Tennis Match Prediction Based on Wavelet-BP Neural Network Model

Yikun Shi*

School of Resources and Environmental Engineering, Anhui University, Hefei, China, 230601

* Corresponding Author Email: x52214033@stu.ahu.edu.cn

Abstract. In sports, assessing a player's performance during a match and the impact of momentum on the outcome of the match are key issues, but are often difficult to measure accurately. In this study, we first preprocessed the match data into serve and non-serve data and then developed a Wavelet-BP neural network model to analyze the player's performance at specific times of the match. It was finally obtained that Carlos Alcaraz performed best in the interval between [2500,5000] and [7000,8000] (s) of the match, and performed poorly after 10000 seconds. Meanwhile, by quantifying the momentum into indicators such as winning streak and whether to serve or not, and verifying its correlation with the match results by chi-square test, the results show that momentum has a significant effect on the match results. This study provides a new method to assess players' performance in a match and quantify the effect of momentum on the match outcome. This is important for improving tactical strategies in sports competitions, enhancing team competitiveness, and increasing the spectacle of the game.

Keywords: Wavelet-BP Neural Network Model, Chi-square test, Momentum.

1. Introduction

The Wimbledon Gentlemen's final saw Spanish rising star Carlos Alcaraz defeat Novak Djokovic, in an emphatic victory that ended Djokovic's unbeaten run at the Wimbledon Open since 2013. The match was very close throughout, with the two men fighting it out. So, this match caused us to think about some things. Momentum is defined as "strength or force gained by motion or by a series of events" in one dictionary. In sports, a team or individual may believe they have the "strength/force" or momentum during a contest, although it can be challenging to quantify this feeling. Even if momentum exists, it is not immediately clear how different things that happen during the match work to influence or produce it.

Many models exist for predicting matches. Li H. (2020) to keep abreast of the development of sports events and adjust the strategy in the process of the game promptly, the traditional back propagation neural network (BPNN) algorithm is improved [1], and adaptive BPNN is utilized to construct the game prediction model, and the improved adaptive BPNN algorithm has a small prediction error after rolling prediction with a high degree of accuracy and reliability [2]. Hodge V J et al. (2019) Optimized by machine learning model and feature engineering, their model has an accuracy of 85% within 5 minutes after the start of the match [3]. Rahman M A. (2020) proposed a deep neural network framework to predict the outcome of soccer matches using datasets including rankings, team performance, etc. The study showed that the framework performed well in predicting the 2018 FIFA World Cup matches with an accuracy of 63.3% [4]. Guan S, and Wang X. (2022) chose a BP neural network model to construct a soccer match prediction model and proposed a neural network-based gray fuzzy limit learner prediction combination model [5]. Through the prediction and analysis of soccer league data, the research results show that the model can effectively improve the prediction accuracy and system performance [6]. Zhang Q, Zhang X, and Hu H (2022) proposed a long short-term memory (LSTM) network based on the attention mechanism for sports match prediction [7]. The model improves the prediction performance by focusing on key information such as player performance and match progress [8].

Despite the progress made in the above studies, there are still some shortcomings. First, some studies may not have fully considered the quality and completeness of the data, resulting in the

predictive accuracy of the model being affected. Second, some studies have not fully explored the stability and interpretability of the model and lacked the interpretation and validation of the model results. In addition, the empirical data of some studies may be too limited and lack comprehensive consideration of different events, scenarios, and factors. In sports, a team or player may feel that they have momentum, or "power/strength" in a match/game, but it is difficult to measure this phenomenon. Evaluating the performance of players in a game and the impact of momentum on the outcome of the game is one of the key issues in predictive modeling in sports, but it is often difficult to measure accurately [9]. Future research needs to consider how to fully utilize and integrate these difficult-to-quantify factors to improve models' predictive power and utility.

In this paper, the study not only used the Wavelet-BP neural network model to analyze the game data, but also combined with the chi-square test to validate the momentum, and this comprehensive analysis method can more comprehensively understand and explain the formation mechanism of the game results, and this model combines the advantages of Wavelet analysis for data degradation and feature extraction, and the advantages of BP neural network for pattern recognition and prediction [10] so that it can be more effective in analyzing and predicting the game data.

