WILEY | Hindawi

Retraction

Retracted: Prediction and Analysis of the Physical Test Scores Based on BP Neural Network and Principal Component Analysis Algorithm

Security and Communication Networks

This article has been retracted by Hindawi following an investigation undertaken by the publisher [1]. This investigation has uncovered evidence of one or more of the following indicators of systematic manipulation of the publication process:

(1) Discrepancies in scope

(2) Discrepancies in the description of the research reported

(3) Discrepancies between the availability of data and the research described

(4) Inappropriate citations

(5) Incoherent, meaningless and/or irrelevant content included in the article

(6) Manipulated or compromised peer review

The presence of these indicators undermines our confidence in the integrity of the article's content and we cannot, therefore, vouch for its reliability. Please note that this notice is intended solely to alert readers that the content of this article is unreliable. We have not investigated whether authors were aware of or involved in the systematic manipulation of the publication process.

In addition, our investigation has also shown that one or more of the following human-subject reporting requirements has not been met in this article: ethical approval by an Institutional Review Board (IRB) committee or equivalent, patient/participant consent to participate, and/or agreement to publish patient/participant details (where relevant).

Wiley and Hindawi regrets that the usual quality checks did not identify these issues before publication and have since put additional measures in place to safeguard research integrity.

We wish to credit our own Research Integrity and Research Publishing teams and anonymous and named external researchers and research integrity experts for contributing to this investigation.

The corresponding author, as the representative of all authors, has been given the opportunity to register their agreement or disagreement to this retraction. We have kept a record of any response received.

References

[1] J. Qu, "Prediction and Analysis of the Physical Test Scores Based on BP Neural Network and Principal Component Analysis Algorithm," *Security and Communication Networks*, vol. 2022, Article ID 5210810, 12 pages, 2022.

WILEY | Hindawi

Research Article

Prediction and Analysis of the Physical Test Scores Based on BP Neural Network and Principal Component Analysis Algorithm

Jiale Qu [iD]

Sports Institute, Inner Mongolia University, Hohhot 010021, China

Correspondence should be addressed to Jiale Qu; 111983018@imu.edu.cn

College students' physical fitness test can comprehensively evaluate students' physical health level from many aspects such as students' body shape, quality, function, and sports ability. Through the analysis and prediction of college students' physical fitness test results, it can be used to understand the actual physical quality of college students, formulate the evaluation standard of students' physical health level, and assist physical education teachers to formulate reasonable teaching plans. For students with different physical quality, different training plans can be formulated to improve their actual physical quality. Aiming at this background, this paper puts forward a method based on BP neural network and principal component analysis algorithm of college students test score prediction algorithm, the BP neural network, and principal component analysis of machine learning algorithm successfully applied to physical test comprehensive performance prediction, implements the college students test transcript accurate prediction, and can effectively assist PE teachers to develop reasonable teaching plan. For students with different physical quality, different training plans can be formulated to improve their actual physical quality.

1. Introduction

In 2019, the arrival of novel coronavirus made us deeply aware of the importance of the human self-immune system. Strengthening physical exercise to improve the own immunity can effectively prevent the invasion of bacteria and viruses. The novel coronavirus outbreak is a warning to all of us humans, especially the younger generation of college students today. With good physical quality, people can put more energy on daily life, study, and work, which is also the basis of contemporary college students' study and life, but also one of the necessary elements to ensure the all-round development of students. College students are the future pillars of our country, and whether they can grow up healthily is related to the future of our nation. The Ministry of Education requires colleges and universities to carry out physical fitness tests for college students every year and report the test data, so as to have a comprehensive understanding of the physical development of college students today [1].

College students' physical fitness test will comprehensively evaluate students' physical health level from the aspects of students' physical form, quality, function, and sports ability, which is a very effective educational means to promote the healthy development of students' physical fitness and encourage students to actively carry out physical exercise and also the evaluation standard of students' physical health level. As the main base for cultivating high-quality talents, colleges and universities are facing a harsh test [2]. It is an important task of physical education teaching in colleges and universities to cultivate college students' interest in physical exercise and make them participate in sports consciously and move forward to lifelong sports. At present, exercise has become an important measure used in many countries to improve the physical health level, and maintaining a moderate amount of daily exercise is beneficial to the health of people of all age groups. Exercise can effectively enhance physical fitness and improve psychological quality; at the same time the human muscles will be more developed and more powerful. Colleges and universities have always paid too much attention to the intellectual development factors of college students and ignored their physical condition. With the improvement of people's daily living standards, the material life of college students has become richer, thus with lack of exercise, and some students'

physical quality has decreased. Unideal physical quality will not only affect the study and life of college students, but also affect their future work in the society.

The analysis and prediction of the physical fitness test results can be used to understand the actual physical condition of college students. The annual physical fitness test can make students deeply aware of their own physical fitness changes. Physical fitness test can make college students deeply realize the importance of physical exercise and encourage them to strengthen exercise to improve their physical health level. Physical fitness test is an important means for teachers to understand students' physical quality. The results of college students' physical fitness test can help the sports management department of universities to set up scientific and reasonable courses and develop the most effective training mechanism. Therefore, this paper uses the physical fitness test data over the years for analysis, and the analysis results are used to assist physical education teachers to develop reasonable teaching plans. For students with different physical quality, different training plans can be formulated to improve their actual physical quality.

Based on this background, in order to realize the prediction and analysis of college students' physical education test scores, this paper proposes an algorithm combining BP neural network and principal component analysis (PCA), which establishes a dynamic model, and realizes the analysis and prediction of physical education test scores. The full text is divided into four chapters. Chapter 1 introduces the research background and research necessity and the chapter arrangement of the paper; Chapter 2 mainly analyzes some research work on the test and analysis of sports performance. The feasibility of BP neural network and principal component analysis in sports exercise test was also discussed; Chapter 3 mainly introduces the theory and modeling process of BP neural network and principal component analysis. A model and algorithm based on BP neural network and principal component analysis algorithm is proposed; Chapter 4 mainly involves the specific implementation of the model and algorithms proposed in Chapter 3 and applied to the actual physical test performance analysis and prediction, validation of the algorithm, analyzing the experimental results and the errors, and arrival at a conclusion. The experiments show that the BP neural network and principal component analysis algorithm can realize the analysis and prediction of body test results, are consistent with the reality, and have good results.

