

# Research on Professional Tennis Performance Metrics Based on Random Forest and Analytic Hierarchy Process Algorithms

Jinyu Dai<sup>1, #</sup>, Kunyan Hu<sup>1, #</sup>, Ding Zhang<sup>2, #</sup>

<sup>1</sup> Academy of Astronautics, Nanjing University of Aeronautics and Astronautics, Nanjing, 210016, China

<sup>2</sup> Faculty of Physics, Nanjing University of Aeronautics and Astronautics, Nanjing, 210016, China

#These authors contributed equally

**Abstract:** Understanding the dynamics of player performance in professional tennis, particularly how quantifiable metrics such as momentum influence match outcomes, is crucial for advancing sports analytics and enhancing coaching strategies. This study investigates performance metrics and momentum in professional tennis. The research introduced an innovative method of quantifying "momentum" in the sport, analyzing its impact alongside other performance indicators. We applied the Analytic Hierarchy Process (AHP) to determine the relative importance of various performance factors, revealing technical skills as the predominant influence. The study employed Random Forest algorithms to model match outcomes, contrasting predictions with and without momentum as a variable. Our results significantly demonstrated that momentum plays a critical role in the dynamics of match outcomes, challenging the notion that player success sequences are merely random. The robustness of the Random Forest model was further validated through sensitivity analysis, highlighting its effectiveness as a predictive tool in sports analytics. This research not only enhances the understanding of player performance in tennis but also supports the incorporation of momentum in sports performance evaluations.

**Keywords:** Tennis; Momentum; Random Forest; Analytic Hierarchy Process.

## 1. Introduction

In professional tennis, understanding the factors that influence match outcomes is paramount for both coaching strategies and performance improvement[1, 2]. Traditional analyses have predominantly focused on physical attributes and simple match statistics, such as serve speed and error counts[3]. However, the dynamic nature of the sport suggests that other, less tangible factors, such as a player's momentum during a match, also play a crucial role in determining the outcome[4].

In this study, this paper first preprocessed the data and then built a hierarchical analysis framework based on players' match data, which defined the concept of "momentum." Subsequently, to validate its impact on player performance, this research explored the predictive performance of algorithms both with and without the inclusion of momentum. Finally, we conducted a sensitivity analysis centered around the Random Forest algorithm.

## 2. Analyzing Momentum and Performance Metrics in Tennis

### 2.1. Data Preprocessing and Preparation

The data for this study comes from [https:// www.comap.com/contests/mcm-icm](https://www.comap.com/contests/mcm-icm). The dataset from 23 years of Wimbledon men's singles matches is utilized to understand and analyze match dynamics. The initial focus is on comprehending the specifics of the provided datasets: Wimbledon\_featured\_matches.csv and data\_dictionary.csv, which detail each match point, including scores, servers, scorers, and other pertinent data. In preprocessing, outliers are eliminated, and duplications are resolved by integrating data for similar elements like point\_victor, p1\_points\_won, and

p2\_points\_won, reflecting athletes' performance more clearly. Key factors are selected from the dataset for in-depth research and analysis. The study then plans to generalize and validate the developed model using data from both men's and women's singles matches at the 2021 U.S. Open and Wimbledon.

Before the model is built, it is necessary to understand the concept of "momentum," which refers to the advantage or force gained by one side in a game as a result of a series of events. To quantify momentum, specific metrics defined in the next chapter are selected to evaluate a player's momentum.

### 2.2. Tennis Player Performance Evaluation Criteria

For the evaluation criteria of players, we assess the following aspects: the winning situation of the match plate points, the player's technique, the player's mentality, and the endurance. This chapter will focus on the following aspects [5].

Scoring a point on a serve revolves primarily around two situations: the situation of serves and defense serves. The former is considered the most direct evaluation metric in tennis, where the serving side typically holds an advantage. As a server, a player can control the pace of the game and maintain offensive initiative, while minimizing errors to keep the game close. Thus, the effectiveness of a serve can be evaluated based on its speed, position, and depth. On the other hand, an opponent's serve often sets the pace of the match. Failing to effectively handle your opponent's serve can make it challenging to seize the initiative. The depth and power of a return are critical in assessing a good defensive serve. By successfully defending against the serve, one can diminish the opponent's advantage.

