Neural Computing and Applications (2022) 34:2525–2541
https://doi.org/10.1007/s00521-021-05930-x

S.I: COGNITIVE-INSPIRED COMPUTING AND APPLICATIONS

Optimization analysis of football match prediction model based on neural network

Shuo Guan[1] · Xiaochen Wang[2]

Received: 21 November 2020 / Accepted: 12 March 2021 / Published online: 31 March 2021

Abstract

How to build a football match prediction model and use scientific methods to solve the prediction problem has become a key point in the application of artificial intelligence in the sports industry. In this paper, we choose a BP neural network model that is powerful in processing nonlinear data to perform research. According to the demand, this paper constructs a gray fuzzy prediction model based on neural network, a gray extreme learning machine prediction model, and a gray fuzzy extreme learning machine prediction combination model based on neural network. Moreover, this paper tests the neural network model by comparing actual results with predicted results. In addition, by predicting and analyzing the football league data, this article tests the three models in terms of match result prediction accuracy, data processing speed, data transmission accuracy, match analysis scores, etc., and uses statistical analysis methods to process data, and uses intuitive statistical graphs to obtain the processing results. The research results show that the gray fuzzy extreme learning machine prediction combination model based on neural network constructed in this paper can retain the advantages of a single model and effectively improve the prediction accuracy of the model and the performance of the system.

Keywords Neural network · Prediction model · Football match · Prediction result

1 Introduction

People's research on artificial neural networks (ANN) began in the 1940s. After decades of development, ANN has penetrated into intelligent control, pattern recognition, computer vision, adaptive filtering, adaptive signal processing, nonlinear optimization, knowledge processing, sensor technology and biomedical engineering, etc., and has achieved gratifying results. In 1990, at the International Gaming Conference, a mathematical model for predicting the outcome of a team's next match based on the average goal rate per football match was first proposed. Later, with the development of football and commercial operations, the method of predicting football matches was gradually updated [1].

✉ Xiaochen Wang
wangxc@syty.edu.cn

1 Wushu and Dance Department, Shenyang Sport University, Shenyang 110102, China

2 Management and Communication Department, Shenyang Sport University, Shenyang 110102, China

With the progress of the times and the popularization of the Internet, the industrialization of football has developed rapidly. It is currently the world's outdoor sports with the highest production value, the greatest influence, and the highest public attention, and its achievements in the field of sports are far ahead of other sports. At present, nearly 300 million fans nationwide follow the Chinese Football Super League. Moreover, the average ratings of domestic broadcasts of Chinese Super League matches reached 1.78%. In addition, 26 million fans go to the scene every year to watch the game. The average number of attendees in the Super League in a season is 18,600, ranking first in Asia. In Asia, the Chinese Football League is broadcasted in more than 50 countries and regions, and more than 1,000 news media and nearly 10,000 related professional sports reporters report on the Chinese Football League [2].

Since the twenty-first century, people's living standards have improved tremendously, and sports-related undertakings have also advanced rapidly and steadily. With the rapid development of artificial intelligence today, how to build a football match prediction model and use scientific methods to solve the prediction problem has become a topic of interest to experts and scholars. Various types of football matches have outstanding similarities in some respects. Therefore, in theory, it is possible to find the law from a large number of football matches to find a way to judge the level of victory or defeat. Based on a large amount of football game data, this paper establishes a football game win-loss prediction model and uses the BP neural network method to train and predict the model. On the one hand, it enhances the fans' awareness of technology and football, and on the other hand, it effectively promotes football knowledge and culture [3].

2 Related work

When completing the establishment of the BP neural network prediction model, the literature [4] first initialized the position vector, used the improved particle swarm algorithm for optimization, took the initial value with the smallest mean square error as the optimization target, and used the obtained global optimal position as the initial weight of the BP neural network. Finally, it used the training model to predict the age of genetic hypertension. The experimental results show that the prediction error of the improved neural network is controlled within 1%, which is significantly improved compared to the 5% of the traditional neural network. The literature [5] used particle swarm to optimize BP neural network and establish a gesture recognition model. The quantum particle swarm algorithm replaces the activation function used by the standard BP network to obtain the optimized BP neural network weights and thresholds. The algorithm is simple and does not depend on the initial weights. Moreover, the BP neural network trained by the quantum particle swarm algorithm has fast convergence speed and high learning efficiency. In dealing with some high-dimensional and complex problems, the algorithm can ensure that the algorithm can quickly and effectively converge to the optimal solution. In addition, the BP neural network trained by the quantum particle swarm algorithm can reach 95% accuracy in processing gesture recognition problems, but the network cannot adjust the learning rate adaptively. When establishing the stock forecast model, in order to improve the forecast effect and increase the accuracy and forecast speed, the literature [6] used the cuckoo algorithm to optimize the neural network and compare it with the neural network optimized by the genetic algorithm and the neural network optimized by the particle swarm algorithm. The literature [7] found that the average relative error and average absolute error obtained by the neural network optimized by the cuckoo algorithm during testing are better than the other two algorithms. Aiming at the initial weight optimization problem of BP neural network in the application of short-term traffic flow, the literature [8] also proposed a short-term traffic flow prediction model in which the cuckoo search algorithm optimizes the parameters of the BP neural network. This model effectively improves the accuracy of the BP neural network in the application of short-term traffic flow prediction, and more accurately reflects the changing trend of short-term traffic flow. In addition to common swarm intelligence optimization algorithms, domestic and foreign scholars have also optimized the BP neural network model through other algorithms. Aiming at the problem that the standard BP algorithm is easy to fall into a local minimum when the gradient descent maximum value method is used to solve the problem, and the phenomenon that it is difficult to avoid the slow convergence rate caused by the oscillation in the network training learning, the literature [9] comprehensively used the advantages of additional momentum and variable learning rate to improve it and establish an improved BP network, that is, GDP prediction model. The prediction results of the standard BP algorithm and the improved BP algorithm are similar to the actual value, but the average relative error of the latter is small, the number of learning iterations is relatively small, and the time-consuming is 1/10 of the former [10]. In practical applications, an appropriate learning rate can not only improve the error accuracy but also reduce network oscillation [11]. The literature [12] proposed a license plate location method based on the AB neural network model, which uses the area histogram and rectangular sliding window method to locate the license plate for the first time, and then uses the area average method and the average jump method to accurately locate the license plate. Finally, it used 480 car pictures to conduct experiments and successfully located the license plate positions of 474 pictures. Aiming at the problem of low prediction accuracy of the previous tax prediction models, the literature [13] integrated the idea of Adaboost algorithm into the BP neural network for tax prediction. The literature [14] used wavelet technology and back-propagation neural network to build a W-BPNN model to predict air pollutants (PM10, SO2, NO2, etc.). According to the comparison of experimental results, it is found that the W-BPNN model has better prediction performance and prediction accuracy than the BP neural network model with a single index.

Bayesian model: The probability of both sides winning the game is determined by various uncertain factors such as the strength of both sides, including the number of historical battles and records of the two teams, as well as offensive and defensive capabilities. These three factors are used to compare the relative strength of any team that ultimately scores more than the competing team. Other factors related to team strength may also affect the winning rate, but the Bayesian model believes that their influence is very small [15]. Fuzzy comprehensive evaluation: The fuzzy comprehensive evaluation method analyzes the factors that are likely to attract everyone's attention in the football game (such as the number of goals, the number of shots, the weather temperature, the overall state of the team) as an index to predict the game. Then, it determines the evaluation weight, applies the maximum membership method, determines the number matrix, uses the data processing system to calculate the searched data, and performs the final uncertainty calculation. Fuzzy comprehensive evaluation requires too many indicator dimensions, and early data acquisition and data cleaning are the most difficult points of this method as a football match prediction [16]. Data mining: In the prediction of the victory or defeat of a football match, a number of representative teams are arbitrarily selected and classified, and attributes are selected as decision attributes. By constructing decision trees, pruning decision trees and extracting decision trees. Perform attribute deletion, attribute conversion, and generalization of continuous attribute data into interval values, and then make reasonable predictions on the outcome of the game according to various rules [17].The literature [18] uses the decision tree C4.5 algorithm to establish a football prediction model. The model designs a data set containing two attribute values, which are the scores of the past five rounds and the team's overall goal rate, and then divides the results of the game into three categories: win, tie, and loss. Through the C4.5 algorithm, a decision tree is generated based on this data set. The article believes that it is difficult to judge the result of a tie in a football match. When the result of the match is a tie, it is believed that the predicted result is basically consistent with the actual situation. Finally, the results of a certain round of the Premier League were tested. There were 5 matches that were completely consistent, 4 were basically the same, and 1 was inconsistent. In addition to the above methods, other experts and scholars have also made some other attempts. The literature [19] used the logistic regression method to analyze the data (including points, goals and losses, etc.) of the Italian Football League match. After the principal component analysis took out the interference factors, the logistic regression model was established to predict the outcome of the game and the prediction was accurate. The rate reaches more than 65%. Considering that the odds of the game is the result of the lottery purchaser's consideration of various influencing factors on the outcome of the game, the data on the end of the game is the result of combining various influencing factors. The literature [20] used Python to obtain the easy-to-obtain handicap data of the English First Division competition and selects the Logistic regression model for modeling according to the characteristics of the data. Finally, the accuracy of the obtained model prediction was verified, and the final accuracy of the model prediction was 70.1%. Through the in-depth analysis of the influencing factors of each game, the literature [21] divided the prediction of the game situation into two types: one is the situation where the initial input influence factors are complete, and the other is the situation where the initial input influence factors are incomplete. Aiming at the feature that the evidence theory can still calculate the uncertainty of the data under the condition that the initial input data source is missing, it combined the fusion formula of the evidence theory to calculate the uncertainty of the football match result in these two situations. Moreover, it provided the general method steps and algorithms for calculating the uncertainty of victory, tie, and defeat in the game result.

