International Journal for Housing Science and Its Applications
Publish August 5, 2025. Volume 46, Issue 3 Pages 4013-4023

https://doi.org/10.70517/ijhsa463341

Research on subgroup computational analysis and intelligent recognition of students' sports performance data during physical education teaching process

Yong Wang[1], Xu Wang[1] and Zongshuai Hao[1,*]
[1] Department of Physical Education, Cangzhou Normal University, Cangzhou, Hebei, 061001, China
Corresponding authors: (e-mail: czsfxytyx@163.com).

Abstract This study proposes an intelligent analysis method for students' movement performance data by integrating binocular vision technology, BP neural network and subgroup computation. Through dual-camera stereo correction and block matching algorithms, high-precision depth information and motion data are acquired, a three-layer BP neural network model based on correlation analysis to screen key indicators is constructed, the nonlinear mapping ability is optimized by combining Sigmoid function, and data normalization and sample grouping are used to improve the model generalization performance. The performance indicator (PI) clustering algorithm is further introduced to narrow the differences of dependent variables within the cluster, and the model training accuracy and efficiency are significantly improved by base adjustment and scaling. The multidimensional analysis showed that the physiological indicators such as morning pulse, blood pressure and blood oxygenation were significantly and dynamically associated with the athletic performance. Compared with the least squares support vector machine and hybrid genetic neural network, the proposed method shows better prediction stability and real-time performance in eight physical measurements, especially in standing long jump (99.41%) and BMI (99.53%), with an average accuracy of 98.2%, which is more than 10% higher than that of the traditional method, and the single prediction time is only 1.3 seconds, which verifies the combination of real-time performance and accuracy of the proposed method. real-time and accuracy.

Index Terms binocular vision technology, BP neural network, swarm computation, motion performance analysis

I. Introduction

School physical education plays an important role in promoting students' physical and mental health and enhancing their physical fitness. However, the traditional teaching mode is no longer able to meet the increasingly diversified needs of students. For example, manual records and static tests are mostly used in the process of physical education teaching, which lacks the tracking of students' dynamic performance and long-term data, leading to difficulties in adjusting the teaching content and insufficient personalized guidance [1], [2]. In addition, in terms of teaching assessment, the application of intelligent equipment and big data technology has yet to be introduced, and the assessment feedback mechanism is simple, so it is difficult for students to obtain targeted advice through the assessment results, and it is not possible to effectively improve the learning effect [3]. In recent years, with the promulgation of policies such as education informatization, the application of information technology in the field of education has been vigorously promoted. Students' physical education teaching, as an important part of education, has also been gradually modernized and scientized with the support of information technology [4]. The application of information technology not only improves the teaching efficiency, but also provides students with more diversified learning methods, promotes the modernization and scientization of physical education teaching, and makes physical education teaching more open and innovative globally [5]-[7].

In modern physical education teaching, information technology is not only used for the display of teaching content, but also widely used in the real-time collection and analysis of students' sports data [8], [9]. With the introduction of wearable devices and sensor technology, etc., teachers are able to obtain students' sports performance data more easily and generate detailed assessment reports, which help teachers and students understand the learning progress and effect more intuitively, and promote the feedback and adjustment of teaching [10]-[13]. At the same time, the combination of deep learning algorithms to analyze students' sports performance, identify common errors and provide targeted feedback helps to improve the accuracy and relevance of teaching, reduces the burden on teachers, and provides new possibilities for physical education teaching [14], [15].

This study proposes a comprehensive method system integrating binocular vision technology, BP neural network and subgroup computing. The article focuses on data acquisition, model construction and optimization, specifically including: student action image acquisition and processing: through dual-camera stereo correction and block matching algorithms, high-precision depth information is acquired to provide a reliable data base for subsequent analysis. BP neural network model for predicting athletic performance: Based on correlation analysis, key training indicators are screened, a three-layer BP network model is constructed, Sigmoid function is used to optimize the nonlinear mapping ability, and data normalization and sample grouping are used to improve the generalization performance of the model. Clustering computation and algorithm optimization: Introducing performance index (PI) to achieve data clustering, narrowing the differences of dependent variables within the cluster, and combining base adjustment and scaling to significantly improve the model training accuracy.