2. State of the Art

Today's society is full of challenges and competition, so the society has put forward higher requirements for the basic quality of talents. There are always many problems in how to cultivate high-quality talents with all-round development in colleges and universities [2]. The important task of physical education in colleges and universities is to cultivate the interest of college students through physical exercise, so that they can consciously participate in sports and maintain it for life. In many countries, exercise has become an important tool to improve physical health.

The study of how to enhance students' physical quality is a great project by all countries around the world. The United States is a very developed country in science, technology, and economy, and it is also one of the earliest countries in the world to pay attention to the study of national physical fitness. In the mid-twentieth century, the AAHPER Association took the lead in developing an efficient physical education evaluation index, adding four items: standing long jump, sit-up, softball throwing, and 50-meter return running to the student physical fitness test. The physical fitness study in the United States combines the physical education courses in colleges and universities and implements a very distinctive fitness program in every school. Japan is the country with the most complete and earliest survey and research data on adolescent physique. Since 1898, Japan has accumulated a large number of data and conducted research. Japan is recognized as one of the more authoritative countries in the world in the field of testing and evaluation of adolescent physical fitness. The Japanese people are leading the way in studying adolescent body shape and bone development, combining their physical fitness with study, life, and work. In particular, Europe has established a committee called "urofield," whose main task is to coordinate the test management of student physical fitness in various countries, check the results of each test, and recommend the whole process of the test and the final results of the test [3]. British physical monitoring in colleges and universities started later than United States and Japan, but its monitoring level has reached the world advanced level, especially in single project testing. The use of national sports monitoring standards, the students' body composition and exercise ability evaluation and monitoring, has reached the degree of automation. The Australian Football League has successfully applied the neural network algorithm in machine learning to the selection of tournament players, predicting the athletic ability of the participants, including anthropometric, mental, and skill estimation, and selecting the best players, thus increasing the possibility of competition victory [4]. The results of this study show that neural networks can better assist recruitment managers to identify talents. Machine learning algorithms have been successfully applied to NBA games to collect historical data of games and predict whether the basketball team can win according to the results of data analysis. It can effectively avoid accidents and adjust the order of game players in time, so as to increase the possibility of victory [5]. Wang platinum et al. used BP neural network algorithm to estimate the development status of folk custom sports events in some parts of China from two aspects of male and female characteristics. The results show that neural network can better predict the future development of traditional sports [6]. Wang Ji'an usage of BP neural network algorithm is relatively stronger, self-fault tolerance and self-training learning ability as well as good advantages, are useful in constructing model, athletes performance prediction, neural network model for sports performance prediction model research provides extremely extensive development space, and neural network in prediction and analysis is very popular [7]. Kerstin Witte uses principal component analysis to study the sports coordination of
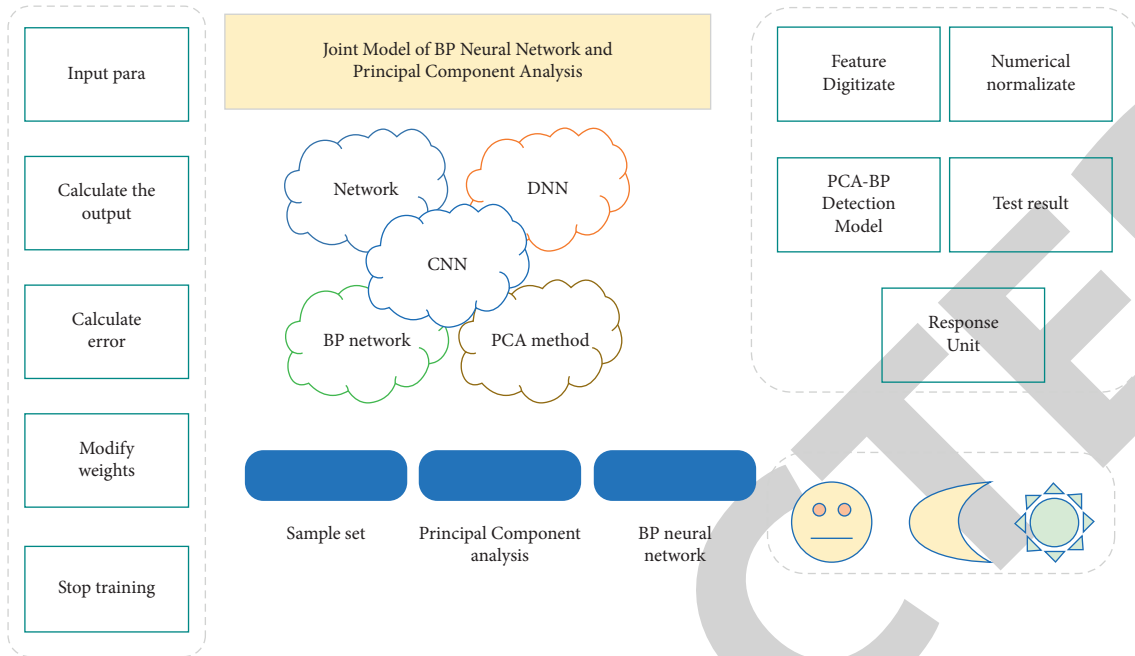
Figure 1: Flowchart of PCA methods.

rehabilitation, triathlon, and horse riding in sports science [8]. The results show that the principal component analysis can effectively characterize people's coordination during movement in a whole. The principal component analysis algorithm is used to decompose the complex movement mode of skiing into the main movement components, determine the standard posture and main movements of the skiers, and assist the coach in training and preaching [9, 10].

