IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received 18 July 2024, accepted 12 August 2024, date of publication 15 August 2024, date of current version 26 August 2024.

Digital Object Identifier 10.1109/ACCESS.2024.3444031

RESEARCH ARTICLE

Rhythms of Victory: Predicting Professional Tennis Matches Using Machine Learning

YILIN LEI[1], **(Student Member, IEEE), AO LIN**[2], **AND JIANUO CAO JR.**[1]
[1]School of Electrical Science and Engineering, Nanjing University, Nanjing, Jiangsu 210023, China
[2]School of Computer Science, Nanjing University, Nanjing, Jiangsu 210023, China
Corresponding author: Yilin Lei (221180156@smail.nju.edu.cn)

ABSTRACT Forecasting the winning matches of professional tennis players has a wide range of practical applications. We introduced an innovative approach to quantify and combine strategic and psychological momentum using the entropy weight method and analytic hierarchy process, and tested its effectiveness. Utilizing data from the Wimbledon Championship 2023, we constructed a support vector machine model to predict the turning point and winner of each point, and optimized it using particle swarm optimization. Our model achieved a significant level of accuracy (96.09% for turning point and 83.52% for predicting the winner) and performed well in different courts and players. Furthermore, we compared its performance with commonly utilized predictive models, including ARIMA, LSTM and BP networks, and found that our model exhibited higher accuracy than other existing models in predicting the point winner. Our study provides a reference for the role of momentum in dynamic matches, and our model can be used to calculate the odds of tennis matches and provide guidance to coaches.

INDEX TERMS Forecasting, support vector machine, machine learning, particle swarm optimization, tennis.

I. INTRODUCTION

In our lives, the outcome of some competitions is often closely related to every step, such as sports contests, elections, and wars [1]. These tournaments are typically referred to as dynamic contests [2]. Researchers found that in a dynamic contest, differences in relative position lead to differences in incentives [3], which can be described as momentum: ''an added or gained psychological power that changes a person's view of him/herself or of others, or others' views of him/her and themselves'' [4]. Momentum, is always discussed by sports fans, coaches, and traders. For example, the momentum strategy (bet on past returns predicting the cross-section of future's), has helped investment banks make great profits [5].

The momentum has also been widely studied in sports. This can be described as a psychological and/or physiological boost from which players can benefit [6]. For example, 76% of basketball players in the NBA believe that their performance is crucially determined by momentum [7], which is the so-called ''hot hand''. A series of studies have shown that ''hot hand'' does exist in basketball [4], which is related to the previous success [8]. Another study shows that ''time out'' is an effective way to disrupt the opposite team by influencing momentum, which reduced its performance of success by 56% compared to the pre-timeout level [9]. Momentum also works in other sports. Another study found that previous successful hit improved the performance of the next hit [10]. Volleyball coaches also believe that momentum can influence individual performance [1]. Researchers also try to predict the performance of football teams, as they do in the stock market [11].

In tennis, momentum also plays an important role in game flow. Through empirical analysis, researchers have found that momentum contributes to the outcomes of the best-of-three tennis contests [12] as well. Nevertheless, studies on the quantitative analysis of momentum are limited and often concentrate on a few factors, such as set score [13]. Researchers have also identified two different types: strategic and psychological momentum [14], whose rationales are different [15]. Strategic momentum arises from the asymmetric future expected prizes created by the different

The associate editor coordinating the review of this manuscript and approving it for publication was Mehul S. Raval.

relative positions of competing agents in a dynamic contest [2], [16]. By contrast, changes in psychological momentum are due to changes in the perception of agents [8]. Past studies have also often made no distinction between psychological and strategic momentum or considered only one. However, in dynamic contests they potentially coexist [16].

Consequently, we set three factors in combination with the evaluation methods commonly used in tennis: technique, mistake (Psychological Momentum) and scoring factor (Strategic Momentum). To investigate momentum and construct our model, we employ data from contests in the Wimbledon Championships 2023 (man single) (N = 7,285) within 12 dimensions. We quantified the momentum indicating the performance of players with multiple factors by analytic hierarchy process(AHP)-entropy weight method(EWM) and visualized their performance. To avoid potential bias that probably appears in the EWM [17], we utilized the AHP to optimize the result. We also used Partial Correlation index and Variance Inflation Factor (VIF) to test the robustness of the EWM results.

Additionally, we focused on tennis because it involves only two players, which eliminates the influence of other athletes. However, differences in skill levels could interfere with the experiment. As a result, the side with the higher level is more likely to win, regardless of momentum [18]. Reference [3] used a quasi-experimental situation to address this. In their experiment, tennis players were equally matched at the end of close tie-breaks in the first set and the effect of winning the tie-break on winning the second set was measured. This method does not match the reality of the game, because we are committed to building a model that describes the full game, and this situation cannot be sustained throughout the tournament. Thus, we used score from Association of Tennis Professionals (ATP) to estimate their level in advance, and make a correction for the momentum, which is also considered as Bayesian prior [19].

Based on this, we established a model to predict swings during a match. Researchers have used Markov chains to predict the flow of football games [11]. However, tennis contests do not satisfy the Markov property completely. Players can remember their performance in the last several shots, which can affect the future in some way. In addition, our data had 12 dimensions. A Markov model with these dimensions may be too complex to satisfy real-time prediction requirements Thus, we decided to use SVM. Some researchers also chose SVM to predict [20], but they did not perform sufficient hyper-tuning because they used random hyperparameter tuning, which resulted in performance degradation. Thus, we used PSO to find the hyperparameter, which has a fast convergence speed while ensuring accuracy [21]. Finally, we tested the generalization performance of the model on other Australian Open Tennis Championships (man single) (N=2,236) and Wimbledon Championships (woman single) (N=1,557). Our model also performed well ($\geq$90% accuracy). We also applied our model to predict the point winner and achieved satisfactory results($\geq$ 80% accuracy).

In brief,we present a new model to transform high-dimensional input data into one dimension (i.e. momentum). based on the AHP and EWM, which is a combination of strategic and psychological momentum. We then built a PSO-SVM model that can predict momentum changes in the flow of games and the winner of every points, which could make profits for betting companies. In addition, this will also provide a reference for the coaches to make decisions during the game. Ultimately, the efficacy of our model was confirmed through cross-validation procedures and by juxtaposing its performance against that of alternative models.