3 Prediction algorithm research

The development law of most sample data will fluctuate greatly due to the influence of many unknown random factors. At this time, if a single prediction method is used for prediction, satisfactory prediction accuracy is often not obtained. However, the combined prediction method can better make up for the shortcomings of the single prediction method, and generally can get better forecast results. The algorithm proposed in this paper is a combination of gray prediction algorithm and extreme learning machine algorithm, which has good prediction accuracy and prediction speed.

3.1 Grey prediction algorithm

The types of gray prediction mainly include: topological prediction, which predicts changes in system behavior; extrapolative series prediction, which predicts system behavior sequence; system prediction, which respectively predicts the dynamic relationship of each element of the system and connects the elements into a comprehensive system model according to the element relationship. The prediction model is currently the most widely used model in gray prediction. The $GM(1,1)$ prediction model is a first-order linear differential equation about series prediction. It accumulates the original time series data by time to form a new time series, and then uses the solution of the first-order linear differential equation to approximate the development law of the new time series. Some scholars have proved that the change law of the original time series data obtained by the method of approximating the first-order linear differential equation solution is an exponential law. Therefore, the $GM(1,1)$ prediction model has a good prediction effect when the original time series data shows exponential changes. Moreover, the $GM(1,1)$ prediction model has a small demand for original data and can predict its systematic rules for at least 4 original data. In addition, it has the characteristics of good short-term prediction effect and stable prediction results [22].

We set $x^{(0)}$ as the original data series, $x^{(0)}=\left[x^{(0)}(1),x^{(0)}(2),\cdots,x^{(0)}(n)\right]$, and the gray generation number series as $x^{(1)}$, $x^{(1)}=\left[x^{(1)}(1),x^{(1)}(2),\cdots,x^{(1)}(n)\right]$. The relationship between $x^{(0)}$ and $x^{(1)}$ is as follows:

$$x^{(1)}(k)=\sum_{i=1}^{k}x^{(0)}(i);\ k=1,2,\cdots,n \tag{1}$$

It is called an accumulative generation, denoted as AGO (accumulating generation operator).

By accumulating the original data for r times, the following sequence is obtained:

$$x^{(r)}(k)=\sum_{i=1}^{k}x^{(r-1)}(i) \tag{2}$$

According to the above formula, the accumulation from the $r-1$th time to the rth time is calculated to obtain:

$$x^{(r)}(k)=\sum_{i=1}^{k}x^{(r-1)}(i)+x^{(r-1)}(k)=x^{(r-1)}(k-1)+x^{(r-1)}(k) \tag{3}$$

$$x^{(r)}(k)=\sum_{i=1}^{k}x^{(r-1)}(i)=\sum_{i=1}^{k}\left(\sum_{j=1}^{i}x^{(r-2)}(j)\right) \tag{4}$$

The modeling principle of the grey prediction model is as follows [23]:

Grey system is mainly based on the development of discrete sequence to establish function fitting differential equation. Among them, $GM(1,1)$ with the first-order differential equation as a model is a commonly used gray system training model, and its expression formula is:

$$\frac{dx}{dt}+ax=u \tag{5}$$

From the derivative definition, we know:

$$\frac{dx}{dt}=\lim_{\Delta t\to 0}\frac{x(t+\Delta t)-x(t)}{\Delta t} \tag{6}$$

When Δt takes the minimum limit to infinitely close to 1, there is a limit formula as follows:

$$x(t+1)-x(t)=\frac{\Delta x}{\Delta t} \tag{7}$$

It is written in discrete form as:

$$\frac{\Delta x}{\Delta t}=x(k+1)-x(k)=\Delta^{(1)}(x(k+1)) \tag{8}$$

This means that $\frac{\Delta x}{\Delta t}$ is a cumulative subtraction of $x(k+1)$, so $\frac{\Delta x}{\Delta t}$ is the equivalent value of the binary combination of $x(k+1)$ and $x(k)$, then the binary combination of $x(k+1)$ and $x(k)$ is an even pair, denoted as $[x(k+1),x(k)]$.

Therefore, we can define a mapping from $[x(k+1),x(k)]$ to $\frac{\Delta x}{\Delta t}$ [23]:

$$F:[x(k+1),x(k)]\to\frac{dx}{dt} \tag{9}$$

If we define $R(t)$ to be the background value of $\frac{dx}{dt}$ at time t (the corresponding value of x), then each $\frac{dx}{dt}$ has a pair of background values $R(t)$ corresponding to it. Now, the first-order differential equation formula $\frac{dx}{dt}+ax=u$ is solved, which is taken as the linear expression of x, u and $\frac{dx}{dt}$. It can be obtained that in this linear combination, in a short time, when $\Delta x=1$, the variable $x(t)\to x(t+\Delta t)$ is approximately equal, and there will be no sudden change. Then, the binary combination can be averaged and calculated [24]:

$$z(t)=\frac{1}{2}[x(k+1)+x(k)] \tag{10}$$

3.2 Extreme learning machine algorithm

The extreme learning machine is a single hidden layer feedforward neural network. The input weight and hidden layer deviation of the extreme learning machine are randomly generated, and the output weight is obtained through analysis and calculation. The extreme learning machine is a single hidden layer feedforward neural network. The input weight and hidden layer deviation of the extreme learning machine are randomly generated, and the output weight is obtained through analysis and calculation. The extreme learning machine uses a simple least square method to determine the best output weights, which overcomes the difficulties of traditional nonlinear optimization problems, such as: input weights and hidden layer deviations, and the determination of the optimal value of output weights. This means that as long as the scale of the weight is small enough, the user does not need to consider all input weights and hidden layer deviations, but only needs to consider the output weights. This is completely different from traditional iterative learning methods.

We assume that the extreme learning machine has k hidden layer neurons, the sample data is m, and the activation function is v. The mathematical expression of the activation function v is shown in the following formula [26]:

$$\sum_{j=1}^{k}v_j\left(w_j,b_j,x_i\right)\beta_j=t_i,\ i=1,2,\cdots,n \tag{11}$$

In the formula, w_j is the input weight vector connecting the jth hidden neuron and the input layer neuron, β_j is the output weight connecting the jth hidden neuron and the output layer neuron, and b_j represents the deviation of the jth hidden node. If the estimation data sample of the extreme learning function is zero error, we will get $\sum_{i=1}^{n}\|y_i - t_i\| = 0$, that is, there are w_j, β_j, b_j to make $\sum_{j=1}^{k} v_j(w_j, b_j, x_i)\beta_j = t_i, i = 1, 2, \cdots, n$. The structure of the extreme learning machine with multiple input and single output is shown in Fig. 1:

The mathematical expression of the activation function v can be simply written as the following:

$$H\beta = T \tag{12}$$

Among them, H is the hidden layer output matrix, and the specific form is as follows:

$$H\begin{bmatrix} v(w_1, b_1, x_1) & & v(w_k, b_k, x_1) \\ \vdots & \cdots & \vdots \\ v(w_1, b_1, x_n) & & v(w_k, b_k, x_n) \end{bmatrix}_{n\times k} \tag{13}$$

It has been confirmed that there is no need to adjust the input weight vector and hidden bias. Once the parameter values are randomly assigned at the start of learning, the matrix H actually remains unchanged.

(1) If the activation function is infinitely differentiable, when the number of hidden neurons is equal to the number of training samples, namely $k = n$, the parameters of the hidden nodes can be randomly assigned, and the output weight is determined by the generalized inverse matrix analysis of the matrix H. Therefore, the extreme learning machine can estimate data samples with zero error [27].

