International Journal for Housing Science and Its Applications
Publish August 6, 2025. Volume 46, Issue 3 Pages 6745-6756

https://doi.org/10.70517/ijhsa463579

Research on multivariate statistical analysis and optimization adjustment model of students' physical fitness data in college physical education teaching

Zhirong Zhao[1,*]
[1] Physical Education College, Luoyang Normal University, Luoyang, Henan, 471934, China
Corresponding authors: (e-mail: luoyangzzr@163.com).

Abstract With the deepening of physical education reform, students' physical fitness training faces problems such as non-standardized movements and single training method, and the use of data analysis and model construction can effectively improve the effect of physical fitness training and help improve the quality of physical education teaching in colleges and universities. In this study, principal component analysis is used to reduce the dimensionality of students' physical fitness data, and a training action detection model is constructed by combining time evolution modeling and multi-case learning to standardize students' physical fitness training actions. Sixty sophomore students from a university were divided into experimental and control groups, and the experimental group applied the action detection model designed in this paper to carry out physical fitness training. The results showed that: the detection accuracy of the three physical training movements (pull-up, 50-meter run, sit-up) were kept above 90%; after the experiment, the indicators of lung capacity (4383±661.29 ml) and sit-up (45±9.07) of the experimental group were significantly better than those of the control group (3317±1307.35 ml, 36±14.28); through the principal component comprehensive evaluation modeling, the comprehensive physical fitness qualities of the experimental and control groups were 2.37 and 1.75, respectively, with significant differences. The study confirms that the physical training movement detection model based on multivariate statistical analysis can effectively identify and correct students' training movements, improve the accuracy of training, promote the overall improvement of students' physical fitness level, and provide new ideas for the reform of physical education teaching in colleges and universities.

Index Terms college physical education, physical fitness data, multivariate statistical analysis, principal component analysis, training movement detection, optimization adjustment model

I. Introduction

Today, the government issued the document “China Education Modernization 2035”, which clarifies the general direction of education reform and puts the establishment of a high-quality education system at the core. In the construction of this system, accelerating the construction of smart schools and promoting the integration of smart education to promote the development of physical education have become important initiatives [1], [2]. This not only reflects a deep understanding of the importance of physical education in the overall education pattern, but also highlights the determination to use modern technology to improve the quality of physical education [3], [4]. With the fast-paced life of modern society and the popularization of information-based education, students tend to neglect physical exercise under the pressure of schoolwork and the burden of learning. And after the epidemic, students' immune system is lowered, leading to the decline of physical fitness and the increase of physical health problems [5]. Cultivating students' physical fitness quality through physical education in colleges and universities can not only improve students' physical health, but also improve students' learning and work efficiency, and enhance their physical and mental health [6], [7]. It can also cultivate students' perseverance and teamwork ability, cultivate students' will quality and positive enterprising spirit, and lay a solid foundation for students' future career development and social life [8]-[10]. It can be seen that the cultivation of students' physical fitness is of great significance in the teaching of physical education in colleges and universities.

However, with the continuous updating of educational concepts in the new period and the improvement of social requirements for talents, physical education teaching in colleges and universities is facing a situation of challenges and opportunities. The main challenge is that the electronic files of students in colleges and universities have not been fully established, the teaching assessment program is single and lacks multi-dimensional integration, and the error of manual records is difficult to ignore, which hinders the improvement of the effect of students' physical education teaching [11]-[14]. Moreover, collective teaching ignores the individual physical differences of students, leading to the rising risk of sports injuries, especially the lack of physical fitness science guidance related to special groups of students, and low effectiveness of physical fitness training [15]-[17]. Based on this, analyzing students' physical fitness data can provide guidance for students' scientific physical training.