II. Motion data acquisition and BP neural network clustering prediction model construction

II. A. Processing of student movement image acquisition

The student's movements are acquired using two cameras which can acquire color information about the student and also depth information about the student. The first step in collecting images of student movements with two cameras is to calibrate the cameras with the goal of depicting stereoscopic information about the student in the form of a two-dimensional plan view. Once calibration is complete, stereo adjustments need to be made between the two cameras to ensure that the pictures taken by the left and right cameras are on the same horizontal plane. To perform stereo calibration of the binocular cameras, Bouguet's method was used. If the rotation matrix and translation vector of the binocular camera are represented by and respectively, in order to minimize the degree of aberration of the reprojection between the left and right binocular cameras and to ensure that the planes of the two cameras are in a coplanar state, Bouguet's stereoscopic adjustment method transforms the composite matrix that is adjusted for both the left and right bicavities. The specific expression is:

$$\begin{cases} r_l = R^{1/2} \\ r_r = TR^{-1/2} \end{cases} \tag{1}$$

where r_l and r_r are the rotation matrices of the left and right cameras, respectively. By applying Eq. (1), the two camera planes will be placed at the same position on the same plane. To ensure their effective alignment on this plane, computing the pole matrix R_{rect} is necessary to find infinitely many polynomials. The formula for R_{rect} is as follows:

$$R_{rect} = \begin{bmatrix} h_1^T \\ h_2^T \\ h_3^T \end{bmatrix} \tag{2}$$

where z is used to describe the transpose operation of the matrix; h represents the poles. On the direction h_1, corresponding to different rows of the translation vector T, there are multiple polygonal shapes, and this relationship is represented by equation (3). h_2 is mutually orthogonal to the direction of the camera's optical axis and represents the direction vector of the image plane, and this relationship is derived by Eq. (4) on h_1 and h_2. Between the points h_1 and h_2, h_3 is obtained by processing the orthogonality, which is described by equation (5):

$$h_1 = T / \|T\| \tag{3}$$

$$h_2 = [-T_y T_x 0] / \sqrt{T_x^2 + T_y^2} \tag{4}$$

$$h_3 = h_1 \times h_2 \tag{5}$$

The optical axis of the camera is defined by the horizontal and vertical translation vectors T_x and T_y, which can be combined with Eq. (2) R_{rect} to obtain the stereo correction matrix for binocular cameras as follows:

$$\begin{cases} R_l = R_{rect} r_l \\ R_r = R_{rect} r_r \end{cases} \tag{6}$$

According to equation (6), stereo adjustment of the binocular camera was accomplished to ensure accurate alignment of the image lines. On this basis, stereo coordination was performed to obtain deeper images of the student. The corresponding algorithm of block was used to complete the stereo balancing. Block matching algorithm is used to accomplish stereo matching. Based on the views of other cameras, the closest block to the delineated image block is acquired as a way to realize the acquisition of the student's motion image.

II. B.BP neural network model for predicting athletic performance

The depth information and motion data acquired through binocular vision technology provide multi-dimensional feature parameters for the input of BP neural network. And section 2.2 further screens key indicators based on these parameters to construct a prediction model, realizing the transition from data acquisition to intelligent analysis.

II. B. 1) Screening of model independent variables

Due to the different strengths of association between various quality training indicators and sports performance, the number of influences on the prediction of sports performance will be different. It is necessary to screen out the quality training indicators that have a greater impact on students' sports performance. Using the historical data information of students included in the State General Administration of Sports from 20023 to 2024, the correlation analysis of each quality training index of students and sports performance was done, and the respective correlation coefficients were calculated (r), and the results of the correlation coefficients between the quality training indexes and the sports performance are shown in Table 1.It can be seen in Table 1 that the BMI value of students, the lung capacity, the 50-meter run, The correlation coefficients between 9 quality training indexes and sports performance, such as standing long jump, sitting forward bend, sit-up/pull-up, 800m/1000m running, are all larger, and these 9 quality training indexes are selected as the predictors of students' sports performance.

Table 1: The correlation coefficient between quality indicators and sports performance

Quality training indicators	Correlation degree
BMI	0.8336
Vital capacity	0.8966
50-meter run	0.8712
Standing long jump	0.9106
Sit forward bend	0.7192
Sit-ups	0.9332
Pull-ups	0.8346
800-meter run	0.8428
1000-meter run	0.7289

II. B. 2) BP Neural Network

BP neural network, i.e., a multi-layer error forward feedback neural network, belongs to the error back propagation algorithm. It consists of an input layer, an output layer and a number of hidden layers, and each of its layers has a number of nodes, each node represents a neuron, the upper nodes and the lower nodes are connected through the weights, and the nodes between layers are connected using full interconnection, and there is no correlation between the nodes within each layer.The composition of the three-layer structure of the BP neural network is shown in Fig. 1.

