Unveiling Momentum Dynamics in Tennis

Xiangqi Meng[1*], Lei Tian[1] and Lingchen Zhang[2]

[1] School of Business, Hohai University, Nanjing, China

[2] School of Earth Sciences and Engineering, Hohai University, Nanjing, China

*1453946095@qq.com

Abstract. This study investigates the impact of momentum in tennis competition through a multifaceted approach. We introduce the Momentum Balance Index (MBI), derived via Principal Component Analysis (PCA) from 42 indicators, identifying 13 significant factors influencing player momentum. MBI analysis reveals a strong correlation between peak momentum and score leads, indicative of player condition and winning likelihood. Spearman correlation analysis validates momentum's link to enhanced performance and victory probabilities, challenging randomness in match outcomes. Additionally, an ARIMA-based prediction model, refined through logistic regression, forecasts momentum shifts within matches, offering real-time tactical insights for coaches. Receiver Operating Characteristic (ROC) analysis confirms the model's robust predictive capability, with AUC values exceeding 0.5 across three matches. This comprehensive approach sheds light on momentum dynamics in tennis, providing valuable insights for coaching and strategic decision-making.

Keywords: Principal Component Analysis, ARIMA One-Factor, Logistic Regression, Receiver Operating Characteristic.

1. Introduction

In tennis, momentum is recognized as a key psychological and physical state that affects the outcome of a match. As a match progresses, a player may experience a rise or fall in momentum, which is often associated with successive points scored or lost. Although athletes are widely discussed in sport psychology and sport performance analysis, their specific role and quantification in tennis have not been fully investigated.

Indeed, the scoring profile of athletes is often a more complex overall system in reality. Therefore, a model that can quantify momentum and predict its impact on the match outcome is important for coaches to formulate tactics and players to adjust their game strategies. The intricacies of scoring patterns and their relation to momentum shifts in tennis are exemplified by discussions in major tennis tournaments, highlighting the sophisticated nature of athletic performance in this domain.

2. Literature Review

With regard to the study of momentum in sports, there is a body of literature that explores its effects on athletes' performance [2-3]. However, correlation studies for tennis matches are relatively few and have focused on qualitative analysis [4]. In terms of quantitative analysis, some studies have attempted to measure momentum through statistical methods, but these studies often lack insight into the psychological and physiological mechanisms behind momentum changes. In addition, research on how momentum analysis can be utilized to guide race strategy is more limited [5].

The Theory of Challenge and Threat States in Athletes (TCTSA) is based on cognitive appraisal theory by Jones and his colleagues, which categorizes appraisal results into two categories: challenge and threat, and combines athletes' physiology, emotional responses, motivational patterns, goal setting and other factors with athletic performance [2]. This theory describes the relationship between athletes' physiological, emotional responses, motivational patterns, goal setting and other factors and athletic performance in more detail. Typically, when challenged, athletes are more positive and more likely to perform well, whereas when threatened they are more negative and more likely to perform poorly. However, not all emotions are positive in challenging situations. Emotions in the challenging state should be characterized as positive or moderately negative [6-7], whereas emotions in the threatening state are characterized by a more homogeneous negative state and sometimes positive emotions, which can mitigate the negative effects of threat on technical performance and transform the individual from a dangerous to a challenging positive state [8]. This state is similar to that expressed in terms of momentum or "momentum" as given in the question.

In our work, we construct a mathematical model to quantitatively analyze the "momentum" of players in tennis matches and discuss its application in actual matches. First, we analyze and build an evaluation model based on the available data set. It is used to evaluate how well any player performs at a given moment in any match. Second, based on the available data and the model developed in the first question, we apply it to different matches in the dataset to illustrate the link between an athlete's success in a match and the athlete's "momentum" during the match. Next, based on the existing analysis, our paper builds a suitable predictive model to predict the trend of the athletes' "momentum" during the competition, analyze the factors related to the momentum shift, and based on the results of the analysis, provide reasonable suggestions for the upcoming new competitions against different opponents. Finally, we analyze and validate the established prediction model, test its ability to predict fluctuations in matches, evaluate its generalization ability to different matches and give reasonable suggestions [9].