Multivariate statistical analysis is a statistical method used to analyze the effects of multiple independent variables on one or more dependent variables at the same time. It can help researchers explore the relationship between multiple variables, predict the value of the dependent variable, perform factor analysis, etc. It can be used to analyze demographic data, survey data, psychological data, educational data, etc. [18]. In the field of education, multivariate statistical analysis methods are also widely used. It can be used to find the connection between students' performance and other variables through the method of educational data analysis, and through principal component analysis, to find the obvious relationship between students' performance and variables such as study time, family background, participation in extracurricular activities, etc., and to guide teachers to develop teaching plans, optimize the learning environment, and improve students' academic performance [19], [20]. This is an effective method for the analysis of students' physical fitness data, combined with the optimization and adjustment model, to adjust students' physical fitness and provide assistance for physical education teaching.

Physical education teaching is an indispensable and important part of the education system in colleges and universities, and plays an important role in cultivating students' physical health and comprehensive quality. In recent years, with the deepening of the reform of physical education teaching in colleges and universities, physical education teaching has gradually changed from focusing on skill teaching to comprehensive development of students' physical fitness. However, in the actual teaching process, due to the large differences in students' basic physical fitness, training movements are not standardized, training methods are single and other problems, resulting in poor physical training results, and it is difficult to meet the requirements of tailored teaching. Especially in the current information technology background, how to effectively use data analysis and artificial intelligence technology to accurately grasp the level of students' physical fitness, optimize training movements, and improve the training effect has become an important problem to be solved in college physical education teaching. Scholars at home and abroad have conducted a series of studies on the application of data analysis in physical education, mainly focusing on the statistical analysis of physical fitness data, the evaluation of training movements, and the prediction of training effect, etc. Wang et al. (2025) applied principal component analysis to the field of quality assessment, which provided a new way of thinking about data reduction and feature extraction.Uwamahoro et al. (2025) realized multivariate through principal component analysis of comprehensive evaluation, proving the effectiveness of the method in multivariate data analysis. Meanwhile, Ren et al. (2025) implemented 3D pose estimation and movement optimization of athletes using C3D technology, which provided technical support for sports training movement detection.Ziegler et al. (2025), on the other hand, proposed a deep multiple example learning framework to improve the detection accuracy of complex data. However, most of the existing studies focus on a single data analysis or movement detection technique, and lack a systematic study that combines multivariate statistical analysis of physical fitness data with training movement detection, especially the practical application research in college sports teaching is still insufficient.

Based on the above research background, this paper proposes a physical fitness training method based on multivariate statistical analysis and optimization adjustment model for college physical education teaching. Firstly, principal component analysis is used to reduce the dimensionality of students' physical fitness data and extract the main influencing factors; secondly, a training action detection model is constructed by combining time-evolutionary modeling and multi-case learning to standardize students' physical fitness training actions; and finally, the validity of the method is verified through comparative experiments. The specific research steps include: first, multivariate statistical analysis of college students' physical fitness data to determine the main influencing factors; second, constructing the key point dataset of physical fitness training movement postures to provide data support for movement detection; third, designing a training movement detection model based on time evolution modeling and multiple example learning to improve the accuracy of movement recognition; and fourth, verifying the optimization effect of the method on students' physical fitness training through experiments. This study will provide theoretical and technical support for physical education teaching in colleges and universities, and help to improve the science and effectiveness of students' physical training.

II. Multivariate statistical methods based on principal component analysis

The main idea of principal component analysis [21], [22] (PCA) is to map the N-dimensional features of the original data matrix onto a completely new M-dimensional data matrix $(M < N)$, and the mapped feature vectors are called principal components. This mapping process essentially selects the features with the vast majority of the variance, i.e., the new data matrix contains the vast majority of the information of the original data matrix. The data is two-dimensional data characterized by x_1, x_2, the data exists quite a lot of information in both dimensions, after mapping, it is changed to two-dimensional data characterized by y_1, y_2, at this time, because most of the information of the data exists in the y_1 features, while the data in the y_2 features relative to y_1 can be ignored, then at this moment can be discarded y_2 features to achieve the effect of data dimensionality reduction.