2. Wavelet-BP Neural Network Model

2.1. Wavelet analysis cycle model

Wavelet analysis periodic modeling is a commonly used analytical method in finance that combines wavelet analysis and periodic modeling. First, the time series data are decomposed by wavelet decomposition to detect the cyclical components in them.

Table 1. Pseudo-code for wavelet analysis

Algorithm 1 Wavelet Analysis Periodic Model
Input: Time series data :D.
1: Perform wavelet decomposition on D to obtain Wavelet coefficients WCC: WCC =wavelet-transform(D).
2: Analyze periodic patterns to obtain Periodic features PF: PF = Analyzation (WCC).
3: Choose an appropriate parameter P of periodic model for fitting: P =Fitting (PF).
4: Predict future periodic patterns Results = Prediction(P)
Output: Prediction results: Results.

This approach combines the strengths of wavelet analysis to efficiently handle non-stationary and non-linear financial time series data and provide insights into cyclical variations and forecasting capabilities. Its pseudo-code is shown in Table 1.

$$\boldsymbol{W_f}(\mathrm{a}, \boldsymbol{b}) = \int_{-\infty}^{+\infty} \boldsymbol{f}(\boldsymbol{t}) \frac{\mathbf{1}}{|\boldsymbol{a}|} \boldsymbol{\psi}^* \left(\frac{\boldsymbol{t} - \boldsymbol{b}}{\boldsymbol{a}} \right) \boldsymbol{dt} \tag{1}$$

where $W_f(\mathrm{a}, b)$ is Wav et coefficients, which represent localized features of the signal f(t) at time point b and scale a. $\psi^*(\mathrm{t})$ denotes the complex conjugate of the wavelet basis function ψ(t). **a** determines the resolution of the detection of signal details, with a larger corresponding to lower frequencies or slower variations and a smaller corresponding to higher frequencies or faster variations. and $\boldsymbol{b}$ is Specify at which point in time to perform the analysis under this scale.

$$\boldsymbol{Envelope_t} = \frac{\mathbf{1}}{\boldsymbol{N}} \sum_{\boldsymbol{i=t-N+1}}^{\boldsymbol{t}} |\boldsymbol{W_f}(\mathbf{a}, \boldsymbol{b})| \quad (2)$$

where $Envelop$ denotes the value of the envelope at time t. N indicates the size of the moving average widow. And $|W_f(\mathrm{a}, b)|$ is denotes the absolute value wavelet coefficients.

2.2. BP neural network neural model

BP neural network is a multi-layer feed-forward neural network trained according to the error back-propagation algorithm, which can iterate and repair its weights according to the relationship between the input and output variables, to estimate the exact functional relationship finally. The network structure is shown in Figure 1.