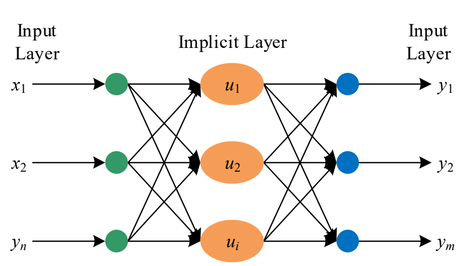

Figure 1: Neural network

BP neural network adopts the Sigmoid (S strong bending) type differentiable function, namely: is strictly incremental, can make the output show a better balance between linear and nonlinear, so it can realize the arbitrary nonlinear mapping between input and output, and it is suitable for medium- and long-term prediction. It has the advantages of good approximation, fast calculation speed and high accuracy. Meanwhile, its theoretical basis is solid, the derivation process is rigorous, the resulting formula is symmetrical and beautiful, and it has strong nonlinear fitting ability. It is suitable for the processing of small data volume.

II. B. 3) Modeling of BP neural network algorithm

The establishment of the BP neural network algorithm model generally requires the completion of the following seven steps: (1) the collection and grouping of sample data; (2) the preprocessing of data; (3) the determination of the number of hidden layers; (4) the determination of the number of nodes in the hidden layer; (5) the training of the neural network; (6) the determination of the initial weights of the network; and (7) the testing of the network algorithm model's performance and generalization ability.BP neural network algorithm model The establishment steps are shown in Fig. 2.

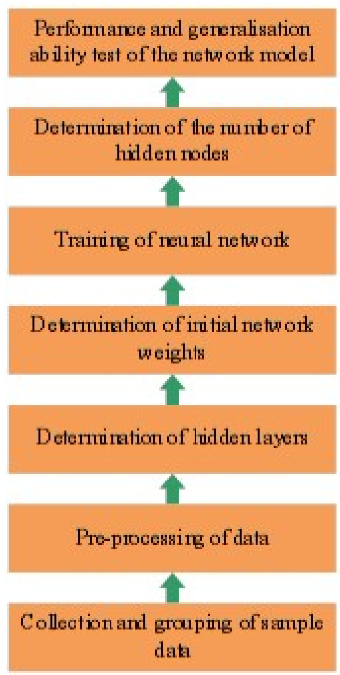

Figure 2: Steps for establishing the BP neural network algorithm model

MATLAB has a neural network toolbox, which provides enough modeling commands and functions to greatly simplify the difficulty and steps of human modeling.

(1) Collection and grouping of sample data

The athletic performance of students in 7 sports of a university in the 2022-2024th session was collected as a total sample. And it is determined that the 7 sports scores of the 2022nd session are the training samples, the 7 sports scores of the 2023rd session are the test samples, and the 7 sports scores of the 2024th session are the test samples. Set the original matrix as p, first of all, the data will be processed to the unit level of seconds, and set the matrix of the behavior of a project of different sessions of the value, listed as a session of the value of different projects. Form a matrix of 7×7.

$$P=\begin{bmatrix} 87.07 & 77.67 & 85.22 & 100 & 76.60 & 67.59 & 87.16 \\ 93.04 & 84.43 & 77.42 & 75.13 & 62.12 & 99.66 & 82.95 \\ 69.41 & 86.55 & 93.66 & 100 & 69.29 & 61.61 & 78.21 \\ 81.34 & 91.96 & 65.56 & 86.17 & 59.93 & 88.62 & 100 \\ 77.52 & 96.68 & 99.37 & 97.24 & 97.78 & 100 & 83.96 \\ 98.15 & 94.00 & 100 & 92.33 & 61.80 & 100 & 99.71 \\ 75.04 & 63.64 & 71.58 & 74.25 & 100 & 87.10 & 64.65 \end{bmatrix} \tag{7}$$

(2) Data preprocessing

Since the hidden layer of the BP network generally adopts the Sigmoid transfer function, it is necessary to preprocess the data on the original input matrix P to obtain the matrix P1. First, ensure that the value of the input data falls between 0-1. After transposing the matrix P, find the maximum value of each column and divide each column in the matrix with its maximum value. P1 is obtained, as shown in Equation (8). Secondly, the variables of training samples, test samples and test samples as well as the corresponding data are provided for the training of the neural network, i.e., 1-5 rows of matrix P1 are taken as training samples P10; 2-6 rows of matrix P1 are taken as test samples P20; and 3-7 rows of matrix P1 are taken as test samples P30.

$$P1=\begin{bmatrix} 0.867 & 0.722 & 0.821 & 1.000 & 0.794 & 0.657 & 0.871 \\ 0.925 & 0.899 & 0.715 & 0.755 & 0.613 & 0.993 & 0.861 \\ 0.704 & 0.855 & 0.943 & 1.000 & 0.702 & 0.604 & 0.806 \\ 0.818 & 0.897 & 0.704 & 0.845 & 0.801 & 0.909 & 1.000 \\ 0.788 & 0.961 & 0.992 & 0.936 & 0.950 & 1.000 & 0.807 \\ 0.959 & 0.920 & 1.000 & 0.919 & 0.605 & 1.000 & 0.992 \\ 0.753 & 0.653 & 0.701 & 0.733 & 1.000 & 0.869 & 0.680 \end{bmatrix} \tag{8}$$

(3) Determination of the number of hidden layers and nodes

A neural network model with implied layers is actually a linear or nonlinear regression model. It is generally believed that increasing the number of hidden layers can reduce the network error. Of course, it will also complicate the network, and increase the training time of the network and the tendency of “overfitting”. Therefore, a 3-layer BP network (i.e., with one hidden layer) is used in this study.