(2) When $k \ll n$, the matrix H becomes a non-square matrix. A specific $\tilde{\beta}_j$ can be determined by the following formula:

$$Min.\|H(w_j, b_j, x_i)\tilde{\beta}_j - y_i\| \tag{14}$$

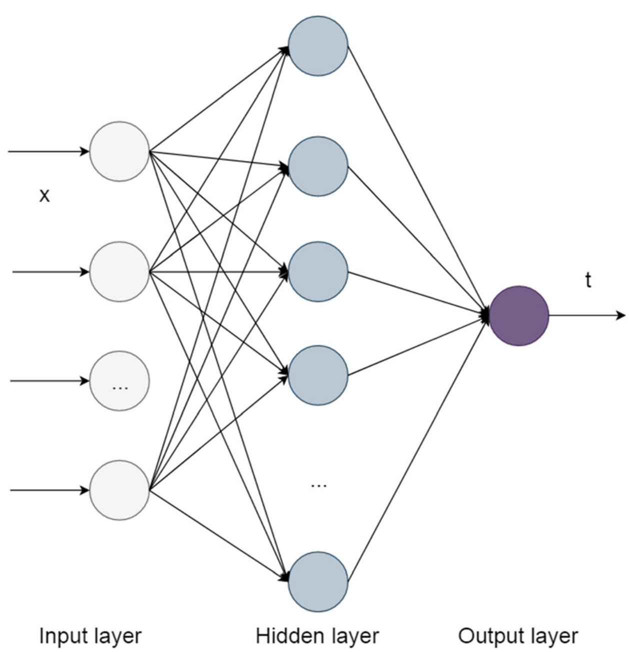


Fig. 1 Structure diagram of the extreme learning machine with multiple input and single output

The above formula is equivalent to solving the minimum value of the following function:

$$Min.E = \sum_{i=1}^{n}\left[\sum_{j=1}^{k}\|v_j(w_j, b_j, x_i)\tilde{\beta}_j - y_i\|\right]^2 \tag{15}$$

Therefore, because w_j and b_j have been given, according to the mathematical expression of the activation function v, the output weight can be calculated with the following formula:

$$\tilde{\beta} = H^{+}T \tag{16}$$

Among them, H^{+} is the Moore–Penrose generalized inverse of matrix H. There are many ways to calculate H^{+}, and the singular value decomposition method is commonly used.

The extreme learning machine has theoretically proved that it can get satisfactory fitting values, and it has good generalization performance and absolutely fast training speed. The only work left by this algorithm for the user is to select the activation function and the number of hidden neurons, which makes the algorithm easy to use.

The activation function has a great influence on the neural network, including its convergence speed and training accuracy, so the choice of the activation function is extremely important in the application of the neural network. There are many types of activation functions for extreme learning machines. The commonly used activation functions are as follows:

(1) Threshold function:

$$v(x) = \begin{cases} 1 & x \geq 0 \\ 0 & x < 0 \end{cases} \tag{17}$$

(2) Sigmoid function:

$$v(x) = \frac{1}{1 + \exp(-\alpha(x - \beta))} \tag{18}$$

(3) Radial basis function:

$$v(x) = \exp(x^2) \tag{19}$$

(4) Triangular basis function:

$$v(x) = \begin{cases} 1 - abs(x) & -1 \le x \le 1 \\ 0 & otherwise \end{cases} \quad (20)$$

4 Combination prediction method

The combined prediction method is a prediction method that uses two or more different single prediction methods to predict the same prediction problem. The combination prediction method can be a combination of several quantitative methods, a combination of several qualitative methods, or a combination of several quantitative and qualitative methods. Moreover, the combined prediction method comprehensively utilizes the information provided by different single prediction methods. Compared with the single prediction method, the combined prediction method improves the accuracy of prediction [28].

We assume that the sample sequence value of the prediction object is x_t, where $t = 1, 2, \cdots, N$ and N is the sample sequence size of the prediction object. Moreover, we assume that there are m single prediction methods, and the m single prediction methods are used to predict the prediction object. The prediction result of the ith single prediction method for the prediction object is f_i, $i = 1, 2, \cdots, m$. After that, we combine the predicted values of these m single prediction methods to obtain the final prediction result $\hat{x}_t$ of the predicted object x_t. $\hat{x}_t$ is expressed as: $\hat{x}_t = \phi(f_1, f_2, \cdots, f_m)$. Among them, $\phi(\cdot)$ is a function determined by the combination rule of the combination prediction method. According to the functional properties of $\phi(\cdot)$, the combination prediction method can be divided into linear combination prediction method and nonlinear combination prediction method.

The combined prediction value $\hat{x}_t$ of the linear combination prediction method of the predicted object sequence x_t satisfies the following formula [29]:

$$\hat{x}_t = \sum_{i=1}^{m} l_i f_i \quad (21)$$

Among them, l_i is the combined weight coefficient of the ith single prediction method in the combined prediction method, and satisfies:

$$\sum_{i=1}^{m} l_i = 1, l_i \ge 0, \ i = 1, 2, \cdots, m \quad (22)$$

(1) Equal weighted average combination prediction method.

The combination prediction method whose combination weight coefficient is the reciprocal of the number of single prediction methods is called the equal weight average combination prediction method. The calculation formula of l_i is as follows:

$$l_i = 1/m, i = 1, 2, \cdots, m \quad (23)$$

The characteristic of this method is that each single prediction method is regarded as equally important, and the combined weight coefficients of m single prediction methods are completely equal.

(2) Combination prediction methods related to adaptability.

This method first sorts the individual prediction methods in increasing order according to the prediction error, and the calculation formula of l_i is as follows:

$$l_i = \begin{cases} 1 - \sum_{i=2}^{m} l_i & i = 1 \\ 1/2^{m+1-i} & i \ne 1 \end{cases} \quad i = 1, 2, \cdots, m \quad (24)$$

This method pays more attention to single prediction methods with small prediction errors and gives more weight to single prediction methods with small prediction errors.

(3) Covariance optimization combination prediction method.

This method assumes that the variances of the prediction errors of the prediction results of m single prediction methods are $\delta_1, \delta_2, \cdots, \delta_m$ respectively, and the calculation formula of l_i is as follows:

$$l_i = \left[\delta_i \sum_{j=1}^{m} \frac{1}{\delta_j}\right]^{-1} \quad i = 1, 2, \cdots, m \quad (25)$$

The variance of the prediction error of the combined prediction method is

$$Var(e) = \sum_{i=1}^{m} l_i^2 \delta_i \quad (26)$$

The covariance optimization combination prediction method makes the variance of the prediction error of the combination prediction method obtain a minimum value under the constraint condition of $\sum_{i=1}^{m} l_i = 1, l_i \ge 0, \ i = 1, 2, \cdots, m$.

The determination function $\phi(\cdot)$ of the combined prediction value $\hat{x}_t$ of the prediction target sequence x_t is a nonlinear function, and the combined prediction method is a nonlinear combined prediction method. Common forms of nonlinear combination prediction methods are:

(1) Weighted geometric average combination prediction method.

In the weighted geometric average combination

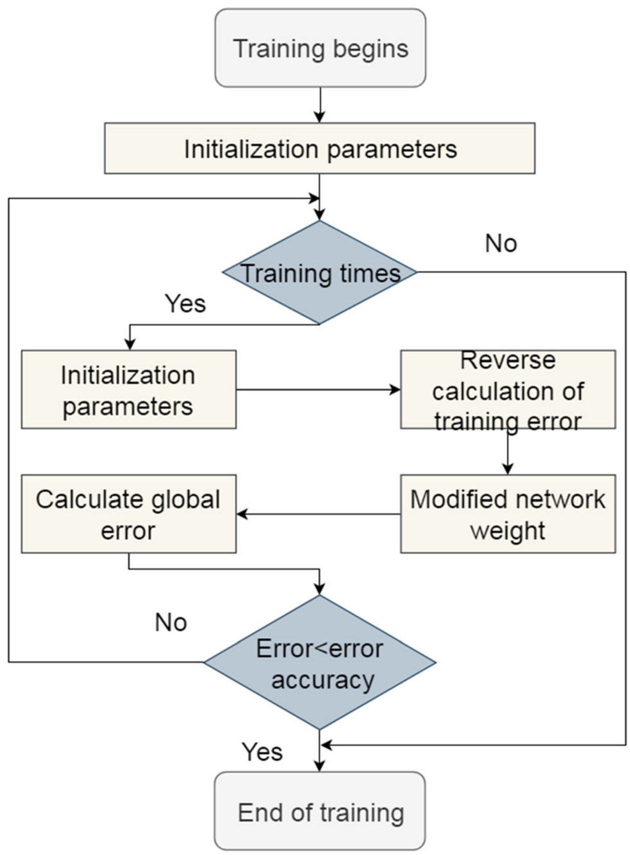


Fig. 2 BP neural network training flowchart

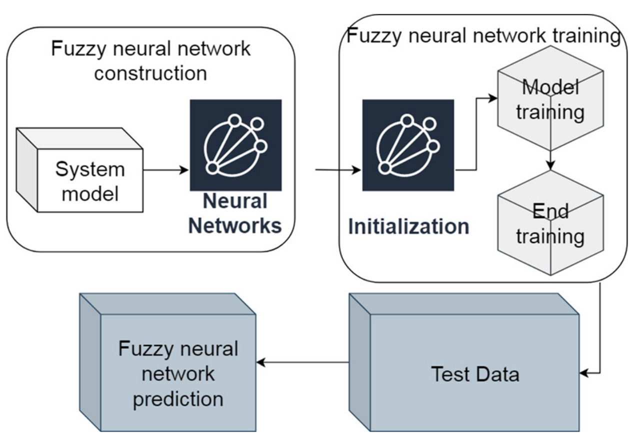


Fig. 3 Algorithm flow chart of grey hybrid fuzzy neural network prediction

prediction method, the calculation formula of the combination prediction value $\hat{x}_t$ of the combination prediction method of the prediction object x_t is as follows:

$$\hat{x}_t = \prod_{i=1}^{m} f_i^{l_i} \tag{27}$$

Among them, l_i is the combined weight

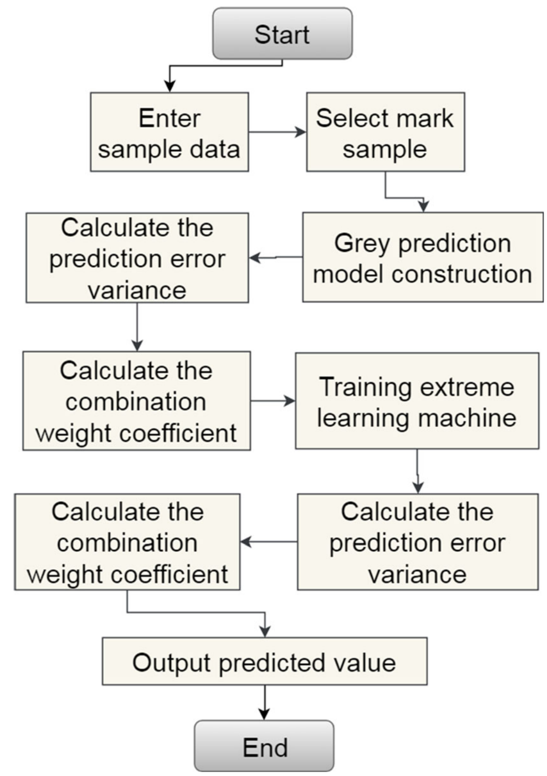


Fig. 4 Algorithm flowchart of grey extreme learning machine

Table 1 Sample data to be processed

	1	2	3	4	5
Home and away	1.01	0.00	1.01	0.00	1.01
Team goals	1.77	1.82	1.41	1.41	1.41
Team shots	11.87	12.32	13.53	12.52	12.73
Team shot on goal	5.05	4.65	4.24	3.84	5.25
Team passes	452.23	446.22	465.81	439.75	408.44
Team corners	4.55	3.84	4.44	4.04	4.04
Team offside	0.76	1.01	1.41	1.82	2.02
Team steals	13.13	12.12	12.93	13.94	11.72
Team possession rate	49.74	52.52	55.55	55.95	54.14
Team competitive status	0.71	0.91	0.81	0.81	0.81
Opponents' goals	1.26	0.40	1.21	0.81	1.01
Opponent shots	18.18	11.72	8.48	13.94	9.70
Opponent shot positive	6.31	3.43	3.23	6.46	5.05
Opponent passes	386.58	454.90	332.49	410.87	329.66
Opponent corners	5.81	4.24	3.03	4.44	6.06
Opponent offside	2.53	1.21	1.82	2.42	1.01
Opponent steals	15.40	18.38	14.34	11.92	15.96
Opponent possession rate	54.79	52.52	41.81	56.16	42.02
Rivalry	0.51	0.30	0.20	0.40	0.40
Win value	1.01	0.51	1.01	0.51	0.51

Fig. 5 The total distribution diagram of sample data to be processed

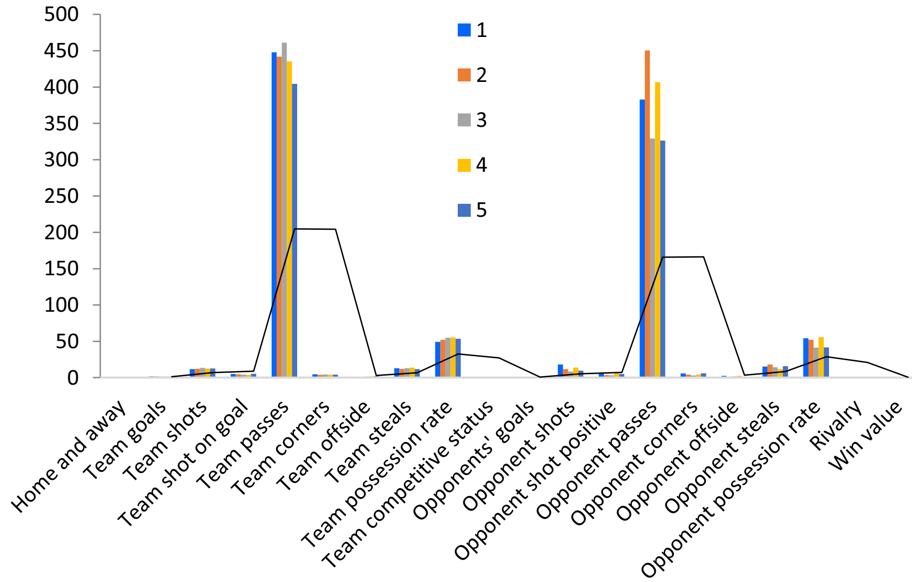
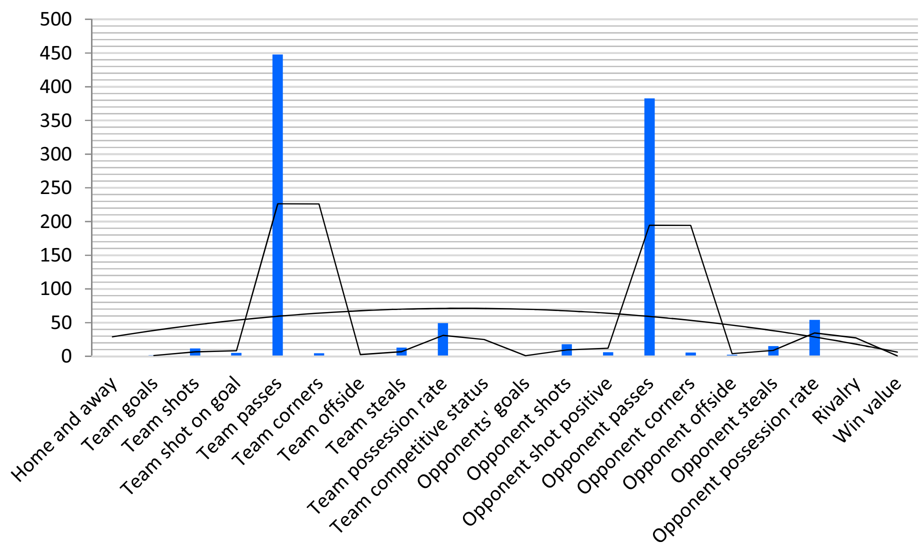


Fig. 6 The distribution of the first round of sample data



coefficient of the ith single prediction method in the combined prediction, and satisfies $\sum_{i=1}^{m} l_i = 1, l_i \geq 0, i = 1, 2, \cdots, m.$.

(2) Weighted harmonic average combination prediction method.