The skills of tennis players can be notably assessed by

examining the number of high-end balls they serve during a match. For instance, a player's ability to serve unreturnable balls, execute untouchable counter serves, skillfully catch balls at the net, and score points on match points are all significant indicators of their skill level. These elements not only reflect their technical prowess but also their tactical acumen and mental strength under pressure.

The mindset of tennis players is a crucial factor in their performance and the overall outcome of the match. A stable and positive mindset aids players in staying focused and confident, enabling them to handle challenges and pressures effectively during the game. Non-mandatory errors and match-point emotions are key aspects to consider. Non-mandatory errors, such as double faults or balls hitting the net, occur without direct pressure from the opponent. These errors serve as important indicators of a player's technical consistency and mental concentration. On the other hand, the mental quality of players during critical moments, such as match points, often determines the game's outcome. Players with strong mental skills can maintain their composure under pressure and perform better.

In tennis, a player's endurance is a crucial factor that influences their form. Endurance primarily pertains to a player's ability to sustain high-intensity exercise throughout a game, significantly impacting their performance and results. This aspect is often assessed by measuring the distance a player runs during a point, expressed in meters. The distance traveled during a match serves as a key indicator of their endurance and overall form. This movement distance not only relates to a player's stamina but also affects their explosiveness and decision-making ability during the game.

### 2.3. Hierarchical Analysis to Determine Indicator Weights

When we determined the weights of the main four dimensions, we chose AHP as the method to comply with the weighting coefficients of all indicators in the evaluation system.

AHP method is called hierarchical analysis, is a combination of qualitative and quantitative decision-making analysis methods, mainly through the complex research object is divided into several levels and several indicators, two and two indicators for comparison, clear evaluation of the indicator system of the importance of each indicator of the percentage of the evaluation system, to clarify the evaluation system of the weight coefficient of each item. In this paper, the athletic performance of a tennis player needs to be evaluated concerning several objectives and attributes. Therefore, from a methodological point of view, the hierarchical analysis method is more suitable for this study. The hierarchical analysis method is used to analyze the significance of each index at different levels in the evaluation system of tennis players' sports performance, construct a suitable judgment matrix, test the consistency, and scientifically evaluate the weights of each type of index at all levels in the system of tennis players' sports index.

First, we construct a good judgment matrix, which is a subjective method of determining the weights of the indicators, the judgment matrix is constructed as follows:

After constructing the judgment matrix  $A_{n \times n}$ , calculate its eigenvalues and eigenvectors, find the largest eigenvalue  $\lambda_{max}$  and the eigenvector corresponding to the largest eigenvalue  $T$ . Further conducting consistency tests for judgment matrices.

**Table 1.** The construction meaning of the matrix

Scale	Connotation
1	Two factors are of importance.
3	The former is slightly more important than the latter.
5	The former is significantly more important than the latter.
7	The former is more strongly important than the latter.
9	The former is extremely more important than the latter.
2,4,6,8	Intermediate values of the above adjacent judgments.
RECIPROCAL (MATH.)	If the ratio of the importance of factor i to that of factor j is $a_{ij}$ , then the ratio of the importance of factor j to that of factor i is $a_{ji}=1/a_{ij}$

$$CI = \frac{\lambda_{max} - n}{n - 1} \quad (1)$$

The following table shows that according to the dimension n of the judgment matrix, RI can be judged.

**Table 2.** Criteria for judging matrix dimensions

n	1	2	3	4	5	6	7	8	9
RI	0	0	0.58	0.9	1.12	1.24	1.32	1.41	1.45

Consistency calculation ratios are as follows:

$$CR = \frac{CI}{RI} \quad (2)$$

If  $CR < 0.1$  passes the consistency test, otherwise the consistency test fails. The eigenvector corresponding to the largest eigenvalue can be obtained, T.

$$T = [t_1, t_2, t_i \dots t_n] \quad (3)$$

The weight vector is the W.

