2(✉)</sup>, Jaafar Zubairu Maitama<sup>3</sup>, Murtala Mohammed<sup>2</sup>,  
and Ram Gopal Raj<sup>1</sup>

<sup>1</sup> Department of Artificial Intelligence, Faculty of Computer Science and Information  
Technology, University of Malaya, Kuala Lumpur, Malaysia

<sup>2</sup> Department of Computer Science, Faculty of Science, Yusuf Maitama Sule University,  
Kano, Nigeria

uharuna@yumsuk.edu.ng

<sup>3</sup> Department of Information Technology, Faculty of Computer Science and Information  
Technology, Bayero University, Kano 3011, Nigeria

**Abstract.** Predicting outcomes of football matches is among the rapid growing area of research due to the interest of large number of people, and the stochastic nature of the results. Many researches have been conducted to predict the outcomes of football matches. Statisticians predict the football match outcomes to showcase their skills whereas Operational researchers use the prediction to experiment with various effects of football tournament design. This research is aimed at determining feature and classifier combination that would provide the best English premier League football outcomes prediction accuracy. Despite the fact that feature and classifier combination might effects the football outcomes prediction accuracy, however, based on our knowledge, no other researches had considered to determine which possible feature and classifier would produce a good English Premier League football prediction accuracy. We proposed to use multiple machine learning classifiers and features combination to determine which feature and classifier combination would produce high football prediction accuracy. Experiments were conducted using various feature and classifier combinations including previous football match records of English premier league for two seasons. The results of all feature and classifier combinations were compared to determine the one that would achieve the highest accuracy. The result of the experiments show that using home team factor, away team, twenty two (22) first players from the two teams, and goal difference together with K-NN machine learning classifier achieved the highest accuracy of 83.95%. In addition, the result of the best feature and classifier combination was compared with the result of the existing English Premier League football prediction accuracy and found that our result achieved improvement over existing results.

**Keywords:** Features · Football result · K-NN · Machine learning techniques

## 1 Introduction

Football which is frequently referred to as soccer is the most popular sport that attracts a large number of fans throughout the world [1, 2]. The football sport can either be

domestic or international leagues. Examples of domestic leagues are English Premier League (EPL), Liga Super Malaysia (LSM), Australian Football League (AFL), etc. Example of most famous international tournament is the world cup. For domestic league, each team has to play with each other both at home and away. Football has three (3) possible outcomes, win, lose or draw [3]. Each match win carries three (3) points, one point (2) for draw, and nil point (0) for lose. The team with highest number of points at the end of the season of the league will be crown the champion. A quite number of leagues exist, and there is an estimate of 108 professional soccer leagues in the world. English premier league is the top league in United Kingdom (UK). It is the most popular league that attracts many followership across the globe, investors, etc., than the other leagues in the world which was reported to be watched on television for about nineteen (19) hours by approximately 4.6 billion audiences [1].

Predicting outcomes of football match is challenging and difficult due to its high dependence on many interactive factors that cannot be easily interpreted, such as refereeing subjectivity, key players, home team, coaching strategy, field dimension, distance between the two teams, etc. Study had shown that machine learning (ML) can be used in football outcomes predictions with better accuracy than the natural human expert [4]. The machine learning aims at determining the value of an unknown sample through learning from the already known dataset [5, 6]. Examples of such machine learning techniques are ANN, K-NN, Genetic Programming algorithm (GPA), SVM, Naïve Bayes (NB), and many others [7–9].

Selecting both feature and classifier plays an important role in determining the success of football match outcomes prediction accuracy. Good choice of features and classifiers tends to improve the prediction accuracy, while wrong choice of features and classifiers tends to decrease the prediction accuracy [10, 23]. Previous work by Cui et al. [11], an ensemble concept was proposed, where Genetic programming together with the majority voting function were used to predict the outcomes of English Premier League Matches. The number of features used in this experiment was six (6), and they included the dynamics profile factor, the ranking factor, the home advantage factor, the strength factor, the infirmity factor and the odds factor. The system achieved the overall accuracy of 75%.