Our study developed a rapid and efficient predictive model utilizing momentum, a novel approach not extensively explored in existing literature. Furthermore, through comparative analyses, we have identified momentum's efficacy in enhancing specific prediction models, thereby offering valuable insights into the potential benefits of incorporating momentum in sports-related forecasting.

The remainder of this paper is organized as follows: we reviewed previous studies. Then, we introduced our process of quantifying momentum (step1,step2)and test its robustness (step3). Subsequently, we established a model to predict the game flow and the winner based on our PSO-SVM model, and tested it on different competitions (step 4). Figure1 may describe these intuitively and precisely.

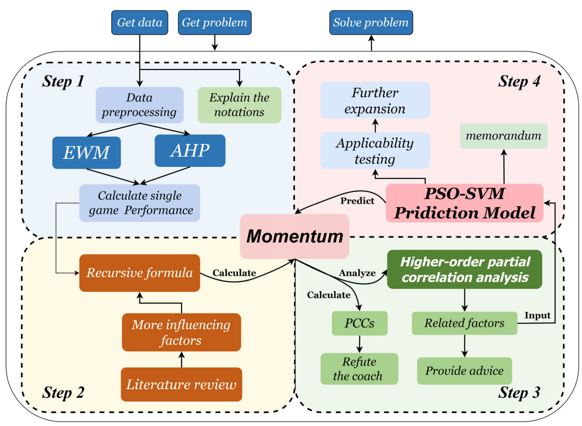

FIGURE 1. Flow chart of forcasting.

II. LITERATURE REVIEW

A. PRELIMINARY STUDY OF MOMENTUM

For a long time, momentum was a vague concept in minds, such as "success breeds success." [4] provides a clear definition: 'an added or gained psychological power that changes a person's view of him/herself or of others, or others.' Subsequently, researchers proposed that two different kinds of momentum might exist: strategic momentum and psychological momentum (PM) [14]. By estimating data from baseball, basketball, and hockey, researchers controlled for teams' relative strength by considering each team's regular season winning percentages and experience in championship series. They find that the effects of strategic considerations are "...in magnitude compared to home-field advantage" [22]. In contrast, another group found that it contributed to the contests [12].

B. STRATEGIC MOMENTUM

Strategic momentum arises from asymmetric future expected prizes created by the different relative positions of competing agents in a dynamic contest [2], [16]. The perfect equilibria of races lead to a situation in which the leader should exert greater effort than the follower [23], which is also proven by [24] in best-of-n contests. For example,in chess, gaining an advantage by capturing the opponent's pieces results in an imbalance in the anticipated outcome of victory for each player. However, other researchers proposed the 'hare-tortoise' heuristic model, where the 'trailing contestant need more effort to catch up, while the leading contestant may slack off [25]. Researchers have also investigated the existence of a momentum as anticipated by game theory models through empirical analysis. Reference [22] estimated a structural model of strategic behavior in best-of-7 contests with data from the US championship series, finding that momentum in these contests is absent, indicating that teams strive to perform at their highest level in every game in which they participate. Subsequently, experiments were conducted in a controlled environment. Reference [26] supported a negative momentum effect ('hare-tortoise' heuristic model). However, [27] found the evidence of traditional strategic momentum. Reference [28] indicate that it plays a positive role in assisting corporate decision-making, meaning that how it works is often controversial.

C. PSYCHOLOGICAL MOMENTUM

In addition to the strategic factors that can create a bridge between prior and subsequent performance, the literature focuses mainly on psychological momentum (PM). Similarly, using chess as a case in point, the side at a disadvantage frequently experiences heightened pressure, resulting in an increased likelihood of errors. Reference [29] proposed that PM is "a state of dynamic intensity marked by an elevated or depressed rate of motion, grace, and success". Then, [30] conceptualized PM in a multi-dimensional model of momentum, which is called "momentum chain". Subsequently, [31] evaluated an antecedents-consequences PM model by employing hypothetical competitive tennis scenarios, which illustrates momentum or non-momentum situations themselves, their influence on performance inferences, subjects' perceptions of momentum and the influence of momentum or non-momentum situations on performance inferences. The results indicated that score configuration exerted an obvious influence on perceptions of momentum. By conducting a descriptive analysis, [32] defined PM as 'success breeds success'. Reference [19] suggest that" the timeline of past performance influences future performance. In other words, the PM suggests that the winner of the prior round has "momentum" going into the next round. While the descriptive analysis of PM (versus strategic momentum) is informative, it does not provide strong evidence of when either form of momentum exists or is dominant. Reference [27] showed that PM had no effect at. However, most researchers believe this is effective.

D. MOMENTUM IN TENNIS

Researchers have also investigated the role of momentum in tennis. Reference [33] suggest that the success-breeds-success model (use PM)explains the tennis data extremely well. Reference [14], separated psychological momentum from strategic momentum, which appears to be the primary factor triggering a performance increase after a converted break point. Reference [12] provided a theoretical and empirical analysis of both strategic and psychological momentums. Their results indicate that the existence of psychological reversal: in every set following a non-tie net score, the winner of the previous set underperforms compared with their predictions. The concepts introduced by the Stats Perform AI group serve as an example [34]. The breaking point has played a significant role in the overall context, despite the insignificant score it brings in itself. However, [3] have different result: they find that momentum effect does not work for women, whereas it exists for male players. Reference [35] also show that momentum will be inconspicuous if the skill of players is controlled [35]. Although there is growing evidence of the existence of momentum, it is still debatable, and further studies are required.

E. PREDICTING MODEL

Predictive models based on momentum are widely used in the financial field, such as stock and futures trading [5], [24]. Such studies have also been extensive in sports. For example, researchers use the Markov chain to predict the result of football games [15]. In tennis, [36] used Markov chains in the 2003 Australian Open. Reference [37] showed that SVM performed better than other models in predicting the winner with 69.6% accuracy. While [20] used a variety of machine learning methods, such as SVM and random forest, to predict the winner,with the highest accuracy of 78% achieved, by random forest.

III. MATERIALS AND METHODS

Momentum is a psychological concept that does not have a specific value. However, if we want to apply this to forecasting, we must quantify it. We then established an SVM prediction model based on it, which was optimized using PSO.

A. DATA PROCESSING

First, we supplemented the missing value by comparing several reliable sources. Subsequently, in cases where certain values remain absent, due to their irregular distribution and lack of discernible pattern within the dataset, they are deemed to be missing at random. Consequently, Multiple Imputation by Chained Equations (MICE) is employed as a method for imputing these missing values. Thus, we obtained a complete data set.

