Article

Predicting Wins, Losses and Attributes' Sensitivities in the Soccer World Cup 2018 Using Neural Network Analysis

Amr Hassan [1],* , **Abdel-Rahman Akl [2]** , **Ibrahim Hassan [3]** and **Caroline Sunderland [4]**

[1] Department of Sports Training, Faculty of Sports Education, Mansoura University, Mansoura 35516, Egypt
[2] Faculty of Physical Education-Abo Qir, Alexandria University, Alexandria 21913, Egypt;
    abdelrahman.akl@alexu.edu.eg
[3] Faculty of Physical Education, Zagazig University, Zagazig 44519, Egypt; ibrahimhassan@fopem.zu.edu.eg
[4] Department of Sport Science, Sport, Health and Performance Enhancement Research Centre,
    School of Science and Technology, Nottingham Trent University, Nottingham NG11 8NS, UK;
    caroline.sunderland@ntu.ac.uk
* Correspondence: amrahh@mans.edu.eg; Tel.: +43-65-0687-0621

check for updates

Abstract: Predicting the results of soccer competitions and the contributions of match attributes, in particular, has gained popularity in recent years. Big data processing obtained from different sensors, cameras and analysis systems needs modern tools that can provide a deep understanding of the relationship between this huge amount of data produced by sensors and cameras, both linear and non-linear data. Using data mining tools does not appear sufficient to provide a deep understanding of the relationship between the match attributes and results and how to predict or optimize the results based upon performance variables. This study aimed to suggest a different approach to predict wins, losses and attributes' sensitivities which enables the prediction of match results based on the most sensitive attributes that affect it as a second step. A radial basis function neural network model has successfully weighted the effectiveness of all match attributes and classified the team results into the target groups as a win or loss. The neural network model's output demonstrated a correct percentage of win and loss of 83.3% and 72.7% respectively, with a low Root Mean Square training error of 2.9% and testing error of 0.37%. Out of 75 match attributes, 19 were identified as powerful predictors of success. The most powerful respectively were: the Total Team Medium Pass Attempted (MBA) 100%; the Distance Covered Team Average in zone 3 (15–20 km/h; Zone3_TA) 99%; the Team Average ball delivery into the attacking third of the field (TA_DAT) 80.9%; the Total Team Covered Distance without Ball Possession (Not in_Poss_TT) 76.8%; and the Average Distance Covered by Team (Game TA) 75.1%. Therefore, the novel radial based function neural network model can be employed by sports scientists to adapt training, tactics and opposition analysis to improve performance.

Keywords: football; match analysis; neural network; sport analytics

1. Introduction

The continual innovation, development and use of technology within sport has resulted in large data sets from soccer. World Cup matches are available to analyze and explain player's and team's performance. The increased accessibility of detailed match data is largely due to the progression made in semi- and fully-automated tracking [1–4]. Noticeably, the technical and tactical match performance of professional soccer players, has been shown to be affected by several different contextual, situational and positional attributes [5–10]. Thus, the continuous movement dynamic and the corresponding

movement within the context of the match produces thousands of different values that reflect in total what the player, players, team or teams do together as a whole during the game.

In the last decade, there has been a plethora of research describing and analyzing the match performance data from soccer World Cups using several approaches such as multivariate analyses and machine learning [1–4,11–15], passing networks based on space, time and the multilayer nature of the game [16] or based on spatial and temporal entropy related to football teams and their players by means of a pass-based interaction [17] and social network analyzes to study the interaction between a player and their teammates (for example a ball passing network) through graph theory to assess the structural and topographical characteristics of personal interactions between team members [18]. This type of descriptive research provides important information that can be used to improve training and adapt tactics, however analyses, such as machine learning, can identify performance indicators, whether physical or technical that may predict what will occur during the match [19–22]. Furthermore, for performance attributes and events to be meaningful, they need to be relevant to the team's success such as the winning of matches [23]. Such analysis needs an amount of data which is an outcome from different sensors, cameras and analysis systems and needs an intelligent solution to invest in these big data which are usually linear and nonlinear in nature. In context, the linear data structure is a collection of datum items having similar types and the non-linear data structure is a type of data structure whereby datum elements are present at various levels [24].

In this regard, the use of artificial neural networks (ANNs) has increased considerably in recent years for the identification, classification and prediction of performance within soccer [25–32]. In particular, predictability has become possible according to players information through the analyses of data from past-recorded matches. For instance, the machine learning algorithms that have been employed within soccer research are K-Nearest neighbors [33,34], Neural networks [25–28,30,35,36], Decision Trees [37–39] and Bayesian networks [29,31,32,40,41].

There is limited research that has employed neural network approaches to data from the FIFA World Cup 2018. Specifically, Correa, et al. [42] used statistical simulations to analyze the progressive results of the tournament and predict the teams most likely to play in the finals and the team with the highest probability of becoming champions. In addition, Groll, et al. [43], investigated three different modelling approaches applied to data from the World Cups between 2002 and 2014 to predict performance in the 2014 World Cup. A hybrid random forest model, which combines random forests with Poisson ranking methods, was the best performing model. Lastly, Clemente [44], assessed the associations between different neural network measures such as group clustering or dyad reciprocity and performance outcomes (wins, goals scored and conceded and shots) to detect the variations within network measures between winners and losers. The results demonstrated that studied general networks were not sensitive enough to predict overall match outcome. Thus, as there is a dearth of research using player's and team's behaviors, the application of ANNs to predict World Cup performance is warranted.

The ease of availability of attributes/data from matches has meant that research using these performance indicators has become frequent. Artificial neural networks have become increasingly popular in the last few years to model and estimate characteristics that cause a team to win or lose a match, or predict the score of a particular match [22,45,46]. However, the modelling of match data to date has not been able to understand and interpret the vast amount of data available and simplify it in an objective manner. In addition, the difficulty of determining the actual contribution value of the complex attributes on match results, has meant that transferring such match attributes into the training process has been neglected so far in sport-related scientific literature.

Therefore, the current study is the first to provide an application model of a learning neural network that can use official data analyzed by FIFA. The approach employed has not previously been used in the literature for match prediction. Further, there is novelty in the ability to determine the percentage effect of the matched attributes on the results objectively using the ANN model.