To sum up, all countries attach great importance to the study of college students' physique. College students are the future and hope of a country, and their physical condition is a problem that we should pay common attention to. The analysis and research of machine learning data in the field of sports are widely involved both at home and abroad, which provides great help to people's life, physical health, and national competitions. In order to analyze and deal with the historical data of college students' physical fitness test scores, classify management students and put forward targeted training plans suitable for students' own physical quality, so as to assist physical education teachers to better develop reasonable teaching plans. This paper analyzes the historical data of the current college students' sports performance test, draws samples, and uses radar maps to visualized data, which intuitively reflects the distribution of each student's scores. Then, the correlation analysis and principal component analysis methods in the machine learning algorithm are used to preprocess the physical fitness test data, eliminate the mutual influence between the attributes of the data, and cover the original attribute information with the new principal component attributes, and the analysis process is simplified and the prediction time is saved. Finally, the BP neural network algorithm was used to establish the college student sports performance prediction model, and the model was used for performance analysis and prediction, so

that the physical fitness test data over the years could play a greater role and provide better help for teachers to make teaching plans.

3. Methodology

3.1. Principal Component Analysis Method. One of the biggest challenges in data processing is data multiple-complexity. Principal component analysis is a commonly used technique used to reduce the dimensionality of datasets to explore and simplify some kind of complex relationship between variables. Principal component analysis was first proposed by Pearson in 1901 and developed by Hotelling in 1933, with the main idea being to represent the majority of the original variables with a few components via dimensionality reduction techniques [11]. The flow of the algorithm using the principal component analysis is shown in Figure 1.

It can transform multiple original variables with strongly correlated properties into several unrelated variables, and the calculated several variables without correlation are the main components [12]. The goal of the principal component is to use a smaller set of unrelated variables instead of a large number of correlated variables, while retaining the raw data information as much as possible. The principal component is a linear combination of the original variables, whose model is shown in Figure 2.

The principal component reflects most of the characteristics of the original variables and can remove strong correlations between the original variables. The principal component is a linear combination of the original variables, with the first one explaining the most variance to the first known variable, while the second one ranks second for the original variable variance interpretation and is orthogonal
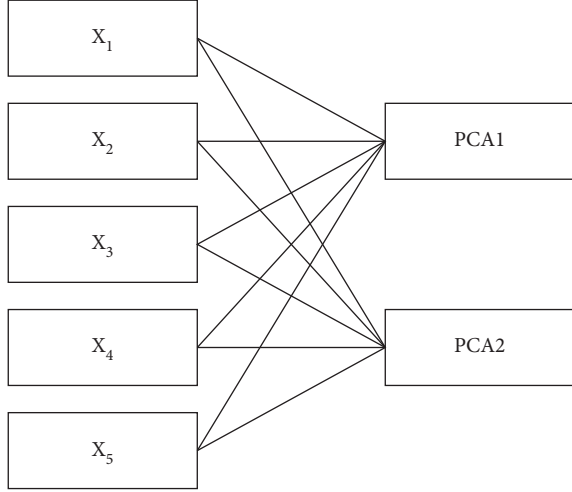
Figure 2: Methods and model for principal component analysis.

to the first principal component, that is, completely irrelevant [13]. And so on, the remaining principal components are all related in orthogonal relationships. It is assumed that there are P variables in the original data, respectively, x_1, x_2, ..., x_p, and the new p mutually independent principal component variables are formed through linear combination. The mathematical model expression is as follows:

$$\begin{cases} y_1 = u_{11}x_1 + u_{12}x_2 + ... + u_{1i}x_i + ... + u_{1p}x_p \\ y_2 = u_{21}x_1 + u_{22}x_2 + ... + u_{2i}x_i + ... + u_{2p}x_p \\ \vdots \\ y_i = u_{i1}x_1 + u_{i2}x_2 + ... + u_{ii}x_i + ... + u_{ip}x_p \\ \vdots \\ y_p = u_{p1}x_1 + u_{p2}x_2 + ... + u_{pi}x_i + ... + u_{pp}x_p \end{cases}. \quad (1)$$

The model changes to the matrix form as shown in formula:

$$y = \begin{bmatrix} u_{11} & u_{12} & \cdots & u_{1p} \\ u_{21} & u_{22} & \cdots & u_{2p} \\ \vdots & \vdots & \vdots & \vdots \\ u_{p1} & u_{p2} & \cdots & u_{pp} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_p \end{bmatrix} = Ux. \quad (2)$$

The sum of the squares of the principal component coefficient is 1, as shown in formula:

$$u_{i1}^2 + u_{i2}^2 + \ldots + u_{ip}^2 = 1. \quad (3)$$

Next, in order to obtain the principal component value of y_p, calculate the principal component coefficient. The covariance matrix is first calculated from the raw data as shown in equation

$$cov = \frac{1}{n} \sum_{i=1}^{n} (x_1 - \overline{x})(y_1 - \overline{y}). \quad (4)$$

Since the data has been normalized processed, the variance of the original data s^2 should be one. See formula

$$s^2 = \frac{1}{n} \sum_{i=1}^{n} (x_1 - \overline{x})^2 = 1. \quad (5)$$

Change formula (5) to (6),

$$\sum_{i=1}^{n} (x_1 - \overline{x})^2 = ns^2 = 1. \quad (6)$$

The formula is

$$r = \frac{\sum_{i=1}^{n} (x_i - \overline{x})(y_i - \overline{y})}{\sqrt{\sum_{i=1}^{n} (x_1 - \overline{x})^2} \sqrt{\sum_{i=1}^{n} (y_i - \overline{y})^2}}$$
$$= \frac{\sum_{i=1}^{n} (x_i - \overline{x})(y_i - \overline{y})}{n} = cov. \quad (7)$$

From formula (7), the correlation coefficient matrix of the original data is actually equivalent to the covariance matrix. The eigenvalues of the covariance matrix represent the variance of the principal components, while the eigenvalues and principal component coefficients μ_{pp} are calculated by the correlation coefficient matrix and the eigenvalues y are calculated by $y = x\mu^T$. Theoretically according to the principal component contribution to choose less principal components instead of the original data, generally selected contribution is 90% of the number of principal components. This will inevitably lead to missing data, resulting in inaccurate experimental results. So in this paper, we will choose the same original number of principal components, so we will not throw the original data information and also can remove the strong correlation between the original variables on the model training. Next, the neural network model was established using eight unrelated new variables, which increased the persuasion of the model accuracy, while excluding the influence of the data factors on the model establishment and parameter optimization.

