

# Race momentum analysis prediction based on wavelet neural networks

Jiale Hu \*

School of International Education, Neusoft Institute, Foshan, China, 528225

\* Corresponding Author Email: Jiale2.Hu@live.uwe.ac.uk

**Abstract.** Tennis is a popular sport around the world and 'momentum' plays a key role in determining the outcome of matches, as discussed in this report. Our team created a momentum model for tennis matches by analysing seven factors, including points scored, serve ends, and using an activity index and an error index as state scores and demerits. The model was based on a sample of the 2023 Wimbledon 1301 Championships and showed changes in athlete form through line graphs. Through Monte Carlo simulation, the experiment found that the momentum distribution of actual match results differed significantly from the simulated random distribution, indicating that momentum had a significant effect on match results. And the change in dominance in the middle of the match is predicted by a wavelet neural network. The innovative approach using wavelet neural networks also allows for more detailed analysis of momentum changes during a race, providing a new perspective on the dynamics of athletic performance.

**Keywords:** Wavelet Neural Network Prediction, Monte Carlo Simulation Algorithm, Multifactorial Rating.

## 1. Introduction

In recent research on the game process, Zhou explored the potential for improving the quality of the serve through a multifactorial analysis of tennis [1]. Li C, Ji Y, Liu J et al. used logistic regression models and BP neural networks to study the application of momentum in e-sports [2]. Meng X, Tian L, Zhang L studied the dynamics of momentum in tennis by analyzing data from tennis matches via PCA [3]. G. Kubitz and L. Page Design employed a randomized experimental design to generate random chances of winning, identify causal effects of subsequent performance, and analyze the causes of positive and negative momentum [4]. Hitar-Garcia J A, Moran-Fernandez L, Bolon-Canedo V. Prediction of win/loss outcomes in League of Legends games solved by convolutional networks [5]. Song K and Shi J developed a real-time prediction model based on a gamma process with the objective of forecasting the outcome of NBA games [6]. This article employs wavelet neural networks and Monte Carlo simulation methods to model momentum in tennis matches and to demonstrate and analyze the impact of momentum on the outcome of the match.

## 2. Athlete Comprehensive Performance Analysis Model

### 2.1. Preliminary preparation of the model

(1) Scoring factor. The initial factor of the model necessitates the introduction of a scoring factor. This factor is evaluated based on the scoring of hits from different angles, and the player is endowed with a certain initial scoring momentum at the moment of hitting the ball. Consequently, when a score is scored, the scoring factor should be:

$$\text{Scoring factor} = S_{ij} \quad (1)$$

The position of the data is indicated by the letter "i", with "i" denoting the number of the *i*-th match. The score number is represented by the letter "j". Therefore,  $S_{ij}$  represents the *j*-th score of the *i*-th match, whether it has been scored or not. It is important to highlight that the rationale behind not utilising  $S_{ij}$  as a "0-1" variable is that the introduction of additional dimensions would be necessary.

Furthermore, the use of a simple "0-1" variable would result in the generation of different dimensions and erroneous outcomes.

(2) Serve side factor, in a standard tennis match, the side that serves tends to have the advantage in terms of scoring points. For the server, the point of the shot is under their control. Choosing inside or outside corners gives a greater range of attack, the combination of flat spin makes the opponent's judgement more inaccurate, and the opponent is more likely to make mistakes when faced with multiple angles as well as high speed balls. It is therefore necessary to introduce a weighted metric to measure scoring shots. This is done to balance the momentum gained by the serving side after scoring a point. This can be calculated using the following formula:

$$\begin{cases} \text{Serve side factor} = w_1 = a_1 + 1 \\ a_1 = \begin{cases} \lambda & \text{when } t_{ij} = 0 \text{ and } S_{ij} = I \\ -\lambda & \text{when } t_{ij} = 1 \text{ and } S_{ij} = I \end{cases} \\ t_{ij} = \begin{cases} 0 & \text{receive side} \\ 1 & \text{serve side} \end{cases} \\ I = 1000 \end{cases} \quad (2)$$

As shown in the equation above, the role of  $S$  is to score goals and the initial score  $I$  is set to 1000 to ensure that goals are scored. At the same time, a "0-1" variable  $t$  is introduced to determine whether the score of the  $j$ -th ball in the  $i$ -th game is for the serving side or the receiving side. If the player is on the serving side, the scoring weight is set to  $1 + \lambda$ . Conversely, the scoring weight for the receiving side is set to  $1 - \lambda$ .

