



مجلة جامعة سبها للعلوم البحتة والتطبيقية
Sebha University Journal of Pure & Applied Sciences

Journal homepage: www.sebhau.edu.ly/journal/index.php/jopas



Soccer Team Performance Forecasting using artificial neural network

*Mabrouka Amhamed Al-Shebany, Jazyza Moftah Amshaher, Hana Mohammed

Department of Computer, Faculty of Sciences, Sirte University, Libya

Keywords:

Artificial Neural Network (ANN)
Forecasting model

ABSTRACT

Recently a large effort was spent on forecasting the outcome of sporting events. Due to forecasting perspective, the presence of competition introduces particular modeling challenges, which in turn limit the applicability of standard techniques. The objective of this study is to create a soccer team performance-forecasting model based on Artificial Neural networks that is capable of forecasting soccer players' performance depending on teams' history and behavior in previous matches as an input. The proposed model was trained and tested using a dataset including the features of Egypt Telecommunications club 15 years soccer team participating in the Egyptian Football Association Youth Dorian. Simulation results indicated that the proposed model could be classified as a stable predication model especially for soccer team's status and performance, achieving high accuracy rate up to 95%.

نموذج للتنبؤ بأداء فريق لكرة القدم عن طريق الشبكة العصبية الاصطناعية

*مبروكة امحمد الشيباني و جازية مفتاح امشهر و هناء محمد الهماي

قسم الحاسوب - كلية العلوم - جامعة سرت، ليبيا

الكلمات المفتاحية:

الشبكة العصبية الاصطناعية
نموذج للتنبؤ

المخلص

تم مؤخرا في مجال تطور العلم المتعلق بالشبكات العصبية الاصطناعية بذل الكثير من الجهود للتنبؤ بنتائج الأحداث الرياضية. ونظراً لمنظور التوقعات أو التنبؤ فان وجود المنافسة يقدم تحديات نموذجية معينة تحد بدورها من إمكانية تطبيق التقنيات القياسية. تهدف هذه الدراسة إلى إنشاء نموذج للتنبؤ بأداء فريق كرة القدم على أساس اصطناعي عصبي قادر على التنبؤ بأداء لاعبي كرة القدم اعتماداً على تاريخ الفريق وسلوكه في المباريات السابقة كمدخل في عملية التنبؤ. تم تطبيق النموذج المقترح واختباره باستخدام مجموعة بيانات من نادي كرة القدم المصري لفترة 15 عاماً لكرة القدم في دوري الشباب المصري لكرة القدم دوريان. أظهرت نتائج المحاكاة أنه يمكن تصنيف استخدام النموذج المقترح كنظام للتنبؤ بأداء فريق كرة القدم كنموذج تنبؤ مستقر خاصة لحالة وأداء فريق كرة القدم، مما يحقق معدل دقة عالية يصل إلى 95%.

Introduction

Many different types of sports are existed for great human entertainments. These sports raise challenges between different teams, clubs and the public besides countries themselves. Soccer is rounded witch, once you played or watched a soccer match you will be addicted to it. Soccer is the queen of not only team sports but also all sports all over the world. Soccer is a sport that is played on a large field divided into two halves [1] [2]. Goals are located on opposite sides of the field, and each team tries to kick the ball through the other team's goal while defending its own goal.

In official games, the teams have 11 players per side, one of whom is a goalkeeper whose sole responsibility is to guard the goal. The

goalkeeper may touch the ball with any part of his body, while the other players cannot touch it with their hands and arms. Players typically hit the ball with their feet, knees, and head. Frequently, soccer is played informally where there are fewer players on each team and there may not be a goalie [1] [3].

Official games run for 90 minutes but informal games run for as long as the players feel like playing. At the end of a game, the team that has scored the most goals wins, and the game is declared a tie if both teams scored an equal number of goals [4] [5].

On each team, players that run around near the enemy's goal are designated as forwards, and they are primarily responsible for

*Corresponding author:

E-mail addresses: malshebane@yahoo.com, (J. M. Amshaher) jazamoftah@su.edu.ly, (H. Mohammed) hanaa.elhemali@yahoo.com

Article History : Received 04 October 2020 - Received in revised form 12 May 2021 - Accepted 17 June 2021

scoring. Players that are usually located close to their own goal are designated as defenders, and they are primarily responsible for staying with enemy forwards and preventing them from scoring [1] [4].

