Transactions on Computer Science and Intelligent Systems Research
ISSN: 2960-1800, eISSN: 2960-2238 | Volume 5
AIDML 2024

A Study of Momentum in Tennis Based on Multiscale Momentum-Success Test Model and Swings Prediction Model

Yiyi Lin [1, *], Yuhang Xiang [2], Yuhan Huang [2]

[1] School of Electrical Engineering & Automation, Wuhan University, Wuhan, China, 430072

[2] School of Remote Sensing and Information Engineering, Wuhan University, Wuhan, China, 430072

* Corresponding author: 13479935255@163.com

Abstract. In tennis, "momentum" is one of the most important factors affecting the results of the game. First, this paper establishes the Multiscale Momentum-Success Test Model and calculates the proportion of the winning side's momentum at four different scales (score, game, set and match). The proportions of the four scales are 70.3%,78.1%,84.6%,93.5%. Therefore, it can be proved that momentum plays a role in the game. Also, the fluctuation and success in the game are not random. Afterward, this paper establishes the Support Vector Machine Model (SVM) and the Feed-Forward Neural Network Model (FNN). Three correlations are analyzed for factors such as scoring ratio, serve, consecutive scores (lost points), highlight scores, major lost points, and physical condition. It is concluded that the serve, consecutive scores (lost points), highlight scores, and major lost points will play a role in the occurrence of Swing Points. By the two models, we create the Swings Prediction Model. The prediction accuracy of SVM and FNN Model are 84.85% and 67.4%. Finally, based on the momentum changes, match suggestions can be made for the coaches.

Keywords: Tennis; Momentum; Swing; SVM; FNN.

1. Introduction

In the 2023 Wimbledon Gentlemen's final, 20-year-old Spanish rising star Carlos Alcarazdefeated 36-year-old Novak Djokovic. This loss at Wimbledon was Novak Djokovic's first since 2013 and ended a brilliant run for the best player in Grand Slams. In this remarkable battle, there were several surprising swings and roundabouts in this game. In the first set of the match,Djokovic won 6-1 easily. However, he ended up losing the match.

The seemingly dominant side can sometimes have incredible swings for many points or even games and sets, which is often attributed to "momentum [1]". In sports, some players may feel thatthey have a kind of momentum, or "strength/force" during a game. However, this phenomenon is difficult to measure. Further, even if momentum exists, it is not yet clear about exactly how it affects the game.

According to previous studies [2], winning any of the first eight games of the first and/or second set was a significant predictor of success in tennis matches. However, when only results from more competitive matches (when matches extended to nine or more sets) were examined, the 8th, 10th, and 11th games of the first set were significant predictors of winning, while only the 4th game of the second set increased the probability of winning. No gender or ability differences were found. Based on these findings, researchers are advised to be cautious when inferring mental momentum, as these findings depend on the fairness of the competitors in the match, and mental momentum appears to be a highly individualized issue.

However, this study has some limitations. It drew its conclusions through interviews with athletes and did not use any data or modeling, which is subjective and not rigorous enough. To fill this gap, this paper uses the data from Wimbledon 2023 Gentlemen's singles matches after the second round and build several models to research this topic.

2. Model

2.1. The Multiscale Momentum-Success Test Model

To prove the momentum plays a role in the game, in this paper, the Multiscale Momentum-Success Test Model is created. Firstly, the paper uses the pre-processed tennis match data and define some indicators, which include objective indicators (W_{object}), general indicators (W_{common}), excellent indicators (W_{good}), and error indicators (W_{bad}). Then we use the relevant formula to express the momentum and substitute the corresponding indicators to find out the current total momentum. The flowchart for the model is shown in figure 1.