However, at a single point, much of the data is discrete, or Boolean, which is unfavorable for data processing. Using such data can lead to irrational answers or unusual swings. Moreover, some factors, such as severe diversity, can not be measured at a single point. We adopted the sliding window method to describe these variables effectively. At the same time, it can also reduce the number of zero values,for as long as there is a non-zero value in the window, all the data in it is non-zero after processing. Thus, we set the window length to half of a game (three points), which was used in table tennis [38]. However, when evaluating the diversity of serve, we need a larger window, because there are four types of serve widths and two serve depths. Finally we set this to the length of a game(six points).

Additionally, the previous points commonly contribute less to the momentum. Therefore, instead of taking the average, we processed the data using the following formula1 to describe this situation,W_i^n represents the value after processing and N_i^i is the original value. Additionally, throughout this paper, the n subscript indicates that this value is at n.

$$W_i^n = \sum_{n-N}^{n} \frac{N_i^i}{(n-i+1)!} \tag{1}$$

For further calculation, we use z-score, S means standard standard deviation of variable "x",Z_x^n is the value of x at point n after z-score standardization.

$$Z_x^n = \frac{W_x^n - \frac{\sum_1^n W_x^n}{T}}{S_x} \tag{2}$$

Subsequently, to avoid overfitting, the dimensions of the data were reduced. Before we calculate their weight, using AHP, we divide the 12 variables into three factors: technique factor, mistake factor and scoring factor.The scoring factor confers uneven advantages to athletes, with those achieving high scores being inherently positioned closer to success. As a result, this phenomenon is categorized as strategic momentum. Technique and mistake factors have minimal influence on scores, but they serve as indicators of the player's present condition, thereby exerting a certain degree of influence on the game dynamics. As a result, these factors are categorized as components of psychological momentum.

Now, we can calculate their weights using EWM. It is based on information entropy proposed by Shannon [39]. To calculate the weights, for each dimension:

$$f_{ij=}\frac{r_{ij}}{\sum_{i=1}^{n} r_{ij}} \tag{3}$$

where m is the number of countries selected for calculation. The function of information entropy, e, is calculated

TABLE 1. Data structure of momentum.

Momentum Type	Factors	Variables	Notations
Psychological momentum	Technique	Running Distance	γ_7
		Serving Diversity	γ_2
		Net-pit-won	γ_5
		ACE	γ_6
		Winner Shot	γ_4
		Break-pit-won	γ_1
		Serve Speed	γ_3
	Mistake	Unforced Error	β_1
		Double Fault	β_2
Strategic Momentum	Scoring	Recent Scoring	α_1
		Point Winner	α_2
		Game Winner	α_3

as follows:

$$e_j = -ln(\frac{1}{n})\sum_{i=1}^{m} ln(f_{ij}) \tag{4}$$

Subsequently, we calculate the weight in each dimension, W_e, the weights of each indicator, are calculated as follows.

$$W_{E_j} = \frac{1-e_j}{m-\sum_{i=1} e_j} \tag{5}$$

Thus, we determine the weights of three factors. By applying the same method to other factors and calculating the average of the weights of the two players, we can determine their weights: (Table2,3,4):

TABLE 2. Technique factor.

Variable	Result
γ_1	0.259
γ_2	0.088
γ_3	0.087
γ_4	0.084
γ_5	0.091
γ_6	0.357
γ_7	0.034

TABLE 3. Scoring factor.

Variable	Result
α_1	0.106
α_2	0.344
α_3	0.550

TABLE 4. Mistake factor.

Variable	Result
β_1	0.402
β_2	0.598

After that we can derive the weight of factors in the same way: (table5). Now that we have the weights for each factor,

TABLE 5. EWM results in 3 factors.

Variable	Result
P_1	0.403
P_2	0.112
P_3	0.485

we can calculate the momentum. First, we built a basic model that adds all the factors by weight, F is the value of factors:

$$S_i^n = P_1F_{i1}^n + P_2F_{i2}^n + P_3F_{i3}^n$$
$$F_{ij}^n = \sum \omega_{jk}x^n \tag{6}$$

However, this model is unfair to the player who returns the serve, because he or she is less likely to win the point. Therefore, the scoring factor of the server must be limited. To balance out the advantages that is irrelevant to momentum brought about by serving, we have decided to add a correction to the scoring factor when serving, called serving coefficient (α). For example, through the official website of ATP, we found out about the first serve percent at that day of Djokovic and Carlos (out player2 and player1), 65.6% and 64.1% respectively. We can infer the value of α as:

$$\alpha = \begin{cases} \dfrac{1-p_1}{p_1} & serve \\ 1 & return \end{cases} \tag{7}$$

Similarly, we calculated the percentage of serve wins for the other players using the same method. Therefore, we can improve our formula:

$$S_1 = \alpha P_1F_{i1} + P_2F_{i2} + P_3F_{i3} \tag{8}$$

The situation is similar for player 2. To see it more clearly, we figure out the difference. The player with a higher performance score is more likely to behave better. To describe it more precisely, the standard deviation of the absolute value of the difference was calculated. If the difference is within one standard deviation, we considered players to be evenly matched. If the difference was within two standard deviation, we considered that there are obvious difference. For others, we think their performance difference is quite large.

However, other important factors must be considered. At the end of a game, undoubtedly, his momentum, if it exists, will decrease more rapidly, for the server will change [8]. To describe this situation, we introduced a correction in the last round of each game (ζ). In the formula, $\bar{S}$ is the average momentum when serving or receiving a ball. It works only when the server changes:

$$\zeta = \left\{ \bar{S}_{serve} - \bar{S}_{return} \right. \tag{9}$$

In addition, researchers have found that a player who wins a close first set tie break is, on average, more likely to win the next game [40]. We believe that this will give players momentum. Thus, we set a factor (TB) to stimulate this phenomenon, which we call "winner correction". Its values are shown in Table 8,10.

Finally, researchers also find the momentum at one point is related to the momentum before [6]. To describe this, we introduce recursion and the genetic index (Δ), which is also shown in table8,10. Thus, the final momentum formula is obtained:

$$M_1^n = TB\alpha P_1F_{11}^n + P_2F_{12}^n + P_3\,F_{13}^n - \zeta + \Delta M_1^{n-1} \tag{10}$$

To make the effect of momentum more intuitively, we calculated the differences in their performance score (figure2). Player 1 is Carlos and player 2 is Djokovic. We also mark every time player 2 (Novak Djokovic) wins the point under the X-axis. Additionally, we figure out the standard deviation and define those points whose difference is larger than half of it as "Advance(AD)". From that we can also observe some interesting things: the player who is in "advance" is more likely to win the next point (marked by red squares).