3. Modeling

The flowchart below illustrates the entire modeling process.

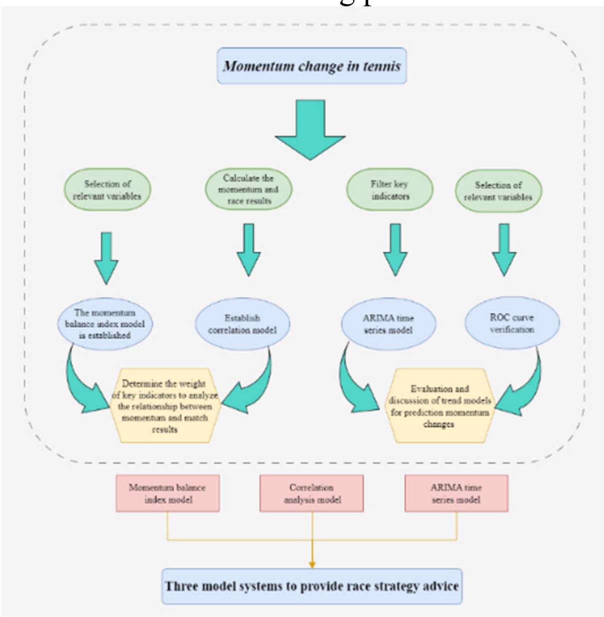

Figure 1 Flow chart

3.1 Building an Evaluation Model for Player Performance Prediction

3.1.1 Pre-processing of data

Before the formal modeling, the known data are first analyzed and processed, as can be seen from the dataset, 42 different indicators are given for the state of each athlete at any moment, which are both numerical and categorical variables, here the data are firstly pre-processed appropriately, and the retrieval reveals that there is no vacant value in the dataset, here the categorical variables are firstly processed so that they can be transformed into numerical variables for the subsequent modeling. can be transformed into numerical variables for easy subsequent modeling, here a dummy variable approach is used to transform these categorical variables into numerical variables, for categorical variables with two levels, a dummy variable is created with a value of either 0 or 1. For example, if the categorical variable is gender (male/female), a variable can be created that has a value of 1 when the individual is a male, and a value of 0 when the individual is a female.

3.1.2 Principal Component Analysis

Establishment of model based on momentum balance

From the above analysis, it was found that for the representation of momentum in tennis matches, it can be regarded as a manifestation of a player's ability to win, and most studies on racquet sports have focused on scoring results rather than scoring sequences and other events that may trigger positive or negative momentum.

For the sake of clarity, the index "Momentum Balance Index in Tennis" was constructed using the question of whether the scoring outcome of a match is related to the outcome of the previous point, two or three points, as well as to matches within the previous scoring range, such as scores, double faults, winners, and errors: The Momentum Balance Index (MBI) in tennis describes the degree of momentum a player has at any moment in a tennis match. It can be understood that the larger the Momentum Balance Index (MBI) a player has, the higher the degree of momentum the player has at that moment.

Here we can define the degree of upward tendency of the momentum of the tennis tele mobilizer: $f_i, i = (1,2,\ldots,n)$, The upward tendency of the momentum of the tennis telemobilizer indicates the upward size index of the degree of change of the momentum of different athletes in different matches, and the principal component analysis will be used to analyze some of the changes in the state of the current athletes who have a favorable tendency$m_i, = (1,2,\ldots,n)$,The upward tendency of the momentum of the tennis telemobilizer indicates the upward size index of the degree of change of the momentum of different athletes in different matches, and the principal component analysis will be used to analyze some of the changes in the state of the current athletes who have a favorable tendency: $f(x)$ To describe it, the upward tendency of the momentum of the far mobilized tennis ball is expressed by the following equationf_i:

$$f_i = f(x) + m_i, i = (1,2\ldots,n) \tag{1}$$

Then the upward propensity of the momentum that has not begun to mobilize far from the tennis ball f_i:

$$f_i = m_i, i = (1,2,\ldots,n) \tag{2}$$

The same can be defined for the tendency of a tennis player to have a decrease in momentum as: $g_i, i = (1,2,\ldots,n)$ The tendency of tennis players to decline in momentum reflects the downward size index of the degree of change in the momentum of different athletes in different matches, and the same principal component analysis will be used to show that some of the current athletes have a detrimental tendency to the degree of change in the state: $n_i, i = (1,2,\ldots,n)$ At the same time, in order to reflect the influence of the different momentum states obtained by the tennis player at different moments of the match, the function was used: $h(x)$ To describe it, the tendency of a tennis player to drop momentum is expressed by the following equationg_i:

$$g_i = h(x) + n_i, i = (1,2,\ldots,n) \tag{3}$$

Then the propensity of the unstarted tennis player to decrease in momentum g_i:

$$g_i = n_i, i = (1,2,\ldots,n) \tag{4}$$

It can be defined in terms of the momentum balance index possessed by the athlete: $\partial_i, i = (1,2,\ldots,n)$ To describe the tendency of the tennis player's momentum balance, while the momentum balance index is related to the tendency of the tennis telemobilizer's momentum to rise f_i and the tendency of the tennis player's momentum to fall g_i, which can be expressed by the following equation:

$$\partial_i = \frac{f_i}{g_i} = \frac{f(x)+m_i}{h(x)+n_i}, i = (1,2,\ldots,n) \tag{5}$$

Then the momentum balance index possessed by the unstarted athlete ∂_i:

$$\partial_i = \frac{m_i}{n_i}, i = (1,2,\ldots,n) \tag{6}$$

Combining the above functional relationships, a functional equation for the sign condition can be established as follows:

$$f(x) = \frac{1}{2}(1 - \frac{1}{e^{\alpha x}}) \quad (7)$$

$$g(x) = \frac{1}{2}(1 - \frac{1}{e^{\beta x}}) \quad (8)$$

Determination of weights by principal component analysis

In our research, we utilize Principal Component Analysis (PCA) to effectively extract pivotal indicators from a comprehensive set of 42 variables. The strength of PCA lies in its ability to simplify complex data, ensuring interpretability without significant loss of information, and pinpointing the most impactful variables.

3.1.3 Selection of indicators

Via Principal Component Analysis, we select the given data in the number of ACE, double serve errors, the number of first serve goals, the number of second serve goals, the first serve goal rate, the number of first serve scores, the number of second serve scores, the number of first serve scores, the number of second serve scores, the number of serve scores, the recovery of the break point, the number of unforced errors, the number of serves of these 13 indicators as a model to build the indexes, which the number of ACE, the number of first serve goals, the number of first serve scores, the number of first serve scores The six indicators, including the number of ACEs, the rate of first serve, the number of first serve points, the number of serve points, the number of service points, and the number of service points recovered, are used as favorable trend indicators, denoted as $F_{b_i}, F_{b_i} = 1,2,\ldots,6$, while the seven indicators, including the number of double serve errors, the number of second serve points, the number of second serve points, the rate of second serve points, the number of unforced errors, and the number of serves, are used as unfavorable trend indicators, denoted as $F_{hi}, F_{hi} = 1,2,\ldots,7$. In order to facilitate the design of the model these thirteen indicators are labeled as:

Table 1 Indicator display table

Variable name	Variable	Variable calculation formula
ACE	X1	
double errors	X2	
first serve goals	X3	
second serve goals	X4	
first serve percentage	X5	First serve goals/Total service goals
first serve points	X6	
second serve points	X7	
first serve points percentage	X8	Number of games won/goals scored
second serve percentage	X9	
serve percentage	X10	
break point saved,	X11	
unforced errors percentage	X12	
serve percentage	X13	

We use Correlation analysis to measure the linear correlation between 13 variables. When 0.8 is taken here, it indicates that the two variables have strong linear correlation, and no absolute value of the 13 variables is greater than 0.8, so there is no strong correlation coefficient between the 13 variables, and all of them are included in the model and included in the next analysis.