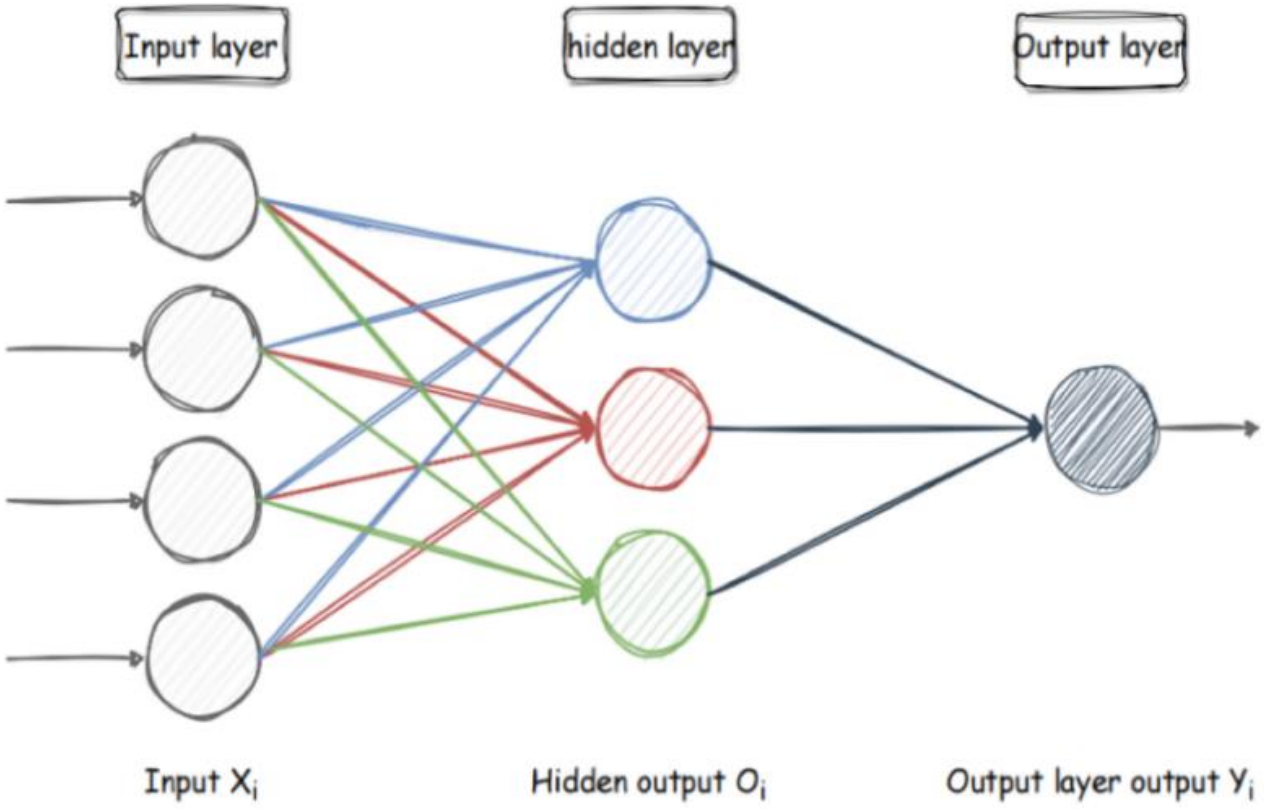

Figure 1. BP neural network neural model schematic

(1) The BP neural network is linked by different node coefficients. When connecting weights and weights are positive, it indicates that the current link is exciting. Conversely, if the link coefficient is negative, the link state is a state of suppression.

(2) The input signal and the linear signal are the combination of the signals for each input signal.

(3) The function of the nonlinear activation function: making the neuron output signal within a certain range.

2.3. chi-square test

The Chi-Square Test is a statistical method used to determine whether there is a significant difference between observed and expected frequencies. It is usually used to analyze the relationship between two or more categorical variables, especially in columnar tables is more widely used, the main process is first to establish the hypothesis, develop a table of observed and expected frequencies, calculate the chi-square statistic, determine the degrees of freedom, find the critical value, carry out the hypothesis test, and make a conclusion. The Chart of chi-square test flow is shown in Figure 2.