Forecasting in soccer plays a major role in our world in two ways. The first is in fans, team managers, sponsors, media and punters who bet on online platforms widespread demand for professional advice regarding the results of sporting events is met by a variety of expert forecasts, usually in the form of recommendations from tipsters [5][6]. In addition, betting odds offer a type of predictor and source of expert advice regarding sports outcomes. The second is purely scientific where experts make use of their knowledge about soccer in enhancing teams performance but this represent only 25% of sports forecasting uses[6] [7].

Due to forecasting perspective, the presence of competition introduces particular modelling challenges, which in turn limit the applicability of standard techniques, e.g. regression and discriminate analysis. So a dedicated forecasting algorithm is needed to get accurate model of the outcomes from different competitive events. Most predictions or forecasts derived from experts or statistical models are more accurate than those made by experts based on informed judgment are. Statistical models may outcome more accurate output than human judgments as they employ objective criteria that lookout against bias and the distinctive analysis of data. However, sometimes-statistical models cannot capture non-quantitative factors. As a result, judgmental forecasting may do a better job, not only in less routine and more uncertain situations, but also in integrating qualitative factors into the forecasting process.

Because, of the increasing number of sport competitions in different mode such as soccer, swimming, ridding, racing, etc. Forecasting of events outcome increases challenges between players, coaches and fans also enhances coaches' game plans through analysing win and lose reasons. Starting from this point through this thesis, we will introduce a prediction model capable of forecasting soccer players' performance using artificial neural networks, based on teams' history and behavior in previous matches as input in the forecasting process. This paper is divided into 5 sections. Section 2 includes a detailed discussion on forecasting in soccer introducing the previous work. Section 3 covers the proposed model. Section 4 describes system analysis, implementations and simulation results. Finally, conclusions and future work are given in section 5.

1. FORECASTING IN SOCCER: RELATED WORK

Through this section we will introduce some of the latest studies in soccer forecasting models as in [8] Jasmine Siwei Xu et. al., were concerned with examining the efficiency of using bookmakers' odds as forecasts of soccer match outcomes of the English Premier League. Base on the results of the profit regression model, they concluded that online bookmakers' odds are efficient forecasts of soccer match outcomes, but even greater efficiency could be achieved by incorporating the impact exudes by the number of yellow cards the home team receives from its last game. However, under the two betting strategy examined, the same model fails to maintain its effectiveness in forecasting the result of soccer match outcomes in the following season.

Constantinou and Fenton [9], despite the massive popularity of probabilistic (association) football forecasting models, and the relative simplicity of the outcome of such forecasts, there is no agreed scoring rule to determine their forecast accuracy. Moreover, the various scoring rules used for validation in previous studies are inadequate since they fail to recognize that football outcomes represent a ranked (ordinal) scale.

Boshnakov et. al., in [10] presented a Weibullinter-arrival times based count process and a copula forecasting model for association football scores. The proposed model was tested against simpler Poisson distribution-based model indicating that an improved fit to data compared to previous models and results in positive returns to betting is achieved by the proposed model.

Corona et. al., in [11] proposed a Bayesian approach for simulating the UEFA Champions League results under alternative seeding regimes. They indicated that the changes in 2015 tended to increase the uncertainty over progression to the knock-out stage, but made

limited difference to which clubs would contest the final.

Koopman and Lit in [12] developed a dynamic multivariate model based on the score of the predictive observation mass function for a high-dimensional panel of weekly match results for the analysis and the forecasting of football match results in national league competitions. Match results from six large European football competitions was considered and validated for a period of seven years for each competition. The results indicated that the proposed model for pairwise counts delivers the most precise forecasts and outperforms benchmark and other competing models.

Egidi & Torelli in [13] developed a broad comparison of four possible Bayesian models, focusing on model checking and calibration and then using Markov chain Monte Carlo replications to explore the predictive performance over future matches. Finally, we can say that predicting the outcome of soccer matches is a very challenging issue. There is a high level of randomness in matches and player's capabilities that make it difficult to obtain consistently accurate predictions. In the next section, a new technique of soccer team performance forecasting based on artificial neural network, will be described in details.