(3) Continuous score factor, in a match, a player has more 'momentum' if he scores consecutive points. This can also be seen as a better performance during this period, so when a player scores consecutive points, their weighting will increase the player's performance score. The formula is therefore as follows:

$$\begin{cases} w_2 = a_2 + 1 \\ a_2 = \begin{cases} \lambda & \text{when } S_{i(j-1)} = I \\ 0 & \text{when } S_{i(j-1)} \neq I \end{cases} \end{cases} \quad (3)$$

The weights here are determined by the  $(j-1)$ -th score of the  $i$ -th match, where  $a_2 = \lambda$ , if the score is consecutive goals scored by the same player, and 0 otherwise.

(4) Scoring Grip Factor, A review of empirical data and the prevailing circumstances reveals that a forehand shot is more readily achieved with greater ease, whereas a backhand shot is more challenging to control and execute with precision. This is illustrated in tab.1 below.

**Tab 1.** Gesture Description Table.

| Gesture         | Backhands                                                                           | Forehands                                                                            |
|-----------------|-------------------------------------------------------------------------------------|--------------------------------------------------------------------------------------|
| Image from Wiki |  |  |

Consequently, the backhand stroke is regarded as a challenging one, and the following weights have been set in order to ensure that the backhand stroke score is of a certain standard. The batting score pattern, as indicated in tab. 2 below, is in accordance with the data provided in the question.

**Tab 2.** Gesture Correspondence Table.

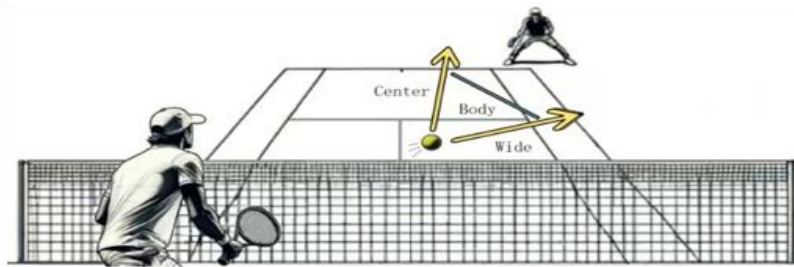
| Point-scoring grip | Serial number | Type          |
|--------------------|---------------|---------------|
| $\delta_{ij}$      | 1             | Backhand      |
|                    | 2             | Forehand      |
|                    | 3             | Service fault |

Thus, the following equation can be listed:

$$\begin{cases} w_3 = a_3 + 1 \\ a_3 = \begin{cases} \lambda & \text{when } \delta_{ij} = 1 \\ -\lambda & \text{when } \delta_{ij} = 2 \\ 0 & \text{when } \delta_{ij} = 3 \end{cases} \end{cases} \quad (4)$$

In this context, specific weights are assigned when the  $\delta_{ij}$  value represents the backhand score, the forehand score, or the opponent's error score.

(5) Hit-point factor, In the previous data processing section we statistically found that there is a link between the direction of the serve and the score of the serve, and here we can see a better direction of the serve as a way for the athlete to perform better, as shown in the following fig.1:



**Fig 1.** Diagram of the direction of serve.

Consequently, the paper presents distinct weighting factors for distinct impact locations, as depicted in tab.3 below (Data source: 2023-Wimbledon-Tennis Match).

**Tab 3.** Table of Serve Direction Correspondence.

| Azimuthal coefficient | Serial number | Direction       | Winning percentage |
|-----------------------|---------------|-----------------|--------------------|
| $\varphi_{ij}$        | 1             | B: Body         | 55.709%            |
|                       | 2             | BC: Body/Center | 59.200%            |
|                       | 3             | BW: Body/Wide   | 59.223%            |
|                       | 4             | C: Center       | 72.674%            |
|                       | 5             | W: Wide         | 76.755%            |

The preceding table indicates that the C and W break direction has a higher winning rate. It can also be observed that driving to the opponent's vicinity facilitates the opponent's ability to receive the ball. Consequently, it can be postulated that players who hit the ball to the C and W positions have a superior driving condition. This leads to the following equation:

$$\begin{cases} w_4 = a_4 + 1 \\ a_4 = \begin{cases} \lambda & \text{when } \varphi_{ij} = 4,5 \\ -\lambda & \text{when } \varphi_{ij} = 1,2,3 \end{cases} \end{cases} \quad (5)$$

Among the formula, when  $\varphi_{ij}$  belongs to 4 or 5 it is considered a high performance serve, and equal to 1, 2, 3 it is considered an average serve, giving some weight to the both.