The current study aimed to suggest a different approach to process the big data resulting from the use of devices, technological tools, and sensors to track and analyze football players 'performance. It was done by predicting the World cup 2018 match outcomes based on the match attributes that have been collected using FIFA technological tools during the match using a suggested neural network model. The second aim was to estimate the sensitivity of the match attributes affecting the chances of winning or losing using the neural network model.

2. Methods

The concept of forecasting is to train the model to predict the match outcome based on the match attributes that have been collected after the match has finished. So, the study is carried out through the following steps:

1. The match analyses.
2. The ANN designing, training, testing.
3. The match attributes' sensitivities analysis.

2.1. Study Sample

Fifty-seven matches were officially analyzed by FIFA during the World Cup Finals 2018 using a recently validated TRACAB®system (position accuracy 0.09 m RMSE; total distance difference 0.42 ± 0.60%) [47]. According to FIFA, the analysis provider used a real-time optical tracking system (25 frames per second including the player position coordinates X, Y in the playground as Pythagorean Theorem and the frame time) to collect the match attributes that constitute the input parameters of the ANN prediction model. Such a system has been successfully used in previous literature [1,13,15,48]. Seventy-five match attributes (from the official match score sheet) were selected to use in the ANN model.

Those match attributes were from the physical indicators (distances covered, speed and sprints ... etc.) and technical indicators (number of passes, passes received, pass success, fouls, shots ... etc.) and are shown in the appendix. All matches that finished with 1 side wining were included (n = 55), matches that were drawn (n = 9) were not considered (n = 7) or were randomly selected for use in the validation phase (n = 2). The goals scored and conceded were not included due to their direct correlation with match outcome, which negatively affects the validity of the results. The match attributes nominal data were transformed into numerical values to be understandable for the ANN model, including technical and physical performance-related variables (see Appendix A), to quantify the match performance of the teams [46].

2.2. The ANN Model Construction

A commercial artificial neural network software NeuroDimension® [49], was used to predict and classify both teams in a match into two classes either win (class 1) or loss (class 2). The artificial neural network applied was a radial basis function (RBF neural network) [49]. Radial basis functions originated as a variant of the artificial neural network in 1988. Radial basis functions usually have 3 layers of the neural network; the input layer, a hidden layer that performs the functions of radial stimulation while the output layer applies the weighted hidden unit outputs. The input in the RBF network is nonlinear while the output is linear. So, RBF networks are able to model complex designations [50].

The construction of the RBF model relies on the design of the neural network elements, their arrangement, as well as the parameters that affect those [51]. The model contained three layers: an input layer, a hidden layer of processing elements (PEs) and an output layer as shown in Figure 1.

The network standardizes the input data values to obtain a high accuracy using the root mean square error evaluation method (RMS) [52]. The Root mean square error (RMSE) is the square root of

the mean square error (MSE). In the input layer, all match attributes are selected as input data while the desired data are coded to be 0 for a loss and 1 for the win.
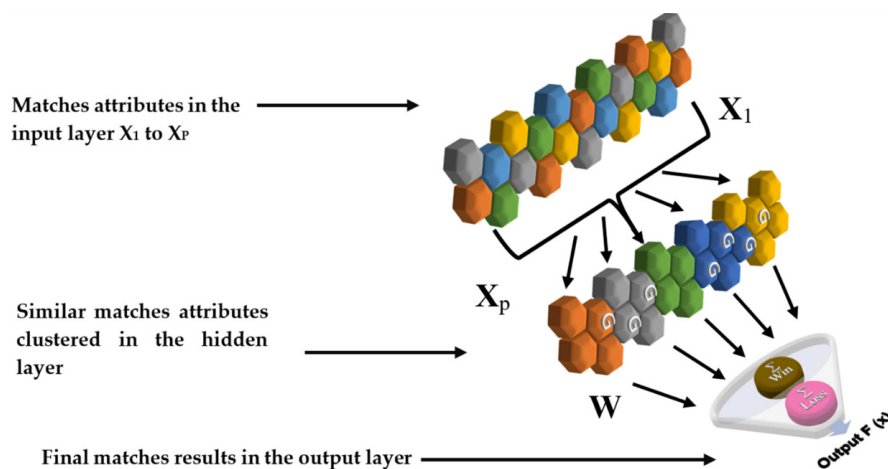
Figure 1. Radial Basis Function (RBF) network. G is Gaussian function, W is weights.

The ANN model cleans the input data automatically. The purpose of data cleaning is to remove noise, inconsistent data and errors in the training data [53]. The number of the input Processing Elements (PEs) was set to 75 (equal to the match attributes) and the output PEs was one (representing either win or lose). The weight regularization option was "weight decay," with values often on a logarithmic scale between 0.01 and 0.001 as a default option.

The model later synthesizes the input data to create the RBF clusters using an unsupervised algorithm (Gaussian classifier) optimization process. Each cluster represents one RBF neuron and one type of playing variable. This neuron compared the similarity of its center with the input vector, with the amount of similarity representing the algorithm input value [54–58]. The competitive learning setting for each single PE is set as Standard Full. This allows the single-layer linear network to group and represent the match attributes in lie neighborhoods of the input space which are represented by a single output PE calculated using the weights according to Formula (1). The metric measures method was set as Euclidean [24].

$$W_{2^*}(n+1) \; = \; W_{2^*}(n) + \eta\,(X(n) - W_{2^*}(n)) \tag{1}$$

where (*W*) is a weight from the synapse component that feeds the Gaussian Axon, (η) is the learning rate for each weight, (*n*) is the gradient and (*i**) is the PE that wins the competition.

The weights update mode is set to batch mode of the supervised learning. Data is randomly separated into three groups as training (65%), cross-validation (15%) and testing data sets (20%). The results were corroborated to each other by making the training; cross-validation and testing data sets include the same data ranges [19,59].

The increasing Mean Square Error (MSE) option was selected as termination during the cross-validation. This means the training iterations will automatically stop when the error value falls under the determined threshold value. Linear TahnAxon option was selected as an activation function interrelating the input and output parameters. The Levenberg–Marquardt (LM) algorithm was chosen as the learning rule to minimize the MSE of the ANN model.