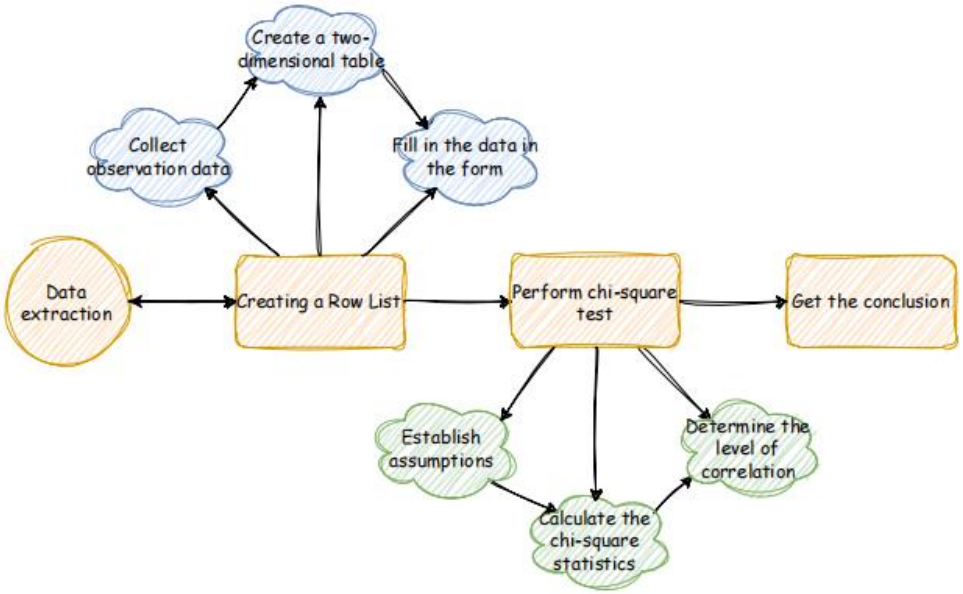

Figure 2. Chart of chi-square test flow

$$\chi^2 = \sum \frac{(O_i - E_i)^2}{E_i} \tag{3}$$

where O_i is denotes the i-th observed frequency.E_i is denotes the i-th desired frequency.

3. Results

3.1. Wavelet analysis cycle model

Wavelet analysis has a wide range of applications in the fields of signal processing, data analysis, and pattern recognition, which can help extract useful information from signals and improve signal processing. In the 2023 Wimbledon Gentlemen's final, a wavelet sub-model was used for the Carlos Alcaraz race data to be processed. The following figure 3 is obtained.

As we considered the great influence of serve and non-serve on the score of the game, we prevented this factor from affecting our analysis of the original data, so we analyzed the data separately. Finally, the two types of data were reduced according to the chronological order to obtain the final data processed by wavelet analysis as shown in Figure 4.