3.2. BP Neural Network. Artificial neural networks, also known as neural networks, have originated in neurophysiology [14]. Neural network is composed of a large number of neural cells, which is a simplified and abstract with simulation of the human brain. Neural network is a kind of machine learning intelligent algorithm to imitate the brain; it has its own unique nonlinear information processing ability, which can be stored in the neural network through adaptive and autonomous learning for information, repeated learning and training, applied in intelligent control, image recognition, combination optimization computing, and speech recognition, etc. Neural networks contain many types, such as perceptron, BP neural network, radial base network, self-organized mapping network, etc.

Among the numerous models of neural networks, the BP neural network model is the most widely used representative neural network. It is a supervised learning algorithm based on the error-square MSE as the objective function and a multilayered network with weights trained on the nonlinear differentiable function [15]. The basic structure of the BP neural network is shown in Figure 3.
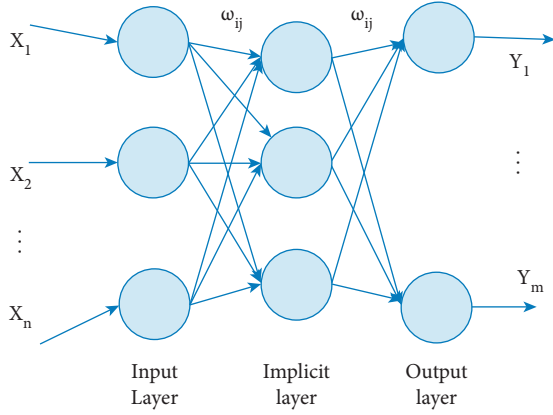
FIGURE 3: B P Neural network basic model structure.

The BP neural network mainly consists of three parts: input layer, implied layer, and output layer, where X_i refers to the input value of the BP neural network, W_{ij} and W_{jk} are the weight of the neural network, and Y_i refers to the output value of the BP neural network. It can also be seen from the above figures that, in the BP neural network model, the network input value is taken as the independent variable of the function, and the network output value is taken as the dependent variable of the function.

The algorithm flow of BP neural network is shown in the figure, mainly including forward propagation and backpropagation. The specific implementation process is as follows.

(1) Forward propagation of the data:

Internet patterns between neural networks are formed by the interconnections of neurons, and the initial weights between each connection are randomly assigned by the computer. The forward propagation stage of the data signal refers to the process of the original data signal passing through the implied layer from the input layer to the output layer; that is, the output of the upper node serves as the input of the lower node.

As shown in Figure 4, each neuron cell has a corresponding computational weight. The output value of the input layer in the implied layer is obtained by the input value, connection weight, and threshold value, and the calculation method is shown in formula

$$\text{net1}_k = \sum_{i=1}^{n} w_{ik} y_i + b_k = w_{1k} y_1 + w_{2k} y_2 + \ldots w_{nk} y_n + b_k. \tag{8}$$

During the process of prediction, applying the activation function processing can obtain better prediction accuracy. There are many kinds of activation functions, such as step function, Sigmoid function, tanh function, and ReLU function. This paper uses the Sigmoid function to activate the output information. The output value of the input layer is activated by the activation function to (f (net1) k). The implied layer family is obtained by formula

$$z_k = f\left(\text{net1}_k\right) = \frac{1}{1 + \exp\left(-\text{net1}_k\right)}. \tag{9}$$

Next, the implied layer data acts as the input layer to pass the data to the output layer. The output value net2_j is the connection weight v_{kj} between the hidden layer value and the hidden layer and the output layer. The weighted sum plus the threshold is still obtained as shown in formula

$$\text{net2}_j = \sum_{k=1}^{n} v_{kj} z_k + b_j = v_{1j} z_1 + v_{2j} z_2 + \ldots + v_{nj} z_n + b_j. \tag{10}$$

Formula (11) activates the output value to obtain the data for the final output layer.

$$z_k = f\left(\text{net2}_j\right) = \frac{1}{1 + \exp\left(-\text{net2}_j\right)}. \tag{11}$$

(2) Error backpropagation

When the signal is transmitted to the output layer, the error function is used to detect whether the training process of the neural network ends. The neural network is stopped by satisfying the error function limit value or reaching the set maximum number of iterations. Training is stopped when the output error function is less than the predetermined value. If the condition is not met, the error is backpropagated. The error function (E) is used to measure the error size between the actual output code and the desired output O_j, as shown in formula

$$E = \frac{1}{2}(d - o)^2 = \frac{1}{2} \sum_{j=1}^{n} \left(d_j - o_j\right)^2. \tag{12}$$

The error signal from each layer was used to adjust the weights of connections between neurons. Equation (13) simulates the process of error backpropagation.

$$E = \frac{1}{2} \sum_{j=1}^{n} \left(d_j - o_j\right)^2 = \frac{1}{2} \sum_{j=1}^{n} \left(d_j - f\left(\sum_{k=1}^{n} v_{kj} z_k\right)\right)^2. \tag{13}$$

The error decreases along the gradient by constantly adjusting the connection weights and thresholds. After calculating the change value of the weight connection value Δv_{kj} between the implied layer and the output layer, update each connection weight and see formulas (14) and (15).