In the weighted harmonic average combination prediction method, the calculation formula of the combination prediction value x_t of the combination prediction method of the prediction target sequence x_t is as follows:

$$\hat{x}_t = \left[\sum_{i=1}^{m} l_{\mathrm{i}}/f_{\mathrm{i}}\right]^{-1} \tag{28}$$

Among them, l_i is the combined weight coefficient of the ith single prediction method in the combined prediction, and satisfies $\sum_{i=1}^{m} l_{\mathrm{i}} = 1, l_{\mathrm{i}} \geq 0, i = 1, 2, \cdots, m.$

Since the forecast is an estimate of the law of future development, there must be a certain gap between the forecast result and the objective actual value, that is, the forecast error. The forecast error is closely related to the accuracy of the forecast, and the forecast error gives a measure of the accuracy of the forecast. Obviously, the larger the prediction error, the lower the prediction accuracy; conversely, the smaller the prediction error, the higher the prediction accuracy. There are many prediction error index measurement methods used to analyze prediction accuracy, mainly including the following:

(1) Mean absolute error (MAE)

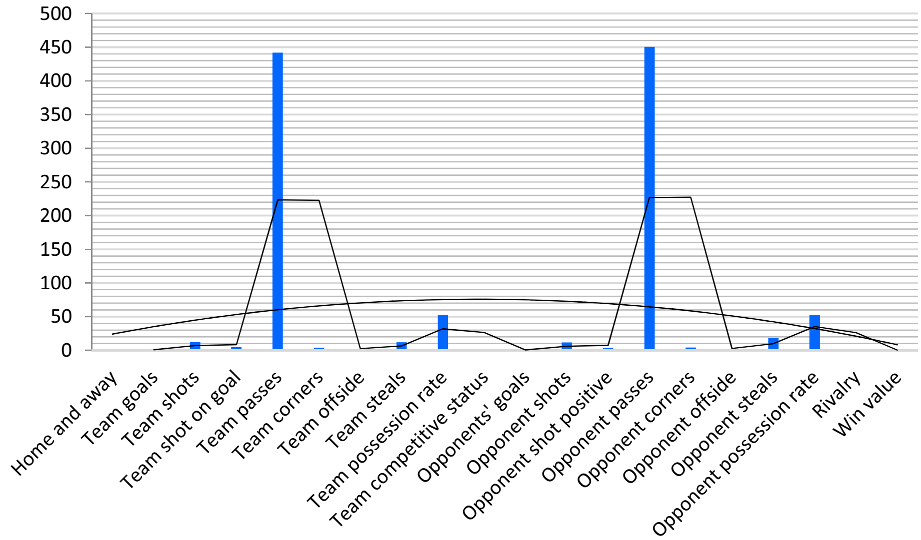


Fig. 7 The distribution of the second round of sample data

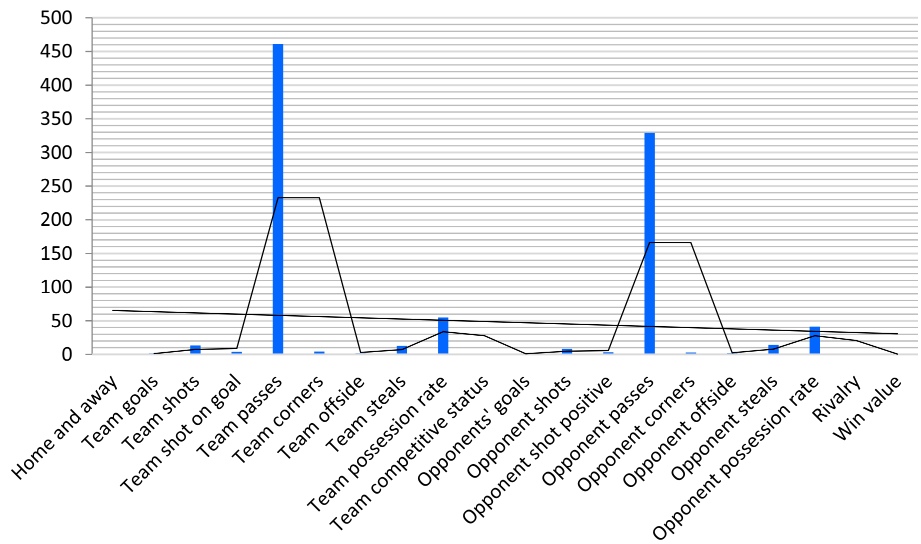


Fig. 8 The distribution diagram of the third round of sample data

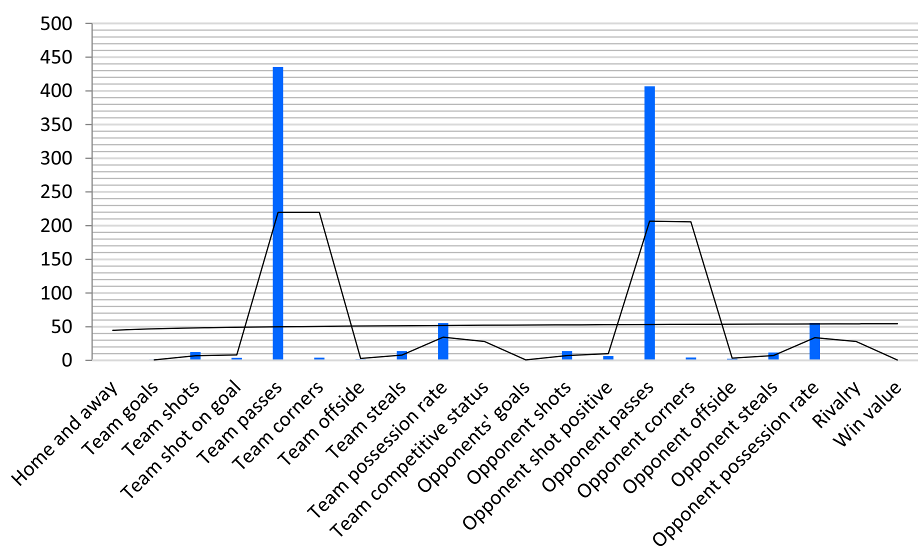


Fig. 9 The distribution diagram of the fourth round of sample data

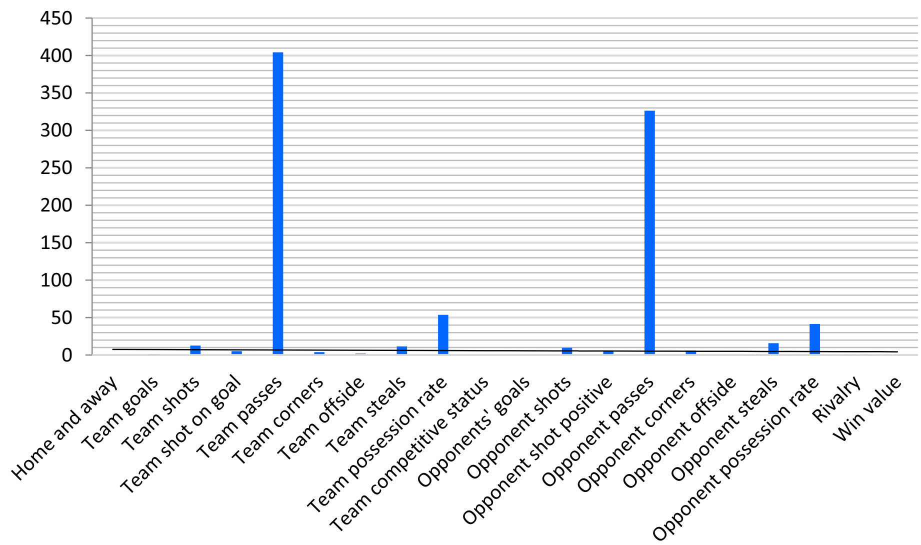


Fig. 10 The distribution diagram of the fifth round of sample data

Table 2 Sample data to be trained

	1	2	3	4	5
Home and away	1.010	0.000	1.010	0.000	1.010
Team goals	0.676	0.684	0.611	0.611	0.611
Team shots	0.955	0.957	0.962	0.958	0.959
Team shot on goal	0.883	0.872	0.860	0.845	0.888
Team passes	1.009	1.009	1.009	1.008	1.008
Team corners	0.869	0.845	0.866	0.852	0.852
Team offside	0.414	0.505	0.611	0.684	0.712
Team steals	0.961	0.957	0.960	0.963	0.955
Team possession rate	0.997	0.998	0.998	0.998	0.998
Team competitive status	0.707	0.909	0.808	0.808	0.808
Opponents' goals	0.576	0.245	0.563	0.434	0.505
Opponent shots	0.974	0.955	0.934	0.963	0.943
Opponent shot positive	0.908	0.826	0.815	0.910	0.883
Opponent passes	1.008	1.009	1.008	1.008	1.008
Opponent corners	0.899	0.860	0.803	0.866	0.904
Opponent offside	0.765	0.563	0.684	0.756	0.505
Opponent steals	0.968	0.975	0.965	0.956	0.969
Opponent possession rate	0.998	0.998	0.994	0.998	0.995
Rivalry	0.505	0.303	0.202	0.404	0.404
Win value	1.010	0.505	1.010	0.505	0.505

$$MAE = \frac{1}{N}\sum_{t=1}^{N}|x_t - \hat{x}_t| \quad (29)$$

MAE is the mean absolute error. Since the prediction error may be positive or negative, in order to avoid the cancellation of positive and negative errors, the average of the absolute value of the error is adopted.