2. ANN SOCCER TEAM PERFORMANCE FORECASTING MODEL

The proposed model clarified in figure 1, mainly consists of an artificial neural network trained by a dataset describing soccer team performances. To start with the proposed model, we will divide it into two parts the first involves dataset while, the second involves the ANN.

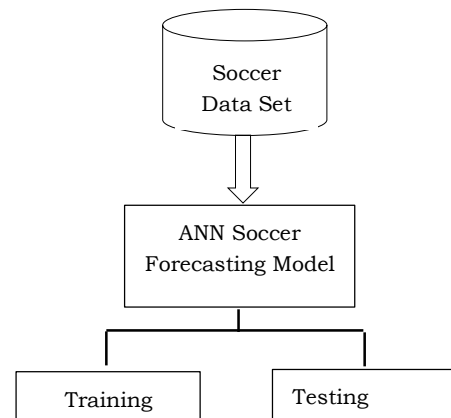


Fig. 1: Soccer Teams Performance Forecasting Model.

2.1. Data set

Dataset was gathered from Egypt Telecommunication Company club 15 year's old soccer juniors team. The dataset was a result of analysing the team performance through 20 matches through the Egyptian Junior Sectors League season 2018/2019.

The team performance was evaluated depending on the number of successful and unsuccessful offensive and defensive tactics in different playground parts. A dataset of 20 matches was used in training and testing the neural network. Table 1 and figures (2 to 7) demonstrate an example of the data containing Individual offensive tactical performance, Playground parts, Successful performances and Unsuccessful performance.

Table 1: A Sample Set of the soccer dataset features.

| # | I.O.T.P | P.p | S.P | U.P |
|---|-----------------------------|-----|-----|-----|
| 1 | Direct short passing | of | 51 | 11 |
| | | mi | 56 | 14 |
| | | de | 35 | 21 |
| 2 | Direct long passing | of | 28 | 16 |
| | | mi | 34 | 12 |
| | | de | 36 | 25 |
| 3 | Receiving and short passing | of | 24 | 5 |
| | | mi | 53 | 28 |
| | | de | 32 | 3 |

I.O.T.P: Individual offensive tactical performance.

P.p: Playground parts.

S.P: Successful performances.
 U.P: Unsuccessful performances.