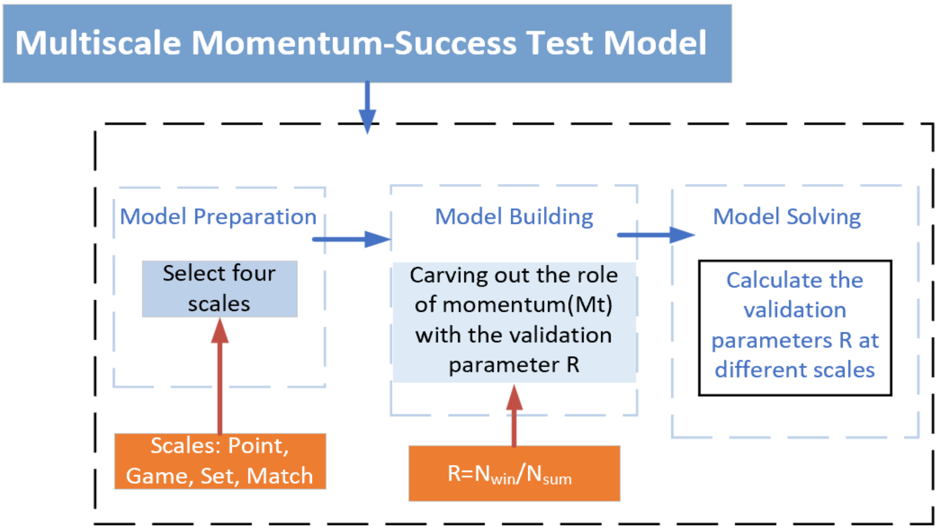

Figure 1. Flowchart for Multiscale Momentum-Success Test Model

2.2. The Swings Prediction Model

To predict the swings of the match and find out the factors with the highest correlation, this paper selects a match and builds the Swings Prediction Model. Some relevant elements of the model are defined as follows. The flowchart for the model is shown in figure 2.

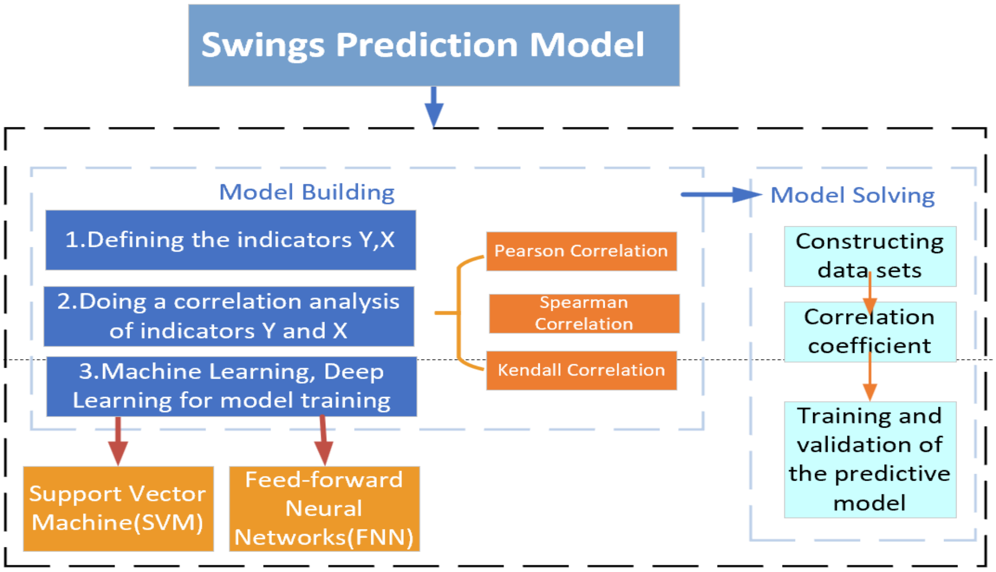

Figure 2. Flowchart for Swings Prediction Model

For the output, prediction, and verification of machine learning later, we define a Y, where Y_A and Y_B represent the Y values of players A and B, respectively. The specific definition is as follows: If player A wins and B loses, then Y_A=1 and Y_B=-1; If player B wins and A loses, thenY_A=-1 and Y_B=1.

After that, the paper identified several potentially relevant indicators and defined them specifically as follows. They are scoring ratio indicator X_1, serve indicator X_2, consecutive points scored (or lost) indicator X_3, highlight scoring indicator X_4, significant missing score indicator X_5, physical condition indicator X_6. X_1 reflects the difference between the scores of two players and is calculated as follows (letplayer A score P_1 and player B score P_2):

$$X_{A1} = \frac{P_1}{P_1 + P_2}; X_{B1} = \frac{P_2}{P_1 + P_2} \tag{1}$$

X_2=1 when the player is on the serve side and X_2=0 when he is not on the serve side. Forexample, if player A serves and B does not serve, there are: $X_{A2} = 1$; $X_{B2} = 0$. The indicator X_3 is equal to 3 (or -3) if at least 2 consecutive scores (or lost points) on ballshave occurred before the current moment. If there were no consecutive scores before the current moment it is recorded as 0. When the five highlights of ACE, Winner, Forehand and Backhand Strike, and Victory onBreak Point occur before the current moment, let H_{A4} and H_{B4} be the sum of the number ofthese highlights of A and B players respectively. The formula for calculating the highlight score index is as follows:

$$\mathrm{X}_{B4} = \frac{H_{B4}}{H_{A4} + H_{B4}} \tag{2}$$

Where X_{A4}, X_{B4} initial value is 50%; if the denominator is 0 (i.e., so far there is no highlight score) then X_{A4}, X_{B4} also equal 50%.