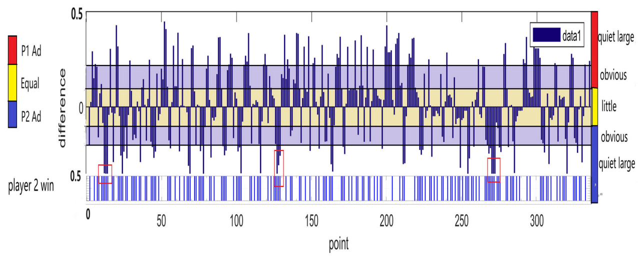

FIGURE 2. Momentum difference.

B. ROBUSTNESS TEST

To test rationality, we determined the correlation index to prove that it is meaningful.Similarly, we define a swing(SW_i) as a momentum change exceeding half of the previous data's standard deviation, and calculate the number of times fluctuations occur. If we use σ_i to represent the standard deviation, and p_i indicates the average win rate in the window i:

$$SW_i = \epsilon(P_i - 0.5\sigma_i) \tag{11}$$

In the equation, ϵ represents the step function. We then tested the correlation between them. The momentum follows a normal distribution and are obviously continuous. Therefore, we used the Pearson correlation coefficient.

$$R_{x,y} = \frac{cov(x,y)}{\sigma_x\sigma_y} \tag{12}$$

After calculation, we obtained the Pearson correlation coefficient between the difference in momentum and the winner, 0.6739, indicating that the correlation between them is.

Alternatively, we can also describe it by simple statistical means: We assume that the person in "advance" wins the point. Finally, our assumption works at 87% of all points, indicating that the player with higher momentum is more likely to win the point.

Additionally, determining the factors that seem to be the most related to swings can explain our weights. We used partial correlation, which describes the correlation between two variables excluding the influence of others. For the two-factor case, partial correlation (δ) is defined as follows.

$$y^{\perp} = y - \sum_{yz}\sum_{zz}^{-} 1z$$

$$x^{\perp} = x - \sum_{xz}\sum_{zz}^{-} 1z$$

$$r_{xy\cdot z} = \delta_{y^{\perp}x^{\perp}} = \frac{con(y^{\perp}, x^{\perp})}{\sqrt{var(x^{\perp}, y^{\perp})}} \tag{13}$$

The situation is similar for ternary and above:

$$r_{i,j,l_1,l_2\ldots l_g} = \frac{r_{i,j,l_1,l_2\ldots l_{g-1}} - r_{i,l_1,l_2\ldots l_g} r_{i,l_1,l_2\ldots l_{g-1}}}{\sqrt{(1 - r^2_{i,l_1,l_2\ldots l_{g-1}})(1 - r^2_{j,l_1,l_2\ldots l_{g-1}})}} \tag{14}$$

where r is the Partial correlation index. $r_{i,j,l_1,l_2\ldots l_g}$ is the higher-order radial correlation betweenx_i and x_j. We evaluated the partial correlation of 12 variables and test them by p-value, which indicates how likely it is to observe the data if there is no effect or relationship.

P-value can be calculated by permutation test. In our study, the objective function is the higher-order partial correlation index;

Permutation test used the Monte Carlo method to simulate whether the test data were normally distributed. Thus,these values were obtained. Generally, we consider a correlation index below 0.1 to be hardly irrelevant (ir), 0.1 to 0.3 to be weakly correlated (w), 0.3 to 0.5 to be moderately correlated (M), and above 0.5 to be strongly correlated (S).

TABLE 6. Partial correlation index of variables.

factor	index
Winning percentage (3 points)	0.178 (w)***
the winner of last point	0.063 (ir)***
whether win the game (Only when the last round of each game)	0.367 (M)***
unforced error	-0.107 (w (negative))**
double fault	-0.072 (ir)***
physical exertion	-0.072 (ir)***
serve diversity	0.127 (w)***
net-pt-won	0.194 (w)**
ACE	0.247 (w)**
winner shot	0.153 (w)***
break-pt-won	0.313 (M)***
serve speed	0.024 (ir)***
win points in a row (3 points here)	0.412 (M)***

In the table, the symbol "*" means that the p-value is less than 0.05, "**" indicates that it is less than 0.025, and "***" means that it is less than 0.01.

According to the table, variables with a greater index often have a greater weight. Simultaneously, we conducted variance inflation factor (VIF) test, indicating that our model has no multicollinearity problem, which provides evidence for the rationality of our division. Therefore, we believe that it is convincing to prove that there is a connection between momentum and swings, which means that our momentum is reasonable.

C. PREDICTION MODEL BUILDING

Thus, we have a much simpler and lower dimensional variable, the momentum, that we can use for forecasting. We designed an SVM optimized by PSO to predict what will happen at the next point, including the turning point of momentum, and the result of that point. Similarly, we divide the data into three different states:Player1 advantage, Player 2 advantage, and in little, as shown in figure2), and define the turning point as the point at which the category changes.

1) SUPPORT VECTOR MACHINE

Support vector machines are supervised learning models with associated learning algorithms that analyze data used for classification and regression analysis. In our study, variables are nonlinear, so we use nonlinear SVM with soft-margin. Firstly we should map the two-dimensional sample to a high-dimensional space, so that the points are linearly divisible in a high-dimensional space. Then,if we use x to represent the original sample point, and $\phi(x)$ represents x mapped to the feature and the new feature space to the new vector. The segmentation hyperplane can be expressed as:

$$f(x) = w\phi(x) + b \tag{15}$$

Then, The dual problem for nonlinear SVM can be expressed as:

$$min_{\lambda}[\frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N}\lambda_i\lambda_j y_i y_j(\phi(x_i)\cdot\phi(x_j)) - \sum_{j=1}^{N}\lambda_j]$$

$$s.t. \quad \sum_{i=1}^{N}\lambda_i y_i > 0 \quad \lambda_i \geq 0 \quad C - \lambda_i - \mu_i = 0 \tag{16}$$