2.3. The ANN Training and Testing Procedures

The model was activated after its network building process was completed and first worked on the training data. After the training process phase, the test phase then started. The main aim of the test processing was determining the validity of the RBF model by activating it with the testing dataset.

During the testing, the probe configuration was set as a prediction confusion matrix, providing lists of the real results against the prediction results for each class [60].

To show how the detection threshold in the RBF model affects detections versus false alarms the Receiver Operating Characteristic (ROC) matrix was chosen [49]. The RBF model estimates the error value of the model panel and minimizes the error as much as possible through the predefined amount of repetition and randomized weight values. The test report shows the percentage of success of the neural network model in identifying correctly the cases of a win or loss. Finally, sensitivity to the mean for all matches' attributes were calculated to assess the effect of those attributes on the chance of winning or losing.

2.4. Statistical Analysis

To assess the training and testing process, the RBF training and testing report shows the RMSE which is the average of the errors between the target prediction and the real win or loss results. Additionally, Pearson correlation coefficients between predicted and target results were completed.

3. Results

The differences between real and predicted loss and wins (RMSE) were 2.9% in the training phase and 0.37% in the testing phase. Differences in the RMSE in the loss and win RBF prediction model represent the different correct prediction percentage for each estimated probability versus the RBF model threshold. Pearson correlation coefficients between real and predicted loss and win were r = 0.72 for both. The correct percentages for the prediction model are 72.7% and 83.3% for loss and win, respectively. The estimated probability of positive prediction was used to evaluate the model creativity. The RBF predictor model makes a single prediction about each example win or loss.

Receiver Operating Characteristic (ROC) matrices are used to show how changing the detection threshold affects the loss and win prediction detections versus false alarms. If the threshold is set too high then the RBF model will miss many detections. Conversely, if the threshold is set low then there will be too many false alarms (Figure 2).
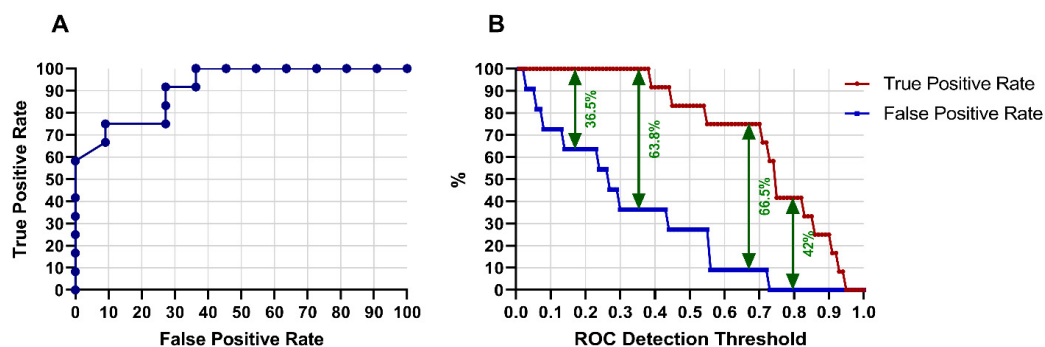
Figure 2. (**A**): Illustrates receiver operating characteristic (ROC curve) and accuracies through thresholds after 100 training iterations improvements. For the testing sets, with pedestrial/non-pedestrial threshold = 1, RBF neural network predicts correctly 91%. (**B**): Comparison between the true and false positive prediction rate.

False Positive Rate is the percentage of false positives of those samples with an actual value of zero. True positive rate represents the percentage of true positives of those samples with an actual value of one. The false discovery rate is the percentage of false positives of those samples predicted as one. The best ROC threshold was 0.70 where there were no more false positive rates and the maximum true positive.

Area under the ROC curve is the sum of the values that are displayed under the ROC chart and represents the overall performance of the RBF model. A measurement of 0.0 would be considered a poor model with no prediction power, while a measurement of 1.0 would be considered a best

model [49,61]. As shown in the Figure 2A the area under curve is 0.91 for all positive rates where the acceptable true positive rate is 76.5% and false positive rate is 10%. Figure 2B shows the ROC area under curve for true/false positives rate.

Based on the high correlation between the real and the predicted results, the radial basis function neural network (RBFNN) sensitivity analysis was produced by the software to assess the effect of match attributes on the loss or win prediction. These results indicated a valuable different contribution for each attribute on the wins and losses. (Figure 3).
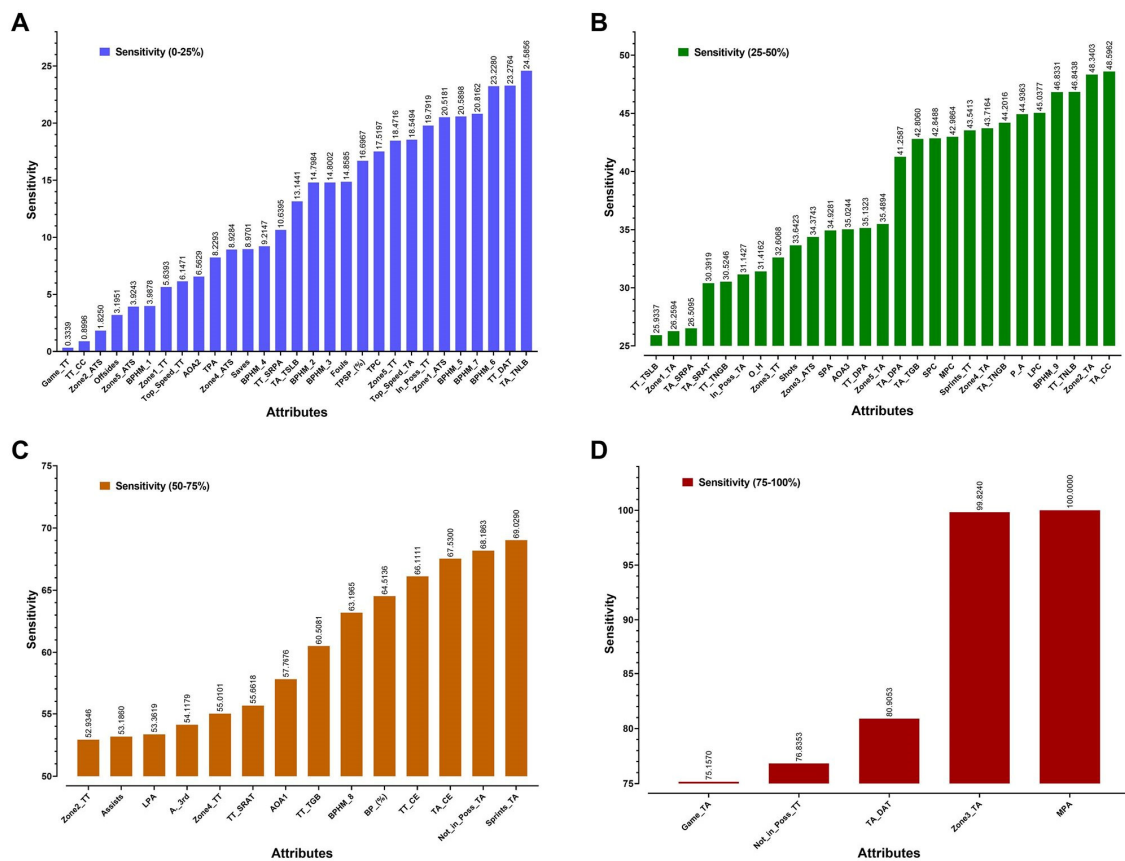
Figure 3. Contribution of match attributes to results (ANN sensitivity analysis); (**A**): sensitivity from 0 to 25%, (**B**): sensitivity from 25 to 50%, (**C**): sensitivity from 50 to 75%, (**D**): sensitivity from 75 to 100%.