When there are five types of major points lost before the current moment, namely, double serve error, unforced error, failure to hit the ball at the net, break point, loss of break point, letL_{A5} and L_{B5} be the sum of the number of these major points lost by players A and B beforethe current moment. The formula for calculating the major point loss indicator is as follows:

$$\mathrm{X}_{A5} = \frac{L_{B5}}{L_{A4} + L_{B4}}, \quad \mathrm{X}_{B5} = \frac{L_{A5}}{L_{A4} + L_{B4}} \tag{3}$$

Where the initial value of X_{A5} and X_{B5} is 50%; if the denominator is 0 (i.e., there have been no significant losses of points so far) then X_{A5} and X_{B5} are also 50%. X_6 reflects the physical strength of the player at the current moment, let D_{A6}, D_{B6} is the sum of the running distance of A and B players up to the current moment, respectively. Theformula for calculating the physical condition indicator is as follows:

$$X_{A6} = \frac{D_{B6}}{D_{A6}}; X_{B6} = \frac{D_{A6}}{D_{B6}}; X_{B6} = \frac{D_{A6}}{D_{B6}} \tag{4}$$

Where the initial values of X_{A6} and X_{B6} are 1.

In summary, these six indicators can be integrated into an eigenvector:

$$X_A = \left[X_{A1}, X_{A2}, X_{A3}, X_{A4}, X_{A5}, X_{A6}\right]; X_B = \left[X_{B1}, X_{B2}, X_{B3}, X_{B4}, X_{B5}, X_{B6}\right] \tag{5}$$

Based on the results above, we conduct correlation analysis [3-5] of Y and X to explore their correlation, and then find out those factors with high correlation, where Y_A, Y_B can only take

$\{\pm 1\}$. The correlation analysis method uses Pearson correlation coefficient, Spearman correlation coefficient and Kendall correlation coefficient.

2.3. Machine Learning and Deep Learning Network Model Training

In this part, the paper applies machine learning [6] and deep learning [7]. The current moment's features are trained to determine whether it is a win or not. It is essentially a dichotomous problem. Based on the correlation analysis the paper inputs several dimensions with correlation in the dataset and divides the training set and testset in the ratio of 7:3 and used two methods--support vector machine and feed-forward neural network, for model training and validation.

The network model is trained using two methods, support vector machine (SVM [8]) and feed-forward neural network (FNN [9]), and validated after training to determine the model accuracy, as described below:

Support Vector Machine (SVM) is a binary classifier that classifies data by solving the maximum margin hyperplane. It provides a clear and powerful approach for solving complex nonlinear problems. In the linearly differentiable case, the optimal classification hyperplane is found for two types of samples, while in the linearly indivisible case, the optimal classification hyperplane is found by introducing relaxation variables and nonlinear mapping, which maps the samples to a high-dimensional space and realizes the linearly differentiable case.

A feedforward neural network (FNN) is a simple artificial neural network where informationis passed in one direction without forming circular connections. It passes forward from the input node to the hidden node (if any) and finally to the output node. There are no loops or loops in the network.

3. Results

3.1. The establishment and results of the Multiscale Momentum-Success Test Model

To prove that "momentum" does play a role in the match and swings in play and runs of success by one player are not random, the paper establishes the Multiscale Momentum-Success Test Model. To better verify the feasibility of the model, this paper uses all the matches, sets, games, andscores of the given 2023-Wimbledon (the data is from Wimbledon 2023 Gentlemen's singles matches) after second round., whose detailed solution steps are as follows.