$$w_i = \frac{t_i}{\sum_{i=1}^n t_i} \quad (4)$$

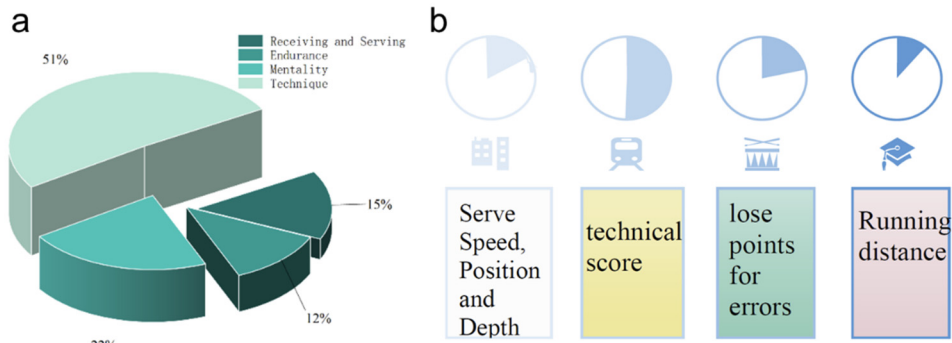
$$W = [w_1, w_2 \dots w_n] \quad (5)$$

Based on the data we collected, we solved the model and obtained the following results, the identity matrix as shown in Eq 6.

$$\begin{bmatrix} 1 & 4 & 3 & 3 \\ \frac{1}{4} & 1 & \frac{1}{2} & 2 \\ \frac{1}{3} & 2 & 1 & 2 \\ \frac{1}{3} & \frac{1}{2} & \frac{1}{2} & 1 \end{bmatrix} \quad (6)$$

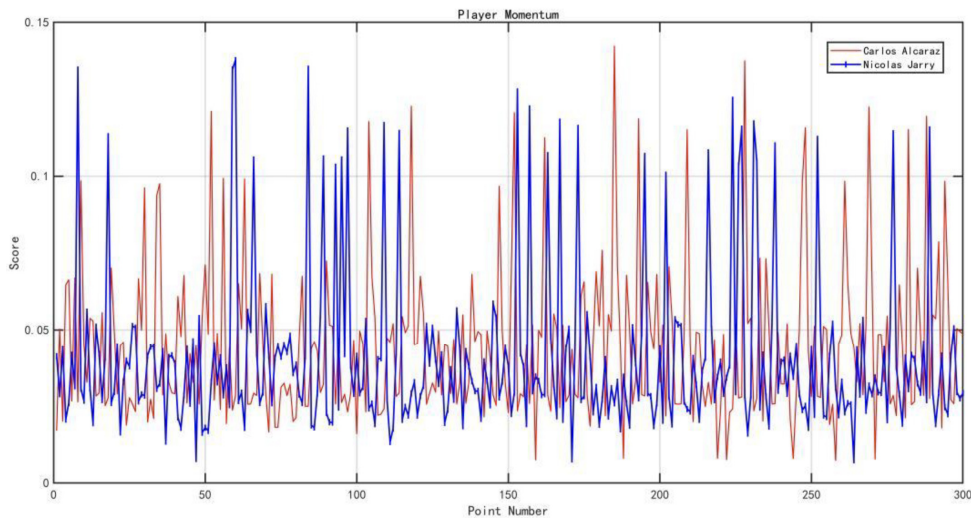
This study visualizes the weights of the selected metrics as shown below:

As shown in Figure 1, it can be seen that the technical factors of the athletes are dominant in assessing their performance ability, with an influence percentage of 51%, while the factors of endurance, mentality, and serve receive also have some influence on the athletes' performance in the game.



**Figure 1.** (a) Weight vector 3D pie chart; (b) Demonstration of the weighting of different indicators.

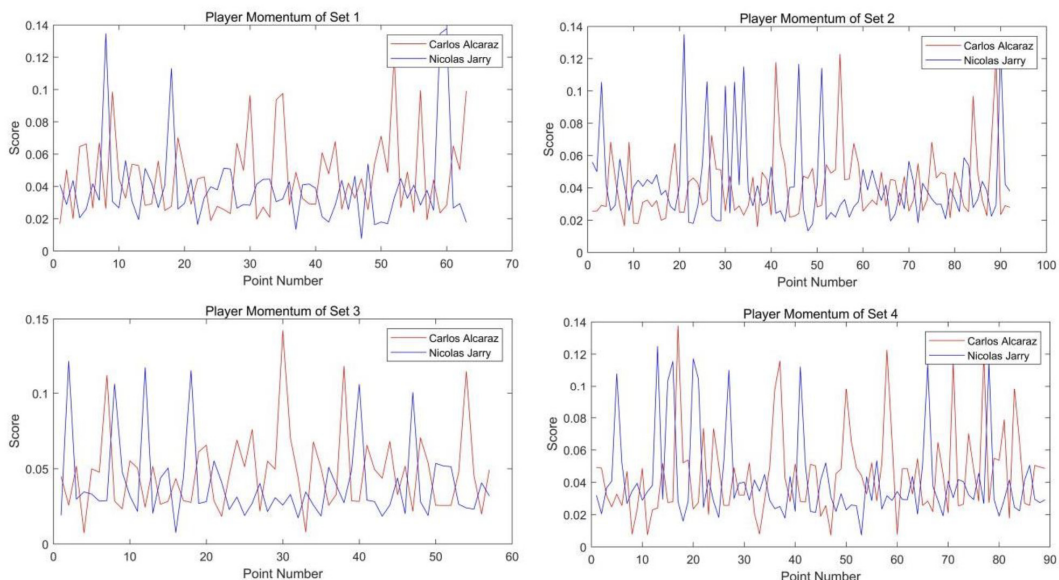
## 2.4. Score Visualization and Result Analysis