4. Discussion

The key finding is that the prediction of match results based on ANN models provides an intelligent method to improve either the match events or the tactical training processes based on the relationship between game attributes and the outcomes in soccer. Previous literature has mainly focused on results prediction based on the location of the match (home or away), prediction dependent on a very low amount of match attributes or based on correlation coefficients between some attributes [45].

The present study aimed to present a method for the real application of such results in the training process and competition process using a wide range of match attributes associated with outcome. To the best of our knowledge, this is the first attempt to objectively quantify the match attributes sensitivity to improve match results or training procedures. ANN approaches for match results prediction have been previously used in research with either limited game attributes or using attributes that are not outcomes from the analysis, for example players' injuries, weather conditions etc. employing ANN and logistic regression (LR) techniques [45]. In contrast with this research, the current study employed an RBF neural network model and successfully predicted the team's results as either loss or win 72.7% and 83.3% of the time, respectively.

The prediction percentage is comparable with the results of other studies. Igiri and Nwachukwu [62], used a data mining tool (Rapid Miner) to predict the results of soccer matches with 75.04% prediction accuracy when the ANN technique and logistic regression were used to build the model. Others have used the Markov Process Model and their results yielded an accuracy of 56% [63]. While the previous studies discuss several merged methods to predict and classify team results, they did not discuss the Estimated Positive Probability, which explains the probability that the network output equaling the ROC Detection Threshold will have a desired output of one. The ROC curve is used to examine the tradeoff between the detection of true positives while avoiding false positives. In Figure 2 the RBF model has the best threshold of 0.70 where the best detection of true positive rate was 76.5% versus false positives of 10% and the best maximum difference rate was 66.5%.

The relationship between sports results and various data elements are directly affected by different tactics and techniques factors. The sensitivity analysis based on ANN have been suggested to predict the results based on available data. Therefore, the employment of the ANN will provide a more objective and credible method to predict performance compared with relying solely on the experience and instincts of experts, which include a high error ratio, or relying on basic statistical data. The ANN sensitivity analysis for players performance and matches attributes will result in more reliable predictions that support the coach and trainer decision-making [64].

The success of the ANN model in match results prediction and in establishing a relationship between the attributes and matches results correspond to the findings of some previous studies except with the value of the attributes' sensitivity [65–67]. The second contribution of using the ANN model is the possibility to draw the effect of each attribute on the results of the matches as shown in Figure 3. The attributes were divided into four groups based upon how powerful they were in predicting success (≤25%, ≤50%, ≤75% and ≤100%). The most powerful attributes for predicting success were the Total Team Medium Pass Attempted (MBA) at 100%, the Average of Distance Covered by the Team where the player's speed ranged between 15 and 20 km/h (Zone3_TA) at 99%, the Team Average ball delivery into the attacking third (TA_DAT) at 80.9%, the Total Team Covered Distance without the Ball (Not_in_Poss_TT) at 76.8% and the Average Distance Covered by the Team (Game_TA) at 75.1%.

These results indicate that teams who employ medium distance passes to maintain possession are more successful. This finding is consistent with previous research that confirmed that the teams who control the game by passing, can patiently move the ball to search for defensive weaknesses [68]. In addition, our results are consistent with previous studies conducted on the same competition, where it was stated that the teams that qualify for the knockout stage are distinguished by their possession of the ball with their direct play and handing the ball in the offensive third of the field [46]. Furthermore, our analysis using artificial neural networks indicated that the distance covered by the team without possession of the ball and the average distance covered by the team as a whole were key attributes for success reaching 76.8% and 75.1%, respectively. This is also consistent with previous research that confirmed that successful teams covered a greater distance at a low speed and maintained possession of the ball with moderate passes [46]. Much of the previous research has focused primarily on high-intensity running with limited attention to low-intensity activity [69,70], perhaps they consider it as a measure of players' physical abilities [71]. Such ANN analysis allows the rationalization of training and provides a deeper understanding of the management of the game, technically and strategically.

The study has limitations as the isolated data from the physical parameters do not allow the control for situational variables, and a drawback of the method presented in this study is the time needed to obtain the match data. The FIFA determination of the match attributes might only be available sometime after the completion of the match or competition. However, the latest TRACAB system does provide data in real-time so if this can be accessed, by the sports scientist, the ANN could be employed the same day allowing feedback between matches. Therefore, the RBF model methods presented for prediction and sensitivity analysis could simplify and improve the training process significantly.