During the process, we used the Gauss core function,which is considered efficient in the vast majority of tasks [41]:

$$k(x_i, x_j) = e^{-\frac{\gamma||x_i - x_j||}{2\delta^2}} \tag{17}$$

Then, we solve optimization problem. We construct a Lagrange function:

$$\begin{aligned} &min_{\omega,b,\epsilon}\ max_{\lambda,\mu}(\omega, b, \epsilon, \lambda, \mu) = \frac{1}{2}||\omega||^2 \\ &+ C\sum_{i=1}^{m}\epsilon_i + \lambda_i\sum_{i=1}^{n} 1 - \epsilon_i - y_i(\omega^T x + b) - \sum_{i=1}^{n}\mu_i\epsilon_i \end{aligned} \tag{18}$$

Then, due to its partial derivative being zero, we obtained the following condition.

$$\begin{gathered} \omega = \sum_{i=1}^{m}\lambda_i x_i y_i \\ \sum_{i=1}^{m}\lambda_i y_i = 0 \\ C = \lambda_i + \mu_i \end{gathered} \tag{19}$$

From these, we can determine the Lagrange multiplier (λ^*) Through the process of SVM operation, we can easily find that the two most important parameters that can be changed: C and γ. To determine suitable values, we used PSO. C in SVM helps control the trade-off between the training error and the margin because it can determine the penalty for misclassified data points during the training process. Specifically, a smaller value allows for a larger margin, potentially leading to more misclassifications on the training data. However, if its value is too small, it will fall into over-fitting easily. γ is another important value used to adjust the degree of influence of core functions in the SVM model and control the nonlinear fitting ability of the model.

2) PARTICLE SWARM OPTIMIZATION

PSO is a bio-inspired algorithm. The effectiveness of this algorithm has been widely verified [42]. The convergence rate of this algorithm is also faster [43] than that of other common evolutionary algorithms,such as genetic algorithm or simulated annealing algorithm (SA). A classical PSO operates as follows (figure3):

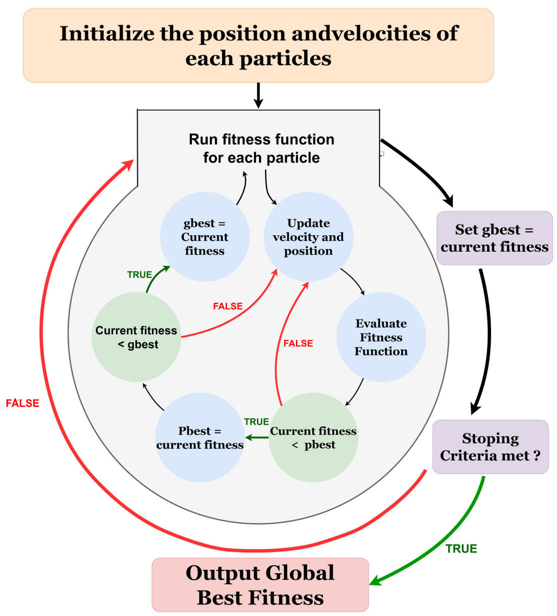

FIGURE 3. PSO procedure.

First, we initialize the swarm and set the historical optimal p_{Best} of an individual as the current location, and the optimal individual in the group as the current g_{Best}. Subsequently, we calculated the fitness values of each particle. If it is better, we update g_Best. and p_{Best} in this way:

$$\begin{aligned} v_{id}^{k+1} &= kv_{id}^{k} + c_1 r_1 (p_{id,pbest}^{k} - x + id^{k}) \\ &\quad c_2 r_2 (p_{d,gbest}^{k} - x + id^{k}) \\ x_{id}^{k+1} &= x_{id}^{k} + v_{id}^{k+1} \end{aligned} \tag{20}$$

This is performed until the number of iterations exceed or the goal is reached.

We then need to relate the parameters in the PSO to the SVM. Clearly, we can use C and γ in our SVM model. In addition, an appropriate function should be built to evaluate the particles. And in our model, we used the function of the accuracy of our SVM prediction:

$$fitness(j) = 1 - svm_{train} \tag{21}$$

where svm_{train} is the SVM accuracy. The lower the fitness index, the better the performance of our model. We also set PSO parameters.After optimizing parameters using the Monte Carlo method, we finally found the best parameter combination. (Table 7)

TABLE 7. PSO hyperparameter settings.

parameter	value	meaning
c_1	1.5	Individual cognitive abilities of particles
c_2	1.7	Social cognition ability of particles
$maxgen$	100	Maximum iterations
$sizepop$	20	Particle population
k	0.6	inertia
WV	1	Elastic coefficient of velocity
WP	1	Elastic coefficient of position
$pop_c max$	10	The maximum change of SVM parameter c
$pop_c min$	0.1	The minimum change of SVM parameter c
$pop_g max$	10	The maximum change of SVM parameter γ
$popgmin$	0.1	The maximum minimum of SVM parameter γ
$pop_T Bmax$	1	The maximum change of TB γ
$pop_T Bmin$	0.01	The maximum minimum of TB γ
$pop_\Delta max$	1	The maximum change of Δ γ
$popg_\Delta min$	0.01	The maximum minimum of Δ γ

In this way, we can determine the best parameters of our SVM.However, PSO faces the problem that it can easily fall into local optimization. To solve this problem, we use the Monte Carlo method.

Finally, the value calculated by PSO. The results are shown in the table8:

TABLE 8. PSO result.

Parameter	Value	Meaning
γ	16.64	data 2
C	0.10	Core function influence
TB	1.25	Winner correction
Δ	0.72	Genetic index

The fitness index, which illustrates the value of fitness function, is as follows. After iterations, the value of the function stabilizes at 0.072. It converges in round 18, quite quickly. In sports competitions, the convergence speed of the model will bring unique advantages, as it allows for adjustments to the model during the competition.

Finally, we apply the trained model to the test set.We used 6-fold cross-validation. Our model performs well in both sets, reaching an average accuracy of 92.78% in training set and 96.11% in test set.

TABLE 9. Model performance on training and test set.

set	accuracy	recall rates	precision	F_1 score
training set	0.9178	0.8540	0.8353	0.8346
test set	0.9611	0.9172	0.7333	0.8148

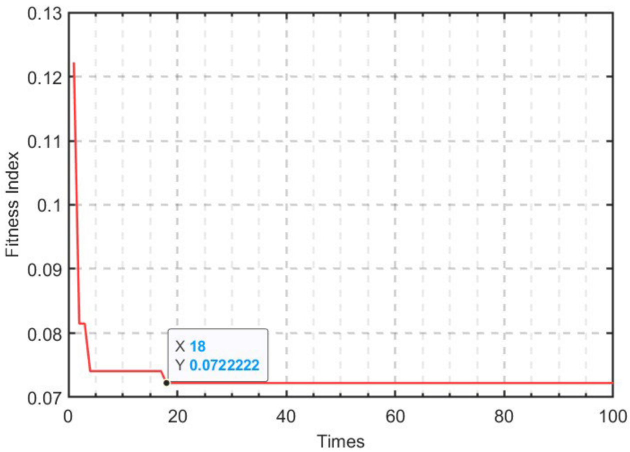

FIGURE 4. Fitness.