Predicting outcomes of football matches has proven to be difficult and challenging to implement [3, 11]. According to Hucaljuk and Rakipovic, such complexity was as a result of a presence of many uncertainties that cannot easily be interpreted. Many researches have been conducted to predict the outcomes of a football match event. Despite the fact that the researches showed that competing interaction improves the prediction accuracy [12], however, based on our knowledge, it had never been indicated which feature and classifier combination would produce high football match outcomes prediction accuracy. Additionally, the fact that the current best English Premier League prediction accuracy is 75%, still there is the need for improvement since it is not up to 100%. The major contribution of this study is that veterans and novice researchers can come up with a new method for predicting different sporting events while considering other uncertain features.

The rest of the paper has been organized as: Sect. 2 contains a literature review, Sect. 3 contains a methodology of the research, and Sect. 4 encompasses Results, Discussion and evaluation. Finally Sect. 5 concludes the research.

## 2 Literature Review

Predicting football match outcomes has been repeatedly studied for decades. For example, in 1982, Maher was first to use a statistical approach to predict the football match outcomes, where Poisson distribution was used to predict a result based on the goals scored by both home and away teams.

Most of the previous approaches predict the football outcomes based on the number of goals conceded for each team, while the obtained match results will determine the actual champion winner [13–15].

Current football prediction approaches determine the outcomes of football matches directly based on one of the three distinct parameters [3, 11]. These parameters include win (w), lose (L) or draw (D). There are two categories of football outcomes prediction techniques, namely football prediction using statistical techniques and prediction using machine learning techniques.

### 2.1 Predictions Based on Statistics Techniques

In sporting event prediction, football outcome prediction is a popular issue among researchers. Many researches have been conducted to predict football scores.

Poisson distribution technique has been widely accepted. A technique that uses Univariate and Bivariate Poisson distribution was proposed in the study of Maher (1982) to reflect the defensive and offensive capabilities of both the two playing teams. Another approach which is more complex than the one proposed by Maher was proposed by Dixon and Coles [13], where the “correction factor” was applied to the independent Poisson model to enhance the performance of the system. Another more powerful approach which uses Bivariate Poisson distribution with more complex likelihood function was proposed. The covariance between the goals scored by two teams was added in the research, such that proposed by Tsionas [14].

In another study, an extremal statistics (ES) was used to analyse the distribution of number of goals that are scored by Home teams, away teams and the overall number of goals in the match in domestic football games from 169 countries between 1999 to 2001 [15].

Besides that, Monte Carlo analysis (MCA) was also proposed to generate a sequence of simulated match outcomes for different teams based on Zero persistence assumptions [16].

### 2.2 Predictions Based on Machine Learning Techniques

Ong and Flitman [4] proposed an artificial neural network to predict the binary outcomes of a sporting event, namely winner of an Australian Football match where the datasets from 1992 to 1994 were used for training purpose. Comparisons were made between

neural networks, logistic regression (LR), and human experts (HE). The result shows that neural networks outperformed both the Logic regression and natural human expert in prediction accuracy. While the logic regression outperformed human expert. Finally, the authors concluded that machine learning techniques can be used to replace human experts in football prediction. One shortcoming with this approach is that the system could not predict the draw outcome.

Rotshtein et al. [18] proposed the use of Fuzzy Logic (FL), Genetic Algorithm (GA) as well as Neural Turning to predict the football results of championship in Finland for eight (8) years from 1994 to 2001. Twelve (12) features were used in this research. These features include the results of five (5) previous matches of the two teams, and the results of the two (2) previous matches between the two playing teams. The results were compared with other classical time series. The fuzzy logic was found to decrease the number of experimental data due to the expert knowledge.

Framework for sport prediction was proposed by Min et al. [19] which combined Bayesian network along with Rule-based reasoning. The result was found to provide more reasonable results in predicting World cup result by simulating various strategies together with their subjective information. However, the research does not consider the domain that has insufficient expert knowledge. That is, the system was solely dependent on expert knowledge.