As soon as the match data is available, the calculation of the RBFNN is done in a very short time. Moreover, calculated attributes contributions will be estimated accurately. However, the error due to the sensitivity model will be acceptable when we consider the effective attributes in terms of the match results. Another challenge might be the difficulty for trainers to use ANNs. Performance analysts and statisticians must be qualified to use such software created by sport scientists in cooperation with computer specialists to optimize the benefits available to enhance soccer performance. Comparison with the few papers that have focused on such an approach is not possible as the rigorous statistical analysis of the important variables was not similar. Future research needs to assess and determine the efficacy of the ANN presented in enhancing performance and to employ a multivariate and integrative approach with physical, positional, technical and tactical variables.

5. Conclusions

In the current study, we provide an approach to predict the win, loss and the sensitivity of match attributes from the soccer World Cup 2018. Using the commercial software, NeuroDimension®(NeuroDimension, Gainesville, F.L., USA) [49], a supervised neural network based on Radial basis function using 75 attributes presented in the official match report by FIFA for the World Cup was produced. The match attributes were selected and trained within the network training model. The match results were selected as a target output. After the neural network training phase finished the test phase was performed. About 72.7% of the loss results could be predicted as desired results, while 83.3% could be predicted as desired win results. In order to optimize the benefit of analysis of the attribute, the match attributes' sensitivities were analyzed. The original neural network model successfully presented the importance of each attribute for the match results. Summarizing, the study clearly demonstrates that the novel neural network can be used to predict results of matches, effective attributes and consequently useful to support trainers' and coach's decisions during the match and improve tactical training in team soccer and similar types of sports. It offers an objective approach to manage training and matches pertaining to match attributes' sensitivities and thus avoids coach's self-impressions.

6. Practical Applications

This study provides a deep insight into the effects of match attributes on the outcome of matches. It is also possible to understand the interrelationship between different match variables that are difficult to find statistically. The medium passes and delivery into the attacking third in addition to the movement without the ball and the running at medium speeds have proven important attributes that determine the results of matches, in agreement with similar studies. Depending on the important matching features as strong or weak forecasting factors, new tactical play methods can be identified and created, or against the perceptions presented by the competitor, to ensure that the best physical and technical capabilities of the players are employed. Considering the limitations of the study, it is possible to monitor the important matching features in the training sessions or matches and re-evaluate the nature of the performance and tasks assigned to the players in order to develop the style of play. Results must be taken with caution because of the small sample size from the FIFA World Cup. This restriction also imposes further research to consider proposing other tools and means that contribute to exploring and examining play situations and their relationship to physical performance and match characteristics with strong or weak influence on the outcome of matches alike.

Appendix A. Abbreviations of Dataset

Attributes	Abbreviation
Total Team delivery into the attacking third	TT_DAT
Total Team solo runs into the attacking third	TT_SRAT
Total Team delivery into the penalty area	TT_DPA
Total Team solo runs into the penalty area	TT_SRPA
Total Team tackles gaining the ball	TT_TGB
Total Team tackles not gaining the ball	TT_TNGB
Total Team tackles suffered losing the ball	TT_TSLB
Total Team tackles suffered not losing the ball	TT_TNLB
Total Team clearances completed	TT_CC
Total Team clearances attempted	TT_CE
Team Average delivery into the attacking third	TA_DAT
Team Average solo runs into the attacking third	TA_SRAT
Team Average delivery into the penalty area	TA_DPA
Team Average solo runs into the penalty area	TA_SRPA
Team Average tackles gaining the ball	TA_TGB
Team Average tackles not gaining the ball	TA_TNGB
Team Average tackles suffered losing the ball	TA_TSLB
Team Average tackles suffered not losing the ball	TA_TNLB
Team Average clearances completed	TA_CC
Team Average clearances attempted	TA_CE
Activity Time Spent Zone 1	Zone1_ATS
Activity Time Spent Zone 2	Zone2_ATS
Activity Time Spent Zone 3	Zone3_ATS
Activity Time Spent Zone 4	Zone4_ATS
Activity Time Spent Zone 5	Zone5_ATS
Distance Covered Team Total all Zones	Game_TT
Distance Covered Team Total Zone 1	Zone1_TT
Distance Covered Team Total Zone 2	Zone2_TT
Distance Covered Team Total Zone 3	Zone3_TT
Distance Covered Team Total Zone 4	Zone4_TT
Distance Covered Team Total Zone 5	Zone5_TT
Distance Covered Team Average all Zones	Game_TA
Distance Covered Team Average Zone 1	Zone1_TA
Distance Covered Team Average Zone 2	Zone2_TA
Distance Covered Team Average Zone 3	Zone3_TA
Distance Covered Team Average Zone 4	Zone4_TA
Distance Covered Team Average Zone 5	Zone5_TA
Team Total Top speed	Top_Speed_TT
Team Total Sprints	Sprints_TT
Team Average Top speed	Top_Speed_TA
Team Average Sprints	Sprints_TA
Total Team Long Pass Completed	LPC
Total Team Long Pass Attempted	LPA
Total Team Medium Pass Completed	MPC
Total Team Medium Pass Attempted	MPA
Total Team Short Pass Completed	SPC
Total Team Short Pass Attempted	SPA
Total Team Pass Completed	TPC
Total Team Pass Attempted	TPA
Total Team Passing success percentage	TPSP_(%)
Ball Possession	BP_(%)
Ball Possession Heat Map Defense Field Left	BPHM_1
Ball Possession Heat Map Defense Field Middle	BPHM_2
Ball Possession Heat Map Defense Field Right	BPHM_3
Ball Possession Heat Map Middle Field Left	BPHM_4
Ball Possession Heat Map Middle Field Middle	BPHM_5
Ball Possession Heat Map Middle Field Right	BPHM_6
Ball Possession Heat Map Attack Field Left	BPHM_7
Ball Possession Heat Map Attack Field Middle	BPHM_8
Ball Possession Heat Map Attack Field Right	BPHM_9
Attack Origin Area Left	AOA1
Attack Origin Area Middle	AOA2
Attack Origin Area Right	AOA3
Total Team Covered Distance Ball Possession	In_Poss_TT
Total Team Covered Distance Not in Ball Possession	Not_in_Poss_TT
Team Average Covered Distance Ball Possession	In_Poss_TA
Team Average Covered Distance Not in Ball Possession	Not_in_Poss_TA
In opposite half	O_H
In attacking third	A._3rd
In penalty area	P_A
Description of Zone Activities	
Zone 1	0–7 km/h
Zone 2	7–15 km/h
Zone 3	15–20 km/h
Zone 4	20–25 km/h
Zone 5	> 25 km/h

