Quantitative Analysis of Tennis Ball Momentum Utilizing a Fusion Model of EMA and Bessel Curves

Pengfang Gao *, Nuolin Yu

School of Science, Shandong Jianzhu University, Jinan, China, 250101

* Corresponding Author Email: encounter117026@163.com

Abstract. Targeting the challenges in quantifying the momentum transmission mechanism during tennis matches and the constraints imposed by traditional models' linear assumptions, this study developed an innovative dynamic model. By integrating logistic regression, exponential moving average (EMA), and Bessel curves, the model captures recent performance trends through EMA weighting while pinpointing momentum turning points using the second derivative of the Bessel curve. When tested with 2023 Wimbledon men's final match data, the model demonstrated a prediction success rate of 78%, surpassing traditional methods by 12%. Moreover, it enables real-time identification of momentum shifts during critical match phases. This advancement not only offers a scientific foundation for enhancing tactical decision-making in tennis but also propels the intelligent evolution of sports analysis, bridging theory and practical application seamlessly. This paper integrates these techniques to provide profound insights into player dynamics and match progression, laying the foundation for more precise predictions and strategic planning in competitive sports.

Keywords: Momentum, Logistic Regression, EMA.

1. Introduction

In sports science, as a typical turn-based sports event, the dynamic game process of tennis contains complex nonlinear dynamics. The traditional analytical paradigm based on empirical observation faces significant limitations in quantifying the momentum conduction mechanism and characterizing the interaction of multi-dimensional technical indicators such as break efficiency, serve stability, and net control. Therefore, using data analysis technology to conduct objective, comprehensive, and in-depth analysis of tennis matches has become a necessary trend and development direction.

Chan Timothy C. Y. et al. pointed out through causal inference that changes in serving rules would affect the scoring probability, and the serving advantage was derived from structural rules rather than random factors, which provided theoretical support for the logistic regression model that gave significant weight to the "server" [1]. Xia et al. built a decision tree model by using factor analysis and found that player momentum difference fluctuations were significantly correlated with match results [2]. Chen et al. built a regression prediction model based on LightGBM, and concluded through SHAP analysis that the player's travel distance is a key factor affecting the match result [3]. Zhan et al. deepened the momentum analysis framework, identified 10 key indicators through factor analysis and combination-weighted methods, and built a nonlinear regression model to evaluate the impact on real-time momentum [4]. Lv Y et al. Used the entropy weight method to build a dynamic momentum visualization model and found that momentum fluctuations were significantly positively correlated with the game results [5]. Through correlation analysis and two-step clustering, Cui Yixiong et al. revealed the differentiated performance characteristics of professional female tennis players in different Grand Slam events and found that relative quality significantly affected serve efficiency, breakpoint success rate, and net performance, indicating that baseline competition is still the core competition mode of women's Grand Slam [6]. By combining SHAP, Yuan Zhang can explain artificial intelligence technology and multi-factor analysis framework, reveal the interactive impact of players' physical conditions, geographical characteristics, court types, and skill differences on tennis match results, and provide actionable tactical optimization suggestions for players and coaches based on normative analysis [7]. By analyzing break point conversion and exogenous interruption (such as hawk-eye challenge) in professional tennis matches, Meier et al. found that the probability of players winning after a break was significantly increased, but interruption would weaken this effect, indicating that psychological momentum rather than strategic momentum is the key factor driving performance improvement, providing a new perspective for the study of momentum mechanism in dynamic competitions [8]. Wei Yang et al. developed a soccer match outcome prediction model based on player performance metrics, achieving AUC, F1, and accuracy scores of 0.8597, 0.6973, and 0.7965 respectively, which validates the method's feasibility and provides a novel approach for sports analytics in dynamic team competitions [9]. Armstrong et al. compared match load profiles in padel, singles, and doubles tennis, revealing that singles impose higher cardiovascular and physical demands, while padel features greater shot variety and effective playing time. These findings highlight the distinct physiological and tactical dynamics across racquet sports [10]. The study by Martínez Gallego et al. shows experienced doubles teams use offensive strategies to shorten points, reducing momentum swings. It links team dynamics to match tempo, offering insights for real-time analysis [11].