Bayesian networks was proposed by Joseph et al. [20] to predict the outcomes of English premier league played by Tottenham Hotspur in two seasons namely 1995/1996 and 1996/1997. The classifiers used in this experiment includes naïve Bayes, K-NN, Hugin BN and Expert BN. Seven (7) features were used in the experiment which revealed that Expert BN outperformed the Naïve Bayes, K-NN and Hugin BN classifiers with accuracy of 59.21%. The limitation of this technique is that the datasets used were relatively very small. Similar approach was proposed by Hucaljuk and Rakipovic [3] to predict the outcomes of English Champion league matches by using large sets of classifiers. The research contained all features and some other classifiers used in other studies [20, 21], but with little increment in classifiers. The accuracy of 60% was achieved in this study, which was better than the result found in the previous study [20].

Recently, another machine learning approach called ensemble concept using genetic programs system was proposed to predict the outcomes of English Premier League Matches [11]. The number of features used in this study was six (6), and this includes the dynamics profile factor, the ranking factor, the home team advantage factor, the strength factor, the infirmity factor and the odds factor. The system was evaluated by comparing the results with that of Artificial Neural Networks. The result achieved by a single Genetic-Program (GP) was 68.8%, which was almost similar with that of referenced paper (Artificial Neural Network which accuracy of 70%). However, after combining the results of GP-generated with that of Majority Voting function (MVF), the accuracy of the whole system was increased to 75%. Finally, the authors concluded that combining decisions of a number of classifiers provides better improving performance than just a single decision. Major limitation of this prediction technique is that it is very complex and time consuming as many classifiers might be involved in the prediction process.

All of the techniques discussed above have only directly considered the use of a number of features together with a particular machine learning classifier in trying to

predict a football match result for a particular league, with the aim of achieving higher accuracy. However, in reality, identifying appropriate feature and classifier combination is an important factor in determining high prediction accuracy result of football matches.

### 3 Proposed Methodology for a Football Match Outcomes Prediction

#### 3.1 Data Sampling

During every English Premier League season, exactly twenty (20) teams participated in the competition, where each team played twice with one another – one match at home, while the second away. A complete season of English Premier League consists of three hundred and eighty (380) matches. For the purpose of this research, data for two seasons were used: 2011–2012 and 2012–2013, which is equivalent to seven hundred and sixty (760) different matches which have been extracted from the following prominent English Premier League websites:

- <http://www.premierleague.com>
- <http://www.bbc.co.uk/sports>
- <http://statto.com/football/stats/england/premierleague>.

#### 3.2 Features Used and Process of Feature Selection

Selecting relevant features is significant in providing accurate prediction accuracy as it affects the performance of machine learning classifiers [23].

For the purpose of this research, various combinations of the following thirty-three (33) features were selected and used in the study. These selected features include twenty-two (22) player names for both home and away teams, Home Team, Away Team, Average Home Team Goals per Game, Average Away Team Goals per Game, Home Team Rank, Away Team Rank, Home Team Attack, Away Team Attack, Home Team Defence, and Goal Differences.

A sequential Forward Selection (SFS) technique was adopted for the feature selection, due to its relatively low computational burden of [24]. This SFS is based on the algorithm, called a greedy search algorithm, which determines an optimal set of features for extraction by first starting from an empty set and then adding a single feature that increases the values of the chosen objective function in the superset in sequence to the subset. Pseudocode for SFS is shown as below:

Feature set initialization

$$F_0 = \{\emptyset\}; i = 0$$

Select the next best feature

$$x = \arg \max[J(F_i + x)].$$

where  $x \neq F_i$

Update the feature set

$$F_{i+1} = F_i + x.$$

While  $i < d$

$$i = i + 1$$

Go to step 2

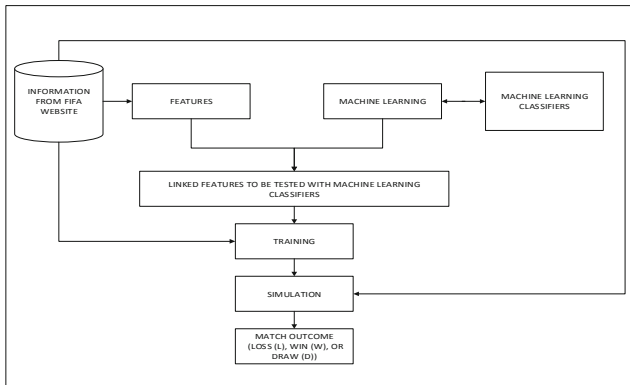
### 3.3 Machine Learning Classifiers

The following machine learning classifiers were used in the experiments:

- Logistic Regression
- SVM
- Random Forest
- K-NN
- Naïve Bayes

Each of the above mentioned classifiers was trained and tested with different feature combinations to determine the best combination.