However, because there are fewer turning points in the data set, even a high accuracy rate may be invalid. According to statistics, the proportion of turning points in data set is 35.31%. Using Bayes Formula (22), we can calculate the probability that a point predicted to be a turning point will actually be:

$$P(B|A) = \frac{P(B)P(A|B)}{P(B)P(A|B) + P(\bar{B})P(A|\bar{B})} \tag{22}$$

The probabilities were 85.69% (training set) and 86.33% (test set). Clearly, our predictions are reasonable.

We then tested the generalization performance of the model, mainly on different courts and players (table11). Our model performed well in different competitions.

D. PREDICTING THE WINNER

However, people often tend to care more about the outcome of a game than the process. Therefore, we built a similar model to predict wins and losses. We use the SVM-PSO model with similar loss function (equation21) and the same initial parameters (Table7). Finally, we obtained the following parameters (Table10). Obviously, there is not much difference from the results of turning points. Simultaneously, to test the performance of our model, we compared it with other commonly used models. Among them, the order of autoregressive terms (p), the order of differencing (d), and the order of moving average terms (q) in the ARIMA model are 2, 1, 2 respectively. The parameters of LSTM consist of a 14-dimensional input size, a 30-dimensional hidden size, and two hidden LSTM layers. The BP network's batch size has been set to 500, and the learning rate is 0.1. These parameters were fine-tuned through the utilization of PSO.

The model exhibited superior performance in the study, demonstrating higher accuracy compared to alternative models(Table12). Furthermore, it is noteworthy that the training cost of SVM (0.63s) is comparatively lower among the effective models(10.85s for LSTM and 3.73s for BP network). These findings suggest that the momentum-based SVM model is highly effective in forecasting the results of tennis matches.

TABLE 10. PSO result for predicting the winner.

Parameter	Value	Meaning
γ	15.93	Penalty coefficient
C	0.12	Core function influence
TB	1.24	Winner correction
Δ	0.69	Genetic index

IV. RESULT

We used data from Wimbledon Championships (man single) (N=7,285), Wimbledon Championships (woman single) (N =1,557), and Australian Open Tennis Championship (man single) (N=2,236).

Through EWM and AHP, we determined the weight of each indicator and calculated a reasonable momentum value, the validity of which was proven by correlation analysis. Based on this, we built a SVM to predict the game flow. After optimization, the performance in predicting who had the upper hand was performed as follows (table11):

TABLE 11. Predict the player favored by the match.

Match	Accuracy	Recall Rates	Precision	F_1 Score
Wimbledon Championships (man single)	92.78%	85.40%	83.53%	83.46%
Wimbledon Championships (woman single)	96.88%	78.04%	88.89%	83.12%
Australian Open Tennis Championship (man single)	96.09%	91.89%	79.07%	80.85%

We also used our model to predict the winner, which is more practical. We also use different models that are employed in prediction, including LSTM (Long Short Term Memory), BP (Back Propagation) network and ARIMA (Autoregressive Integrated Moving Average) model. Among all method we tested, ARIMA contributes little to prediction (but the time series is stable), and our SVM model performs the best, though its performance declines. The specific data are presented in table12.

TABLE 12. Predict the winner(Accuracy).

Match	ARIMA	LSTM	BP Network	SVM
Wimbledon Championships (man single)	57.1%	77.43%	76.50%	**83.52%**
Wimbledon Championships (woman single)	53.3%	78.82%	72.12%	**84.28%**
Australian Open Tennis Championship	51.2%	75.61%	74.27%	**81.65%**

V. DISCUSSION

We quantify momentum using EWM and build a predictive model based on it, which has good performance. Our model suggests that ACE, break-pit, and scoring contribute most to momentum, which is consistent with previous research [6], [38], [44]. These factors also tend to receive more attention in competitions. However, it is unclear whether their importance in match brings attention or whether their attention makes them psychologically powerful.

Additionally, our SVM model outperformed commonly used prediction models. This achieved by the advantages of SVM in handling small samples and multi-dimensional data.

Making predictions using momentum instead of raw data also contributed to the performance of our model, because momentum is more stable compared to raw data, such as scoring and ACE. Furthermore, the accuracy of our BP neural network saw an enhancement of approximately ten percentage points following the implementation of momentum for forecasting (previously at approximately 65%), suggesting that the utilization of momentum for prediction purposes proves to be notably efficacious.Simultaneously, the reduced dimensionality of momentum results in decreased computational expenses. Furthermore, the parameter tuning using PSO has a considerable contribution to the performance improvement of our algorithm. Compared with manual tuning, the accuracy of our algorithm is improved by 10.52% on average.

VI. CONCLUSION

This study proposes a new method to define momentum and develop a novel approach for predicting professional tennis. Our model is built on SVM and momentum [14], indicating that both strategic and mental momentum influence matches and can be used to make predictions. The application of momentum improves the accuracy of tennis prediction model, making it more advantageous for practical applications.

One of the major contributions of our study is that we proposed a new method to quantify momentum and use it to make predictions. Compared with the traditional method (directly based on player performance) [20], [36], [37], our accuracy has been greatly improved. Additionally, the duration required for processing in our model is notably reduced in comparison to that of the models being compared. At the same time, our momentum takes into account the differences between different athletes (ATP score and success rate of serving), which gives our model good generalization performance.

However, our model has some shortcomings. For example, at the beginning of matches, our model's performance declined, which may be due to changes in the state of players in different matches. Pre-match calibration method could help solve this problem [24]. Additionally, the average time required to train our model was around 0.7 seconds. According to ATP, the average time per round is approximately 6.7 seconds, which suggests that it is possible to provide additional training based on the most up-to-date data available. Therefore, we need to improve the efficiency of our model to leave more time for the other steps.