The paper uses point, game, set, match as the scale respectively, counts the number of points (P_{win}, G_{win}, S_{win}, M_{win}) in which the winning side has a momentum greater than the losing side in the previous moment, and according to the total number of points (P, G, S, M), calculates the validation parameter R1, R2, R3, R4 under the point, game, set, match scales, the formula is as follows:

$$R_1 = \frac{P_{win}}{P}, R_2 = \frac{G_{win}}{G}, R_3 = \frac{S_{win}}{S}, R_4 = \frac{M_{win}}{M} \tag{6}$$

The results of the four validation parameters are shown in table 1:

Table 1. The Results of The Four Validation Parameters

Validation parameters	Results	Validation parameters	Results
R_1	70.3%	R_3	84.6%
R_2	78.1%	R_4	90%

Based on the results given above, the following conclusions can be drawn: first, the momentum credibility by R_1, R_2, R_3, R_4 is 1.04, 1.16, 1.26, and 1.39. Second, according to the relative size relationship of R_1, R_2, R_3, R_4, some innovative conclusions are obtained. In the above results,

$R_4>R_3>R_2>R_1$, the momentum calculated on the scale of MATCH is most capable of predicting the victory of the match, SET follows, GAME and POINT are relatively less accurate.

In summary, it can be proved that "momentum" does play a role in the match and swings in play and runs of success by one player are not random.

3.2. The establishment and results of the Swings Prediction Model

To predict the swings of the match this paper chooses and finds out the factors with the highest correlation, the Swings Prediction Model is built. First, the paper constructs the dataset used for machine learning and deep learning, to ensure the sufficiency of the sample. The paper uses the given 2023-wimbledon all the process of the game, its specific calculation steps are as follows:

Step1. Judge the winning situation of the current point and obtain the response value Y of the winning and losing sides respectively.

Step2. Calculate the score proportion indicator X_1 according to the score before the current point respectively.

Step3. According to the serving side at the current point, which can be determined before the current point occurs, obtain both sides of the serving right indicator X_2 respectively.

Step4. Calculate the consecutive scores (or lost points) indicator X_3 respectively.

Step5. Calculate the highlights of the score indicator X_4.

Step6. Calculate the major lost points indicator X_5.

Step7. Count the running distance before the current point, respectively calculate the physical strength indicator X6.

Step8. Integrate the feature vectors that may lead to swings at each point with the results of this point, in addition to facilitate subsequent training and learning.