### 3.4 Experimental Methodology



**Fig. 1.** Experimental methodology processes

In Fig. 1, a large amount of datasets for two previous seasons of an English Premier League matches, seven hundred and sixty matches (760) have been retrieved from the aforementioned English premier league websites. The two English premier league seasons considered are 2011–2012 and 2012–2013. It is then followed by feature and machine learning classifier selections. And it then followed by combining these selected

features with the selected machine learning classifiers one at a time. These retrieved datasets were divided into two. The first dataset for 2011–2012 season is used for training purpose, while the second dataset which contains information about matches for 2012–2013 season is used for testing purpose. The predicted result could either be home win, draw, or home loss.

Lastly, we then compared the accuracy of our prediction approach in two phases. The first phase is to compare the results of our five (5) classifiers. The second phase is to compare the prediction accuracy of our best classifier with the performance of the existing similar football prediction techniques.

### 3.5 Final Implementation Based on the Experiments

K-NN machine learning classifier calculates the distance between scenarios that exists in the dataset and a query scenario using distance function formula to compute the distance between scenarios, where  $a, b$  scenarios have  $N$  features, in this case, the value of  $N = 25$ , such that  $a = \{H\_Team, A\_Team, Goal\_Diff, Player1a, Player2a, \dots, Player11a, Player1b, Player2b, \dots, Player11b\}$ , and  $b = \{H\_Team^1, A\_Team^1, Player1a^1, Player2a^1, \dots, Player11a^1, Player1b^1, Player2b^1, \dots, Player11b^1\}$ .

Where  $d(a, b)$  can be obtained in two ways:

1. Using absolute distance formula,

$$dA(a, b) = \sum_{i=1}^{N=25} |a_i - b_i|$$

2. Using Euclidean Distance formula,

The procedure is as follows:

Step 1: Set  $z \leftarrow 380$ .

Step 2: Store the output of the  $p$  nearest neighbours to the query scenario  $q$  in vector  $r = \{r^1, \dots, r^p\}$  by repeating the following loop  $p$  times:

- i. Go to the next scenario  $s^i$  in the data set, where  $i$  is the current iteration within the domain  $\{1, \dots, z\}$
- ii. If  $q$  is not set or  $q < d(q, s^i): q \leftarrow d(q, s^i), t \leftarrow o_i$
- iii. Loop until the end of the data set (i.e.  $i = z$ )
- iv. Store  $q$  into vector  $c$  and  $t$  into vector  $r$ .

Step 3: Calculate the arithmetic mean output across  $r$  as follows:

$$\bar{r} = \frac{1}{p} \sum_{i=1}^p r_i$$

Step 4: Return  $\bar{r}$  as the output value for the query scenario  $q$ .

## 4 Results, Discussion, and Evaluation

### 4.1 Presentation of Experimental Results

In this section of the chapter, the results obtained from the experiments conducted are presented and described.

Empirical evidences indicated that match location (either team is playing at home or at the opposition home), match status (whether team was losing, winning or drawing) and the quality of opposition features have been reported to be among the most significant influences on football match performance [25–27]. Similarly, the current position of the team in the League ranking and the average number of conceived and the scored goals per game features have been selected in the research of Hucaljuk and Rakipovic [3].

Experiment 1:

In this experiment, thirty eight (38) matches played by Manchester United team have been considered in order to start with a small scale number of datasets. Therefore, eight (8) distinct features have been considered in the experiment as follows:

Home Team, Away Team, Home Team rank, Away Team Rank, Home Team attack, Home Team Defence Rank, Away Team Attack Rank, Away Team Defense Rank.

After applying our proposed five (5) Machine Learning Classifiers, we have obtained the result shown in Fig. 2.