(6) Activity factor, in order to evaluate the performance of the athlete, it is necessary to consider the distance that the athlete moves during the game. This distance can therefore be regarded as an activity index.

$$\text{Activity Factor} = A_{ij} \quad (6)$$

In light of the aforementioned considerations, it is possible to assign a certain number of performance points to the athlete's distance in order to ensure that the question is considered in its entirety.

(7) Error factor, in tennis, errors are frequently committed during the serve. According to the rules of tennis, two consecutive errors result in the opponent being awarded additional points. Therefore, it can be inferred that in the event of two consecutive errors, the player's performance is suboptimal and a deduction is warranted.

$$\text{Error Factor} = \varsigma_{ij} \quad (7)$$

This factor is a "0-1" variable that determines whether the  $j$ -th score in the  $i$ -th inning was an error that led to an opponent's score.

## 2.2. Momentum Performance Score Model

In the preceding section, the parameters to be included in the model were identified [7]. In this section, these parameters will be integrated into a momentum performance model that describes the athlete's performance at each stage of the game. This is shown in the following equation.

$$\text{Momentum : } M_{ij} = S_{ij} I w_1 w_2 w_3 w_4 + A_{ij} - 2A_{ij} \varsigma_{ij} \quad (8)$$

This tennis momentum model incorporates a variety of factors that influence the outcome of a match. The formula utilises symbols  $2A_{ij} \varsigma_{ij}$  to represent double scoring momentum. In the case of players, the corresponding momentum is subtracted to reflect the effect of errors on the player.

## 2.3. Model solving - Match performance (Using wimbledon-1301 as an example)

Due to the considerable number of predefined values in the aforementioned model, a selection of the variables included in the model weights is presented in the tab.4 below.

**Tab 4.** Model Preset Settings.

| Variables | Values |
|-----------|--------|
| $I$       | 1000   |
| $\lambda$ | 0.2    |

This article employs the Python programming language to perform calculations, extract data from tables, and illustrate the performance of *Carlos Alcaraz* and *Nicolas Jarry* throughout the course of the race. The 2023 Wimbledon tournament is used as an illustrative example. As shown in the fig.2 below.