Although the existing research on tennis momentum has made remarkable progress, there are still multi-dimensional limitations: the early research relies on historical data, which is difficult to reflect the technical innovation of modern tennis; the dynamic characterization is limited to linear model, ignoring nonlinear wave characteristics. This approach effectively overcomes the limitations of characterizing nonlinear waves within a linear framework. The interaction between technology and strategic indicators has not been integrated. The mechanism of momentum conduction is not deeply understood. Based on the latest tournament data in 2023, this study reflects the competitive characteristics of modern tennis. Through the introduction of exponential moving average (EMA) and Bessel curve analysis, dynamic weight allocation and nonlinear feature capture are innovatively realized. EMA method solves the defect of the static weight of the traditional model by describing momentum persistence with weighted factors. Bessel curve smoothing technique and second derivative analysis reveal the non-random fluctuation rule of score probability. The dynamic momentum update of the model can be realized by sliding windows to provide immediate support for tactical adjustment. These improvements provide a new way to build a momentum quantization framework that is closer to the real scene of the game.

2. Model building

2.1. Logistic Regression Model

Logistic regression represents a well-established binary classification method, widely applied to predict outcomes such as match victory based on numerical covariates. Unlike multiple linear regression, logistic regression incorporates a logistic transformation to model probabilities of binary dependent variables. Leveraging preselected performance indicators, this study constructs separate logistic regression models for both competitors to estimate their time-specific winning probabilities:

$$\begin{cases} Y_{1t} = \beta_{10} + \sum_{i=1}^{11} X_{1t} + \varepsilon_{1t} \\ Y_{2t} = \beta_{20} + \sum_{i=1}^{11} X_{1t} + \varepsilon_{2t} \end{cases} \tag{1}$$

Where, Y_{1t} and Y_{2t} indicating whether Player 1 and Player 2 win at the current moment.

Assuming that Y is a logical variable, we incorporate the Bernoulli distribution and utilize conditional probability to express the likelihood of the occurrence of a logical variable. This approach helps tackle endogeneity problems, as demonstrated below:

$$\begin{cases} P(Y = 1|X) = F(X,\ \beta) \\ P(Y = 0|X) = 1 - F(X,\ \beta) \end{cases} \tag{2}$$

Where $F(X,\beta)$ is a function defined on [0, 1], and we choose the Sigmoid function, as follows:

$$F(X,\beta)=\frac{e^{x_i^T\beta}}{1+e^{x_i^T\beta}} \tag{3}$$

Finally, calculate X and Y, which denote the winning probabilities of player 1 and player 2 at time t, respectively. Subsequently, we assign weights to the winning probabilities of both players at time t in the following manner:

$$\begin{cases} PW_{1t}=\dfrac{P(Y_{1t}=1|X)}{P(Y_{1t}=1|X)+P(Y_{2t}=1|X)} \\ PW_{2t}=\dfrac{P(Y_{2t}=1|X)}{P(Y_{1t}=1|X)+P(Y_{2t}=1|X)} \end{cases} \tag{4}$$

In the context of the logistic regression model, the overall regression coefficients remain unknown and need to be estimated through relevant sample data. Consequently, after incorporating control variables, we will apply Ordinary Least Squares (OLS) for estimating all regression coefficients. Furthermore, the joint significance of these coefficients will be evaluated using the F-statistic.

By applying the logistic model, we are able to calculate the probabilities of player 1 and player 2 winning at any specific point in time. These probabilities act as measures of the players' performance quality at that moment, capturing their current momentum.

2.2. Exponentially Weighted Moving Average

Generally speaking, the momentum of players in a tennis match is dynamic, requiring us not only to consider a player's performance at a single moment but also to account for the continuity of player momentum and the changing trends during the match.

Match data were split into training and validation subsets for model development and performance evaluation. For illustrative purposes, Match 1302 (N=201 data points) was randomly divided into a 50% training set and a holdout testing set. The model's predictive performance is visualized in Figure 2, demonstrating its ability to generalize across different match phases:

$$\widetilde{\mathrm{PW}_{1t}}=\alpha\mathrm{P}\widetilde{\mathrm{W}_{1,t-1}}+(1-\alpha)\mathrm{PW}_{1,t} \tag{5}$$

Where, α is the smoothing coefficient, the value range is $0<\alpha<1$. We will employ hyperparameter optimization to calculate the optimal smoothing coefficient.