3.1.4 Solving the model and results

On the basis of the data on tennis matches provided in the annex, we determined the weights of the indicators. The weights of these 13 indicators were calculated and the results of the weights are shown in the table 2.

Combining the results of the above analyses of the upward tendency of the momentum of the tennis telemobilizer and the downward tendency of the momentum of the tennis telemobilizer, the following equation for determining the momentum balance index can be obtained:

$$\partial_i = \frac{m_i+f(x_1)t}{n_i+g(x_2)t}, i = (1,2,\dots,n), t \quad (9)$$

Among them:

$$f(x_1) = \frac{1}{2}(1-\frac{1}{e^{\alpha_1 x_{11}}})(1-\frac{1}{e^{\alpha_2 x_{12}}})\cdots(1-\frac{1}{e^{\alpha_6 x_{16}}}) \quad (10)$$

$$g(x_2) = \frac{1}{2}(1-\frac{1}{e^{\alpha_1 x_{21}}})(1-\frac{1}{e^{\alpha_2 x_{22}}})\cdots(1-\frac{1}{e^{\alpha_7 x_{27}}}) (11)$$

With the help of the momentum balance model to calculate the potential energy of the player at each point, taking the race number ' 2023-wimbledon-1403' as an example, we calculate the information of the potential energy.

Table 2 Weighting of different indicators table

Variable	Variable weight	Sig(P) value
X1	0.218	0.000
X2	0.028	0.000
X3	0.515	0.126
X4	0.141	0.021
X5	0.242	0.525
X6	-0.392	0.000
X7	0.042	0.000
X8	-0.312	0.000
X9	0.186	0.000
X10	-0.038	0.532
X11	0.158	0.326
X12	-0.263	0.358
X13	0.226	0.000

We plotted the potential energy versus score for player one and player two in Figure 5.

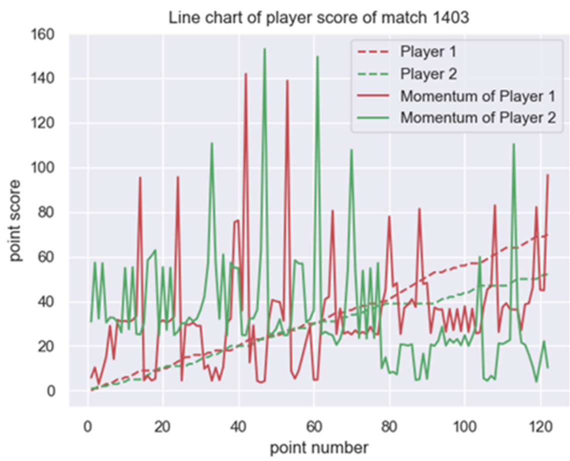

Figure 2 No. 1403 Plot of the change in momentum of the two players in the game

3.2 Analyzing Player Performance and Momentum Dynamics Across Match Data Sets

3.2.1 Define test data

In order to find out if there is a correlation between momentum and match scoring, we need to analyze all the matches and if most of the matches indicate that there is some relationship between momentum and scoring, then we can draw conclusions.

We counted the player's wins and losses in 31 matches and whether the player's average momentum was greater than the opponents in the 31 matches.

3.2.2 Results

In addition to analyzing the relationship between the momentum of all the games and the win or loss of the game, we should also pay attention to whether the players' status is better in the time period with more scores under a single game, we take the data of the first game as an example, and we have counted the win and the average momentum of the players under all the GAMES of the first game, and calculated the Spearman correlation coefficients of the two cases, and the results are as follows.

Table 3 Race Momentum and Race Endings

Situation	Correlation coefficient
All match	0.6325
Match: 2023-wimbledon-1301	0.7638

The Spearman correlation coefficients in both cases were greater than 0.5, indicating that there is a correlation between momentum and player scores.