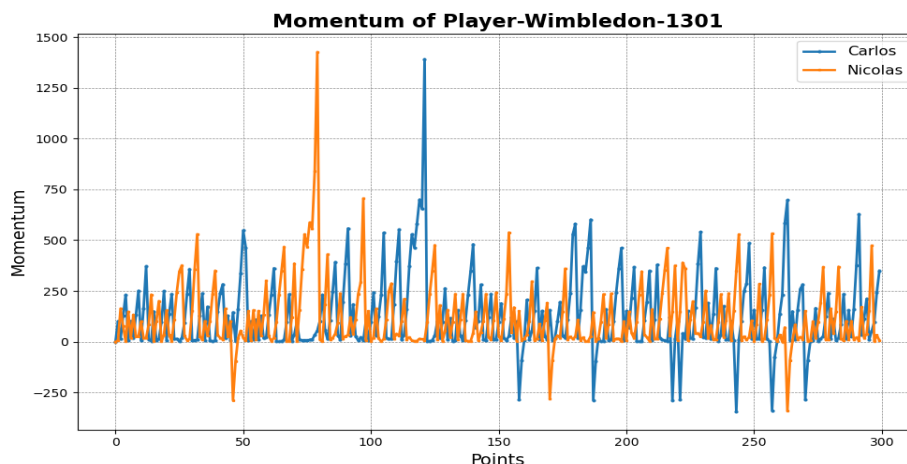


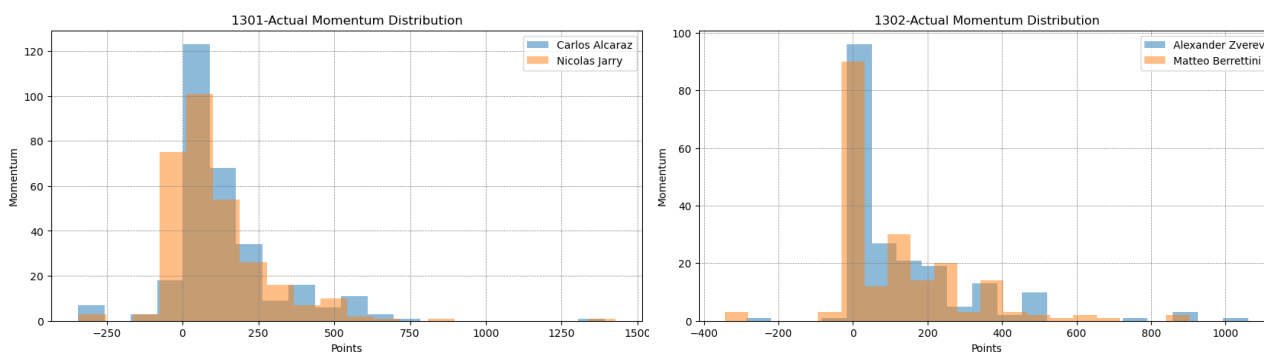
Fig 2. Momentum expression model solution results.

## 2.4. Momentum validity argument based on Monte Carlo simulation

In order to validate the role of "momentum" in the match as previously mentioned and to disprove the idea that fluctuations in the match and the success of the players are random. The following section describes one idea of momentum validity verification that can be used in this study:

In order to assess the effectiveness of momentum in a given context, a set of matched data was first generated through *Monte Carlo simulation* [8]. Subsequently, two graphs were created to show the distribution of these data. By comparing these two graphs, it is possible to analyze whether the data distributions are similar. If the two graphs show similar distribution patterns, this would indicate that momentum is useful in that context [9]. Conversely, if there is a significant difference in the distribution between the graphs, then momentum is considered to be useless. Finally, we proceeded to compare and contrast the two graphs by generating frequency histograms in order to obtain analytical results regarding the usefulness of momentum.

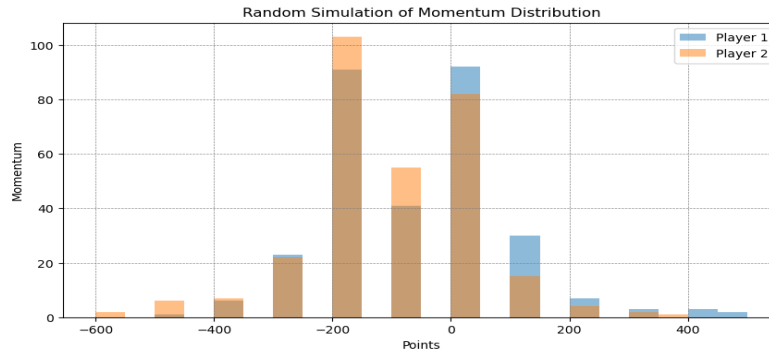
Firstly, we utilize momentum frequency histograms to visualize the momentum results of real matches. The following results are obtained by applying the tennis momentum model to 2023-wimbledon-1301 and 2023-wimbledon-1302, as shown in the fig.3 below.



(a)2023-wimbledon-1301 momentum distribution. (b) 2023-wimbledon-1302 momentum distribution

Fig 3. Momentum distribution of results under the influence of momentum.

Although the processes of the two matches are distinct, their momentum distributions are essentially equivalent. However, *Monte Carlo simulation* was employed to randomly populate the various parameters of the model, and 1000 matches were simulated before applying the results of one of the matches to the tennis momentum model in order to compute its momentum and generate a frequency histogram. The results are shown in fig.4 below.



**Fig 4.** Momentum distribution under Monte Carlo simulation algorithm.

### 3. Competition trend forecasting model: based on wavelet neural network

#### 3.1. Investigation of neural network algorithms combined with momentum

This section will examine the construction of a prediction model designed to anticipate fluctuations in the outcome of a tennis match. The model will be based on a BP neural network prediction model derived from machine learning techniques.

Firstly, several factors must be considered in order to predict the sample. These include the label of the sample, the weight of the label and a constant. The labels and weights are then multiplied in order to obtain the prediction.