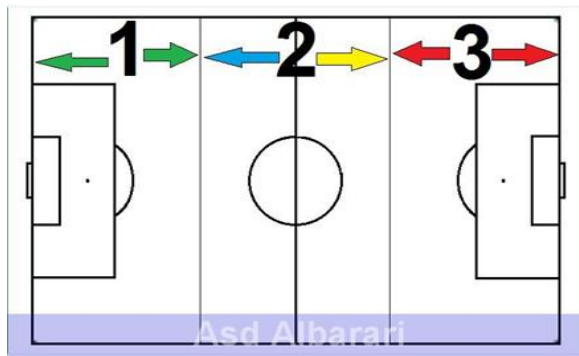


Fig. 2: Soccer Field Parts

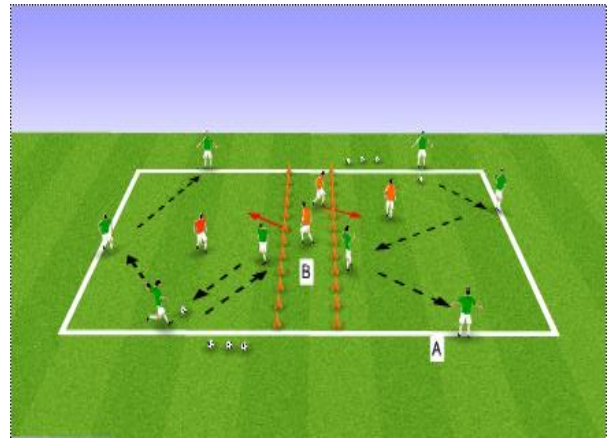


Fig. 6: Receiving and Passing

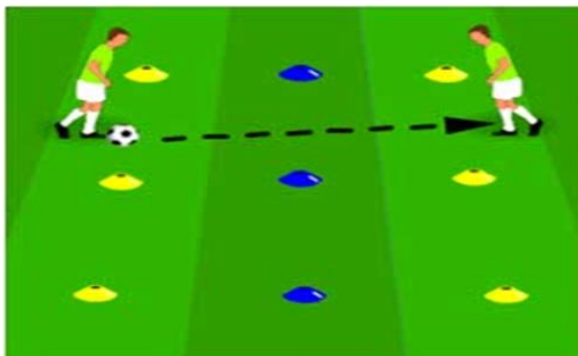


Fig. 3: Direct Short Passing

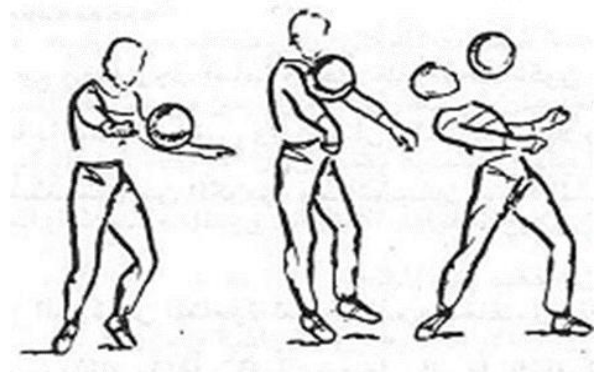


Fig. 7: Ball Absorbing

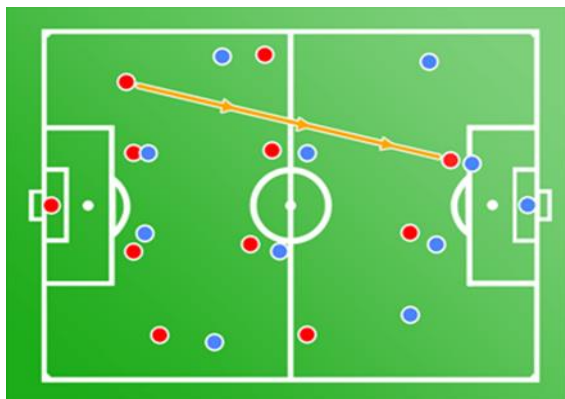


Fig. 4: Direct Long Passing

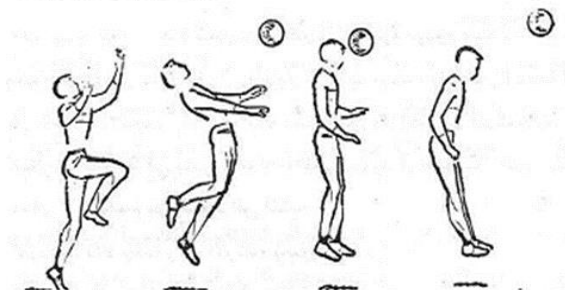


Fig. 5: Head Passing

2.2. ANN Soccer Team Performance Forecasting Model

For team performance forecasting a feed forward neural network with an input layer of 21 neurons and one hidden layer was used. The number of neurons in the layer varied from the 1/2 of the number of inputs to the double number of inputs + 1 which is from 10 to 43 hidden neurons using a LOGSIG activation function.

$$f(x) = \frac{1}{1 + \exp(-x)} \dots\dots\dots(3.1) \quad [14]$$

While, the output layer contains three neurons, one for a loser, one for a winner and the last for draw using a linear activation function. The training algorithm that proved to be the most effective and was used in final computations was the backpropagation algorithm.

$$MSE = \frac{1}{n} \sum (P_e^i - P_a^i)^2 \dots\dots(3.2) \quad [14]$$

Where P_e^i is the success probability of the i th output estimated by the network, P_a^i is the actual probability of the i th output, and n is the total number of proposals in the testing set. In order to start NN predictor phase, we have split the soccer dataset into two parts the first for training, and the rest for network performance testing. A MATLAB environment with the Neural Networks Toolbox used to perform predictor training and testing.