(2) Mean square error (MSE)

$$MSE = \frac{1}{N}\sum_{t=1}^{N}(x_t - \hat{x}_t)^2 \quad (30)$$

MSE is the mean square error. In order to avoid the cancellation of positive and negative errors, the average of the square sum of errors is used. Since the absolute error is squared, the effect of the large error value is amplified.

(3) Root-mean-square error (RMSE)

$$RMSE = \sqrt{\frac{1}{N}\sum_{t=1}^{N}(x_t - \hat{x}_t)^2} \quad (31)$$

RMSE is the root-mean-square error, that is, the square root of the mean square error.

(4) Mean absolute percentage error (MAPE)

$$MAPE = \frac{1}{N}\sum_{t=1}^{N}\left|\frac{x_t - \hat{x}_t}{x_t}\right| \times 100\% \quad (32)$$

MAPE is the mean absolute percentage error. MAPE uses relative error to narrow the error range to [0, 1], so that when measuring the prediction method, there is no need to consider the impact of the prediction object data itself.

5 Construction of prediction models for different football matches based on neural networks

Artificial neural network is composed of a large number of neurons, used to simulate the structure and function of the brain nervous system. The specific steps of a BP neural network learning are described in Fig. 2.

Table 3 Statistical table of the prediction accuracy of the neural network model for football match results

NO	Grey fuzzy model(%)	Grey extreme Learning Machine model(%)	Combination model(%)	NO	Grey fuzzy model(%)	Grey extreme learning machine model(%)	Combination model(%)
1	44	57	76	31	41	47	72
2	46	51	77	32	44	52	81
3	51	45	70	33	40	64	84
4	46	57	70	34	50	57	77
5	55	46	84	35	50	52	71
6	50	61	70	36	52	61	79
7	50	52	67	37	46	46	66
8	52	52	69	38	38	61	84
9	53	48	74	39	57	53	75
10	50	48	80	40	36	53	72
11	44	57	73	41	51	55	84
12	49	46	84	42	48	55	79
13	45	56	80	43	52	48	74
14	50	57	79	44	47	45	82
15	45	58	75	45	46	56	67
16	51	48	82	46	52	63	72
17	54	45	69	47	47	61	79
18	49	46	73	48	47	48	84
19	52	59	80	49	46	59	81
20	38	60	76	50	44	46	83
21	49	55	78	51	46	48	70
22	46	63	81	52	43	48	77
23	39	63	80	53	54	54	67
24	50	54	78	54	41	50	72
25	56	56	74	55	41	60	67
26	57	51	78	56	55	50	74
27	51	64	78	57	38	48	73
28	51	64	79	58	52	65	74
29	37	60	84	59	49	62	84
30	41	53	77	60	55	45	75

Combined with neural network, construct a football prediction model based on neural network. Figures 3 and 4 show the grey fuzzy prediction model based on neural network and the grey extreme learning machine prediction model based on neural network, respectively.

Gray model, referred to as GM model. This model mainly generates gray processing for a historical data sequence that does not have obvious statistical correlation and lacks overall information. Through description, this method makes the factors of gray data from insignificant to obvious, from less to more. Moreover, through the gray system, it mines the internal change law of the data. The GM (1,1) gray model is the most commonly used simple model in the gray model. Its calculation formula is only a single variable differential equation, so the model construction is simple, and the conversion is fast. Its advantages and disadvantages are as follows: (1) The GM (1,1) model method has the advantages of less data required and less computational complexity, so it is only widely used in some specific short-term prediction industries. (2) When the sample size is small, the data distribution is not considered, and the trend of excessive changes is not considered, the gray GM (1,1) model can achieve good accuracy requirements. (3) If the gray model is used in a long-term numerical prediction model, the results that are more consistent with the true value are generally only the first or two predicted values. If the model performs post-prediction, the error of the result is too large, and the result has no reference significance. Because in the gray model, the gray influence accumulated by the original data

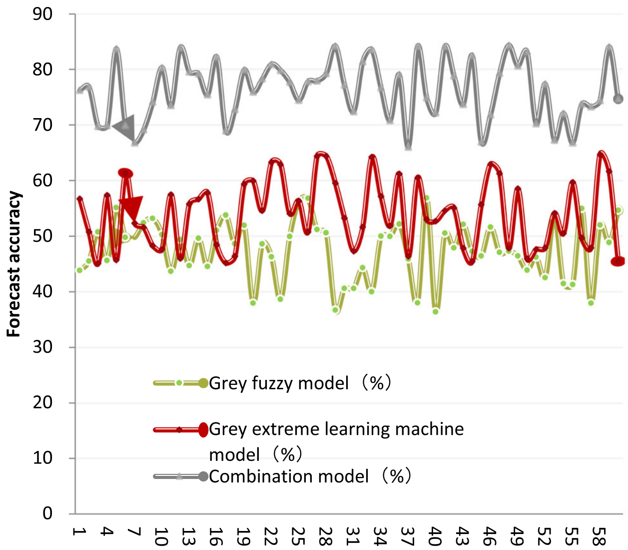


Fig. 11 Statistical diagram of the prediction accuracy of the neural network model for football match results

can only extend to the most recent values. When the prediction time is extended, the influence of gray factors will become weaker and weaker, and the correlation between the results obtained and the historical value will become lower and lower, and the practical significance will be smaller.

The gray extreme learning machine prediction algorithm is mainly composed of two prediction modules: the gray prediction algorithm and the extreme learning machine algorithm. The gray prediction model is suitable for data prediction in the case of small samples. Regarding the prediction problem of large sample data, the prediction effect of the gray prediction model using all the data in the sample set for modeling may be worse than the prediction effect of using a small amount of data before the prediction time point. However, the extreme learning machine is suitable for data prediction in the case of large samples. In general, the more training data, the better the fit of the sample data and the smaller the prediction error. Therefore, the gray extreme learning machine prediction algorithm uses a small amount of sample data before the prediction time point in the sample set to model the gray prediction algorithm and uses all the sample data in the sample set to train the extreme learning machine. After that, it then weights and sums the different prediction results predicted by the two models on the test sample to obtain the final prediction result.

Table 4 Statistical table of performance test results of gray fuzzy prediction model based on neural network

NO	Grey fuzzy model(ms)	Grey extreme learning machine model(ms)	Combination model(ms)	NO	Grey fuzzy model(ms)	Grey extreme learning machine model(ms)	Combination model(ms)
1	197.5	198.6	192.6	21	203.2	169.8	167.0
2	204.9	195.7	172.7	22	223.3	186.4	206.5
3	239.0	173.6	181.8	23	222.7	181.6	177.3
4	204.1	207.8	190.9	24	201.9	168.1	192.2
5	247.0	162.1	179.9	25	186.2	204.5	179.3
6	253.0	199.2	177.2	26	239.5	173.0	205.4
7	221.5	205.0	183.6	27	252.7	167.5	160.2
8	235.0	186.0	183.0	28	198.9	203.2	161.0
9	190.0	189.3	196.3	29	213.9	169.4	174.6
10	252.2	177.2	190.5	30	187.8	167.8	166.9
11	227.8	181.2	206.4	31	247.2	208.1	191.2
12	237.2	165.0	188.4	32	216.2	209.5	199.7
13	209.1	202.5	198.4	33	181.8	176.4	186.1
14	202.5	161.6	192.1	34	181.8	201.6	169.1
15	194.8	180.1	177.0	35	251.9	200.5	203.6
16	241.2	193.4	205.4	36	182.3	174.3	191.6
17	205.9	167.6	171.1	37	183.6	178.9	175.8
18	235.8	160.8	193.0	38	232.3	207.8	203.7
19	206.0	198.6	173.2	39	190.9	194.3	195.9
20	193.9	191.6	171.1	40	187.9	177.2	190.5

Table 5 Statistical table of performance test results of gray extreme learning machine model based on neural network

NO	Grey fuzzy model(%)	Grey extreme learning machine model(%)	Combination model(%)	NO	Grey fuzzy model(%)	Grey extreme learning machine model(%)	Combination model(%)
1	77.7	69.7	92.6	21	78.2	69.3	92.8
2	80.7	73.7	94.9	22	80.1	68.1	95.7
3	79.1	76.0	92.6	23	76.0	70.8	93.7
4	78.0	69.8	95.3	24	78.4	75.7	94.0
5	79.8	70.3	95.9	25	77.9	79.5	95.7
6	82.4	72.3	93.3	26	78.1	76.3	92.2
7	85.3	71.0	92.1	27	83.6	79.6	92.1
8	77.4	73.2	95.0	28	78.7	77.2	92.4
9	81.7	80.7	95.8	29	85.1	76.2	93.4
10	79.5	80.7	92.7	30	75.6	71.2	93.0
11	75.6	72.0	94.2	31	78.3	69.9	93.4
12	85.9	74.8	95.3	32	82.4	79.3	95.9
13	84.3	78.7	94.4	33	85.8	74.3	93.3
14	78.2	78.0	92.5	34	81.2	68.4	92.0
15	82.0	70.9	92.9	35	78.4	70.4	92.4
16	75.8	73.9	92.3	36	87.3	73.4	93.1
17	77.0	81.6	94.4	37	81.8	80.5	92.6
18	76.2	72.2	94.7	38	79.7	72.6	94.6
19	87.4	75.4	93.8	39	76.5	73.3	94.3
20	84.3	77.8	95.8	40	84.1	70.5	94.0

6 Model prediction effect analysis

In this paper, a gray fuzzy prediction model and a gray extreme learning machine prediction model based on neural networks are constructed. In addition, a gray fuzzy extreme learning machine prediction combination model is constructed. This paper compares the prediction performance of the three models. This section takes a football league as an example to analyze and read the data to obtain sample data to be processed. Table 1 shows the distribution of observation data. When it is drawn into a statistical graph, the result is shown in Figs. 5, 6, 7, 8, 9 and 10.