Utilizing the EMA model to calculate the player's winning probabilities considers not only the player's current match situation but also incorporates the performance trend from the previous moment. This results in a more accurate quantification of the player's momentum.

2.3. Bessel curve

Bessel curve is a parameterized polynomial interpolation method based on control points, which realizes the controllable generation of geometric shapes by Bernstein-based functions. In tennis matches, the swings on the court signify significant moments of change in a player's momentum, crucial instances when an advantage shifts from one player to another. Since the data is discrete, momentum transitions can be identified by computing the first-order difference (gradient) of the Bessel curve, as shown below:

$$\nabla B=B(t+1)-B(t) \tag{6}$$

Where, $\mathrm{B(t)}$ represents the value of the Bessel curve at time t.

To gain deeper insights into the existence of swings, this research analyzes the dynamics of momentum shifts in the proximity of these critical transition junctures. A more pronounced shift in momentum correspondingly enhances the probability that such a juncture serves as a potential swing point within the match. To quantify this behavioral pattern, the study employs the second-order difference of the Bessel curve as a quantitative metric to characterize the trend intensity:

$$\nabla^2 B = \nabla B(t+1) - \nabla B(t) \tag{7}$$

If the second-order difference value at a specific transition point is notably large, that point is considered a swing in the match.

Based on the comprehensive analysis above, we define a swing as follows: if the absolute value of the second-order difference of the Bessel curve at that point is greater than the mean absolute value of the historical data second-order difference, then that point is identified as a swing in the match.

2.4. Hypothesis testing

In this study, a dataset comprising randomly selected player match statistics was curated, processed to quantify momentum indicators including swing points and success streaks, and structured into a binary sequence. Guided by statistical hypothesis testing methodologies, a framework was formulated as follows:

H_0 (Null Hypothesis): Momentum does not exert a statistically significant influence on match dynamics.

H_1 (Alternative Hypothesis): Momentum has a notable impact on match outcomes.

Within this analytical structure, empirical rejection of the null hypothesis would provide evidence supporting momentum's significance in shaping match results. To examine this, a Run Test was applied to the swing point and success streak data. A remarkably low total run count in the sequence suggested an elevated frequency of repeated run patterns, underscoring the presence of symbol clustering within the binary structure.

3. Results

3.1. Data Pre-Processing

Firstly, we initiates missing value treatment. This ensures data accuracy and effective model training, as neglecting missing values could impact machine learning efficiency and prediction accuracy. Utilizing the DataFrame.isnull() function in the pandas library of Python, the study identifies and quantifies missing values for each indicator in the dataset. To visually represent the distribution of missing values among different indicators, data visualization techniques, such as the heatmap method, are employed. The result is shown in Figure 1:

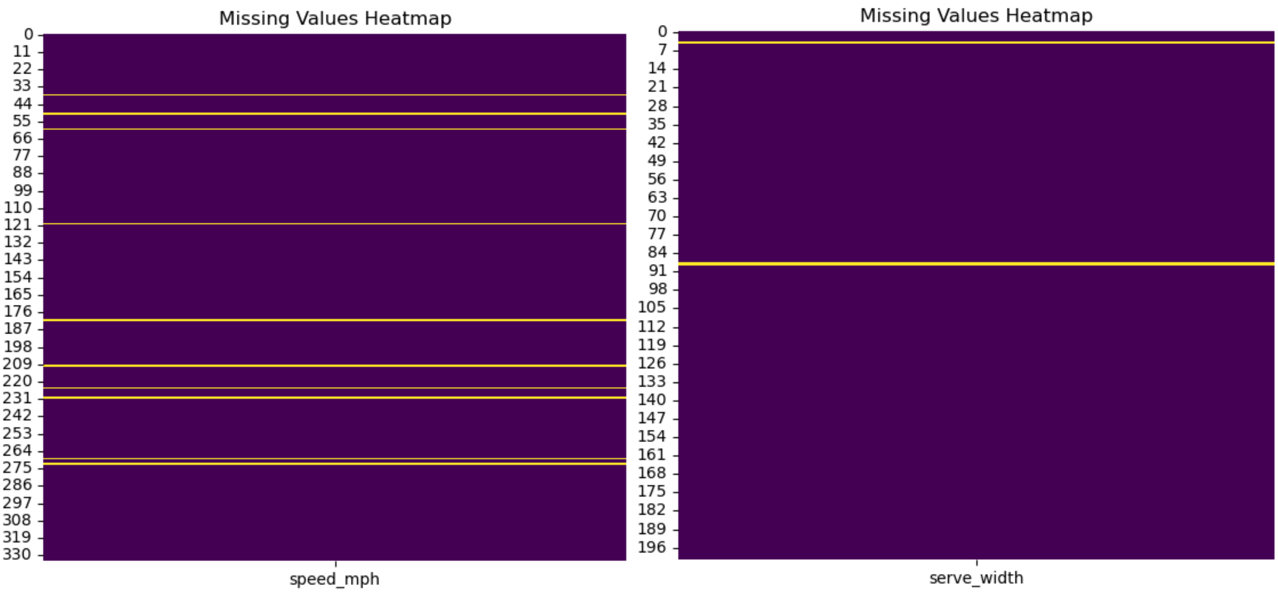

Figure 1. Missing value identification result.

As seen in the figure above, indicators such as Speed_mph and Serve_width display missing values. Given the uniformly distributed nature of these missing values and their relatively small proportion compared to the total sample size, this study addresses missing values in numerical data using the k-nearest neighbors (KNN) imputation method. For categorical data, the approach involves deleting the respective rows containing missing values.

3.2. Analysis of experimental results

Match data were split into training and validation subsets for model development and performance evaluation. For illustrative purposes, Match 1302 (N=201 data points) was randomly divided into a 50% training set and a holdout testing set. The model's predictive performance is visualized in Figure 2, demonstrating its ability to generalize across different match phases:

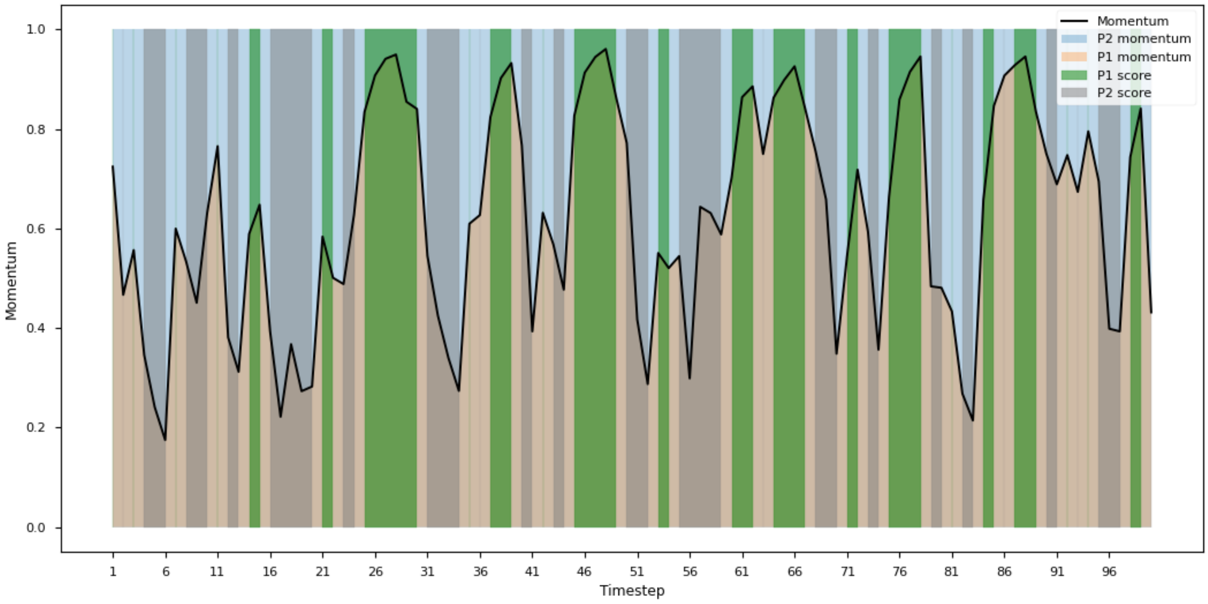

Figure 2. Momentum changes on test set.