The number of nodes in the implicit layer is not only related to the number of nodes in the input and output layers, but also to the complexity of the problem to be solved and the type of transfer function and the characteristics of the sample data. To determine the number of nodes in the hidden layer must meet the conditions: the number of nodes in the access layer and the hidden layer must be less than $N-1$ (where N is the number of training samples) based on the study of all types of training samples, the input is 5, the output is 1, so the number of nodes in the hidden layer is determined to be 3.

(4) Determination of functions and parameters

The activation function of the hidden layer in this study is determined as tansig hyperbolic tangent S type (Tan-Sigmoid) transfer function, namely

$$\tan sig(x)=\frac{e^x-e^{-x}}{e^x+e^{-x}} \tag{9}$$

The activation function of the output layer is determined to be purelin, i.e.

$$purelin(x)=x \tag{10}$$

At the same time, the maximum number of learning iterations of the network is set to be 10000; the learning accuracy of the network is set to be 0.00005.

II. C.Calculation of subgroups

Although the BP neural network has the ability of high-precision prediction, it is prone to the problem of low training efficiency when facing large-scale dispersed data. For this reason, this section proposes the clustering algorithm, which classifies the samples by PI indicators, optimizes the model input structure, and significantly improves the convergence speed and prediction stability of the neural network.

II. C. 1) Segmentation algorithms

A clustering algorithm is proposed for the following two reasons:

(1) When the sample data is large, the number of hidden nodes required to simulate the functional relationship between the independent and dependent variables with the BP model is large and the training time is long. When the samples are divided into several clusters and the corresponding BP network is trained, the amount of information is higher because there are fewer samples within each cluster.

(2) When the values of the dependent variable (the output of the BP model) are more dispersed (i.e., the difference between the maximum and minimum values is large), it is difficult to train the BP model to a high accuracy. If we can propose a clustering algorithm to sort the samples basically according to the size of the dependent variable,

the samples with small differences in the dependent variable data can be divided into a group, so that the purpose of improving the training accuracy of the BP model can be achieved by adopting some processing measures. It should be clarified that the data of the dependent variable can not be utilized when dividing the group, only the data of the independent variable can be utilized. Because the purpose of grouping and training the corresponding BP model is to use them to identify which group the sample set belongs to and to calculate its corresponding sports performance data, whose sports performance data (dependent variable) is to be calculated and unknown.

Considering that the performance of the students' sports data can be considered as an increasing function of the four characteristic parameters mentioned above, if the samples with the smaller of the four characteristic parameters are ranked in the front (and vice versa), the results of the ranking should be basically in order of the sports performance data from the smallest to the largest. We define the following performance index (Performance Index-PI):

$$PI_i = \sqrt{\sum_{j=1}^{4} C_j \left(Z_{ij} - Z_{ij\min}\right)^2} \quad i = 1,2,\cdots,N \tag{11}$$

In the formula:

$$Z_{ij\min} = \min_{i\in[1,N]} Z_{ij} \quad j = 1,2,3,4 \tag{12}$$

C_j is the weight of the j feature parameter, which is used to indicate the speed of change of the motion performance data with the j feature parameter. When the relationship between the change of motion performance data with the j feature parameter can be considered as a linear function, C_j can be taken as its slope. In the following study, we take $C_j\left(j = 1,2,3,4\right)$ to be all 1.

From equation (11), it can be seen that the samples with larger values of PI for all four feature parameters have larger values and correspondingly larger motion performance data. In this way, sorting the samples according to the size of their PI values can achieve the purpose of sorting the samples basically according to the size of the motion performance data. After the sorting result is obtained, the samples can be artificially divided into several clusters in order, and the values of the motion performance data of the samples in each cluster are relatively close to each other. Obviously, the above clustering algorithm does not utilize the motor performance data (dependent variable) data. In addition, the order of the PI values of the clustering algorithm from the smallest to the largest is not necessary to fully reflect the order of the motor performance data from the smallest to the largest, as long as it reflects the general trend, because the order of the samples within the same cluster does not have much influence on the learning process of the BP model.