There are 5 columns of data to be trained, representing the statistics of the Jiangsu Suning competition from the 5th to the 9th round. Moreover, the first 19 rows are the input data of the BP neural network model, and the last row is the output result. The sample data to be trained is obtained by using the inverse cotangent function to convert the data to be processed and normalized, as shown in Table 2.

Finally, the processed data is used as the input of the neural network. The number of learning is set to 10,000 times, and the root mean square error (RMSE) is displayed every 1000 times of learning, and finally the predicted winning rate is obtained. This article uses a professional league game as an example to randomly combine all these games to form 60 sets of data. After that, the gray fuzzy prediction model based on neural network, gray extreme learning machine prediction model, and gray fuzzy extreme learning machine prediction combination model based on neural network constructed in this paper are used to predict the results of these games. The results obtained are shown in Table 3 and Fig. 11.

From the statistical results in Fig. 11, among the models constructed in this article, the gray fuzzy model has the lowest accuracy in predicting the outcome of the game, and the gray extreme learning machine model has a higher prediction accuracy than the gray fuzzy model. The prediction accuracy of the two combined models is basically above 70%, which is a higher probability in football match prediction. It can be seen that the combined prediction model based on neural network constructed in this paper has a good performance in football match prediction.

Next, this article analyzes the performance of the above models in football matches. This article mainly compares the data processing speed predicted by the model, the accuracy of data transmission, and the competition analysis score. The results obtained are shown in Tables 4, 5 and 6. Moreover, this article draws a statistical diagram of data processing speed, data transmission accuracy, and

Table 6 Statistical table of performance test results of combined model based on neural network

NO	Grey fuzzy model	Grey extreme learning machine model	Combination model	NO	Grey fuzzy model	Grey extreme learning machine model	Combination model
1	63.7	76.1	92.3	21	56.8	56.5	87.7
2	62.9	58.6	88.3	22	60.2	59.5	91.2
3	57.8	56.3	89.6	23	52.3	66.8	88.0
4	64.6	65.8	93.2	24	50.8	70.6	89.0
5	59.5	71.4	88.3	25	60.3	76.5	90.1
6	57.8	67.7	93.9	26	49.8	57.5	92.3
7	52.2	66.2	90.8	27	51.0	75.4	91.8
8	50.4	60.8	90.6	28	54.1	58.6	93.8
9	59.9	74.9	93.4	29	50.6	61.3	93.8
10	57.6	63.1	93.6	30	60.0	71.4	92.5
11	59.5	73.6	88.6	31	58.2	64.6	89.7
12	55.2	68.8	90.7	32	52.5	69.8	88.4
13	51.1	63.5	90.8	33	62.1	65.7	89.2
14	53.7	76.5	88.4	34	50.2	70.9	91.3
15	58.0	70.4	88.7	35	62.4	55.2	87.8
16	51.8	60.3	92.7	36	59.4	70.9	88.4
17	54.0	57.4	92.3	37	53.1	72.9	87.6
18	56.1	70.4	88.3	38	49.4	73.3	92.2
19	59.1	63.4	90.8	39	62.1	58.7	92.7
20	51.8	60.5	87.3	40	62.8	59.8	92.1

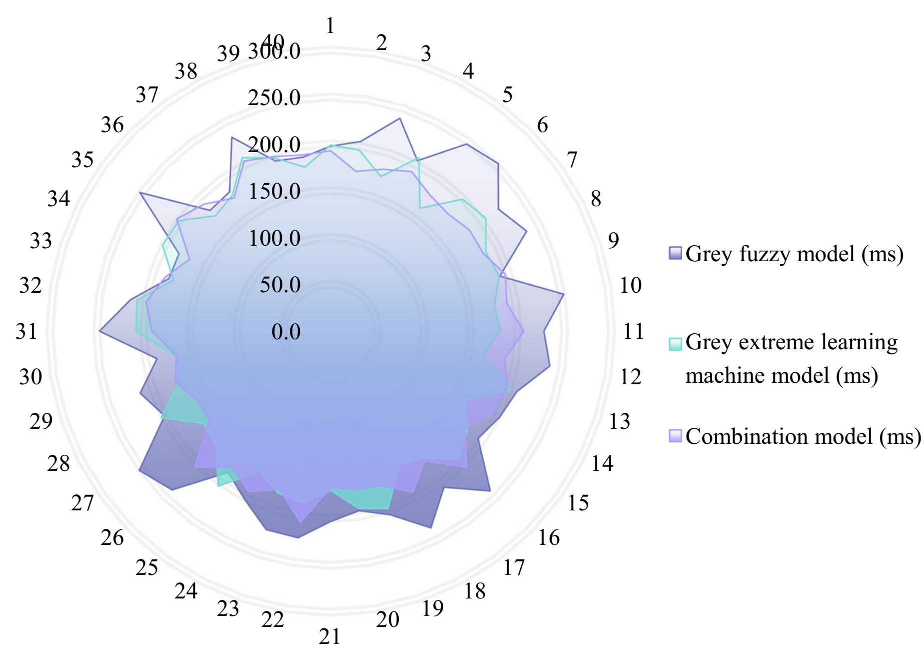


Fig. 12 Statistical diagram of performance test results of gray fuzzy prediction model based on neural network

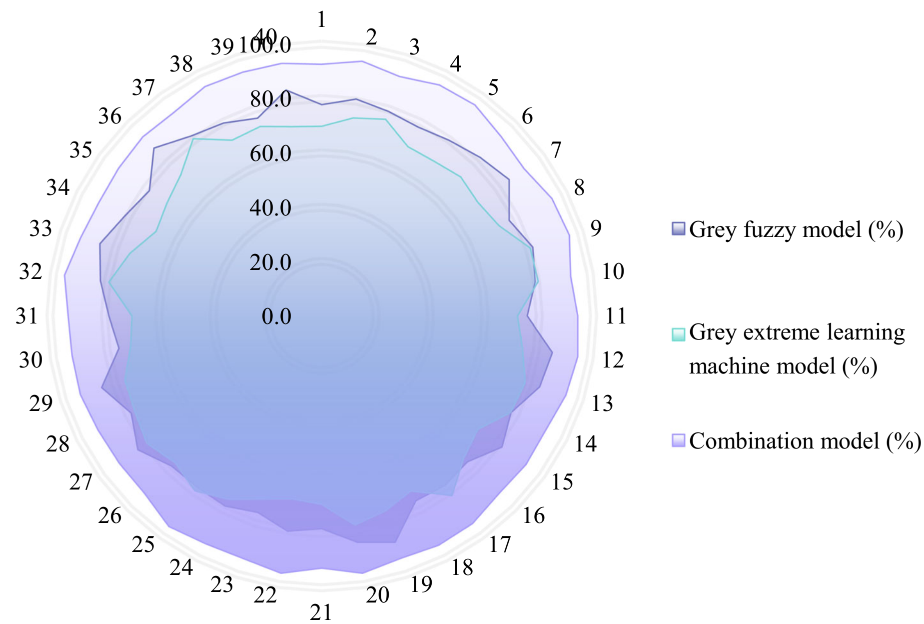


Fig. 13 Statistical diagram of the performance test results of the gray extreme learning machine model based on the neural network