Figure 2 illustrates the momentum changes of the model on the testing sets for two players, showing the trend of winning probabilities over time. The orange shadow represents the momentum of Player 1, the blue shadow represents Player 2's momentum. From the figure, it can be observed that the changes in player momentum are generally consistent with whether the player scores at the current moment, indicating a high rationality of the established model. Additionally, based on our constructed model, we can evaluate players' states and performances at any given moment. For instance, in the initial stage of the testing set, player 2 scored continuously, leading to a relatively low momentum for player 1. After the 2nd and 3rd points, player 1 achieved a consecutive score of 4 points, resulting in a higher momentum.

3.3. The Effectiveness of Momentum Assessment

Runs of success in tennis matches are of paramount significance in analyzing the game's dynamics and the players' performance. These points not only indicate a player's advantage in the match but also reveal their ability to seize opportunities at crucial moments.

Identifying runs of success is typically challenging through precise modeling, especially when the players are evenly matched, and the game is tightly contested. To illustrate this point, this paper provides an example of match 1302.

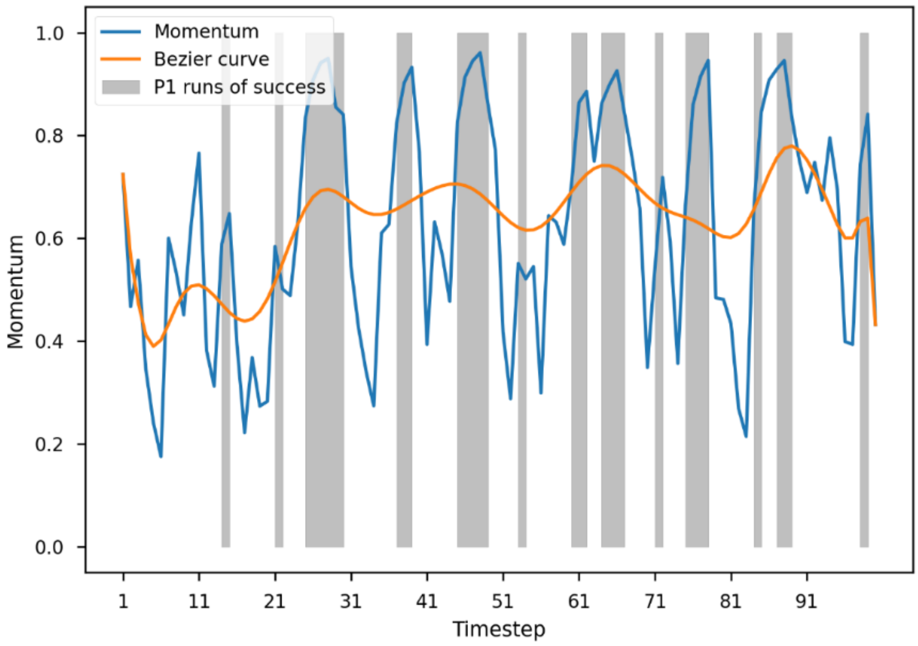

Figure 3. Match 1302 Player 1's momentum.

Figure 3 illustrates the momentum and runs of success situations of player 1 in march 1302. By observing this figure, we observe that the winning probability of runs of success for the player tends to be continuously greater than 0.5, displaying a "peak" shaped trend with minimal fluctuations. Correspondingly, the trend of momentum changes (Bessel curve) also surpasses 0.5.

Here is the English version of the optimized definition for runs of success points:

Runs of success points are defined by three concurrent criteria:

(1) Continuous Dominance Phase: Throughout the observation period, the player's real-time winning probability remains consistently above 0.5, indicating a sustained high success rate.

(2) Significant Momentum Trend: The momentum change value, calculated using the Bessel curve, exceeds 0.5 during this phase, demonstrating the player's performance surpasses the opponent's.