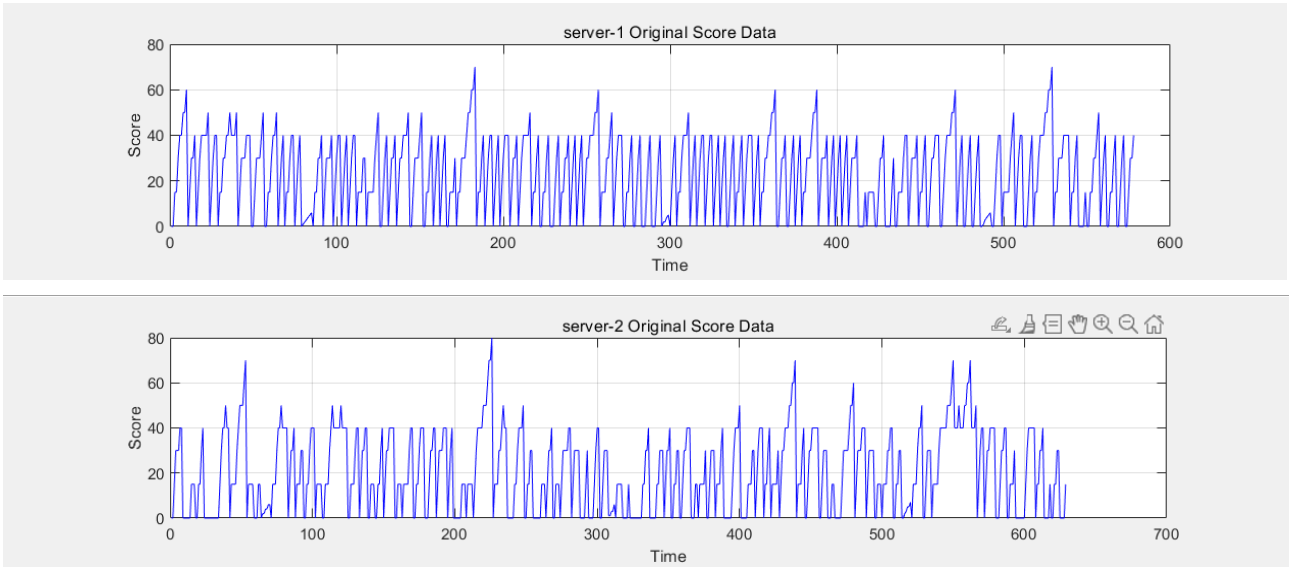

Figure 3. Raw data versus smoothed envelopes

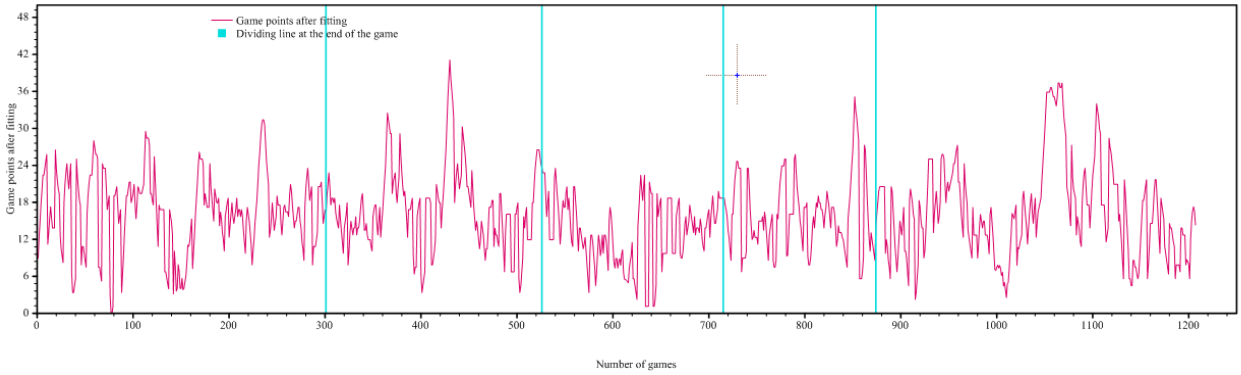

Figure 4. Smooth Envelope Folding Chart

Firstly, the Morlet wavelet transform is used to extract features and fit to the data, and then smooth wavelet coefficients envelope is used for signal processing, data analysis, and feature extraction can help to improve the signal quality, extract useful information, and simplify the process of data analysis. Finally, we can observe from the figure: that there is a rough cycle of data changes, but due to the original data value changes after a game there is a sudden decline in the value, making the value unstable. After the modeling process, the values tend to be smoothed out, paving the way for later analysis.

3.2. BP neural network neural model

The data was imported, the first four games data was used as the training set and the fifth game data was used as the test set for training, and the post-training data was visualized as a three-dimensional scatterplot with a two-dimensional scatterplot with the game time as the x-axis. The final results can be seen in Figure 5.

Meanwhile, according to the formula for accuracy, the accuracy rate is obtained as Accuracy=76.32%. (This is more accurate than the 62.57% accuracy obtained by using the BP neural network directly).