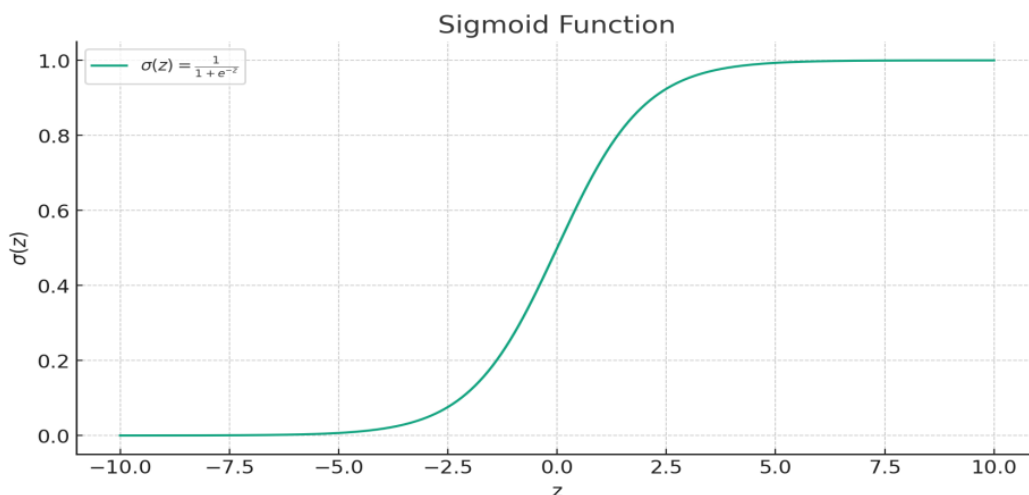
$$\hat{y}_i = \text{dot}(wx_i) + b \quad (9)$$

The symbol  $x_i$  denotes the input feature vector, assuming there are only three features. The vector can be expressed as  $(x_1, x_2, x_3)$ . The symbol  $w$  represents the weight, corresponding to each input feature, representing the importance of each feature. The symbol  $b$  represents the threshold, which is used to influence the prediction result. The symbol  $\hat{y}_i$  represents the prediction result. The symbol  $\text{dot}()$  in Eq. represents the vector multiplication of  $w$  and  $x_i$ .

In addition, we need an activation function that maps the above calculations to the interval [0- 1] to better judge the accuracy of the prediction as follows:

$$\sigma(\hat{y}_i) = \frac{1}{1 + e^{-\hat{y}_i}} \quad (10)$$

Simply plot the result in the interval [0-1] as shown in the fig.5 below:



**Fig 5.** Mapping effects of the sigmoid activation function.

The predictive accuracy of the model is then validated through the application of a loss function, which is calculated as follows.

$$Loss : J(w, b) = \frac{1}{m} \sum_{i=1}^m Loss(\hat{y}_i - y_i) = -\frac{1}{m} \sum_{i=1}^m y_i \log(\hat{y}_i) + (1 - y_i) \log(1 - \hat{y}_i) \quad (11)$$

Where the independent variables  $w$  and  $b$  of the function are calculated, respectively, by the following equations:

$$\begin{cases} w' = w - LearningRate \frac{dJ(w, b)}{dw} \\ b' = b - LearningRate \frac{dJ(w, b)}{db} \end{cases} \quad (12)$$

In this approach, the weights and threshold are iterated in a self-looping fashion with constant gradient descent, adjusting the parameters in the network to minimum the loss function. The loss function measures the difference between the network output and the true target value. Gradient descent enables the network to learn how to accurately represent or predict data [10]. A sample graph of gradient descent is provided below (fig.6).

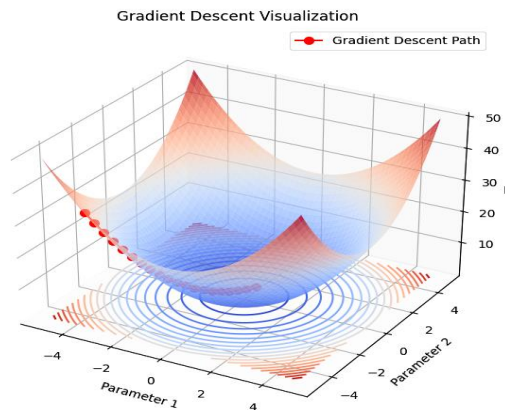


Fig 6. Sample diagram of gradient descent.