Despite the good performance of our model, much work remains to be conducted. Next, we will focus on improving the speed of our model and evaluating the possibility of applying this system to other sports, including sports with two players such as table tennis or badminton, and those with more players, such as volleyball. At the same time, we will also strive to implement a practical odds mechanism based and betting strategy on this model, which will be more effective than the traditional method based on the ELO rating algorithm. Simultaneously, we aim to integrate it with image recognition technology in order to develop a completely automated forecasting system. The generality of our method needs to be improved. It is expected that the method of using momentum for predicting game flow can be used in other models such as LSTM or BP network.

ACKNOWLEDGMENT

The authors would like to thank the Mathematical Contest in Modeling/The Interdisciplinary Contest in Modeling (MCM/ICM) Conference of Consortium for Mathematics and its Applications (COMAP) for part of the original data and article inspiration, also would like to thank RYAN.A from Kaggle for their open access data, also would like to thank Prof. Yvxin Wang and his team for the criticism and revision of this article, and also would like to thank the reviewers and the IEEE Access editors for their valuable suggestions for this article.

REFERENCES

[1] K. A. Konrad and D. Kovenock, "Multi-battle contests," *Games Econ. Behav.*, vol. 66, no. 1, pp. 256–274, May 2009.

[2] D. Cohen-Zada, A. Krumer, and Z. Shtudiner, "Psychological momentum and gender," *J. Econ. Behav. Org.*, vol. 135, pp. 66–81, Mar. 2017.

[3] R. Gauriot and L. Page, "Does success breed success? A quasi-experiment on strategic momentum in dynamic contests," *Econ. J.*, vol. 129, no. 624, pp. 3107–3136, Nov. 2019. [Online]. Available: https://api.semanticscholar.org/CorpusID:108099461

[4] S. E. Iso-Ahola, "Intrapersonal and interpersonal factors in athletic performance," *Scandin. J. Med. Sci. Sports*, vol. 5, no. 4, pp. 191–199, Aug. 1995.

[5] K. Daniel and T. J. Moskowitz, "Momentum crashes," *J. Financial Econ.*, vol. 122, no. 2, pp. 221–247, Nov. 2016.

[6] C. A. Depken, J. M. Gandar, and D. A. Shapiro, "Set-level strategic and psychological momentum in best-of-three-set professional tennis matches," *J. Sports Econ.*, vol. 23, no. 5, pp. 598–623, Jun. 2022, doi: 10.1177/15270025221085715.

[7] T. Gilovich, R. Vallone, and A. Tversky, "The hot hand in basketball: On the misperception of random sequences," *Cogn. Psychol.*, vol. 17, no. 3, pp. 295–314, Jul. 1985, doi: 10.1016/0010-0285(85)90010-6.

[8] S. E. Iso-Ahola and K. Mobily, "'Psychological momentum': A phenomenon and an empirical (unobtrusive) validation of its influence in a competitive sport tournament," *Psychol. Rep.*, vol. 46, no. 2, pp. 391–401, Apr. 1980.

[9] D. Kahneman and A. Tversky, "Prospect theory: An analysis of decision under risk," *Econometrica*, vol. 47, no. 2, p. 263, Mar. 1979.

[10] S. E. Iso-Ahola and C. O. Dotson, "Psychological momentum: Why success breeds success," *Rev. Gen. Psychol.*, vol. 18, no. 1, pp. 19–33, Mar. 2014.

[11] C. López-Serrano, M. P. M. Arroyo, D. Mon-López, and J. J. M. Martín, "In the opinion of elite volleyball coaches, how do contextual variables influence individual volleyball performance in competitions?" *Sports*, vol. 10, no. 10, p. 156, Oct. 2022.

[12] L. Lin, M. Schatz, and D. Sornette, "A simple mechanism for financial bubbles: Time-varying momentum horizon," *Quant. Finance*, vol. 19, pp. 1–23, Nov. 2018, doi: 10.1080/14697688.2018.1540881.

[13] F. C. Mace, J. S. Lalli, M. C. Shea, and J. A. Nevin, "Behavioral momentum in college basketball," *J. Appl. Behav. Anal.*, vol. 25, no. 3, pp. 657–663, Sep. 1992.

[14] P. Meier, R. Flepp, M. Ruedisser, and E. Franck, "Separating psychological momentum from strategic momentum: Evidence from men's professional tennis," *J. Econ. Psychol.*, vol. 78, Jun. 2020, Art. no. 102269, doi: 10.1016/j.joep.2020.102269.

[15] M. Ötting, R. Langrock, and A. Maruotti, "A copula-based multivariate hidden Markov model for modelling momentum in football," *AStA Adv. Stat. Anal.*, vol. 107, nos. 1–2, pp. 9–27, Mar. 2023.