3.3 Constructing Predictive Models for Analyzing Player Momentum Shifts

3.3.1 Screening of key indicators

In order to further analyze the importance of the indicators obtained, the idea of multiple linear regression can be used to analyze the relationship between these indicators and success in a game. Multiple linear regression analysis is the most commonly used statistical method to study the relationship between a variable and multiple independent variables. By analyzing the data above the 13 indicators screened were brought into the binary regression model, set the dependent variable win=1, negative=0, and finally the 13 indicators were analyzed by using one-way logistic regression, the results are shown in Table 4.

Table 4 Race Momentum and Race Endings

Variable	B	Standard error	Wald	Degree of freedom	significance
X1	0.097	0.025	3.203	1	0.074
X2	0.062	0.038	0.489	1	0.458
X3	0.038	0.012	14.132	1	0.000
X4	0.035	0.036	1.260	1	0.265
X5	0.157	0.027	10.941	1	0.000
X6	0.086	0.014	3.825	1	0.001
X7	0.072	0.033	3.112	1	0.020
X8	0.025	0.065	0.587	1	0.004
X9	0.015	0.014	3.264	1	0.078
X10	0.096	0.032	5.624	1	0.000
X11	0.135	0.056	3.684	1	0.000
X12	0.048	0.028	36.074	1	0.265
X13	0.397	0.074	15.532	1	0.153
constant	-23.89	5.231	33.169	1	0.000

Bringing the data from the table into the regression equation gives the logistic regression model as: $Logit(P)=-23.89+0.157X5+0.072X7+0.025X8+0.096X10$.

Where P is the probability of the athlete's winning, X5 is the rate of first serve, X7 is the number of second serve points, X8 is the rate of first serve points, and X10 is the number of serve points.

Finally, it is concluded that the factors affecting the athlete's winning are the four technical indicators: the rate of first serve, the number of points scored on second serve, the percentage of points scored on first serve, and the number of points scored on serve.

3.3.2 Principal component analysis of key indicators

After determining the key indicators, the momentum model established in the first question is utilized to recalculate the magnitude of the athlete's momentum value at each moment, which will not be repeated here since the model established in this question has been established previously.

3.3.3 ARIMA time series

In fact, in order to better describe the process of helping tennis players to adjust their state in the process of the match and improve the match winning rate of the staff, the calculated momentum value can be treated as the dependent variable and the event node of the match as the independent variable, which is obviously an obvious correspondence with a time series, so considering the characteristics of the time series, the momentum prediction model based on ARIMA time series is established here.

The first step is to identify and order the model. It is necessary to judge the categories of AR (p), MA(q) or ARMA(p,q) models and estimate the order p,q. In fact, it all comes down to the order of the model. After the model is graded, the model parameters $\varphi = [\varphi_1, \varphi_2, \cdots, \varphi_p]^T$ and $\theta = [\theta_1, \theta_2, \cdots, \theta_q]^T$ are evaluated.

After the order determination and parameter estimation are completed, the model should be tested to check whether$\varepsilon_t$is stationary white noise. If the test is passed, the modeling of the ARMA time series is completed.

3.3.4 Model solution and results

We take the first game as an example to model and analyze the momentum of the first game. Divided according to the training set and test set 7:3, the selected part of the momentum prediction clip of player 1 and plotted the true and predicted values as shown in the figure, the result conforms to a certain pattern, and we consider this fitting to be successful.

Table 5 Error analysis table

Metrics	Value
MSE	0.0076
RMSE	0.0823
R2	0.9802
MAE	0.0050

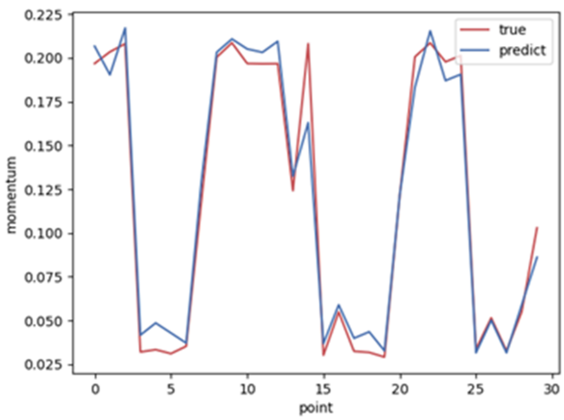

Figure 3 Tests of predictive models

With the help of this model, we can accurately predict the fluctuations in the game, and recalculate the level of importance of the predictors.