$$\Delta v_{kj} = -\eta \frac{\partial E}{\partial \text{net2}_j} z_k. \tag{14}$$
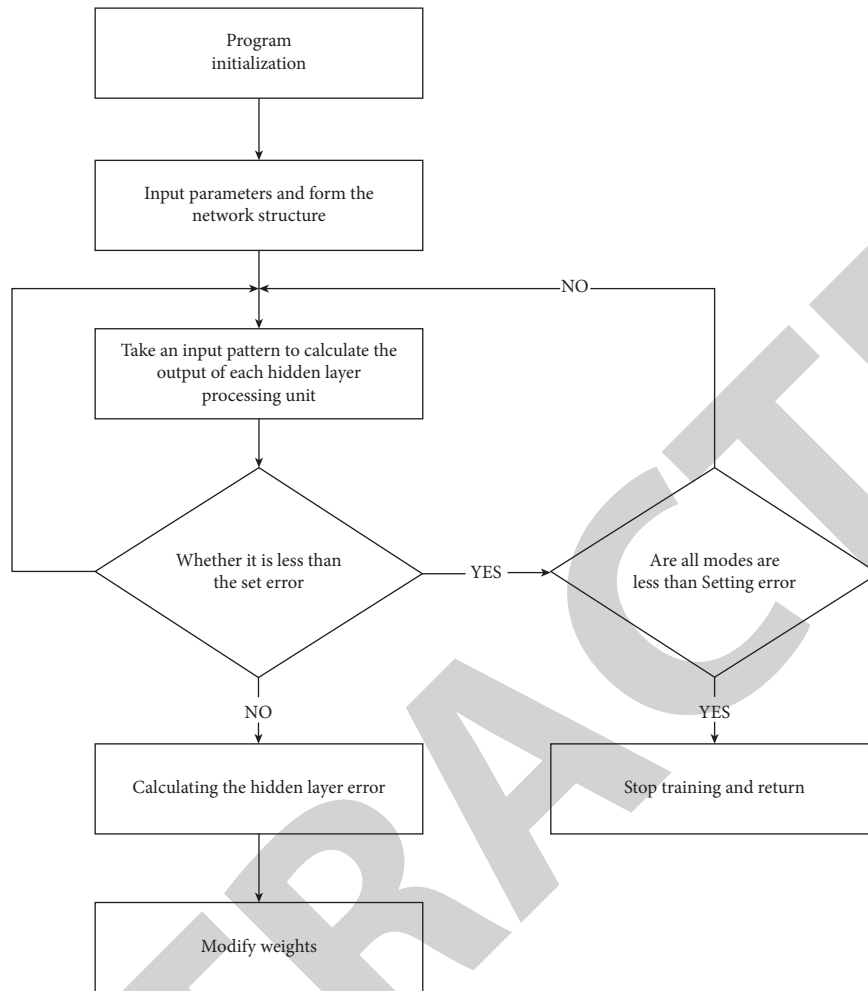
$$v_{kj} = v_{kj} + \Delta v_{kj}. \tag{15}$$

Figure 4: Training process of the BP neural network.

Adjust the connection weight between the input layer and the hidden layer as above. When all weights are readjusted, the signal forward propagation will continue. When the model reaches the convergence criterion, the training is stopped, the model is established, and the model parameters are adjusted to optimize the model. The established model is used to predict the physical fitness test data, calculate the error size between the predicted value and the actual value to verify the feasibility of the model, and then apply the model.

3.3. Predictive Model Based on the BP Neural Network and Principal Component Analysis Algorithm. Physical fitness test data were preprocessed and standardized to eliminate strong correlations between the data using principal component analysis. After transforming the raw data by using the principal component analysis method, the BP neural network is used to build the model. The overall process is as shown in Figure 5.

Before selecting the actual results of college students' sports test as the original data and using principal component analysis, we need to standardize the original data to eliminate the impact of different dimensions of variables on

the analysis results. We then chose the first few principal components where the cumulative variance contribution is no less than 85% after dimensionality reduction as a new learning sample of the neural network.

4. Result Analysis and Discussion

4.1. Overview of the Physical Fitness Test Data. The physical fitness test dataset used in this institute is composed of students in the 2016–2019 academic year. The dataset contains the physical fitness test records of all students in the university within four years, including the measurement results of instruments and equipment, the score of individual tests, additional points, and the final comprehensive score, and the total data of more than 80,000 pieces.

The analysis of the dataset used can be performed by using the data visualization analysis technology, and one of the sampled student information can be represented in Figure 6, and from the figure, the actual physical condition of the student can be even more clearly observed.

The physical fitness tests in the dataset are as follows:

(1) Height (H) and weight (W): Height and weight are the required basic items in the physical fitness test.
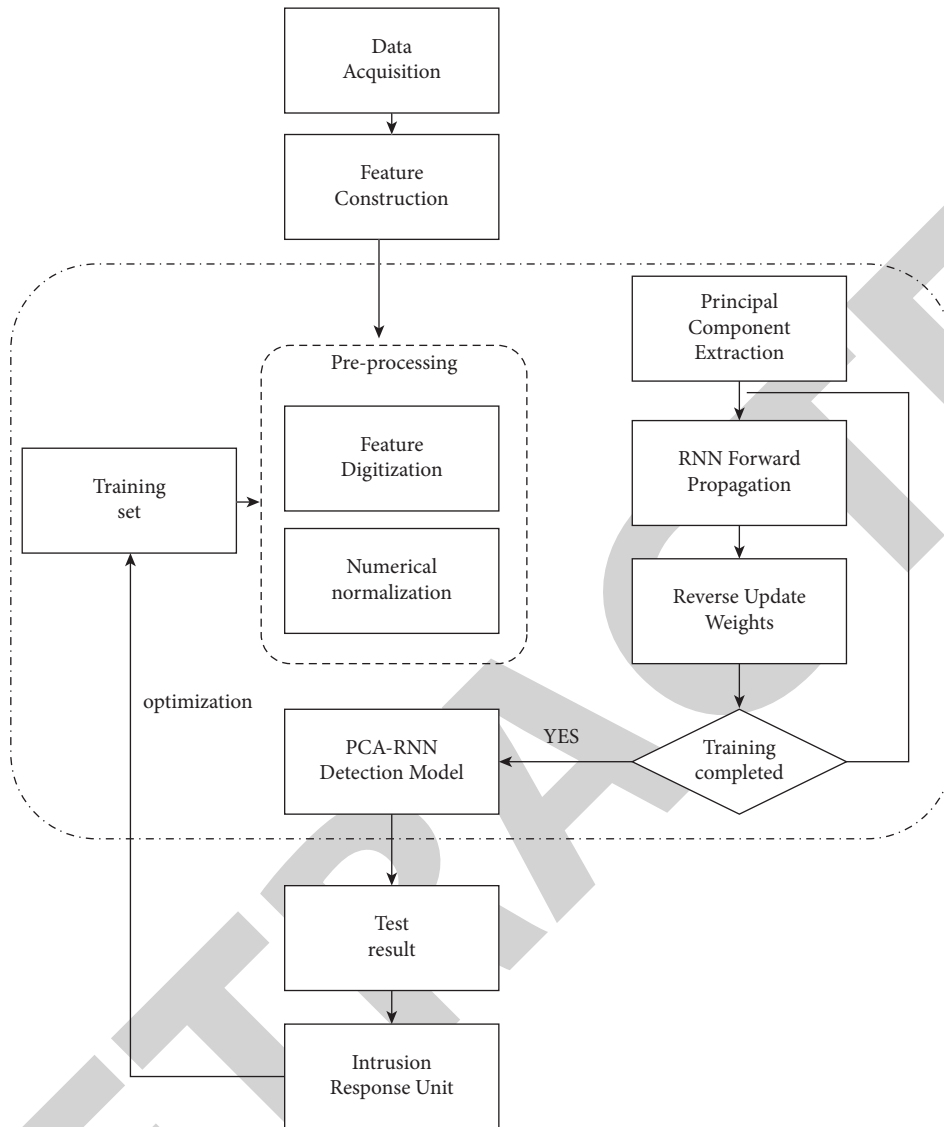
Figure 5: Physical measurement analysis and prediction model based on principal component analysis and BP neural network.