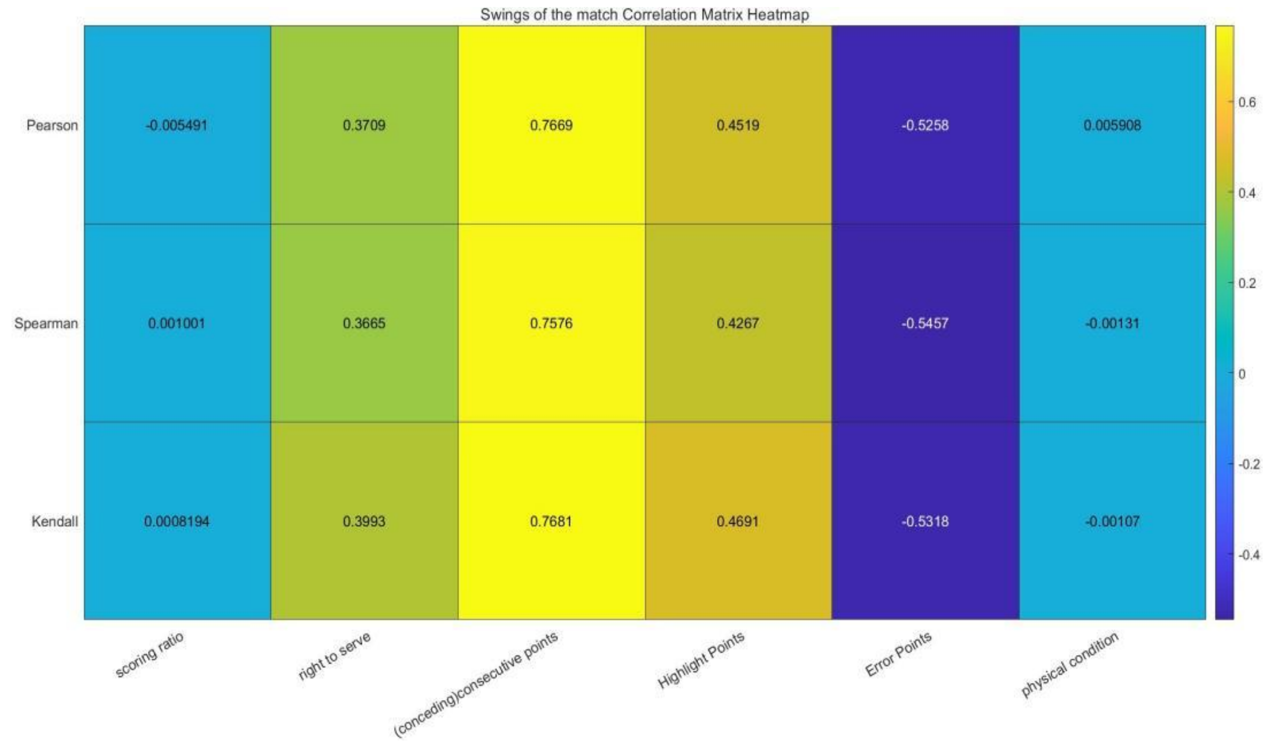

Figure 3. Correlation coefficients between the occurrence of swings and the indicators

Then, according to the dataset generated above, the Pearson correlation coefficient, Spearman rank correlation coefficient and Kendall rank correlation coefficient were calculated as shown in figure 3.

Therefore, several conclusions can be drawn below. First, the largest correlation is the consecutive points scored (or consecutive pointslost) indicator, showing a strong positive correlation, which is very likely to lead to swings. Second, Highlighted points and major points lost have a medium positive and medium negative correlation respectively, indicating that the occurrence of key points or points lost during the match will affect the occurrence of swings to a certain extent. Last, the weak positive correlation between serving and not serving and the occurrence of swings suggests that the mastery of serving power during the match may slightlyaffect the occurrence of Swings.

Based on the correlation analysis above, we input several dimensions with correlation in the dataset we generated and divide the training set and testset in the ratio of 7:3. Then we use two methods (support vector machine and feed-forward neural network) for model training and validation. The validation is shown in figure 4 and 5.

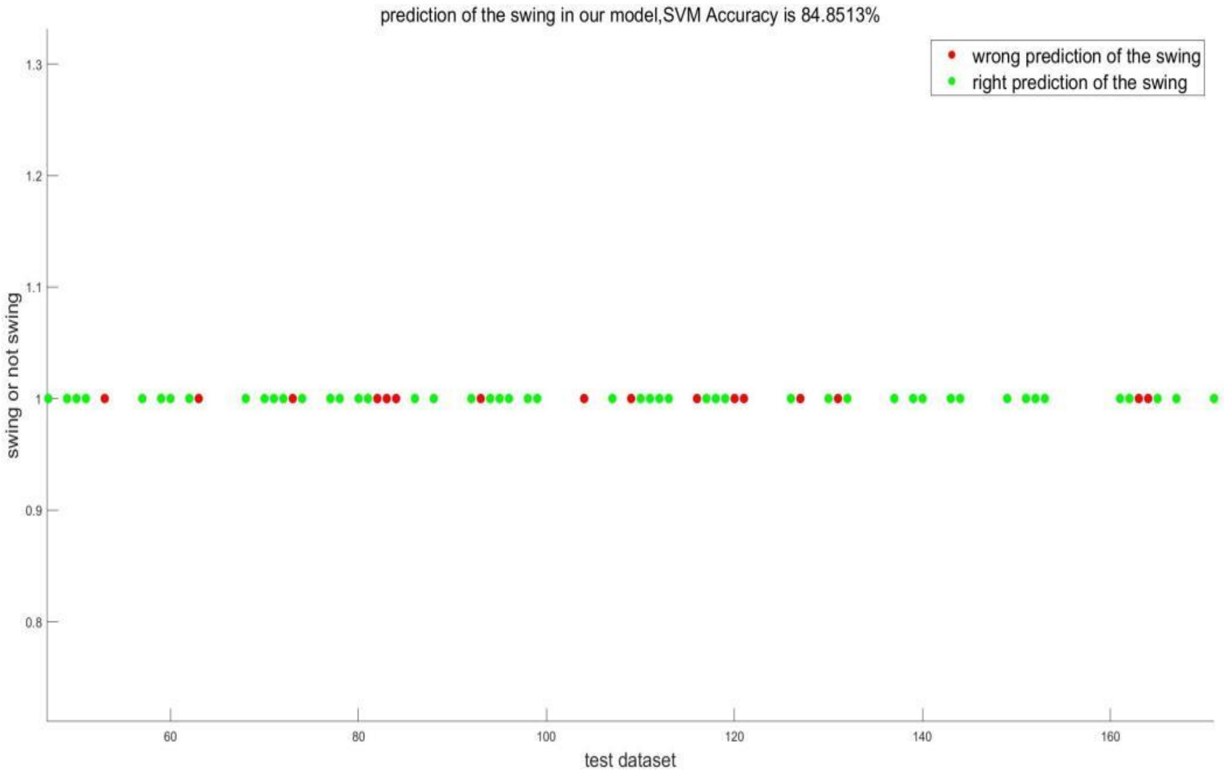

Figure 4. SVM prediction of swing occurrence in the test set

In summary, the following conclusions can be drawn:

First, the model we built has a clear advantage in predicting the occurrence of swing. The trained forward feedback neural network, which predicts whether swing occurs or not with each point, achieves 65%+ accuracy in the training set, validation set, and test set. The accuracy of the trained support vector machine even reaches 84.85% in the test set for each game or each set. Therefore, based on the differences in "momentum" shifts in past matches, we suggest that when dealing with a match against a different opponent, it is important to aggressively amplify one’s own winning streak, awarding highlights, and at the same time adjusting for the fact that one’s opponent is being "swept" by the other.

Figure 5. Confusion matrix during feed forward neural network training

4. Conclusions and outlooks

In conclusion, “momentum” does play a role in the match and swings in play and runs of success by one player are not random. Besides, by building a suitable model, it is possible to predict the occurrence of swing in a particular game. In addition, this paper can generalize the above models. Here are some examples.

First, the paper can generalize the models to different tennis court surfaces [10]. Tennis court surfaces have a great impact on the game of tennis, and different types of surfaces (e.g. hard courts, red clay, grass courts, indoor courts) will have different effects on the style of play as well as the strategy of the game. Second, it can generalize the models to table tennis. Table tennis and tennis are essentially two different ball sports, with significant differencesin sports equipment, court size, technical requirements, rules and styles, and this difference willhave an impact on the accuracy of the model. Furthermore, this paper reviews the models and finds some strengths and weaknesses of them as follows. As for the advantages of the model, the model fully considers the tennis process of a variety of game data and extracts the corresponding features for machine learning or deep learning for training, and finally obtains high prediction accuracy, which can provide guidance for the coachto predict whether the situation will be reversed. As for the weaknesses, the model requires more input parameters.

References

[1] Dietl H, Nesseler C. Momentum in tennis: Controlling the match [J]. UZH Business Working Paper Series, 2017 (365).

[2] Richardson P A, Adler W, Hankes D. Game, set, match: Psychological momentum in tennis [J]. The Sport Psychologist, 1988, 2 (1): 69 - 76.

[3] Yu H, Hutson A D. Inferential procedure based on the weighted Pearson correlation coefficient test statistic[J]. Journal of Applied Statistics, 2024, 51 (3): 481 - 496.

[4] Ali Abd Al-Hameed K. Spearman's correlation coefficient in statistical analysis [J]. International Journal of Nonlinear Analysis and Applications, 2022, 13 (1): 3249 - 3255.

[5] SHAQIRI M, ILJAZI T, KAMBERI L, et al. DIFFERENCES BETWEEN THE CORRELATION COEFFICIENTS PEARSON, KENDALL AND SPEARMAN [J]. Journal of Natural Sciences and Mathematics of UT, 2023, 8 (15-16): 392 - 397.

[6] Badillo S, Banfai B, Birzele F, et al. An introduction to machine learning [J]. Clinical pharmacology & therapeutics, 2020, 107 (4): 871 - 885.

[7] Mathew A, Amudha P, Sivakumari S. Deep learning techniques: an overview [J]. Advanced Machine Learning Technologies and Applications: Proceedings of AMLTA 2020, 2021: 599 - 608.

[8] Cervantes J, Garcia-Lamont F, Rodríguez-Mazahua L, et al. A comprehensive survey on support vector machine classification: Applications, challenges and trends [J]. Neurocomputing, 2020, 408: 189 - 215.

[9] Yu H, Hutson A D. Inferential procedure based on the weighted Pearson correlation coefficient test statistic [J]. Journal of Applied Statistics, 2024, 51 (3): 481 - 496.

[10] DOĞAN İ, REVAN S, ARIKAN Ş. Analysis of tennis competitions on different court surfaces [J]. Turkish Journal of Sport and Exercise, 2021, 23 (1): 60 - 66.