3.4 Evaluating Forecasting Model Performance

3.4.1 Assessment based on AUC and ROC curves

The following graph shows the ROC curves of the prediction models for three different matches, where the size of the enclosed area is equal to the size of the AUC, and it can be found that the AUC values of three of the models are still good, indicating that the models have a certain predictive effect.

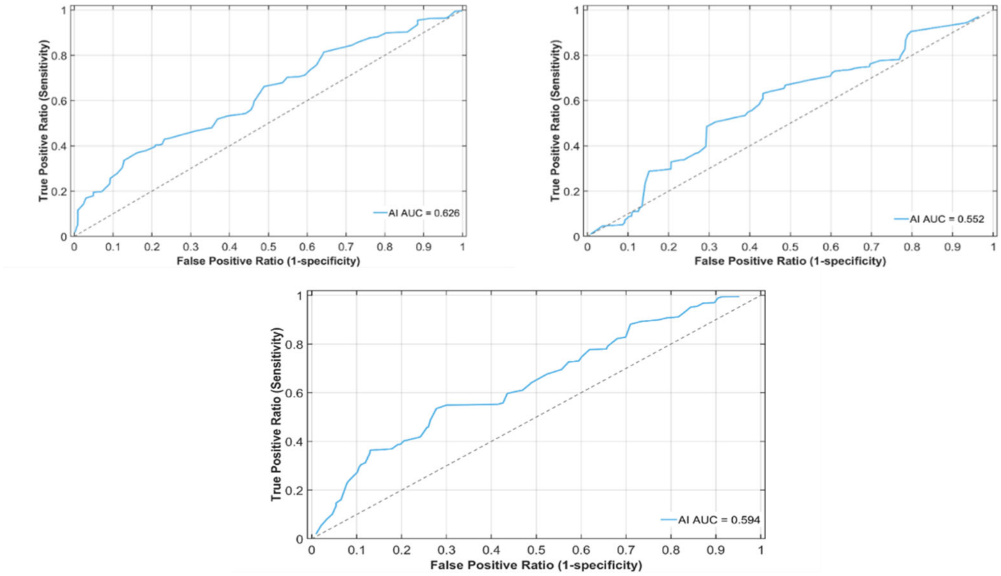

Figure 4 ROC test chart

3.4.2 Generalization and Extension of the Model

The application of the momentum prediction model we have built in tennis has proven its effectiveness in motion analysis and performance enhancement. By analyzing an athlete's motion data, this model is able to predict his or her performance and strategy adjustments during a match, thus providing valuable information to athletes and coaches. Extending this model to other sports such as soccer, basketball, track and field, swimming, cycling, skiing and snowboarding, and baseball will help improve the overall level of competition in these sports.

In basketball, momentum modeling can analyze players' movement efficiency and shot selection, providing coaches with the basis for tactical adjustments. In athletics, momentum modeling can be used to guide athletes on how to distribute their physical strength in order to maintain the best condition until the end of the race. For swimmers, the model can be used to analyze the stroke frequency and strength to optimize the technical movements and improve the performance. In cycling, momentum models are used to assist teams in understanding when to lead. In winter sports, momentum modeling is used to help athletes maintain optimal posture and reduce errors during high-speed skating. In baseball, momentum models are used to analyze the pitching motion of pitchers and the timing of batters' hits to reveal potential weaknesses of opponents.