II. C. 2) Treatment of dependent variables in raw data

After clustering with the algorithm in Section 2.3.1, several clusters with small differences in line losses within the same cluster can be obtained. At this point, the following processing of the dependent variable data of the samples within each cluster can improve the training accuracy of the BP model.

Let the largest dependent variable in an in-group sample be $d_{\max}$ and the smallest be $d_{\min}$. We first subtract a base number b from the dependent variable d_i of the samples in the cohort, and then multiply the dependent variables so processed by a number g, so that the largest of them (i.e., $d_{\max} - b$) is equal to g, i.e., $g = g/\left(d_{\max} - textb\right)$. We take b to be a number slightly less than $d_{\min}$, i.e., $b = d_{\min} - b'$, with b' taken to be a small positive number, and g taken to be a value of slightly less than 1 (e.g., 0.99). The values of the dependent variables so treated are all within $\left[0,1\right]$ and the largest figure is slightly less than 1 and the smallest figure is slightly greater than 0 . That is, the dependent variables d_i are treated as follows:

$$d_i' = (d_i - b)g = g(d_i - d_{\min} + b')/(d_{\max} - d_{\min} + b') \tag{13}$$

In this way, the BP model can be trained by using the data of the sample independent variable processed by the method of the previous section as the input of the BP model and the data of the dependent variable processed by the above method as the output of the BP model.

Obviously, since the dependent variable values of the samples within the same group are all closer, the difference between the data processed above is greatly amplified, which is conducive to improving the learning accuracy of the BP model. However, the output d_i'' at the end of the training of the BP model corresponds to d_i' instead of d_i. From equation (13) (d_i' is replaced by d_i'', and d_i is replaced by $\hat{d}_i$):

$$\hat{d}_i = d_{\min} - b' + d_i''(d_{\max} - d_{\min} + b') / g \tag{14}$$

where $\hat{d}_i$ is the calculated value of the output of the BP model corresponding to d_i.

III. Multidimensional analysis of students' exercise data and correlation of physiological indicators

Based on the data acquisition system constructed in Chapter 2 and the BP neural network model of subgroup optimization, this chapter will rely on the actual students' physical measurement data, and carry out the research from the three levels of data preprocessing, physiological indexes and athletic performance correlation, and model validation, to verify the practical application value of the proposed method.

III. A. Introduction to student exercise data

The research object of this paper is college students' physical measurement data, and this section will focus on the presentation and analysis of the data and data preprocessing.

III. A. 1) Data presentation and analysis

The physical fitness test scores of students in a university were used for data mining and analysis, with a view to providing reference suggestions for efficient student exercise and physical education reform. In order to achieve this purpose, the physical test scores from the first year to the fourth year of a university in 2024 were selected as the research data in this paper. Some of the students' physical test scores are shown in Table 2, which contains information on students' gender and scores of each physical test item, but for privacy protection reasons. In this paper, the privacy information such as students' names and school numbers are replaced by serial numbers. Among them, the physical test scores of male students include BMI scores, lung capacity scores, 50-meter scores, standing long jump scores, sitting forward bend scores, pull-up scores and 1000-meter scores; and the physical test scores of female students include BMI scores, lung capacity scores, 50-meter scores, standing long jump scores, sitting forward bend scores, sit-up scores and 800-meter scores. By mining and analyzing these data, this paper will provide schools with recommendations on physical education reform and provide students with personalized exercise advice to help them better develop their physical health and participate in physical activities.

Table 2: some of the students' physical test results

ID	Sex	BMI score	Vital capacity fraction	50-meter score	Standing long jump score	Sitting forward bend score	Sit-up score	Pull-up score	800-meter score	1000-meter score
1	Male	93	100	81	79	75	-	100	-	100
2	Female	83	100	74	59	91	86	-	90	-
3	Female	100	100	82	77	90	72	-	63	-
4	Male	100	100	86	100	73	-	65	-	91
5	Male	100	100	93	70	64	-	56	-	82
6	Female	99	100	65	72	83	95	-	81	-
7	Female	100	100	70	73	96	97	-	72	-
8	Male	100	100	83	82	76	-	98	-	71
9	Female	85	99	73	90	100	79	-	92	-
10	Female	100	90	72	81	89	80	-	56	-
…	…	…	…	…	…	…	…	…	…	…
37293	Male	100	100	70	84	73	-	68	-	85

III. A. 2) Data pre-processing

There are a large number of incomplete, noisy and inaccurate data in the original data, in order to make it able to meet the data requirements of the classical association rule mining algorithm, it is necessary to integrate, clean up, delete and transform the acquired data. The specific data preprocessing operations are as follows.