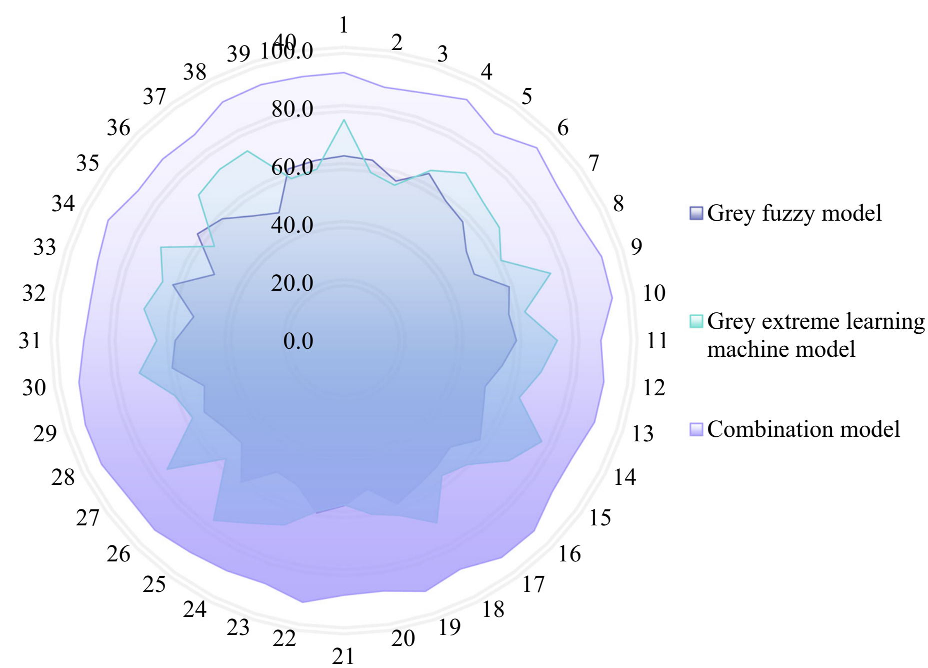


Fig. 14 Statistical diagram of performance test results of combined model based on neural network

competition analysis score. The results obtained are shown in Figs. 12, 13 and 14.

As shown in the above figure, the combined model constructed in this paper performs very well in terms of data processing speed, data transmission accuracy, and match analysis scores for football match predictions, and meets the time requirement. However, the gray fuzzy prediction model and the gray extreme learning machine prediction model are not ideal in terms of data processing speed, data transmission accuracy, and competition analysis scores. It can be seen that the combined model can retain the advantages of a single model and effectively improve the prediction accuracy of the model and the performance of the system.

7 Conclusion

With the continuous development of the commercialization of football, football matches focusing on various big and small matches have become an indispensable part of the lives of football fans, and football quiz is also a topic of discussion among fans. Fans use different methods to predict the outcome of the game, and many experts and scholars at home and abroad are also committed to exploring better football quiz methods. Therefore, predicting the outcome of football matches through different algorithms is a topic of long-term research. This article takes the football match as the research object and analyzes the factors that influence the relationship between the outcome of the football match. Due to the diverse characteristics of these factors, we cannot use a general linear relationship to describe the inherent relationship between them and the outcome of the game. In response to this situation, this article chooses a BP neural network model with powerful processing capabilities for nonlinear data. According to requirements, a gray fuzzy prediction model based on neural network, a gray extreme learning machine prediction model, and a gray fuzzy extreme learning machine prediction combination model based on neural network are constructed. Moreover, this article uses a professional league game as an example to randomly combine all these games to form 60 sets of data. After that, this paper uses the neural network-based gray fuzzy prediction model, gray extreme learning machine prediction model, and neural network-based gray fuzzy extreme learning machine prediction combined model to predict football matches. The results show that the combined model can retain the advantages of a single model and effectively improve the model's prediction accuracy and system performance.

Acknowledgements The study was supported by "2017 Liaoning Province Higher College Basic Scientific Research Project, (Grant No. WQN2017ST03)and (Grant No. WQN2017ST07)".

Declarations

Conflict of interest The authors declared that they have no conflicts of interest to this work. We declare that we do not have any commercial or associative interest that represents a conflict of interest in connection with the work submitted.

References

1. Arabzad SM, Tayebi Araghi ME, Sadi-Nezhad S et al (2014) Football match results prediction using artificial neural networks; the case of Iran Pro League. J Appl Res Ind Eng 1(3):159–179
2. Tümer AE, Koçer S (2017) Prediction of team league's rankings in volleyball by artificial neural network method[J]. Int J Perform Anal Sport 17(3):202–211
3. Igiri CP (2015) Support Vector Machine–Based Prediction System for a Football Match Result[J]. IOSR Journal of Computer Engineering (IOSR-JCE), 17(3): 21–26.
4. Leung CK, Joseph KW (2014) Sports data mining: predicting results for the college football games[J]. Procedia Comput Sci 35:710–719
5. Igiri CP, Nwachukwu EO (2014) An improved prediction system for football a match result. IOSR J Eng (IOSRJEN) 4(12):12–20
6. Bunker RP, Thabtah F (2019) A machine learning framework for sport result prediction. Appl Comput Inf 15(1):27–33
7. Poole VN, Breedlove EL, Shenk TE et al (2015) Sub-concussive hit characteristics predict deviant brain metabolism in football athletes. Dev Neuropsychol 40(1):12–17
8. Katircioglu I, Tekin B, Salzmann M et al (2018) Learning latent representations of 3d human pose with deep neural networks. Int J Comput Vision 126(12):1326–1341
9. Zhang Y, Shen T, Ji X et al (2018) Residual highway convolutional neural networks for in-loop filtering in HEVC. IEEE Trans Image Process 27(8):3827–3841
10. Korotyeyeva T, Tushnytskyy R, Kulyk V (2018) Applying Neural Networks to Football Matches Results Forecasting[C]//2018 IEEE 13th International Scientific and Technical Conference on Computer Sciences and Information Technologies (CSIT). IEEE, 1: 278–282.
11. Dubbs A (2018) Statistics-free sports prediction[J]. Model Assist Stat Appl 13(2):173–181
12. Bačić B (2016) Predicting golf ball trajectories from swing plane: An artificial neural networks approach. Expert Syst Appl 65:423–438
13. Carey DL, Ong K, Morris ME et al (2016) Predicting ratings of perceived exertion in Australian football players: methods for live estimation. Int J Comput Sci Sport 15(2):64–77
14. Polson NG, Sokolov VO (2017) Deep learning for short-term traffic flow prediction. Transp Res Part C: Emerg Technol 79:1–17
15. Strnad D, Nerat A, Kohek Š (2017) Neural network models for group behavior prediction: a case of soccer match attendance. Neural Comput Appl 28(2):287–300
16. Komaris DS, Pérez-Valero E, Jordan L et al (2019) Predicting three-dimensional ground reaction forces in running by using artificial neural networks and lower body kinematics. IEEE Access 7:156779–156786
17. Constantinou AC (2019) Dolores: a model that predicts football match outcomes from all over the world. Mach Learn 108(1):49–75
18. Peterson KD (2018) Recurrent neural network to forecast sprint performance. Appl Artif Intell 32(7–8):692–706
19. Tax N, Joustra Y (2015) Predicting the Dutch football competition using public data: A machine learning approach. Trans Knowl Data Eng 10(10):1–13
20. Cho Y, Yoon J, Lee S (2018) Using social network analysis and gradient boosting to develop a soccer win–lose prediction model. Eng Appl Artif Intell 72:228–240
21. Martins RG, Martins AS, Neves LA et al (2017) Exploring polynomial classifier to predict match results in football championships. Expert Syst Appl 83:79–93
22. Anfilets S, Bezobrazov S, Golovko V et al (2020) Deep multilayer neural network for predicting the winner of football matches. International Journal of Computing 19(1):70–77
23. Cornforth D, Campbell P, Nesbitt K et al (2015) Prediction of game performance in Australian football using heart rate variability measures[J]. International Journal of Signal and Imaging Systems Engineering 8(1–2):80–88

24. Baboota R, Kaur H (2019) Predictive analysis and modelling football results using machine learning approach for English Premier League[J]. Int J Forecast 35(2):741–755
25. Visbal-Cadavid D, Mendoza AM, Hoyos IQ (2019) Prediction of efficiency in Colombian higher education institutions with data envelopment analysis and neural networks. Pesquisa Operacional 39(2):261–275
26. Bataineh M, Marler T, Abdel-Malek K et al (2016) Neural network for dynamic human motion prediction. Expert Syst Appl 48:26–34
27. He T, Mao H, Guo J et al (2017) Cell tracking using deep neural networks with multi-task learning. Image Vis Comput 60:142–153
28. Constantinou A, Fenton N (2017) Towards smart-data: Improving predictive accuracy in long-term football team performance. Knowl-Based Syst 124:93–104
29. Angelini G, De Angelis L (2017) PARX model for football match predictions. J Forecast 36(7):795–807