References

1. Clemente, F.M.; Couceiro, M.S.; Martins, F.M.; Ivanova, M.O.; Mendes, R. Activity Profiles of Soccer Players during the 2010 world cup. *J. Hum. Kinet.* **2013**, *38*, 201–211. [CrossRef]

2. Clemente, F.M.; Martins, F.M.L.; Kalamaras, D.; Wong, P.D.; Mendes, R.S. General network analysis of national soccer teams in FIFA World Cup 2014. *Int. J. Perform. Anal. Sport* **2015**, *15*, 80–96. [CrossRef]

3. Hongyou, L.; Miguel-Ángel, G.; Carlos, L.-P.; Jaime, S. Match statistics related to winning in the group stage of 2014 Brazil FIFA World Cup. *J. Sports Sci.* **2015**, *33*, 1205–1213.

4. Tuo, Q.; Wang, L.; Huang, G.; Zhang, H.; Liu, H. Running Performance of Soccer Players During Matches in the 2018 FIFA World Cup: Differences Among Confederations. *Front. Psychol. Front. Psychol.* **2019**, *10*, 1–6. [CrossRef] [PubMed]

5. Liu, H.; Gomez, M.A.; Goncalves, B.; Sampaio, J. Technical performance and match-to-match variation in elite football teams. *J. Sports Sci.* **2016**, *34*, 509–518. [CrossRef]

6. Memmert, D.; Lemmink, K.A.P.M.; Sampaio, J. Current Approaches to Tactical Performance Analyses in Soccer Using Position Data. *Sports Med.* **2017**, *47*, 1–10. [CrossRef] [PubMed]

7. Redwood-Brown, A.J.; O'Donoghue, P.G.; Nevill, A.M.; Saward, C.; Sunderland, C. Effects of playing position, pitch location, opposition ability and team ability on the technical performance of elite soccer players in different score line states. *PLoS ONE* **2019**, *14*, e0211707. [CrossRef]

8. Rein, R.; Memmert, D. Big data and tactical analysis in elite soccer: Future challenges and opportunities for sports science. *Springerplus* **2016**, *5*, 1410. [CrossRef]

9. Bradley, P.S.; Lago-Peñas, C.; Rey, E.; Sampaio, J. The influence of situational variables on ball possession in the English Premier League. *J. Sports Sci. J. Sports Sci.* **2014**, *32*, 1867–1873. [CrossRef]

10. Lago-Ballesteros, J.; Lago-Peñas, C.; Rey, E. The effect of playing tactics and situational variables on achieving score-box possessions in a professional soccer team. *J. Sports Sci.* **2012**, *30*, 1455–1461. [CrossRef]

11. Göral, K. Passing Success Percentages and Ball Possession Rates of Successful Teams in 2014 FIFA World Cup. *Int. J. Sci. Cult. Sport* **2015**, *3*, 86–95. [CrossRef]

12. Liu, H.; Gómez, M.-A. Relationships between match performance indicators and match outcome in 2014 Brazil FIFA world cup. In Proceedings of the VIII Congreso Internacional de la Asociacion Espanola de Ciencias del Deporte, Cáceres, Spain, 13–15 November 2014.

13. Ribeiro, G.; Thiengo, C.R.; Gimenes, S.V.; Bradley, P.S. The effects of ball possession status on physical and technical indicators during the 2014 FIFA World Cup Finals. *J. Sports Sci.* **2016**, *34*, 493–500.

14. Rumpf, M.C.; Silva, J.R.; Hertzog, M.; Farooq, A.; Nassis, G. Technical and physical analysis of the 2014 FIFA World Cup Brazil: Winners vs. losers. *J. Sports Med. Phys. Fit.* **2017**, *57*, 1338–1343. [CrossRef]

15. Moura, F.A.; Martins, L.E.B.; Cunha, S.A. Analysis of football game-related statistics using multivariate techniques. *J. Sports Sci.* **2014**, *32*, 1881–1887. [CrossRef] [PubMed]

16. Buldú, J.M.; Busquets, J.; Martínez, J.H.; Herrera-Diestra, J.L.; Echegoyen, I.; Galeano, J.; Luque, J. Using Network Science to Analyse Football Passing Networks: Dynamics, Space, Time, and the Multilayer Nature of the Game. *Front. Psychol.* **2018**, *9*, 1900. [CrossRef] [PubMed]

17. Martínez, J.H.; Garrido, D.; Herrera-Diestra, J.L.; Busquets, J.; Sevilla-Escoboza, R.; Buldú, J.M. Spatial and Temporal Entropies in the Spanish Football League: A Network Science Perspective. *Entropy* **2020**, *22*, 172. [CrossRef]

18. Ribeiro, J.; Silva, P.; Duarte, R.; Davids, K.; Garganta, J. Team Sports Performance Analysed Through the Lens of Social Network Theory: Implications for Research and Practice. *Sports Med.* **2017**, *47*, 1689–1696. [CrossRef]

19. Hassan, A.; Schrapf, N.; Ramadan, W.; Tilp, M. Evaluation of tactical training in team handball by means of artificial neural networks. *J. Sports Sci.* **2017**, *35*, 642–647. [CrossRef]

20. Castellano, J.; Casamichana, D.; Lago, C. The use of match statistics that discriminate between successful and unsuccessful soccer teams. *J. Hum. Kinet.* **2012**, *31*, 139–147. [CrossRef]

21. Harrop, K.; Nevill, A. Performance indicators that predict success in an English Professional League One Soccer Team. *Int. J. Perform. Anal. Sport* **2014**, *14*, 907–920. [CrossRef]

22. Lago-Peñas, C.; Lago-Ballesteros, J.; Dellal, A.; Gómez, M. Game-Related Statistics that Discriminated Winning, Drawing and Losing Teams from the Spanish Soccer League. *J. Sports Sci. Med.* **2010**, *9*, 288–293. [PubMed]

23. Higham, D.G.; Hopkins, W.G.; Pyne, D.B.; Anson, J.M. Performance indicators related to points scoring and winning in international rugby sevens. *J. Sports Sci. Med.* **2014**, *13*, 358–364. [PubMed]

24. Principe, J.C.; Euliano, N.R.; Lefebvre, W.C. *Neural and Adaptive Systems: Fundamentals Through Simulations*; Wiley: New York, NY, USA, 2000; Volume 672.