(1) Data integration

The acquired data contains the physical measurement data of undergraduate students from different colleges of the university from the first year of university to the fourth year of university. In order to facilitate data integration and analysis, it is necessary to integrate the physical measurement data of undergraduates from different colleges. In the integration process, the data of each physical measurement item need to be arranged in a certain order in order to facilitate the subsequent data analysis and mining work.

(2) Data cleaning

For incomplete data in the data, if there are a large number of missing data items, the data will be deleted; if there are only a small number of missing data items, the data will be analyzed comprehensively by the average, median and plurality of the data, to make up the information for the missing items and to improve the completeness of the data.

(3) Data Screening

Based on the research goal of mining the correlation between physical test items in this paper, in order to facilitate the standardization of data, the physical test score data were adopted as the transaction attributes, i.e., the data part of the physical test scores were retained, which contained lung capacity scores, BMI (height and weight) scores, 50-meter scores, standing long jump scores, seated forward bending scores, sit-up/pull-up scores, and 800-meter/1000-meter scores, and the data part of the physical test scores were retained. Serial numbers, grades, and specific physical test scores were excluded. Due to the different physical characteristics of boys and girls, there are gender-specific items in the data, which are redundant with the gender attributes, and there are two solutions: the first is to add the gender attributes without distinguishing the gender-specific names of the physical tests, and the second is to remove the gender attributes without distinguishing the gender-specific names of the physical tests, and the latter is adopted in this paper.

III. B. Relationship between students' physical condition and performance

After completing the data preprocessing and preliminary analysis, this section further explores the dynamic associations between students' physiological indicators (morning pulse, blood pressure, and blood oxygen) and their athletic performance, revealing the potential impact of physical status on training effectiveness.

Students' performance is determined by their comprehensive physical condition, and it is difficult to study the effect of one of their health indicators on their performance in practical research. In this paper, the relationship between a single indicator of students' physical fitness and their body-side performance was investigated using the constructed BP neural network.

III. B. 1) Relationship between students' morning pulse and athletic performance

The relationship between their morning pulse and performance was first studied, and the relationship between students' morning pulse and their athletic performance is shown in Figure 3.

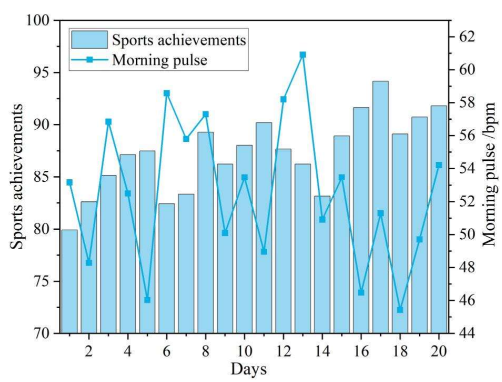

Figure 3: The relationship between students' morning pulse and sports performance

Within the 20-day data sample, if the students' morning pulse was more stable, their exercise performance was also more stable, and even had a tendency to gradually improve; however, if their morning pulse was in a period of instability, the increase in the number of beats per minute of the morning pulse would lead to the instability of the students' performance, and even a significant decline. Analyzing the reasons, it may be due to the students' excessive exercise the day before, resulting in the instability of their morning pulse, which further led to the fact that their physical strength and body condition were not in a perfect state in the subsequent training, so their performance dropped more significantly in the subsequent training.

III. B. 2) Relationship between students' blood pressure and athletic performance

In this paper, the relationship between students' blood pressure and their athletic performance was investigated using the constructed BP neural network, and the relationship between students' blood pressure and athletic performance is shown in Figure 4.

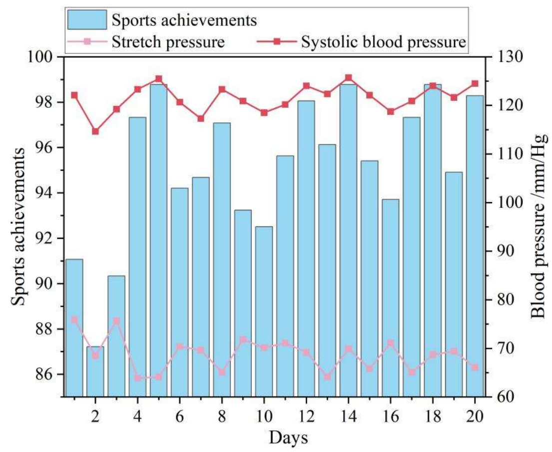

Figure 4: The relationship between students' blood pressure and sports performance

Within the 20-day data sample, there was a closer relationship between students' blood pressure and their grades. The test results showed that when the students' systolic and diastolic blood pressure were both in a lower state, the corresponding performance was poorer, such as when on the second day, both systolic and diastolic blood pressure were at a lower level, and their athletic performance was the worst, 87.17 points, the students' training program needs to be adjusted so that their blood pressure is back to the normal state; comparatively, if the systolic blood pressure was relatively high and the diastolic blood pressure was low, the students' athletic performance is better. When a student's blood pressure is at a critical value, his or her physical state is generally in the best state, and the sports performance achieved at this time is also generally the best performance. For example, on the fifth day, the systolic blood pressure is at its highest level, 125 mm/Hg, while the diastolic blood pressure is at a lower level, and the athletic performance at this time is the best performance.