These two tests can judge whether a person's growth and development tend to be normal. The change of human body size is one of the important criteria reflecting the health of a person's basic body.

(2) Spirometry (VC): Spirometry refers to the maximum deep breath after the maximum deep inhalation. This process represents the maximum functional activity of human lungs at a time and reflects the potential ability of respiratory function in the lung. This activity volume is an important indicator used to evaluate the function of human respiratory system.

(3) 50-meter sprint (S): 50-meter sprint is a common international "displacement speed" test project. It measures students' physical speed quality through short distance and high intensity running, including movement speed and reaction speed.

(4) Standing long jump (SLJ): Standing long jump is a long jump that does not start from a standing position without any support run. This process is mainly used to measure the explosive force of the lower limb muscles of the human body when jumping forward. The explosive force of the human body depends on the strength of the human body itself, which is one of the inaccessible factors in daily life.

(5) Sitting forward flexion (SR): Sitting anterior flexion is used to measure the maximum range of activity of the trunk, waist, and other joints of the human body at rest. The measurement is mainly used to reflect the elasticity and extension of the body's joints, muscles, and ligaments, as well as the level of body flexibility.

(6) Endurance items (SP): Endurance items are measured in middle and long distance running. Due to the different physique of boys and girls, the project distance measurement is different, including 1000 meters for boys and 800 meters for girls. Middle and long distance running is a kind of exercise mode to cultivate students' endurance quality, but also an aerobic metabolism
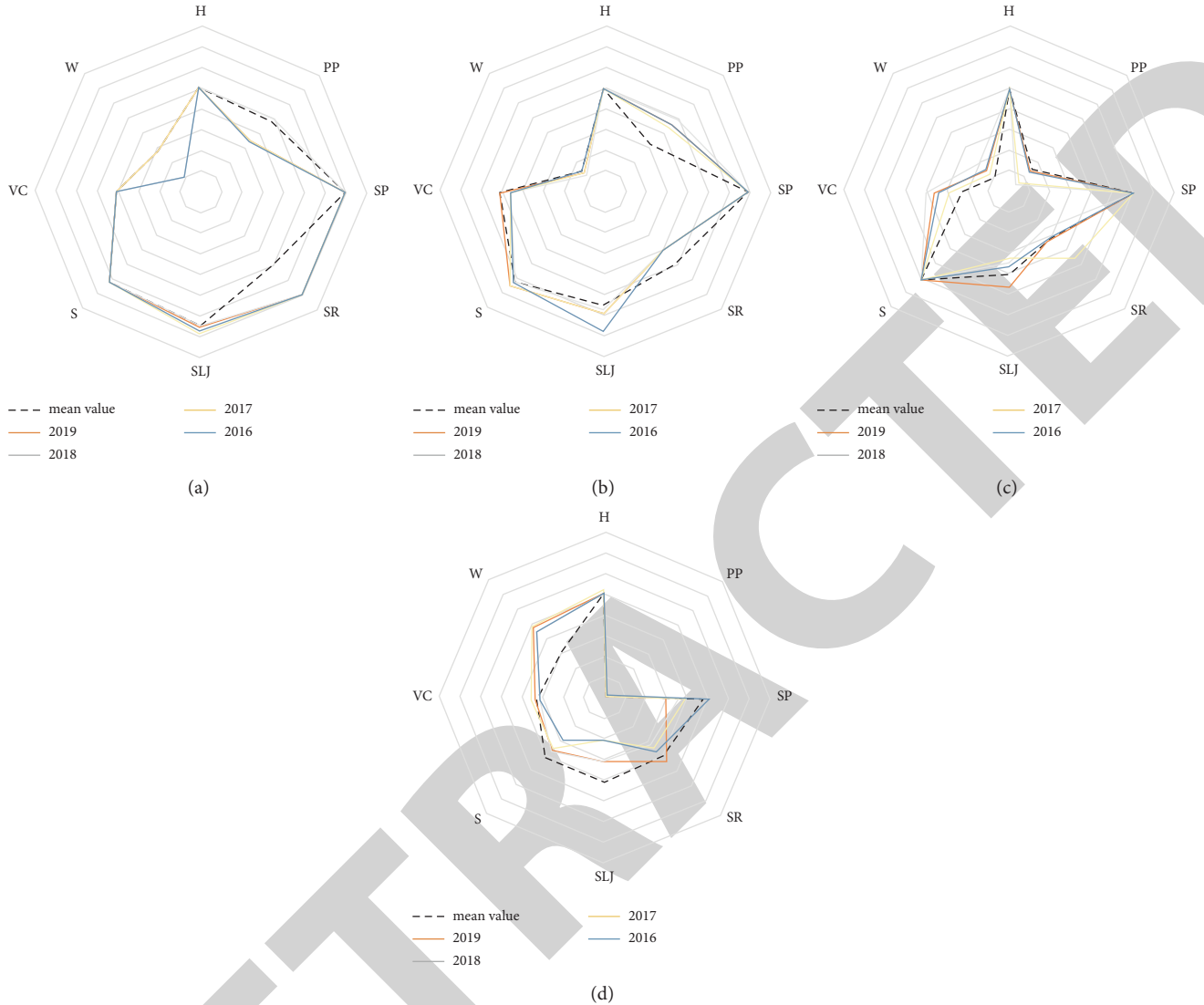
Figure 6: Drawing four-year test data for a sample of students. (a) Students with perfect grades. (b) Students with individual weaknesses. (c) Students reach to the passing score nearly. (d) Students with lowest scores.

project, which is mainly used to test students' endurance and body oxygen supply function.