**Figure 2.** Changes in player momentum as the game progresses

Figure 2 above is a graph of the changing patterns of the two athletes' momentum with the course of the game, which we have visualized from the date of the game. Where the horizontal coordinate represents the course of the game (number of hits in the game) and the vertical coordinate represents the momentum score. Through the curve to analyze the state of the two contestants, the degree of the vertical

coordinate of the curve can be converted into a reflection of the contestant's "momentum", the larger the value of the vertical coordinate, the higher the contestant's "momentum". The smaller value of the vertical coordinate can, to some extent, reflect that the player's own mistakes are more frequent and the state of the game is relatively poor.



**Figure 3.** Player momentum in every game

Figure 3 shows the overall data, the data is relatively large,

and it is difficult to analyze the specific state of the player's

game, so we further decomposition the data, respectively, to study the changes in the player's "momentum" and the state of the game in each game. It can be concluded that there is a certain periodicity in the "momentum" of the two players in each game and the "momentum" of the two players alternates. We analyze that it may be due to the greater influence of the players' techniques, such as the influence of the high-end balls served by the players such as the untriggered balls, untouchable counterattacks, and the points scored at the nets, etc. The "momentum" of the two players in each set can be analyzed as the influence of the players' techniques.

### 3. The Role of Momentum and Prediction Accuracy

The tennis coach is skeptical that "momentum" plays any role in the game, believing that a player's fluctuations and successful runs are random. Therefore, we need to analyze

and verify whether "momentum" has a significant impact on the outcome of the game. After determining the influencing factors, we performed multivariate categorization, and for the first time, we selected the other four factors except "momentum", and analyzed the correctness of the results of the game through the Random Forest algorithm of multivariate categorization; The second time, the influence of "momentum" was added to the four influences, resulting in the correct-ness of the results of the second analysis. The coach's opinion is determined by comparing the correct rates of the two classifications and determining the significance of the impact of "momentum" on the outcome of the game based on the difference in the correct rates.

The outcome of a tennis match is influenced by a variety of elements, and in this part, this study has selected the following factors that have a significant impact on the outcome of the match:

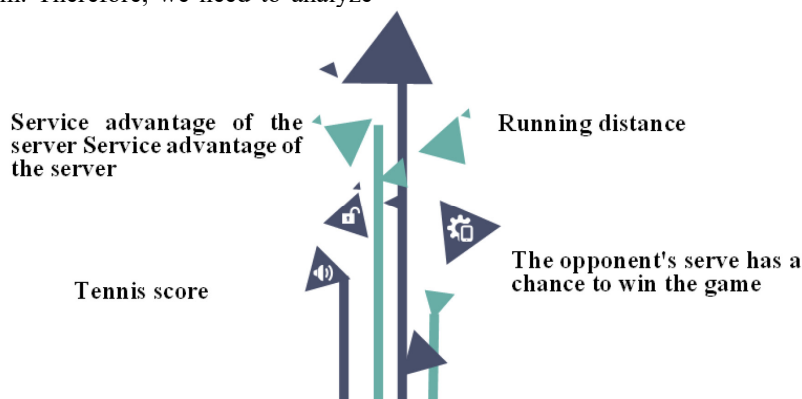


Figure 4. Metrics that influence the game

To achieve the effects described above, this study will utilize Random Forest as a tool. Random Forest is a model based on the Bootstrap resampling technique, composed of multiple decision trees. The indicators in Figure 4 will be used

as independent variables, with match outcomes serving as the dependent variable output. The specific results are shown in Figure 5.