2.2.1. Training

The proposed predictor was ready for data training with 10 matches data records. Network training was performed with the Neural Networks Toolbox in MATLAB environment in two cases first at learning rate (LR)=0.5 and second at LR = 1. In each case attributes were trained and tested through eight classifiers with hidden layer ranging from 10 to 15 hidden neurons at three trails 1000, 1500 and 2000 epochs. Different random initial weights are used in each trial; the network is trained until the mean square error reaches zero.

2.2.2. Testing

The soccer predictor neural network was tested with the rest 10 matches data records. After testing the network each time, the mean square error value (MSE) was registered and the least MSE value will be of the suitable network to our predictor.

3. RESULTS AND DISCUSSION

Figure (8) displays the MSE test results for FFNN predictor at different number of hidden neurons ranges from 10 to 15 neurons not from 10 to 43 as we mentioned in the previous chapter as the MSE of the 16 to 43 hidden neurons predictors gave the same number which make incomparable with others. So we had to eliminate their result and stick to two cases of LR at each case predictor with 10 to 15 hidden neurons were trained at 1000,1500 and 2000 epochs.

Case I: at learning rate = 0.05

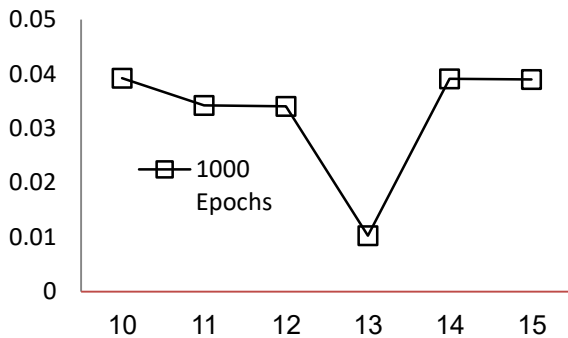


Fig. 8: MSE results for LR=0.5 at 1000 epochs.

Figure (8), illustrated that in the network trained at 1000 epochs the least MSE value is 0.01031 with a 13 hidden neurons, while the maximum MSE value is 0.039226 appeared at 10 hidden neurons.

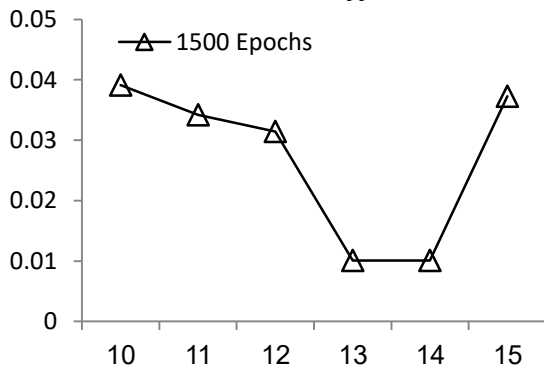


Fig. 9: MSE results for LR=0.5 at 1500 epochs.

Figure (9), illustrated that in the network trained at 1500 epochs the least MSE value is 0.010094 with a 14 hidden neurons, while the maximum MSE value is 0.039121 appeared at 10 hidden neurons.

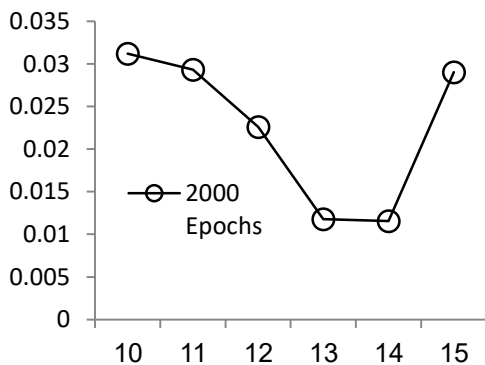


Fig. 10: MSE results for LR=0.5 at 2000 epochs

Finally, Figure (10), illustrated that in the network trained at 2000 epochs the least MSE value is 0.011551 with a 14 hidden neurons, while the maximum MSE value is 0.031211 appeared at 10 hidden

neurons.

Finally, it is determined that the least MSE value is 0.010094 appeared at a predictor with 14 hidden neurons trained at 1500 epochs.