(7) Strength project (PP): Strength test is the number of corresponding items completed in one minute. For the difference of physical requirements between male and female students, boys conduct pull-up project, while girls perform sit-up project. Both pull-ups and sit-ups are methods used to measure muscle endurance and strength. The measured results of the strength test can provide a better understanding of the students' physical muscle strength and endurance.

4.2. Implementation of the Body Measurement Algorithm Based on the BP Neural Network and Principal Component Analysis. The 2016 physical fitness test data were selected to build the model, with 80% of the student sample as training set and the remaining 20% as test set for model evaluation. The scoring criteria and methods are different for boys and

Table 1: Physical fitness test predicted model parameters.

Parameter name	Parameter values
Number of neurons in the input layer	8
Number of neurons in the output layer	1
Number of neurons in the hidden layer	11
Threshold	0.005
Learning rate	0.1
Maximum iteration number	$1.0e10$
Training algorithm	rprop+
Error function	sse
Activation function	Logistic

girls, so the test data were separated to build separate models for prediction. Continuous tuning optimized the model using 80% of the training set, and the final model parameters are shown in Table 1.

As shown in Table 1, the threshold serves as the conditional value for training stopping, which indicates the predetermined value in the error function. The maximum
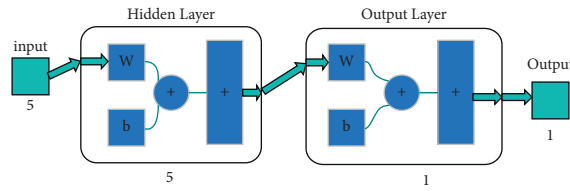
Figure 7: The Matlab neural network toolbox.

Figure 8: Neural network training process.

number of iterations forces the training to stop when the predetermined value is never reached and the iteration cannot stop. The algorithm "rprop+" used for training is a weighted error backpropagation algorithm, namely, the BP neural network algorithm. The error function "SSE" was used to calculate the magnitude of the error at the end of the forward propagation. The activation function uses the parameter "logistic" for the Sigmoid activation function. The number of neurons in the hidden layer is determined by the mean square error (MSE) and formula
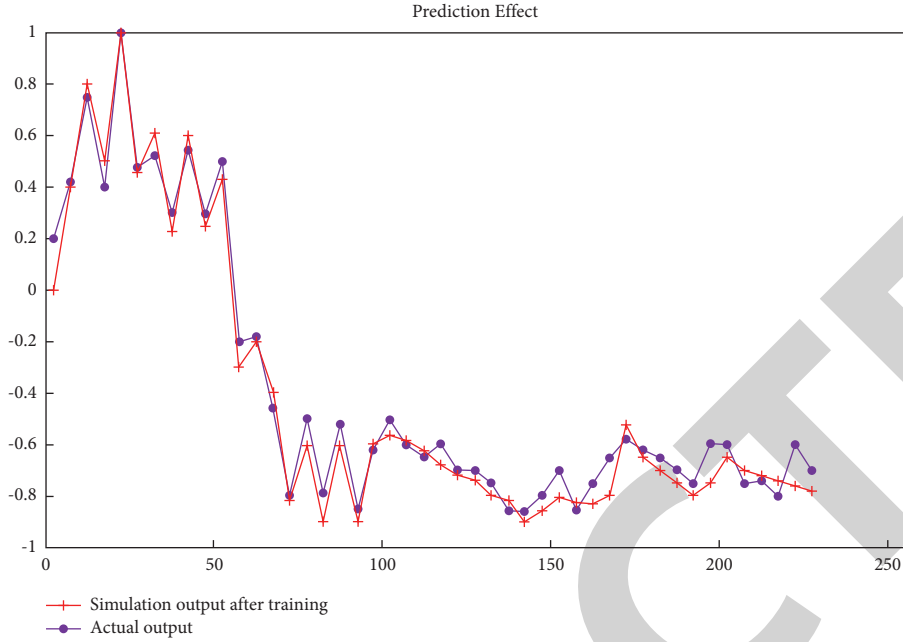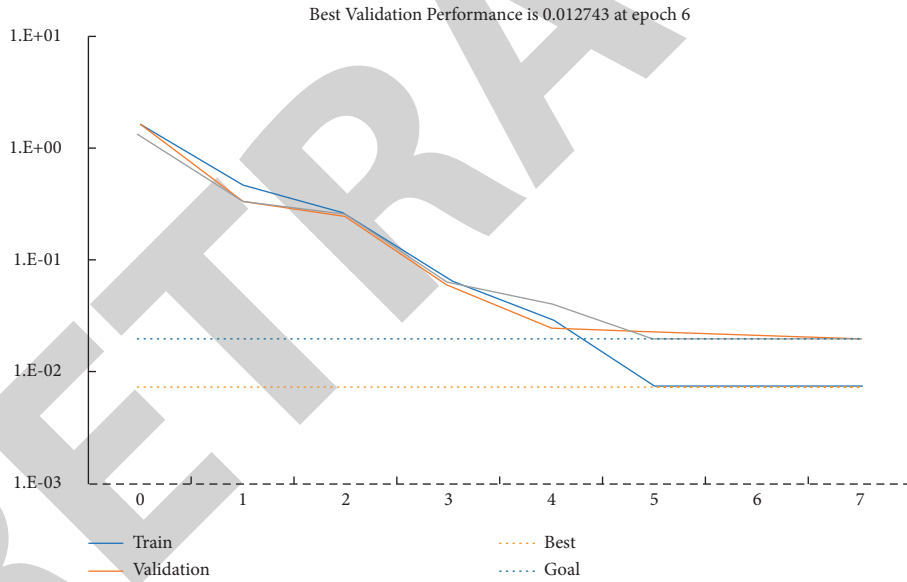
FIGURE 9: Neural network training fitting effect.