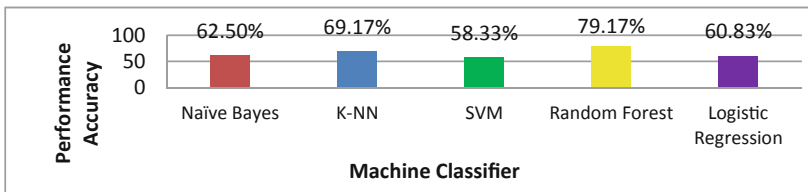


Fig. 2. Results achieved in Experiment-I

Experiment 2:

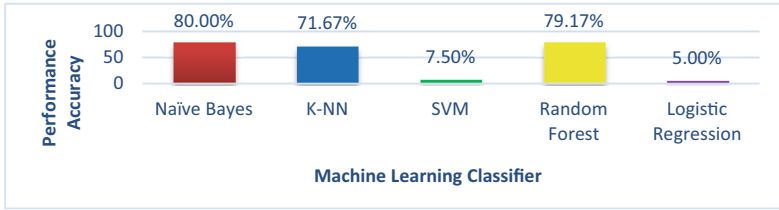
In this experiment, the same sets of dataset as in experiment one (1) were used. It is noted that some features are complimentary: For instance, the average home team goals per game and the average away team goals per game are complimentary with goal difference. Therefore, a technique called feature combination was used, which is a method used in object classification activity which takes the advantage of combining the strength of multiple complementary features to provide a more powerful feature [28].

Adding these features to the features used in the previous experiment, experiment I, the researcher had eleven (11) distinct features, and they were used in this experiment and the subsequent one. These features are as follows:

Home Team, Away Team, Average Away Goals per Game, Average Home Goals Per Game, Home Team Rank, Away Team Rank, Home Team Attack, Home Team Defence, Away Team Attack, Away Team Defence, and Goal Difference.

After our five (5) Machine Learning classifiers have been applied on the aforementioned feature combination, the results illustrated in Fig. 3 have been obtained from this experiment:





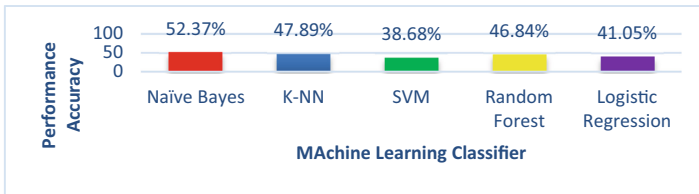
**Fig. 3.** Results achieved in Experiment-II

**Experiment 3:**

Due to the small scale number of testing datasets that were used previously in the two preceding experiments to test the feature and classifier combination, and also because the number of dataset might affect the success of football outcomes prediction [23], the number of testing dataset used here was increased to three hundred and eighty (380) matches played by all the twenty (20) English premier league teams in 2012/2013. Ten (10) features were considered in this experiment. These features include:

Home Team club, Away team club, Average Home Team Goals per Game, Average Away Team Goals per Game, Home Team Rank, Away Team Rank, Home Team Attack, Away Team Attack, Home Team defence, Away Team Defence.

The results are illustrated in Fig. 4 below:



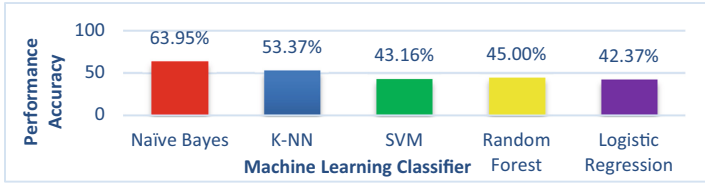
**Fig. 4.** Results achieved in Experiment-III

**Experiment 4:**

As in the previous experiment, experiment 3, three hundred and eighty (380) English premier league football matches played by all the twenty (20) teams for 2012–2013 were also used in this experiment. The only difference here is the addition of twenty two (22) player names and goal difference, making features to become thirty three (33) features. This is because, players are complementary with one another, and also the average home team goal per game and the away team goal per game are both complementary with the goal difference. These features are reported as follows:

Twenty-two (22) player names, Home Team, Away Team, Average Home Team Goals per Game, Average Away Team Goals per Game, Home Team Rank, Away Team Rank, Home Team Attack, Away Team Attack, Home Team Defence, Away Team Defence, & Goal Difference.

Figure 5 has illustrated the results, after applying our five (5) Machine Learning classifiers on the aforementioned feature combinations.