25. Barron, D.; Ball, G.; Robins, M.; Sunderland, C. Artificial neural networks and player recruitment in professional soccer. *PLoS ONE* **2018**, *13*, e0205818. [CrossRef] [PubMed]

26. Clemente, F.M.; Martins, F.M.L. Network structure of UEFA Champions League teams: Association with classical notational variables and variance between different levels of success. *Int. J. Comput. Sci. Sport* **2017**, *16*, 39–50. [CrossRef]

27. Huang, K.Y.; Chang, W.L. A neural network method for prediction of 2006 World Cup Football Game. In Proceedings of the 2010 International Joint Conference on Neural Networks (IJCNN), Barcelona, Spain, 18–23 June 2010; pp. 1–8.

28. Memmert, D.; Perl, J. Game creativity analysis using neural networks. *J. Sports Sci.* **2009**, *27*, 139–149. [CrossRef]

29. Owramipur, F.; Eskandarian, P.; Mozneb, F.S. Football Result Prediction with Bayesian Network in Spanish League-Barcelona Team. *Ijcte Int. J. Comput. Theory Eng.* **2013**, *5*, 812–815. [CrossRef]

30. Perl, J.; Memmert, D. Soccer analyses by means of artificial neural networks, automatic pass recognition and Voronoi-cells: An approach of measuring tactical success. In *Proceedings of the 10th International Symposium on Computer Science in Sports (ISCSS)*; Springer: Cham, Switzerland, 2016; pp. 77–84.

31. Razali, N.; Mustapha, A.; Yatim, F.A.; Ab Aziz, R. Predicting Football Matches Results using Bayesian Networks for English Premier League (EPL). *Iop Conf. Ser. Mater. Sci. Eng.* **2017**, *226*, 1–6. [CrossRef]

32. Wang, Q.; Yao, Y.; Zhu, H.; Hu, W.; Shen, Z. Discerning tactical patterns for professional soccer teams: An enhanced topic model with applications. In Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Sydney, NSW, Australia, 10–13 August 2015; pp. 2197–2206.

33. Cintia, P.; Giannotti, F.; Pappalardo, L.; Pedreschi, D.; Malvaldi, M. The harsh rule of the goals: Data-driven performance indicators for football teams. In Proceedings of the 2015 IEEE International Conference on Data Science and Advanced Analytics (DSAA), Paris, France, 19–21 October 2015; pp. 1–10.

34. Wagenaar, M.; Okafor, E.; Frencken, W.; Wiering, M.A. Using Deep Convolutional Neural Networks to Predict Goal-scoring Opportunities in Soccer. In Proceedings of the 6th International Conference on Pattern Recognition Applications and Methods (ICPRAM 2017), Porto, Portugal, 24–26 February 2017; pp. 448–455.

35. Sahin, M.; Erol, R. A comparative study of neural networks and ANFIS for forecasting attendance rate of soccer games. *Math. Comput. Appl.* **2017**, *22*, 2–12.

36. Winter, C.; Rasche, C.; Pfeiffer, M. Linear vs. non-linear classification of winners, drawers and losers at FIFA World Cup 2014. *Sci. Med. Football* **2017**, *1*, 164–170. [CrossRef]

37. Eggels, H.; Elk, R.V.; Pechenizkiy, M. Explaining Soccer Match Outcomes with Goal Scoring Opportunities Predictive Analytics. In Proceedings of the 2016 European Conference on Machine Learning and Principles and Practice of Knowledge Discovery in Databases, Riva del Garda, Italy, 19–23 September 2016; pp. 1–10.

38. Eryarsoy, E.; Delen, D. Predicting the Outcome of a Football Game: A Comparative Analysis of Single and Ensemble Analytics Methods. In Proceedings of the 52nd Hawaii International Conference on System Sciences, Wailea, HI, USA, 8–11 January 2019; pp. 1107–1115.

39. Wei, X.; Sha, L.; Lucey, P.; Morgan, S.; Sridharan, S. Large-Scale Analysis of Formations in Soccer. In Proceedings of the 2013 International Conference on Digital Image Computing: Techniques and Applications (DICTA), Hobart, Australia, 26–28 November 2013; pp. 1–8.

40. Carpita, M.; Ciavolino, E.; Pasca, P.; Groll, A.; Manisera, M.; Schauberger, G.; Zuccolotto, P. Exploring and modelling team performances of the Kaggle European Soccer database. *Stat. Model.* **2019**, *19*, 74–101. [CrossRef]

41. Liti, C.; Piccialli, V.; Sciandrone, M. Predicting soccer match outcome using machine learning algorithms. In Proceedings of the MathSport International 2017 Conference, Padua, Italy, 26–28 June 2017; p. 229.

42. Correa, M.; Barrera-Causil, C.; Marmolejo-Ramos, F. The next winner of the 2018 FIFA World Cup will be . . . : An illustration of the use of statistical simulation to make a prediction in a complex tournament. *Chil. J. Stat.* **2018**, *9*, 59–65.