Case II: at learning rate = 1

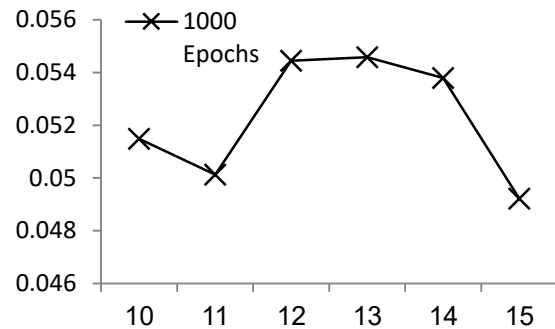


Fig. 11: MSE results for LR=1 at 1000 epochs

Figure (11), illustrated that in the network trained at 1000 epochs the least MSE value is 0.04921 with a 15 hidden neurons while, the maximum MSE value is 0.054578 appeared at 13 hidden neurons.

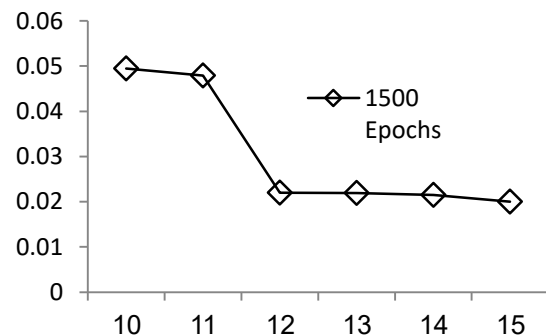


Fig. 12: MSE results for LR=1 at 1500 epochs

Figure (12), illustrated that in the network trained at 1500 epochs the least MSE value is 0.020015 with a 15 hidden neurons while, the maximum MSE value is 0.049465 appeared at 10 hidden neurons.

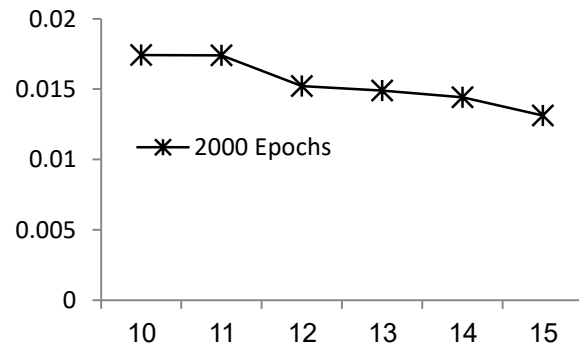


Fig. 13: MSE results for LR=1 at 2000 epochs

Finally, Figure (13), illustrated that in the network trained at 2000 epochs the least MSE value is 0.013125 with a 15 hidden neurons, while the maximum MSE value is 0.017422 appeared at 10 hidden neurons. We can now also determine that the least MSE value at LR=1 is 0.013125 appeared at a predictor with 15 hidden neurons trained at 2000 epochs.

Finally, it is included that the MSE value produced by 14 hidden neurons predictor trained at 1500 epochs at LR =0.5 was the least MSE among all predictor in case I and II. From the previous figures, the researcher can clarify that the prediction error is so small and nearly equal zero. Also, since it is possible to obtain it with a smaller neural network (21 Input- 14 Hidden- 3 Output) with LR=0.5 trained at 1500.

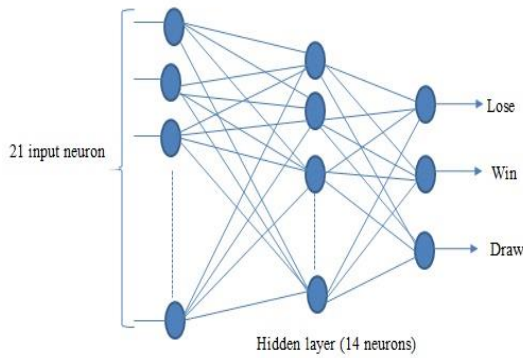


Fig. 14: The proposed model neural network

Table (2) represents the confusion matrix of the soccer predictor, where there was only 1 mispredicting value as the desired output was draw where the real one was win. So the researcher can finally conclude that the proposed model provided 95 % accuracy in predicting the soccer team’s status and performance evaluation through the testing stage which is considered a good result.