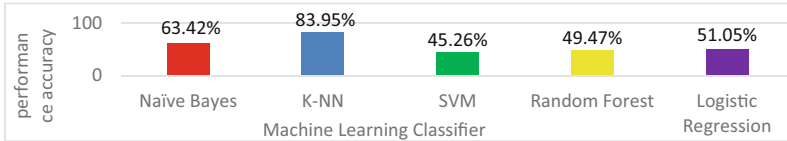
**Fig. 5.** Results achieved in Experiment-IV

**Experiment 5:**

As in the experiment 3 and experiment 4, three hundred and eighty (380) English premier league matches played by twenty (20) teams were used in this experiment. The number of features used in this experiment was twenty-five (25). These features were reported as follows:

Twenty-two (22) Player names, Home Team, Away Team & Goal Difference.

The result is presented in Fig. 6 as shown below:

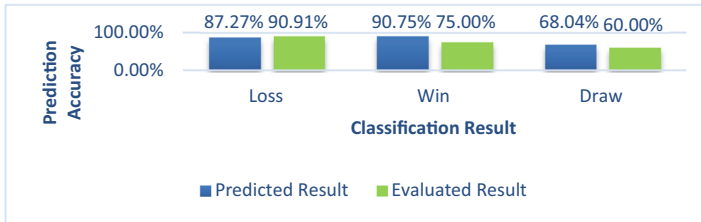


**Fig. 6.** Results achieved in Experiment-V

From the results of all the experiments that have been conducted so far, it is observed that the feature combination used in the second experiment, experiment 2, which includes the home team club, away team club, average away team goals per game, average home team goals per game, home team ranking, away team ranking, home team attacking, away team attacking, home team defence, away team defence and goal difference with the Logistic Regression classifier that were used in Fig. 3 provided the most poorest prediction accuracy result; while the feature combination that was used in the fifth experiment, experiment 5, which includes home team club, away team club, twenty-two (22) players, and goal differences, and K-NN classifier in Fig. 6 produced the highest prediction accuracy. In most cases in this experiments, it is observed that the classifiers find it difficult to predict the draw; this is the facts that when draw occurs, either the home team is better than away team, or vice-versa, making the prediction task to become very difficult due to the occurrence of tie under potential upsets. In other cases, the classifiers also find it difficult to correctly predict the loss; this occurred possibility due to the poor decision made by the team manager (coach) purposely designed to intentionally increase the team chance of losing the match instead of ordering the players actually on the field to deliberately underperform.

For the evaluation purpose, another experiment was conducted, where the features comprises home team club, away team club, twenty-two (22) players, and goal differences together with the K-NN classifier combination were tested but different league was considered, Australian League, in twenty (20) matches played by an Adelaide United

club as with the previous experiment, where the best feature and classifier combination were obtained. The result achieved the overall prediction accuracy of 80.00%. Even though the result was not as good as the result that was achieved in experiment 5, where the best feature and classifier combination were obtained, however, an accuracy which is higher than that obtained from the previous related studies was still obtained.



**Fig. 7.** Comparisons between predicted result and evaluated result

Figure 7 compares the classification rates between the results that we were obtained (experimental results) and the result of the experiment that have been conducted (evaluated results) to measure the effectiveness of the proposed approach. In terms of predicting the loss; the classification rate of the experimental result is 87.27% while the classification rate of the evaluated result is 90.91%. In terms of winning prediction; the experimental result achieves the classification rate of 90.75% while the classification rate for the evaluated result is 75.00%. Lastly, for draw prediction; the experimental result achieves the classification rate of 68.04% while the classification rate for the evaluated result is 60.00%. This shows that the approach works efficiently.

Comparing the accuracy that has been achieved with all the accuracies achieved by the previous techniques, the approach has the highest accuracy.

## 5 Conclusion

In this research article, an approach which was not previously explored was used, determining the feature and classifier combination to predict the outcomes of an English Premier League football matches with a better accuracy than the existing research. The past and the current research efforts in football outcomes prediction were extensively reviewed, determining their strengths, weaknesses, evaluation techniques used and conclusion. A feature and classifier combination which provide better accuracy than other related existing approaches previously used in predicting football match outcomes have been presented.