43. Groll, A.; Ley, C.; Van Eetvelde, H.; Schauberger, G. A hybrid random forest to predict soccer matches in international tournaments. *J. Quant. Anal. Sports* **2019**, *15*, 271–287. [CrossRef]

44. Clemente, F.M. Performance outcomes and their associations with network measures during FIFA World Cup 2018. *Int. J. Perform. Anal. Sport* **2018**, *18*, 1010–1023. [CrossRef]

45. Gevaria, K.; Sanghavi, H.; Vaidya, S.; Deulkar, K. Football match winner prediction. *Int. J. Emerg. Technol. Adv. Eng.* **2015**, *10*, 364–368.

46. Yi, Q.; Gomez, M.A.; Wang, L.; Huang, G.; Zhang, H.; Liu, H.; Wang, L.; Huang, G.; Zhang, H. Technical and physical match performance of teams in the 2018 FIFA World Cup: Effects of two different playing styles. *J. Sports Sci.* **2019**, *37*, 2569–2577. [CrossRef]

47. Linke, D.; Link, D.; Lames, M.; Kerhervé, H.A. Football-specific validity of TRACABs optical video tracking systems. *PLoS ONE* **2020**, *15*, e0230179. [CrossRef]

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

49. NeuroDimension, I. NeuroSolutions for Excel, 7.0.1.0. Available online: http://www.neurosolutions.com/neurosolutions/:nDimensional//4LibertySquare// (accessed on 7 April 2019).

50. Liu, J. *Radial Basis Function (RBF) Neural Network Control for Mechanical Systems: Design, Analysis and Matlab Simulation*; Tsinghua University Press: Beijing, China; Springer-Verlag: Berlin/Heidelberg, Germany, 2013. [CrossRef]

51. Ezeldin, A.S.; Sharara, L.M. Neural Networks for Estimating the Productivity of Concreting Activities. *J. Constr. Eng. Manag.* **2006**, *132*, 650. [CrossRef]

52. Anysz, H.; Zbiciak, A.; Ibadov, N. The influence of input data standardization method on prediction accuracy of artificial neural networks. *Procedia Eng.* **2016**, *153*, 66–70. [CrossRef]

53. Jeatrakul, P.; Wong, K.W.; Fung, C.C. Data Cleaning for Classification Using Misclassification Analysis. *J. Adv. Comput. Intell. Intell. Inform.* **2010**, *14*, 297–302. [CrossRef]

54. Balasubramanie, P.; Florence, M.L. Application of radial basis network model for HIV/AIDS regimen specifications. *J. Comput.* **2009**, *1*, 136–140.

55. Bullinaria, J.A. *Radial Basis Function Networks: Algorithms*; University of Birmingham: Birmingham, UK, 2014; pp. 6–7.

56. Hassan, A.; Schrapf, N.; Tilp, M. The prediction of action positions in team handball by non-linear hybrid neural networks. *Int. J. Perform. Anal. Sport* **2017**, *17*, 293–302. [CrossRef]

57. Magoulès, F.; Zhao, H.-X. *Data Mining and Machine Learning in Building Energy Analysis*; Wiley Online Library: Hoboken, NJ, USA, 2016.

58. Mashor, M.Y. Hybrid training algorithm for RBF network. *Int. J. Comput. Internet Manag.* **2000**, *8*, 50–65.

59. Akl, A.-R.I.; Hassan, A.A. An Artificial Neural Network Approach for Predicting Kinematics in Handball Throws. *Am. J. Sports Sci.* **2017**, *5*, 35–39.

60. Mühlhäuser, M.; Erwin, A.; Alois, F. Constructing Ambient Intelligence. In *Proceedings of Communications in Computer and Information Science*; Springer: Berlin, Germany; pp. 33–36.

61. Khosrow-Pour, D.M. *Bioinformatics: Concepts, Methodologies, Tools, and Applications*, 1st ed.; Khosrow-Pour, D.M., Ed.; Medical Information Science Reference (an imprint of IGI Global): 701 E Hershey, PA, USA, 2013.

62. Igiri, C.P.; Nwachukwu, E.O. An improved prediction system for football a match result. *Iosr J. Eng.* **2014**, *4*, 12–20.

63. Fan, Z.; Kuang, Y.; Lin, X. Chess game result prediction system. In *CS 229 Machine Learning Project Report*; Stanford University: Stanford, CA, USA, 2013; pp. 1–5.

64. Schumaker, R.P.; Solieman, O.K.; Chen, H. Sports knowledge management and data mining. *Aris Annu. Rev. Inf. Sci. Technol.* **2010**, *44*, 115–157. [CrossRef]

65. Kahn, J. Neural Network Prediction of NFL Football Games. 2003. Available online: http://homepages.cae.wisc.edu/~{}ece539/project/f03/kahn.pdf. (accessed on 7 April 2019).

66. Liang, M. Data Mining: Concepts, Models, Methods, and Algorithms M. Kantardzic. *IIE Trans.* **2004**, *36*, 495–496. [CrossRef]

67. Haghighat, M.; Rastegari, H.; Nourafza, N. A review of data mining techniques for result prediction in sports. *Adv. Comput. Sci. Int. J.* **2013**, *2*, 7–12.

68. Hewitt, A.; Greenham, G.; Norton, K. Game style in soccer: What is it and can we quantify it? *Int. J. Performance Anal. Sport* **2016**, *16*, 355–372. [CrossRef]

69. Bradley, P.S.; Lago-Penas, C.; Rey, E.; Diaz, A.G. The effect of high and low percentage ball possession on physical and technical profiles in English FA Premier League soccer matches. *J. Sports Sci.* **2013**, *31*, 1261–1270. [CrossRef]

70. Bradley, P.S.; Sheldon, W.; Wooster, B.; Olsen, P.; Boanas, P.; Krustrup, P. High-intensity running in English FA Premier League soccer matches. *J. Sports Sci.* **2009**, *27*, 159–168. [CrossRef] [PubMed]

71. Bradley, P.S.; Carling, C.; Archer, D.; Roberts, J.; Dodds, A.; Di Mascio, M.; Paul, D.; Gomez Diaz, A.; Peart, D.; Krustrup, P. The effect of playing formation on high-intensity running and technical profiles in English FA Premier League soccer matches. *J. Sports Sci.* **2011**, *29*, 821–830. [CrossRef] [PubMed]