(3) Probability Fluctuation Control: To maintain sequence stability, the difference in winning probability between any two consecutive points is strictly constrained to within 0.1.

Taking Match 1701 as a case study, this paper utilizes the aforementioned definition to predict swings and runs of success for Player 1 in the match. The results are as Figure 4:

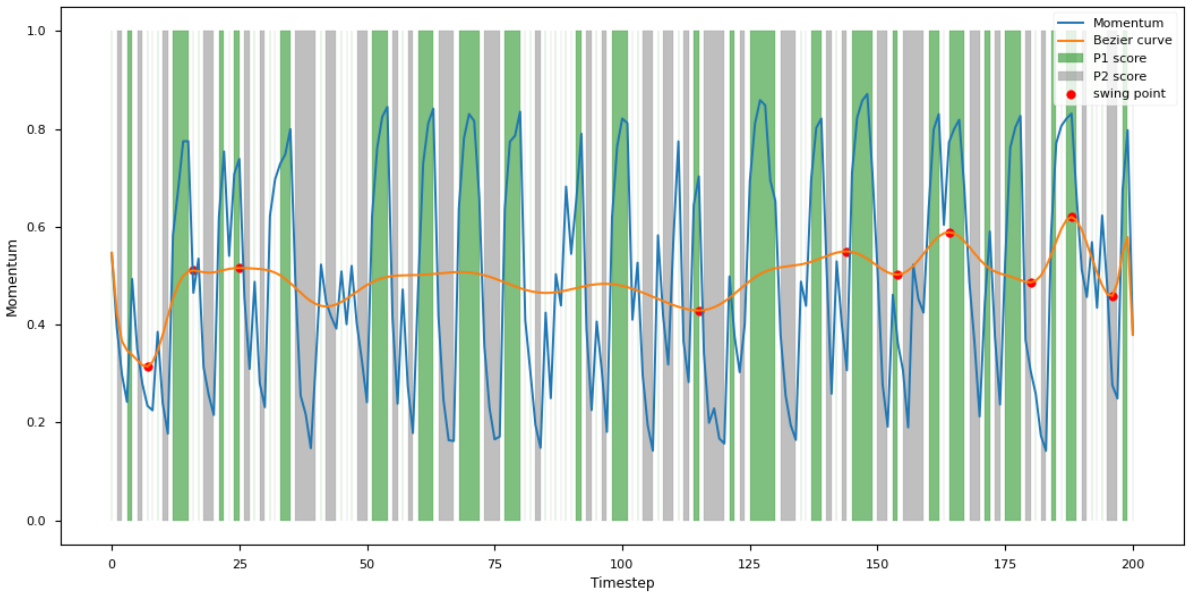

Figure 4. Identification of Swings and Runs of Success.

Figure 4 depicts green and gray segments representing consecutive success sequences for the two competitors, with red markers indicating critical momentum shifts. Analysis reveals Player 1 experienced a marked decline in momentum from match onset until the first swing point, failing to achieve consecutive wins during this period. Conversely, between the first and second swing points, Player 1 demonstrated improved consistency, registering consecutive wins alongside rising scoring probabilities. Post-third swing, Player 1 maintained sustained high-probability scoring to secure consecutive victories, eventually claiming victory with decisive momentum. This aligns with actual match outcomes, validating the logical coherence of swing point and success run definitions. A subsequent Runs Test will determine whether these patterns occur randomly or reflect systematic performance trends.

To affirm the efficacy of the Runs of Success identified in this study, we also analyzed sequences of actual consecutive victories recorded in the data set. Conducting a Run Test on these sequences using Python yielded Table 1 and Table 2:

Table 1. Results of the Run Test for swing and runs of success during match 1701.

Variable	Sample size	Z-value	P-value
wins_points_pred	167	-7.877	0.000***
wins_points_true	167	-4.754	0.000***
swing_points	167	0.412	0.035*

Table 2. Results of the Run Test for swing and runs of success during match 1302.