After conducting series of experiments, the result shows that using home team, away team, twenty-two (22) players and goal difference features helps provide best prediction result. Additionally, the results show that K-NN classifier can be used to predict the outcomes of English premier matches, getting a highest prediction accuracy of 83.95%, which is better than all other classifiers used in the research. This indicates that the combination of home team, away team, twenty-two (22) players, and goal difference along with K-NN classifier has the highest prediction accuracy.

Comparing the prediction accuracy of the best classifier to that used in other similar research in Joseph et al. [20] and Cui et al. [11], it can be concluded that the approach had achieved the best prediction accuracy.

However, the combination of features used in the second experiment, experiment 2, which includes Home Team, Away Team, Average Away Goals per Game, Average Home Goals Per Game, Home Team Rank, Away Team Rank, Home Team Attack, Home Team Defence, Away Team Attack, Away Team Defence, and Goal Difference with both the random forest and logistic regression, respectively, presented to provide the poor draw prediction accuracy. The reason for that might be the facts that when draw occurs, either the home team is better than away team, or vice-versa, causing the prediction task difficult due to the occurrence of tie under potential upsets. In future research, more data and more feature combinations should be used to design a more accurate model.

Indeed, this research was the first to introduce the concept of using feature and classifier combination in the area of football outcome prediction.

## References

1. Dobson, S., Goddard, J.: *The Economics of Football*, Cambridge University Press, London (2011)
2. Dunning, E.G., Maguire, J.A., Pearton, R.E.: *The Sports Process: A Comparative And Developmental Approach*, Human Kinetics Publishers, Champaign (1993)
3. Hucaljuk, J., Rakipovic, A.: Predicting football scores using machine learning techniques. In: 34th International Convention MIPRO, Opatijia (2011)
4. Ong, E., Flitman, A.: Using neural networks to predict binary outcomes. In: International Conference on Intelligent Processing Systems, IEEE, Beijing (1997)
5. Arowolo, M.O., Ogundokun, R.O., Misra, S., Kadri, A.F., Aduragba, T.O. Machine learning approach using KPCA-SVMs for predicting COVID-19. In: Garg, L., Chakraborty, C., Mahmoudi, S., Sohmen, V.S. (eds.) *Healthcare Informatics for Fighting COVID-19 and Future Epidemics*. EAI/Springer Innovations in Communication and Computing. Springer, Cham (2022). <https://doi.org/10.1007/978-3-030-72752-9-10>
6. Abayomi-Zannu, T.P., Odun-Ayo, I., Tatama, B.F., Misra, S.: Implementing a mobile voting system utilizing blockchain technology and two-factor authentication in Nigeria. In: Singh, P.K., Pawłowski, W., Tanwar, S., Kumar, N., Rodrigues, J.J.P.C., Obaidat, M.S. (eds.) *Proceedings of First International Conference on Computing, Communications, and Cyber-Security (IC4S 2019)*. LNNS, vol. 121, pp. 857–872. Springer, Singapore (2020). [https://doi.org/10.1007/978-981-15-3369-3\\_63](https://doi.org/10.1007/978-981-15-3369-3_63)
7. Awotunde, J.B., Ogundokun, R.O., Jimoh, R.G., Misra, S., Aro, T.O.: Machine learning algorithm for cryptocurrencies price prediction. In: Misra, S., Kumar Tyagi, A. (eds.) *Artificial Intelligence for Cyber Security: Methods, Issues and Possible Horizons or Opportunities*. SCI, vol. 972, pp. 421–447. Springer, Cham (2021). [https://doi.org/10.1007/978-3-030-72236-4\\_17](https://doi.org/10.1007/978-3-030-72236-4_17)
8. Odusami, M., Abayomi-Alli, O., Misra, S., Abayomi-Alli, A., Sharma, M.M.: A hybrid machine learning model for predicting customer churn in the telecommunication industry. In: Abraham, A., Sasaki, H., Rios, R., Gandhi, N., Singh, U., Ma, K. (eds.) *IBICA 2020*. AISC, vol. 1372, pp. 458–468. Springer, Cham (2021). [https://doi.org/10.1007/978-3-030-73603-3\\_43](https://doi.org/10.1007/978-3-030-73603-3_43)