FIGURE 10: Performance curves for the BP neural network training.

$$h = \sqrt{m+n} + a. \qquad (16)$$

m is the number of input neurons, h is the number of output neurons, and a is a constant ranging from 1 to 10, so the value of the number of hidden layers h ranges from 4 to 13. Therefore, the model building process is conducted for the number of different hidden layers, and the mean square error (MSE) is used to compare the accuracy of the model prediction. Mean square error (MSE) is calculated by absolute error (AE), and the absolute error is the difference between the actual output value and the model predicted value as shown in formulas (17) and (18).

$$AE = o_i - o_j. \qquad (17)$$

$$MSE = \frac{1}{n} \sum_{i=1}^{n} \left(o_i - o_j \right)^2. \qquad (18)$$

Mean square error (MSE) method effectively avoids the problem that positive and negative errors cannot be added and can be used to evaluate the accuracy of model prediction. The smaller the value of MSE, the stronger the ability of the model to fit the experimental data, but the value of MSE cannot be 0, which proves that our established model is overfitted. Through the size of MSE under

the model of each number of hidden layers, the number of neurons in the hidden layer was finally determined to be 11. The setup of the neural network using Matlab is shown in Figure 7, and the training process is shown in Figure 8.

Then, 80% of the data after the principal component analysis was brought into the function to build the neural network model and evaluate the accuracy of the model using the data from the test set. The test set data not involved in the model training process were brought into the model for prediction, respectively, and finally the prediction results of male and female students were combined to observe the prediction performance of the overall model. After sampling individual samples to evaluate the test results, the model prediction performance is further observed from the overall data. The visual results show that the model prediction performance proposed in this paper is very high, which has certain practical value for the prediction of comprehensive physical fitness test performance in the future.

The sample of 40 students was randomly selected in the test set of the 2016 data, and the difference between the actual value of these 40 students and the predicted value of the model is shown in Figure 9. Shown are the predicted versus actual values of a random sample of 40 students in the test set. The real line in the figure is the actual value, and the dashed line is the predicted value. It is obvious that these two lines have a high coincidence rate, and only individual samples can see the obvious error. The results show that the physical fitness test of the prediction model is very accurate and has a good performance.

Figure 10 shows the performance curve of the neural network training, indicating the variance changes. After several cycles, the network achieves convergence with the mean variance of 0.0070188 and 0.0098638, which are less than the set expected error target of 0.001. The whole curve drops faster, which indicates the appropriate learning rate. Error can reflect the reliability of the prediction results, and the absolute error (AE) value close to 0 means that the prediction is very accurate.

5. Conclusion

College students' physical fitness test mainly evaluates their physical condition and training effect through the test results of various items, which is a very sound and effective strategy and can urge college students to actively participate in sports training. College PE teachers can provide students with the most scientific training plan according to the test results, so as to continuously enhance their own physical fitness. In order to better analyze the physical education test results of college students and to predict them, this paper proposes a comprehensive performance prediction model of physical fitness test to successfully apply machine learning algorithms such as BP neural network and principal component analysis to the prediction of physical fitness test. The performance prediction model to predict the comprehensive performance reduces the performance calculation time and solves the problem of inconsistent scoring standards due to manual calculation over the years. Compared with the traditional calculation method, the physical fitness test results play greatest value. Compared with the annual test results of the student, the child labor can reflect whether the student's physical fitness is improved and whether the corresponding training plan provided by the teacher is reasonable and effective. At the same time, it provides guidance for the adjustment of the training plan and the teachers' scientific formulation of the teaching plan. The results show that the comprehensive performance prediction model is highly accurate and can serve the physical education teaching very well.

Data Availability

The labeled datasets used to support the findings of this study are available from the author upon request.

Conflicts of Interest

The author declares no conflicts of interest.

Acknowledgments

References

[1] B. Zhao, "Performance analysis and prediction of the world's top male athletes in Decathlon," *Sports Culture Guide*, vol. 3, p. 4, 2013.

[2] F. Wang, "Research and analysis of physical fitness test management system of baoshan university," *Yunnan University*, vol. 3, pp. 76–79, 2017.

[3] F. Zhe, *Research on European Sports Development Mode Reform*, East China Normal University, Shanghai, China, 2013.

[4] Duncan, "Data mining in sports," 2010.

[5] B. Loeffelholz, E. Bednar, and K. W. Bauer, "Predicting NBA games using neural networks," *Journal of Quantitative Analysis in Sports*, vol. 5, no. 1, p. 7, 2009.

[6] B. Wang, X. Li, and Y. Sun, "BP neural network model-based urbanization process traditional sports cultural development e," *The Open Cybernetics & Systemics Journal*, vol. 9, no. 1, pp. 3005–3010, 2015.

[7] J. Zhou, W. U. Guilin, and Y. U. Shengsheng, "Algorithm of BP neural network-based face detection," *Computer Engineering*, vol. 30, 2004.

[8] K. Witte, N. Ganter, C. Baumgart, and C. Peham, "Applying a principal component analysis to movement coordination in sport," *Mathematical and Computer Modelling of Dynamical Systems*, vol. 16, no. 5, pp. 477–488, 2010.

[9] B. Natàlia, G. Jacob, J. Casimiro et al., "Cardiorespiratory coordination after training and detraining: a principal component analysis approach," *Frontiers in Physiology*, vol. 7, 2016.

[10] P. Federolf, R. Reid, M. Gilgien, P. Haugen, and G. Smith, "The application of principal component analysis to quantify technique in sports," *Scandinavian Journal of Medicine & Science in Sports*, vol. 24, no. 3, pp. 491–499, 2014.

[11] W. Wang Bin, "Multivariate statistical analysis and R language modeling," Jinan University Press, Guangzhou, China, 2010.

[12] M. Tauno and V. Jaak, "ClustVis: a web tool for visualizing clustering of multivariate data using principal component analysis and heatmap," *Nucleic Acids Research*, vol. 45, pp. W566–W70, 2015.

[13] C. f. Geng, *Application Research on the Identification of Online Game Cheating Behavior Based on Machine Learning*, University of Electronic Science and Technology, Chengdu, China, 2019.

[14] J. Behler, "Constructing high-dimensional neural network potentials: a tutorial review," *International Journal of Quantum Chemistry*, vol. 16, 2015.

[15] L. Han, *Group.Artificial Neural Networking Tutorial*, Artificial Neural Network Tutorial, IGI Global, Hershey, PA, USA, 2006.