III. B. 3) Relationship between students' blood oxygen and athletic performance

In this paper, the relationship between students' blood oxygen levels and their athletic performance was investigated using the constructed BP neural network, and the relationship between students' blood oxygen and athletic performance is shown in Figure 5.

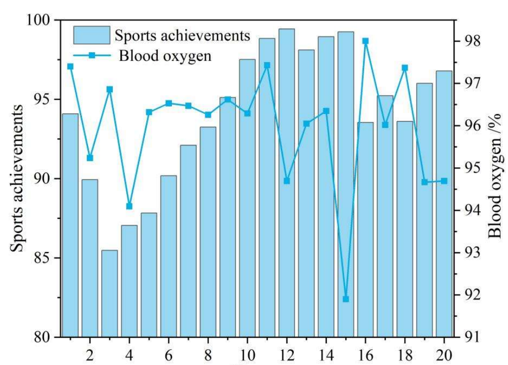

Figure 5: The relationship between students' blood oxygen and sports performance

Within the 20-day data sample, there was also a close relationship between students' oxygen levels and their performance. The results showed that when the students' blood oxygen levels were in a stable state, such as during the 5th-10th days, the performance of the exercise showed a gradual improvement, while the fluctuation of the blood oxygen levels resulted in the same fluctuation of the exercise performance with a lag. When the blood oxygen level of the students increased on the previous day, their performance increased on the following day; if the blood oxygen level of the students decreased on the previous day, their performance decreased on the following day.

III. C. Analysis of accuracy and estimation efficiency

By clarifying the association law between physiological indicators and exercise performance, this section quantitatively evaluates the accuracy and efficiency of the model and compares it with other popular methods to validate the comprehensive advantages of the proposed method.

In order to test the accuracy and estimation efficiency of the proposed method, the proposed method is utilized for comparison with Method 1: Least Squares Support Vector Machine and Prediction Error Correction based method and Method 2: Hybrid Genetic Neural Network based method.

III. C. 1) Accuracy analysis

The accuracy of the three methods for estimating different sports performance is shown in Table 3.

Table 3: The estimation accuracy of different sports achievements

Sports events	OURS	Method 1	Method 2
BMI	99.53	86.70	90.29
Vital capacity	97.01	86.26	88.48
50-meter run	97.57	93.19	87.53
Standing long jump	99.41	93.53	78.14
Sit forward bend	98.29	88.01	89.90
Sit-ups	97.88	85.09	79.47
Pull-ups	97.60	88.08	78.09
800-meter run	97.31	89.37	77.40
1000-meter run	97.02	83.19	77.13

Table 3 demonstrates a comparison of the accuracy of the three methods in estimating performance in eight sports. The data show that the proposed method significantly outperforms the other two methods in all sports. For example, in BMI prediction, the accuracy of this paper's method reaches 99.53%, which is much higher than that of the least squares support vector machine-based method (86.70%) and that of the hybrid genetic neural network-based method (90.29%), and the 99.41% accuracy of this paper's method in the standing long jump event emphasizes the advantage of this method. In addition, the accuracy of this paper's method is higher than 97% for all events, while the accuracy of other methods fluctuates greatly, such as only 77.13% for Method 2 in the 1000-meter run. This indicates that the proposed method significantly improves the prediction stability and generalization performance with the support of nonlinear mapping capability and subgroup optimization strategy.

III. C. 2) Estimated efficiency analysis

The estimation times of the three methods are shown in Table 4. According to Table 4, it can be seen that for different types of sports, the estimation time of student performance of the proposed method is significantly lower than the remaining two methods, and the estimation of student performance is more efficient.