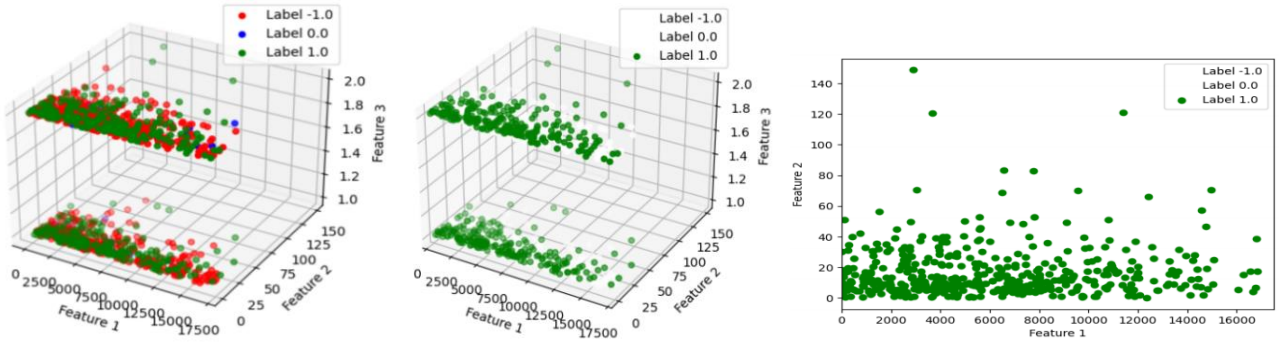

Figure 5. Wavelet-BP Neural Network Model analyzing result diagram

Here category 1 reflects the excellent performance data, -1 is the poor performance data, and 0 is the data at the end of a game. It can be noticed that the green dots are denser when the game is played within 8000 seconds compared to after 8000 seconds, and it is easy to observe that the densest densities are in the intervals of 2500 to 5000 and 7000 to 8000.

Combining the graph and the figure, it can be concluded that the intervals of [2500,5000] and [7000,8000] are the best times for a player's performance during the match. And after 10,000 seconds, the player's performance slips. It can be seen that player C's lack of physical fitness affects the performance of the game.

3.3. Chi-square test

We analyze momentum in an abbreviated quantitative way, where we consider that a player's momentum increases when he or she is on a winning streak and decreases when he or she loses, and

we analyze whether the current momentum affects a win or a loss in the following match. We define momentum as when a player wins a match. To use the data more widely, we consider the momentum of a player winning a game here as plus one, and the momentum of a losing game as minus one, to quantitatively describe the momentum. The data is then manipulated as described above to obtain a chi-square test plot, shown in Figure 6.

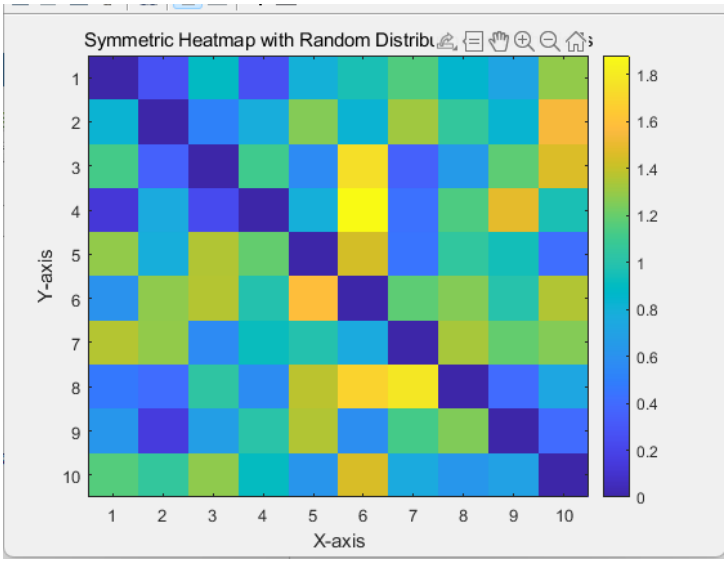

Figure 6. Chi-square diagram

The X-axis marked from 1 to 10 and the Y-axis marked from 1 to 10 may represent the order in which players or teams played, while the shade of the color shows the degree of momentum (winning or losing streak). Yellow blocks may indicate higher positive momentum (more wins), while dark blue blocks may indicate negative momentum (losses). The color change from blue to yellow implies different levels of momentum in different matches.

The chi-square test is used here to determine if there is a statistically significant relationship between the current momentum (winning or losing streak) and the outcome of the next match (victory or defeat). The heat map itself does not directly provide statistical results but rather visualizes the data used for such tests. The color shade of each cell indicates the degree of association between momentum (the sum of 1 and -1) and the outcome in the sequence of matches.

The chi-square test metrics were then obtained and are shown in Table 2.