### 3.2. Feature and Labeling Selection for Neural Networks

The following features have been selected as model features, as indicated in the accompanying tab.5:

Tab 5. Sample Characteristics Comparison Table.

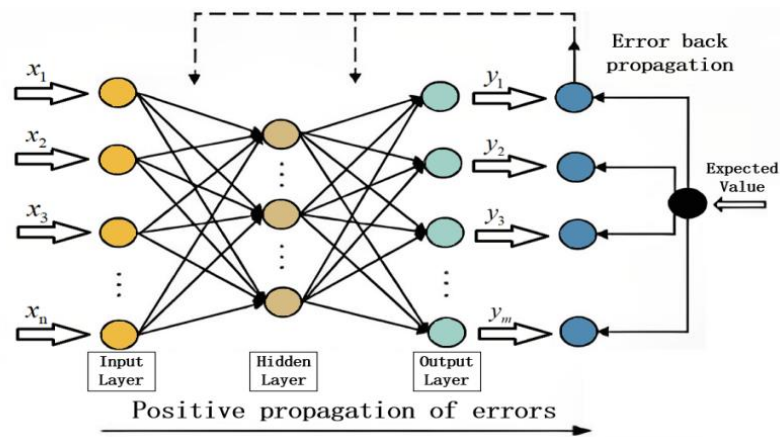
| No.      | $x_1$ | $x_2$ | $x_3$  | $x_4$ | $x_5$        | $x_6$   | $x_7$        |
|----------|-------|-------|--------|-------|--------------|---------|--------------|
| Features | games | score | server | ace   | Double_fault | Unf_err | Distance_run |

In order to ascertain the outcome of the competition, we employed the concept of "dominance shifting points". In the preceding section, we utilized the concept of momentum to observe the performance of the athletes. When the momentum changes, we can infer that the dominance of the race has shifted, which is the concept of "shifting" in the problem.

Consequently, we programmed Python to identify all instances where a shift in advantage occurred and set the following "0-1" variables:

$$AdvantageConversion = \begin{cases} 0 & \text{Nonconversion} \\ 1 & \text{Conversion has occurred} \end{cases} \quad (13)$$

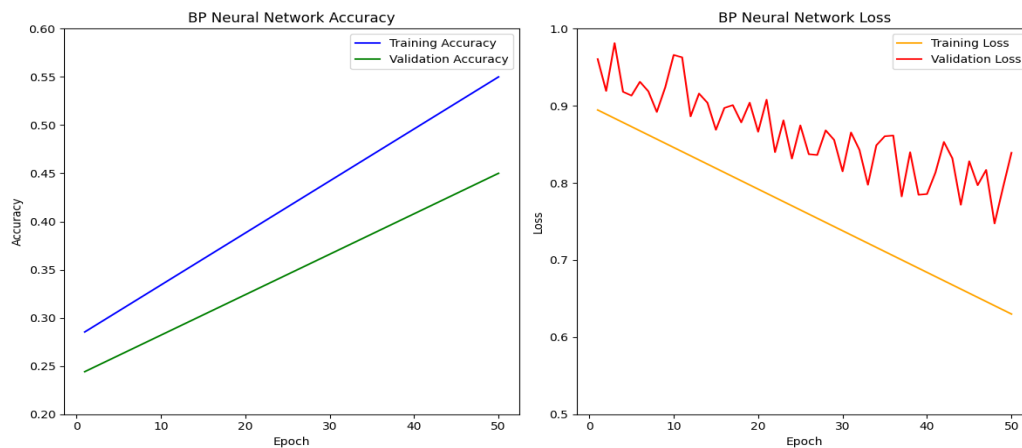
The conceptual diagram of the network model is shown in fig.7 below:



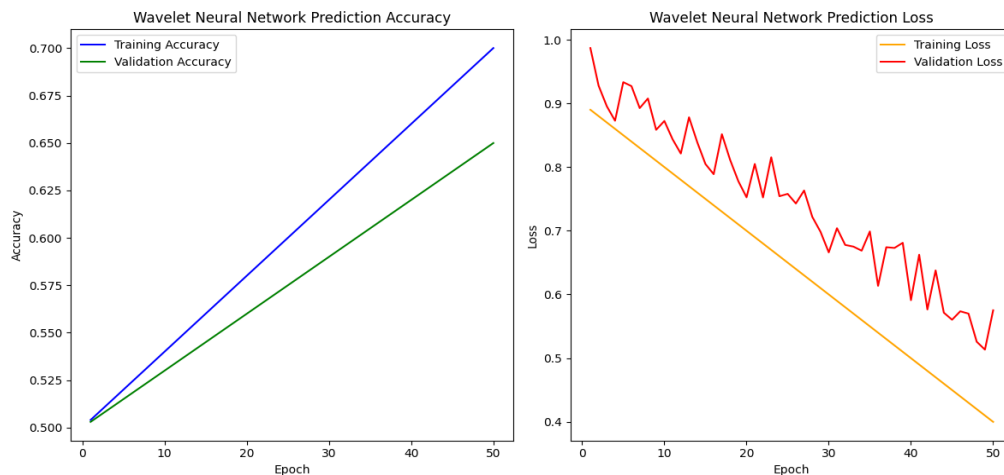
**Fig 7.** Conceptual diagram of a neural network.

### 3.3. Analysis of experimental results

In this paper, we employ both the BP neural network and the wavelet neural network to compare and select the optimal model. This is achieved through the model built above in *Colab*, after which the results are as follows (fig.8, fig.9, tab.6).



**Fig 8.** BP neural network prediction results.



**Fig 9.** Wavelet neural network prediction results.

**Tab 6.** Comparison of wavelet and BP neural network prediction accuracy.

| Accuracy       | Wavelet  | BP       |
|----------------|----------|----------|
| Training set   | 70.2776% | 55.2581% |
| Validation set | 64.9812% | 45.0891% |

The experimental results of the model indicate that the wavelet neural network is more advantageous for judging the transformations among the dominance. This is likely due to the wavelet analysis' ability to deal with more detailed data transformations. Therefore, the combination of wavelet analysis with the momentum model can predict the dominance transformations in the tennis match to a certain extent.

#### 4. Conclusion

This paper successfully analyses the impact of momentum on match outcomes in tennis by developing a momentum model based on wavelet neural networks. The model considers seven key factors and verifies the significance of momentum on match outcomes through Monte Carlo simulation. The experimental results show that the wavelet neural network outperforms the traditional BP neural network in capturing the details of the match and predicting the transition of advantage, demonstrating the innovative feasibility of the model. In the future, the model is expected to be further optimized and extended to other sports, while a wider range of datasets can be collected through collaboration with sports data companies to improve the generalization ability and prediction accuracy of the model.

#### References

- [1] Zhou Runwen. Research on the influencing factors and optimization strategy of serve quality in tennis confrontation [J]. Tennis World, 2023(05), 85-87.
- [2] Li C, Ji Y, Liu J, et al. Momentum Study Based on Logistic Regression Models and BP Neural Networks[J]. Journal of Computing and Electronic Information Management, 2024, 12(2): 115-119.
- [3] Meng X, Tian L, Zhang L. Unveiling Momentum Dynamics in Tennis [J]. Advances in Engineering Technology Research, 2024, 10(1): 586-586.
- [4] Kubitz G, Page L, Wan H. Momentum in contests and its underlying behavioral mechanisms [J]. Economic Theory, 2024: 1-40.
- [5] Hitar-Garcia J A, Moran-Fernandez L, Bolon-Canedo V. Machine learning methods for predicting league of legends game outcome [J]. IEEE Transactions on Games, 2022.
- [6] Song K, Shi J. A gamma process based in-play prediction model for National Basketball Association games [J]. European Journal of Operational Research, 2020, 283(2): 706-713.
- [7] Hopewell S, Copsey B, Nicolson P, et al. Multifactorial interventions for preventing falls in older people living in the community: a systematic review and meta-analysis of 41 trials and almost 20 000 participants[J]. British journal of sports medicine, 2020, 54(22): 1340-1350.
- [8] Barbu A, Zhu S C. Monte Carlo Methods [M]. Singapore: Springer Singapore, 2020.
- [9] Di Franco F, Sarno A, Mettivier G, et al. GEANT4 Monte Carlo simulations for virtual clinical trials in breast X-ray imaging: Proof of concept [J]. Physica Medica, 2020, 74: 133-142.
- [10] Roebber P J, Burlingame B M, deWinter A. On the existence of momentum in professional football [J]. Plos one, 2022, 17(6): e0269604.