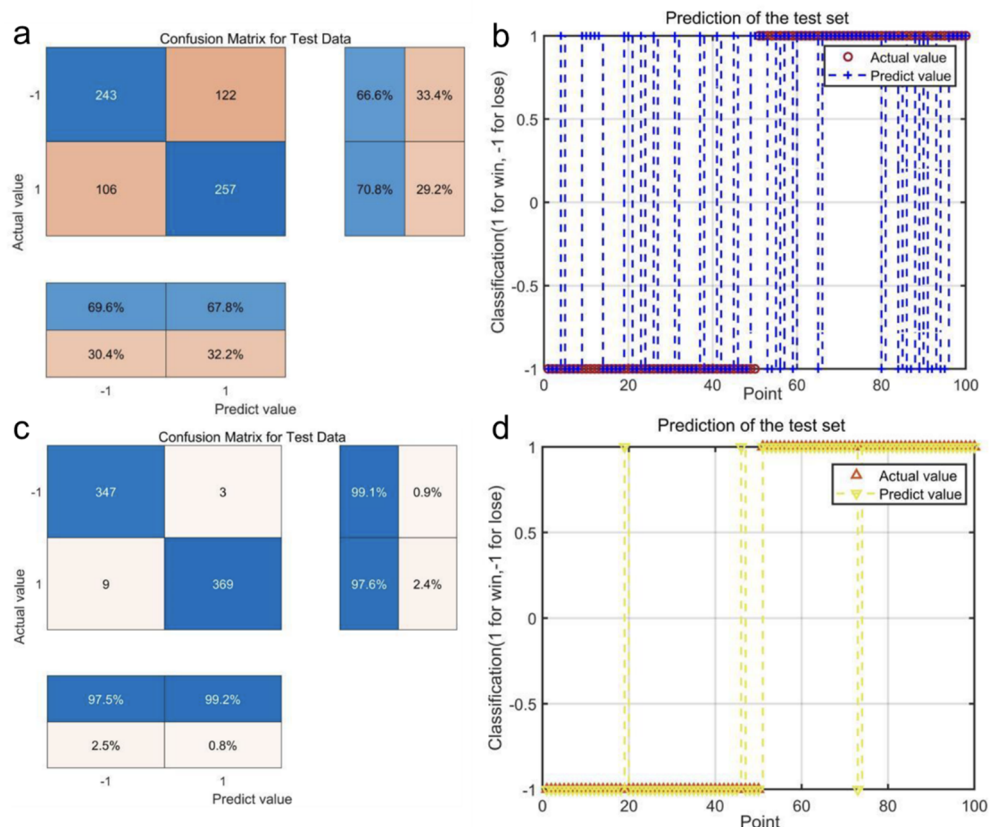


Figure 5. a and b predict without considering momentum, while c and d incorporate momentum into their predictions

Analyzing the results of the first classification, we can conclude that the correct rate of judging the result of the match through the classification of the first four indicators is 66.209%, the correct rate of judging the result of the match as a win is 67.8%, and the correct rate of judging the result of the match as a loss is 69.6%, which is not a very high value, and there is a large discrepancy between the actual value and the predicted value.

In the second classification, the factor of "momentum" was added to the four influencing factors, and we can see from the

results that the correct rate of judging the outcome of the match increased greatly, reaching 98.352%. The error rate of judging the outcome of the match as a loss is 2.5%, and the error rate of judging the outcome of the match as a win is only 0.8%, which is a significant decrease compared to the first classification. The true values also largely match the predicted values, with only a few lines of discrepancy. Therefore, we can conclude that "momentum" plays an important role in predicting the outcome of a game, and that player fluctuations and successful runs are not random.