Promoting this model improves athletic training efficiency and performance, offers coaches scientific decision-making support, enhances game spectacle and audience engagement. As technology advances, momentum prediction models are poised to play a greater role in sports programming, driving innovation and development across the industry. Further, these models can be integrated with existing sports science research, including biomechanics and psychology to offer more holistic guidance for athletes.

4. Model Evaluation and Further Discussion

4.1 Strengths

This study adopts a variety of statistical and forecasting methods, such as principal component analysis, correlation coefficient analysis, ARIMA time series model and ROC curve, and the combined application of these methods improves the comprehensiveness and depth of the analysis.

By constructing a momentum balance index model, the study provides a scientific method to quantify and analyze the momentum changes of tennis players during a match, which is of high practical value for coaches and players.

The ARIMA time series modeling and the validation of ROC curves show that the model has good predictive ability in predicting the momentum changes of athletes, which helps coaches to develop more effective match strategies.

4.2 Weaknesses

The accuracy of the model is highly dependent on the quality and completeness of the input data. If the data is biased or missing, it may affect the predictive effectiveness of the model.

Combining multiple analytics may result in an overly complex model, which may make it more difficult to understand and implement, especially in terms of data preprocessing and model interpretation.

4.3 Further Discussion

Attempts to simplify the model structure to make it more intuitive and understandable while maintaining predictive power to make it more accessible and usable by coaches and athletes.

Test the model across different competitions, athletes and sports to validate its generalization ability and make adjustments based on feedback.

5. Conclusion

In this study, a comprehensive tennis momentum analysis and prediction system was constructed through the comprehensive use of principal component analysis, correlation coefficient analysis, ARIMA time series model and ROC curve validation. The study firstly screened 13 key indicators that have significant influence on tennis players by preprocessing the data, and determined the weights of these indicators using principal com-ponent analysis to construct a momentum balance index model. This model not only reveals the significant relationship between momentum changes and match results, but also provides coaches with a scientific basis for real-time adjustment of athletes' status.

Further, the study established a correlation model based on the correlation coefficient to analyze the correlation between the athlete's momentum value and the win or loss of the game, and confirmed the positive correlation between the momentum value and the athlete's condition and the probability of winning. In addition, through the ARIMA time series model, the study predicted the momentum changes of the athletes during the game, which provided targeted suggestions for the coaches to help the athletes optimize the game rhythm.

Finally, through the validation of ROC curves, the study proved the validity of the pre-diction model, and its prediction ability performed well in different matches. These findings not only provide a new analytical tool for tennis, but also provide a valuable reference for the analysis of athletes' momentum changes in other sports. The study suggests that future work could further optimize the model to improve the prediction accuracy and explore the application of this system to other sports with the aim of providing deeper insights into athlete training and competition strategies.

References

[1] Australian Open. (2023). Retrieved from https://ausopen.com/

[2] J. Arkes and J. Martinez, Finally, evidence for a momentum effect in the NBA, J. Quant. Anal.Sports7 (2011), pp. 12–21.

[3] M. Fry and F. Shukairy, Searching for momentum in the NFL, J. Quant. Anal. Sports 8 (2012),pp. 32–47.

[4] T. Chen and Q. Fan, A functional data analysis approach to model score difference in professional basketball games, J. Appl. Stat. 45 (2018), pp. 112–127.

[5] A.E. Hoerl and R.W. Kennard, Ridge regression: Biased estimation for nonorthogonal problems,Technometrics 12 (1970), pp. 80–86.

[6] R.P. Bunker and F. Thabtah, A machine learning framework for sport result prediction, Appl.

[7] D. Cervone, A. D'Amour, L. Bornn, and K. Goldsberry, A multiresolution stochastic processmodel for

predicting basketball possession outcomes, J. Am. Stat. Assoc. 111 (2016), pp. 585–599.

[8] J. Friedman, T. Hastie, and R. Tibshirani, Regularization paths for generalized linear models viacoordinate descent, J. Stat. Softw. 33 (2008), pp. 1–22.

[9] Onan, A. . (2019). Mining opinions from instructor evaluation reviews: a deep learning approach.