9. Garba, A., et al.: Machine learning model for recommending suitable courses of study to candidates in Nigerian Universities. In: 21st International Conference on Computational Science and Its Applications, pp. 257–271. Springer Cham (2021)
10. Maitama, J.Z., et al.: Text normalization algorithm for Facebook chats in Hausa language. In: 5th International Conference on Information and Communication Technology for the Muslim World, pp. 1–4. IEEE, Kuching (2014)
11. Cui, T., Li, J., Woodward, J.R., Parkes, A.J.: An ensemble based genetic programming system to predict English football premier league games. In: IEEE Conference on Evolving and Adaptive Intelligent Systems (EAIS) (2013)
12. De Pero, R., Amici, S., Benvenuti, C., et al.: Motivation for sport participation in older Italian athletes: the role of age, gender and competition level. *Sport Sci. Health* **5**, 61–69 (2009). <https://doi.org/10.1007/s11332-009-0078-6>
13. Dixon, M.J., Coles, S.G.: Modelling association football scores and inefficiencies in the football betting market. *J. Roy. Stat. Soc.: Ser. C (Appl. Stat.)* **46**(2), 265–280 (1997)
14. Tsonas, E.G.: Bayesian multivariate Poisson regression. *Commun. Stat. Theory Methods* **30**(2), 243–255 (2001)
15. Greenhough, J., Birch, P., Chapman, S.C., Rowlands, G.: Football goal distributions and extremal statistics. *Phys. A* **316**(1), 615–624 (2002)
16. Dobson, S., Goddard, J.: Persistence in sequences of football match results: a Monte Carlo analysis. *Eur. J. Oper. Res.* **148**(2), 247–256 (2003)
17. Tsakonas, A., Dounias, G., Shtovba, S., Vivdyuk, V.: Soft computing-based result prediction of football games. In: First International Conference on Inductive Modelling (ICIM'2002), Lviv, Ukraine (2002)
18. Rotshtein, A.P., Posner, M., Rakityanskaya, A.B.: Football predictions based on a fuzzy model with genetic and neural tuning. *Cybern. Syst. Anal.* **41**(4), 619–630 (2005). <https://doi.org/10.1007/s10559-005-0098-4>
19. Min, B., Kim, J., Choe, C., Eom, H., McKay, R.I.: A compound framework for sports results prediction: a football case study. *Knowl.-Based Syst.* **21**(7), 551–562 (2008). <https://doi.org/10.1016/j.knosys.2008.03.016>
20. Joseph, A., Fenton, N.E., Neil, M.: Predicting football results using Bayesian nets and other machine learning techniques. *Knowl.-Based Syst.* **19**(7), 544–553 (2006)
21. Mitchell, T.M.: *Machine Learning*. McGraw Hill, Burr Ridge, p. 45 (1997)
22. Louzada, F., et al.: A Bayesian Approach to Predicting Football Match Outcomes Considering Time Effect Weight Interdisciplinary Bayesian Statistics, pp. 149–162. Springer, Cham. (2015). <https://doi.org/10.1007/978-3-319-12454-4>
23. John, G.H., Kohavi, R., Pflieger, K.: Irrelevant features and the subset selection problem. In: International Conference on Machine Learning (1994)
24. Doak, J.: *An evaluation of Feature Selection Methods and Their Application to Computer Security*: University of California, Computer Science (1992)
25. Carling, C., Williams, A.M., Reilly, T.: *Handbook of Soccer Match Analysis: A Systematic Approach to Improving Performance*: Psychology Press, Hove (2005)
26. Taylor, J.B., Mellalieu, S.D., James, N., Shearer, D.A.: The influence of match location, quality of opposition, and match status on technical performance in professional association football. *J. Sports Sci.* **26**(9), 885–895 (2008)
27. Tucker, W., Mellalieu, S.D., James, N., Taylor, J.B.: Game location effects in professional soccer: a case study. *Int. J. Perform. Anal. Sport* **5**(2), 23–35 (2005)
28. Gehler, P., Nowozin, S.: On feature combination for multiclass object classification. In: 12th International Conference on Computer Vision, IEEE (2009)