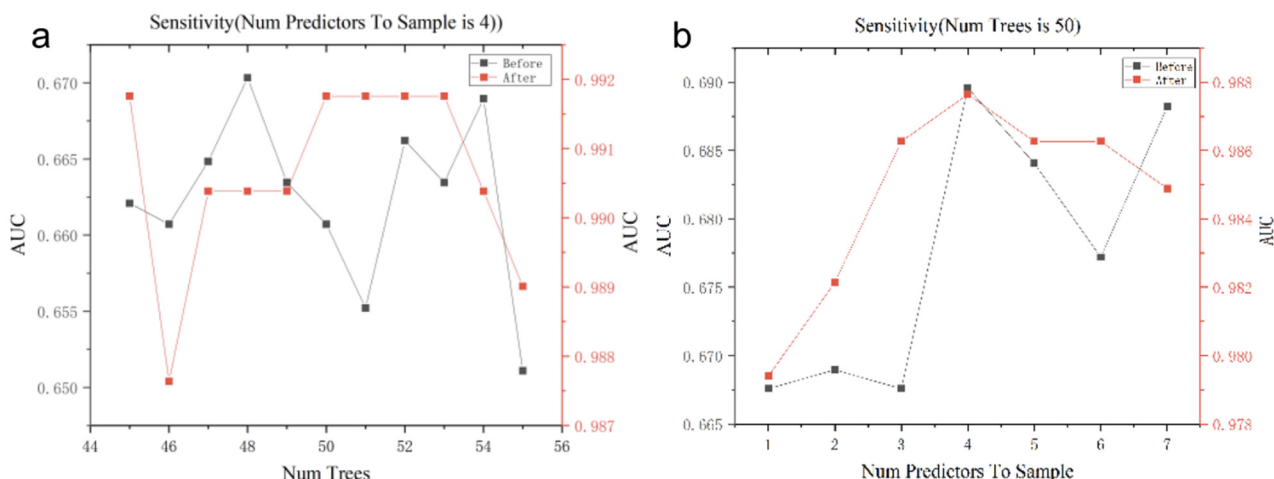


Figure 6. Model parameter validation chart

Furthermore, to verify the impact of model parameters on model performance, this study conducts further sensitivity analysis. As shown in Figure 6, we analyze the sensitivity of the random forest model established in this chapter, firstly, we control the number of features is 4 to keep the same, and analyze the change of the correct rate by changing the value of the feature tree, then we exchange the fixed parameter with the input variable, control the number of features is 50 to keep the same, and analyze the change of the correct rate by changing the value of the number of features, which can be seen that, when the input variable changes, the fluctuation value of the correct rate is less than 0.01 both before and after the factor "momentum" is added. It can be seen that when the input variables are changed, before and after adding the factor of "momentum", the fluctuation value of the correct rate is less than 0.01, which can be judged that the stability of the model we have established is high.

#### 4. Conclusion

This study meticulously analyzed performance metrics and momentum in professional tennis, using data from major tournaments like Wimbledon. The data preprocessing effectively cleaned the dataset, setting a robust foundation for nuanced analysis. Employing the Analytic Hierarchy Process (AHP), technical skills emerged as the most influential factor in player performance, validated by hierarchical analysis. The investigation into the impact of momentum on match outcomes revealed that including it as a variable significantly enhances predictive accuracy. This finding disputes the belief that success sequences in tennis are random, highlighting

momentum's pivotal role in match dynamics. Furthermore, the robustness of the Random Forest model was confirmed through sensitivity analysis, showcasing its potential as a reliable predictive tool in sports analytics. Overall, this research advances the understanding of player dynamics and supports the integration of momentum in sports performance analysis.

#### References

- [1] TONER J, CARSON H J, COLLINS D, et al. The prevalence and influence of psychosocial factors on technical refinement amongst highly-skilled tennis players[J]. International Journal of Sport and Exercise Psychology, 2020,18(2): 201-217.
- [2] ANDERSON E, STONE J A, DUNN M, et al. Coach approaches to practice design in performance tennis[J]. International Journal of Sports Science & Coaching, 2021,16(6): 1281-1292.
- [3] PLUIM B M, JANSEN M G, WILLIAMSON S, et al. Physical demands of tennis across the different court surfaces, performance levels and sexes: a systematic review with meta-analysis[J]. Sports medicine, 2023,53(4): 807-836.
- [4] WEIMER L, STEINERT-THRELKELD Z C, COLTIN K. A causal approach for detecting team-level momentum in NBA games[J]. Journal of Sports Analytics, 2023,9(2): 117-132.
- [5] KRONZER J R. Using Video-Based Deliberate Practice Training Techniques to Improve Serve Return Performance and Self-Efficacy in Collegiate Tennis Players[D]. University of Minnesota, 2020.