[16] D. A. Malueg and A. J. Yates, "Testing contest theory: Evidence from best-of-three tennis matches," *Rev. Econ. Statist.*, vol. 92, no. 3, pp. 689–692, Aug. 2010. [Online]. Available: https://api.semanticscholar.org/CorpusID:57567027
[17] M. Jain, V. Saihjpal, N. Singh, and S. B. Singh, "An overview of variants and advancements of PSO algorithm," *Appl. Sci.*, vol. 12, no. 17, p. 8392, Aug. 2022. [Online]. Available: https://api.semanticscholar.org/CorpusID:251825301
[18] L. Page and J. Coates, "Winner and loser effects in human competitions. Evidence from equally matched tennis players," *Evol. Human Behav.*, vol. 38, no. 4, pp. 530–535, Jul. 2017, doi: 10.1016/j.evolhumbehav.2017.02.003.
[19] R. Gauriot and L. Page, "Psychological momentum in contests: The case of scoring before half-time in football," *J. Econ. Behav. Org.*, vol. 149, pp. 137–168, May 2018.
[20] A. Somboonphokkaphan, S. Phimoltares, and C. Lursinsap, "Tennis winner prediction based on time-series history with neural modeling," in *Proc. Int. MultiConf. Eng. Comput. Scientists*, Hong Kong, 2009, pp. 1–6. [Online]. Available: https://api.semanticscholar.org/CorpusID:15902052
[21] Y. Zhu, D. Tian, and F. Yan, "Effectiveness of entropy weight method in decision-making," *Math. Problems Eng.*, vol. 2020, pp. 1–5, Mar. 2020.
[22] C. Ferrall and A. A. Smith, "A sequential game model of sports championship series: Theory and estimation," *Rev. Econ. Statist.*, vol. 81, no. 4, pp. 704–719, Nov. 1999. [Online]. Available: https://api.semanticscholar.org/CorpusID:4472274
[23] C. Harris and J. Vickers, "Perfect equilibrium in a model of a race," *Rev. Econ. Stud.*, vol. 52, no. 2, p. 193, Apr. 1985. [Online]. Available: https://api.semanticscholar.org/CorpusID:154554420
[24] R. Novy-Marx, "Is momentum really momentum?" *J. Financial Econ.*, vol. 103, no. 3, pp. 429–453, Mar. 2012.
[25] K.-K. Tong and K. Leung, "Tournament as a motivational strategy: Extension to dynamic situations with uncertain duration," *J. Econ. Psychol.*, vol. 23, no. 3, pp. 399–420, Jun. 2002. [Online]. Available: https://api.semanticscholar.org/CorpusID:143080458
[26] Q. Fu, C. Ke, and F. Tan, "'Success breeds success' or 'pride goes before a fall'?: Teams and individuals in multi-contest tournaments," *Games Econ. Behav.*, vol. 94, pp. 57–79, Nov. 2015. [Online]. Available: https://api.semanticscholar.org/CorpusID:12641292
[27] S. D. Mago, R. M. Sheremeta, and A. Yates, "Best-of-three contest experiments: Strategic versus psychological momentum," *Int. J. Ind. Org.*, vol. 31, no. 3, pp. 287–296, May 2013, doi: 10.1016/j.ijindorg.2012.11.006.
[28] T. Keil, Y. Deutsch, T. Laamanen, and M. Maula, "Temporal dynamics in acquisition behavior: The effects of activity load on strategic momentum," *J. Manage. Stud.*, vol. 60, no. 1, pp. 38–81, Jan. 2023. [Online]. Available: https://api.semanticscholar.org/CorpusID:249997643
[29] T. C. Hood and P. Adler, "Momentum: A theory of social action," *Social Forces*, vol. 62, no. 3, p. 846, Mar. 1984. [Online]. Available: https://api.semanticscholar.org/CorpusID:144731249
[30] J. Taylor and A. Demick, "A multidimensional model of momentum in sports," *J. Appl. Sport Psychol.*, vol. 6, no. 1, pp. 51–70, Mar. 1994.
[31] R. J. Vallerand, P. G. Colavecchio, and L. G. Pelletier, "Psychological momentum and performance inferences: A preliminary test of the antecedents-consequences psychological momentum model," *J. Sport Exercise Psychol.*, vol. 10, no. 1, pp. 92–108, Mar. 1988. [Online]. Available: https://api.semanticscholar.org/CorpusID:145574525
[32] S. D. Mago, R. M. Sheremeta, and A. Yates, "Best-of-three contest experiments: Strategic versus psychological momentum," *Int. J. Ind. Org.*, vol. 31, no. 3, pp. 287–296, May 2013.
[33] D. Jackson and K. Mosurski, "Heavy defeats in tennis: Psychological momentum or random effect?" *Chance*, vol. 10, no. 2, pp. 27–34, Mar. 1997.
[34] R. Seidl and P. Lucey, "Live counter-factual analysis in women's tennis using automatic key-moment detection," in *Proc. MIT Sloan Sports Anal. Conf.*, Mar. 2022, pp. 1–7.
[35] J. M. Silva, C. J. Hardy, and R. K. Crace, "Analysis of psychological momentum in intercollegiate tennis," *J. Sport Exercise Psychol.*, vol. 10, no. 3, pp. 346–354, Sep. 1988.
[36] T. Barnett, A. Brown, and S. R. Clarke. (2006). *Developing a Tennis Model That Reflects Outcomes of Tennis Matches*. [Online]. Available: https://api.semanticscholar.org/CorpusID
[37] A. Sekar and D. Pierpaolo. (2019). *Predicting the Winner of a Tennis Match Using Machine Learning Techniques*. [Online]. Available: https://api.semanticscholar.org/CorpusID
[38] H. Song, Y. Li, C. Fu, F. Xue, Q. Zhao, X. Zheng, K. Jiang, and T. Liu, "Using complex networks and multiple artificial intelligence algorithms for table tennis match action recognition and technical-tactical analysis," *Chaos, Solitons Fractals*, vol. 178, Jan. 2024, Art. no. 114343, doi: 10.1016/j.chaos.2023.114343.
[39] H. Yang, F. Jiang, X. Wu, G. Zhao, X. Shi, G. Liu, and M. Wang, "Optimizing the cutting edge geometry of micro drill based on the entropy weight method," *Int. J. Adv. Manuf. Technol.*, vol. 125, nos. 5–6, pp. 2673–2689, Mar. 2023.
[40] L. Page. (2009). *The Momentum Effect in Competitions: Field Evidence From Tennis Matches*. [Online]. Available: https://api.semanticscholar.org/CorpusID
[41] C.-W. Hsu and C.-J. Lin, "A comparison of methods for multiclass support vector machines," *IEEE Trans. Neural Netw.*, vol. 13, no. 2, pp. 415–425, Mar. 2002, doi: 10.1109/72.991427.
[42] H. M. Cartwright, "Swarm Intelligence. by James Kennedy and Russell C Eberhart with Yuhui Shi. Morgan Kaufmann Publishers: San Francisco, 2001. $43.95. xxvii + 512 pp. ISBN 1-55860-595-9," *Chem. Educator*, vol. 7, pp. 123–124, Apr. 2002.
[43] C. A. C. Coello, G. T. Pulido, and M. S. Lechuga, "Handling multiple objectives with particle swarm optimization," *IEEE Trans. Evol. Comput.*, vol. 8, no. 3, pp. 256–279, Jun. 2004, doi: 10.1109/TEVC.2004.826067.
[44] H. Dietl and C. Nesseler, "Momentum in tennis: Controlling the match," *Int. J. Sport Psychol.*, vol. 48, no. 4, pp. 2–5, Jul. 2017.

YILIN LEI (Student Member, IEEE) is currently pursuing the bachelor's degree with Nanjing University, Jiangsu, China. He was admitted to Nanjing University, to study electronic and engineering, in 2022. He is also working on artificial intelligence and medical image recognition.

AO LIN was admitted to Nanjing University, Jiangsu, China, to study computer science, in 2021. Good at programming, currently learning big data processing.

JIANUO CAO JR. is currently pursuing the bachelor's degree. She was admitted to Nanjing University, Jiangsu, China, to study electronic and engineering, in 2023. Her research interests include data analytics and machine learning models.