Table 2: The Confusion Matrix of Soccer Predictor

| Output results | Win | Lose | Draw |
|----------------|-----|------|------|
| Desired output | | | |
| Win | 14 | 0 | 0 |
| Lose | 0 | 3 | 0 |
| Draw | 1 | 0 | 2 |

4. Conclusion

Forecasting in sports plays a major role in our world in two ways. The first is in fans, team managers, sponsors, media and punters who bet on online platforms widespread demand for professional advice regarding the results of sporting events is met by a variety of expert forecasts, usually in the form of recommendations from tipsters. Starting from this point through this study the researcher introduced a prediction model capable of forecasting soccer players’ performance using artificial neural networks, based on teams’ history and behaviour in previous matches as input in the forecasting process. The researcher has implemented the proposed model using MATLAB environment with the Neural Networks Toolbox used to perform predictor training and testing. Experimental results from testing the proposed model were particularly encouraging, showing that this model is capable of achieving accurate forecasting results up to 95% through the testing phase.

References

[1]- B. Boulier , H. Stekler , J. Coburn and T. Rankins, (2009). "Evaluating National Football League Draft Choices: The Passing Game". Working Papers 2009-003, The George Washington University, Department of Economics, Research Program on Forecasting.

[2]- B.Boulier, L. Bryan, H. Stekler and C. Jason, (2010). "Evaluating National Football League draft choices: The passing game". International Journal of Forecasting, Elsevier, Vol. 26(3), pp. 589-605.

[3]- M. Cengiz and N. Murat, (2012). "A Bayesian Computation for the Prediction of Football Match Results Using Artificial Neural Network". International Journal of Scientific Knowledge, Vol. 1(2), pp 25-31.

[4]- W. Young and G. Weckman, (2008). "Evaluating The Effects of Aging for Professional Football Players in Combine Events Using Performance-Aging Curves". International Journal of Sports Science and Engineering Vol. 2(3), pp. 131-143.

[5]- J. Goddard and I. Asimakopoulos, (2004). "Forecasting Football Results and the Efficiency of Fixed-Odds Betting". Article first published online: 21 JAN 2004 DOI: 10.1002/for.877, Journal of Forecasting Vol. 23(1), pp. 51–66.

[6]- M. Dixon and S. Coles, (1997). "Modeling Association Football Scores and Inefficiencies in The Football Betting Market". Applied Statistics Journal, Vol. 46, pp. 265–280.

[7]- M. Spann and B. Skiera (2008), "Sports Forecasting: A Comparison of the Forecast Accuracy of Prediction Markets, Betting Odds and Tipsters" Journal of Forecasting, Vol. 28(1), pp. 55-72.

[8]- J. Siwei Xu, (2011). "Online Sports Gambling: A Look in To The Efficiency of Bookmakers' Odds as Forecasts In The Case Of The English Premier League".

[9]- A. Constantinou, N. Fentona and M. Neil, (2012). "Pi-football: A Bayesian Network Model for Forecasting Association Football Match Outcomes". Working Paper, pp. 1-35.

[10]- Boshnakov, G., Kharrat, T., & McHale, I. G. (2017). A bivariate Weibull count model for forecasting association football scores. *International Journal of Forecasting*, 33(2), 458-466.

[11]- Corona, F., Forrest, D., Tena, J. D. D., & Wiper, M. (2019). Bayesian forecasting of UEFA Champions League under alternative seeding regimes. *International Journal of Forecasting*, 35(2), 722-732.

[12]- Koopman, S. J., & Lit, R. (2019). Forecasting football match results in national league competitions using score-driven time series models. *International Journal of Forecasting*, 35(2), 797-809.

[13]- Egidi, L., & Torelli, N. (2020). Comparing Goal-Based and Result-Based Approaches in Modelling Football Outcomes. *Social Indicators Research*, 1-13.

[14]- M. Hematinezhad, M. Gholizadeh, M. Ramezaniyan, S. Shafiee, and A. Zahedi, (2011). "Predicting the Success of Nation in Asian Games Using Neural Network". Sport SPA Journal, Vol. 8(1), pp. 33-42.