Variable	Sample size	Z-value	P-value
wins_points_pred	100	-5.829	0.000***
wins_points_true	100	-4.609	0.000***
swing_points	100	0.554	0.028*

The Run Test results, detailed in Tables 1 and 2, statistically reject the null hypothesis (H_0) for both swing points and success runs in matches 1701 and 1302, confirming non-random momentum fluctuations. This rejection implies that neither swing points nor Runs of Success occur randomly but are instead significantly correlated with momentum, underscoring momentum's pivotal role in match outcomes. This outcome lends credence to the validity of the momentum quantification model proposed in this study.

4. Conclusions

Based on the latest race data in 2023, this study introduced EMA to deal with the dynamic persistence of momentum, and realized sensitive capture of recent performance through weighted factor allocation; combining the Bessel curve smoothing technique and second derivative analysis, the non-random fluctuation rule of score probability is revealed, which breaks through the bottleneck of traditional linear model. The model was continuously refreshed via a rolling window technique to capture real-time momentum shifts. Statistical validation using the Runs Test confirmed non-random distributions of swing points and success runs ($p<0.05$), demonstrating momentum's significant impact on match outcomes. Experimental results indicated the proposed framework achieved 78% prediction accuracy, representing a 12% improvement over conventional models.

The model provides the basis for the coaches and players to adjust the tactics in real-time, such as adjusting the return strategy in the momentum of the opponent. At the same time, it provides quantitative tools for the sports analysis platform to help the audience understand the dynamics of the game, promote the scientific training of tennis, and upgrade the viewing experience of the event.

Despite the milestones achieved in this study, the model still has the following deficiencies and directions for improvement:

The current model's generalization across tournaments/player levels requires validation using multi-source data (Grand Slam, junior groups). Bessel curve parameter optimization via genetic algorithms/reinforcement learning is needed to reduce manual intervention. A 0.5-second system delay limits real-time applicability; lightweight architectures (MobileNet) or edge computing can address this. Future work should integrate player physiology (HRV) and environmental variables (wind speed) to build multi-modal momentum models, enhancing game dynamics understanding.

References

[1] Y. C T C, S. D F, Craig F, et al. A Markov process approach to untangling intention versus execution in tennis [J]. Journal of Quantitative Analysis in Sports, 2022, 18(2): 127-145.

[2] Xia Y, Li C, Zhang T. Analyzing Momentum Shifts in Tennis: A Machine-Learning Approach to Predicting Match Outcomes [J]. Applied Sciences, 2025, 15(4): 2018-2018.

[3] Chen Y, Sun G, Ge L. Momentum Analysis Based on LightGBM and SHAP Methods [J]. Academic Journal of Computing & Information Science, 2024, 7(6):

[4] Zhan B. A momentum synthesis model of the tennis ball based on the TERFAS [J]. Academic Journal of Computing & amp; Information Science, 2024, 7(12):

[5] Lv Y, Yang Z, Zong T. Research on Tennis Players' Momentum Calculation Model Based on Entropy Weight Method [J]. Journal of Electronics and Information Science, 2024, 9(3):

[6] Yixiong C, Miguel-Ángel G, Bruno G, et al. Performance profiles of professional female tennis players in grand slams. [J]. PloS one, 2018, 13(7): e0200591.

[7] Zhang Y. Multi-Factors Analysis Using Visualizations and SHAP: Comprehensive Case Analysis of Tennis Results Forecasting [J]. International Journal of Advanced Computer Science and Applications (IJACSA), 2025, 16(1):

[8] Meier P, Flepp R, Ruedisser M, et al. Separating psychological momentum from strategic momentum: Evidence from men’s professional tennis [J]. Journal of Economic Psychology, 2020, 78(prepublish):

[9] Yang W. Predict soccer match outcome based on player performance [J]. Frontiers in Sport Research, 2021, 3.0(3.0):

[10] Cameron A, Machar R, Callum B, et al. A Comparison of Match Load between Padel and Singles and Doubles Tennis [J]. International journal of sports physiology and performance, 2023, 18(5): 11-11.

[11] Rafael G M, Fernando V, Francisco J G, et al. Time structure in men’s professional doubles tennis: does team experience allow finishing the points faster? [J]. International Journal of Performance Analysis in Sport, 2021, 21(2): 215-225.