Table 4: The estimated time of the three methods

Sports events	OURS	Method 1	Method 2
BMI	1.62	9.52	8.16
Vital capacity	1.48	9.57	10.08
50-meter run	1.35	8.34	7.52
Standing long jump	1.12	8.16	10.32
Sit forward bend	1.54	7.27	10.09
Sit-ups	1.12	7.63	9.59
Pull-ups	1.09	8.06	7.99
800-meter run	1.27	8.19	7.47
1000-meter run	1.03	8.69	9.13

Table 4 compares the estimation time efficiency of the three methods. This paper's method takes the shortest time among all items, with an average of about 1.3 seconds, while Method 1 and Method 2 take an average of 8.5 and 8.9 seconds, respectively. For example, the prediction of this paper's method on the BMI item took only 1.62

seconds, while Method 1 and Method 2 took 9.52 seconds and 8.16 seconds, respectively. The clustering algorithm optimizes the model training efficiency by reducing the intra-cluster data variability, which makes the BP neural network-based clustering computation method in this paper achieve fast computation while guaranteeing high accuracy, and is especially suitable for large-scale real-time data processing scenarios.

IV. Conclusion

In this study, the analysis accuracy and computational efficiency of students' sports performance data were effectively improved by binocular vision technology and the BP neural network model with clustering optimization. The experimental data show that the clustering algorithm significantly reduces the model training complexity by narrowing the data differences within the cluster, which makes the prediction accuracy reach more than 99% in BMI, standing long jump, etc., and the estimation time consumed is only 1/6 of the other methods.Physiological indexes correlation study shows that the stability of morning pulse, blood pressure and blood oxygen is significantly positively correlated with the athletic performance, which provides a scientific basis for the development of personalized training plan. This provides a scientific basis for the development of personalized training programs. Compared with the least squares support vector machine and hybrid genetic neural network, the proposed method shows significant advantages in nonlinear mapping ability, generalization performance and real-time performance, which provides efficient technical support for the reform of physical education and the management of students' physical fitness and health.

References

[1] Van der Mars, H., McNamee, J., & Timken, G. (2018). Physical education meets teacher evaluation: Supporting physical educators in formal assessment of student learning outcomes. Physical Educator, 75(4), 582-616.

[2] Liu, C. (2021, September). Exploration and analysis of physical education model based on computer multimedia and network data system. In 2021 4th International Conference on Information Systems and Computer Aided Education (pp. 1450-1454).

[3] Gu, Y. (2022). Deep integration of physical education and multimedia technology using Internet of Things technology. Wireless Communications and Mobile Computing, 2022(1), 2556142.

[4] Pengli, L. (2019). Discussion on the Application Effect of Informatization in Physical Education Teaching. In 3rd International Conference on Economics, Management Engineering and Education Technology. Kuala Lumpur: ICEMEET.

[5] Dong, J. (2023). Advancing Health Outcomes in Higher Education: Innovative Strategies for Integrating Physical Education with Informatization Systems under the Great Health Perspective. Journal of Commercial Biotechnology, 28(6).

[6] Deng, C., Feng, L., & Ye, Q. (2024). Smart physical education: Governance of school physical education in the era of new generation of information technology and knowledge. Journal of the Knowledge Economy, 15(3), 13857-13889.

[7] Kashuba, V., Asaulyuk, I., & Diachenko, A. (2017). A modern view on the use of information technologies in the process of physical education of student youth. Journal of Education, Health and Sport, 7(2), 765-775.

[8] Xu, Y., Peng, J., Jing, F., & Ren, H. (2024). From wearables to performance: how acceptance of IoT devices influences physical education results in college students. Scientific Reports, 14(1), 23776.

[9] Arnold, J. F., & Sade, R. M. (2017). Wearable technologies in collegiate sports: the ethics of collecting biometric data from student-athletes. The American Journal of Bioethics, 17(1), 67-70.

[10] Fu, D., Chen, L., & Cheng, Z. (2021). Integration of wearable smart devices and internet of things technology into public physical education. Mobile Information Systems, 2021(1), 6740987.

[11] Wang, B. (2017). Evaluation of sports visualization based on wearable devices. International Journal of Emerging Technologies in Learning (iJET), 12(12), 119-126.

[12] Wu, G., & Zhang, X. (2022). Realization of wireless sensors and intelligent computer aided teaching in physical education and training. Wireless Communications and Mobile Computing, 2022(1), 6415352.

[13] Liu, X., & Jiang, J. (2022). Digital twins by physical education teaching practice in visual sensing training system. Advances in Civil Engineering, 2022(1), 3683216.

[14] Wang, Y., Wang, X., & Hao, Z. (2025). Research on Cluster Analysis and Pattern Recognition of Students' Sports Performance Data during Physical Education Instruction. J. COMBIN. MATH. COMBIN. COMPUT, 127, 9079-9095.

[15] Sheng, Y. (2025). The Relationship between Sports Measurement and Evaluation in Physical Education through Intelligent Analysis and Data Mining. International Journal of High Speed Electronics and Systems, 2540295.