Table 2. Indicators of chi-square test

Chi-square statistic	degrees of freedom	Significance
38.181	7	0.05

In this test we choose the significant level (0.05) because the degree of freedom of the data we screened is 7, so we get the critical value of 20.28 by looking up the table, we calculated the value of 38.181 is greater than this critical value, so the original hypothesis is rejected, so in the case of the significant level of 0.05 there is a significant relationship between the momentum and the player's victory or defeat data.

4. Conclusions

In this article, we analyze the performance of players at specific times of the match and determine that momentum is also an important factor affecting the match. The given data is preprocessed and considering the effect of serve on the player's performance, the data is categorized into serve data and non-serve data in this paper. Then to analyze the player's performance at a specific time of the match, this paper develops a Wavelet-BP Neural Network Model. using BP Neural Network to analyze the data after fitting the wavelet analysis, we finally get that Carlos Alcaraz performs the best in the

interval between [2500,5000] and [7000,8000] (s) of the match, and performs poorly after 10000 s. The data is analyzed by BP Neural Network. Then, to analyze whether momentum affects winning or losing a match, we need to determine the correlation between the two. Since momentum is a concept, we quantified the momentum into the winning streak, whether to serve or not, observed the general relationship in the line list, and finally verified the correlation through the chi-square test, and the final chi-square statistic of 38.181 is larger than the rejection domain of our hypothesis, so the momentum has a significant impact on the game.

The model in this paper can be used in practice to help coaches and athletes better understand the performance trends and key moments of players in a match, to optimize training plans and tactical strategies. Secondly, the model can be used to predict players' performance in matches, which helps spectators and bettors better understand players' competitive status and tournament results, and improves the enjoyment and participation in competitive games. In addition, the establishment of this model can also help to promote research and exploration in the field of sports science and deeply understand the mechanisms and influencing factors of athletes' performance.

For the outlook of this model. We think the model can be improved by adding data dimensions and features, such as information about the player's physical performance index and psychological state, to improve the accuracy and comprehensivenes

s of the model. Secondly, improve the model algorithm and structure, try to use more advanced neural network structure or other machine learning algorithms to improve the performance and generalization ability of the model. In addition, the model can be applied to other sports or competitive scenarios, such as tennis, basketball, soccer, etc., to extend its application scope. Finally, combining real-time data and intelligent decision-making, an intelligent decision-making system can be developed to provide real-time game analysis and tactical advice to coaches and players, helping them make smarter decisions and adjustments during the game.

References

[1] Zhang J, Qu S. Optimization of backpropagation neural network under the adaptive genetic algorithm[J]. Complexity, 2021, 2021: 1-9.

[2] Li H. Analysis on the construction of sports match prediction model using neural network[J]. Soft Computing, 2020, 24(11): 8343-8353.

[3] Hodge V J, Devlin S, Sephton N, et al. Win prediction in multiplayer esports: Live professional match prediction[J]. IEEE Transactions on Games, 2019, 13(4): 368-379.

[4] Rahman M A. A deep learning framework for football match prediction[J]. SN Applied Sciences, 2020, 2(2): 165.

[5] Machado R J, Barbosa V C, Neves P A. Learning in the combinatorial neural model[J]. IEEE Transactions on Neural Networks, 1998, 9(5): 831-847.

[6] Guan S, Wang X. Optimization analysis of football match prediction model based on neural network[J]. Neural Computing and Applications, 2022, 34(4): 2525-2541.

[7] Yu Y, Si X, Hu C, et al. A review of recurrent neural networks: LSTM cells and network architectures[J]. Neural Computation, 2019, 31(7): 1235-1270.

[8] Zhang Q, Zhang X, Hu H, et al. Sports match prediction model for training and exercise using attention-based LSTM network[J]. Digital Communications and Networks, 2022, 8(4): 508-515.

[9] Fry M J, Shukairy F A. Searching for momentum in the NFL[J]. Journal of Quantitative Analysis in Sports, 2012, 8(1).

[10] Dong Y, Fu Z, Peng Y, et al. Precision fertilization method of field crops based on the Wavelet-BP neural network in China[J]. Journal of Cleaner Production, 2020, 246: 118735.