(1) Objective function

Suppose the input data is normalized to $X = [x_1, x_2, \dots, x_n] \in R^{m \times n}$, with m samples and n feature variables. Where $\vec{p}$ is the projection vector and $y_i = \vec{p}^T x_i$, the objective function of PCA is:

$$
\begin{aligned}
&J_{PCA}(\vec{p}) = \max\left(\frac{1}{n}\sum_i (y_i)(y_i)^T\right) \\
&s.t. \\
&\vec{p}^T \vec{p} = 1
\end{aligned} \tag{1}
$$

(2) Matrix decomposition

For the normalized data X, it can be decomposed according to the dimensionality reduction principle of PCA algorithm:

$$X = TP^T + E \tag{2}$$

where $P \in R^{n \times k}$ denotes the load matrix, and by decomposing the covariance matrix C, $T \in R^{n \times k}$ is the score matrix. E is the residual matrix. Covariance matrix decomposition is usually done in two ways: by eigenvalue decomposition and by SVD decomposition.

Algorithm 1. Eigenvalue based decomposition of covariance matrix algorithm

1) Normalize the original data matrix to obtain $X = [x_1, x_2, \dots, x_n] \in R^{m \times n}$.

2) Calculate the covariance matrix:

$$C = \frac{1}{m} XX^T \tag{3}$$

3) Find the first a eigenroots of the covariance matrix C: $\lambda_1 > \lambda_2 > \lambda_3 > \ldots > \lambda_a$, and a is the number of principal components, which a is the eigenvector of the eigenroot pairwise: $p_1, p_2, p_3 \cdots p_a$.

4) Calculate the principal component feature vector $t_i = p_i X^T$ after dimensionality reduction.

Algorithm 2. SVD based decomposition of covariance matrix algorithm

1) Normalize the original data matrix to obtain $X = [x_1, x_2, \dots, x_n] \in R^{m \times n}$.

2) The singular value decomposition of X is processed: $X = U \sum V^T$, which gives u_i, v_i, σ_i, $U = [u_1, u_2, \dots, u_m]$, $V = [v_1, v_2, \dots, v_n]$. Among them:

$$\sum = \begin{bmatrix} \sigma_1 & 0 & \cdots & 0 \\ 0 & \sigma_2 & \cdots & 0 \\ 0 & 0 & \cdots & \sigma_n \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & 0 \end{bmatrix} \tag{4}$$

3) Let $p_i = v_i$ to compute the principal component eigenvector $t_i = \sigma_i u_i$ after dimensionality reduction.

(3) Determine the principal components

Arrange all the eigenvalues in descending order to derive all the eigenvalues $\lambda_1 > \lambda_2 > \lambda_3 > \ldots > \lambda_n$, and the number of principal components is usually derived by the cumulative covariance contribution method (CPV):

$$\sum_{i=1}^{k} \lambda_i \Big/ \sum_{i=1}^{n} \lambda_i \times 100\% \geq \rho \tag{5}$$

where λ is the eigenvalue obtained by C eigendecomposition, and ρ represents the proportion of the cumulative contribution value of the principal component taken out to the contribution value of the total eigenvariable, and its value is generally 85%~95%.

III. Temporal evolutionary modeling and learning action detection models with multiple examples

In this paper, we propose a model framework for physical training action detection based on temporal evolutionary modeling and multi-example learning for correcting irregular actions during students' physical training in order to improve the effectiveness of students' physical training.

The structure of the training movement detection model framework is shown in Figure 1. Video clips whose temporal intersection and concatenation ratios (tIoU) with real action annotations exceed a certain threshold are selected as training samples using a sliding window approach, and then temporal evolution modeling and C3D feature extraction are performed on each training sample, followed by inputting the features into a deep multi-case learning network and training the network with a multi-task loss function. The network can not only classify video clips, but also regress video clips on front and back boundaries.

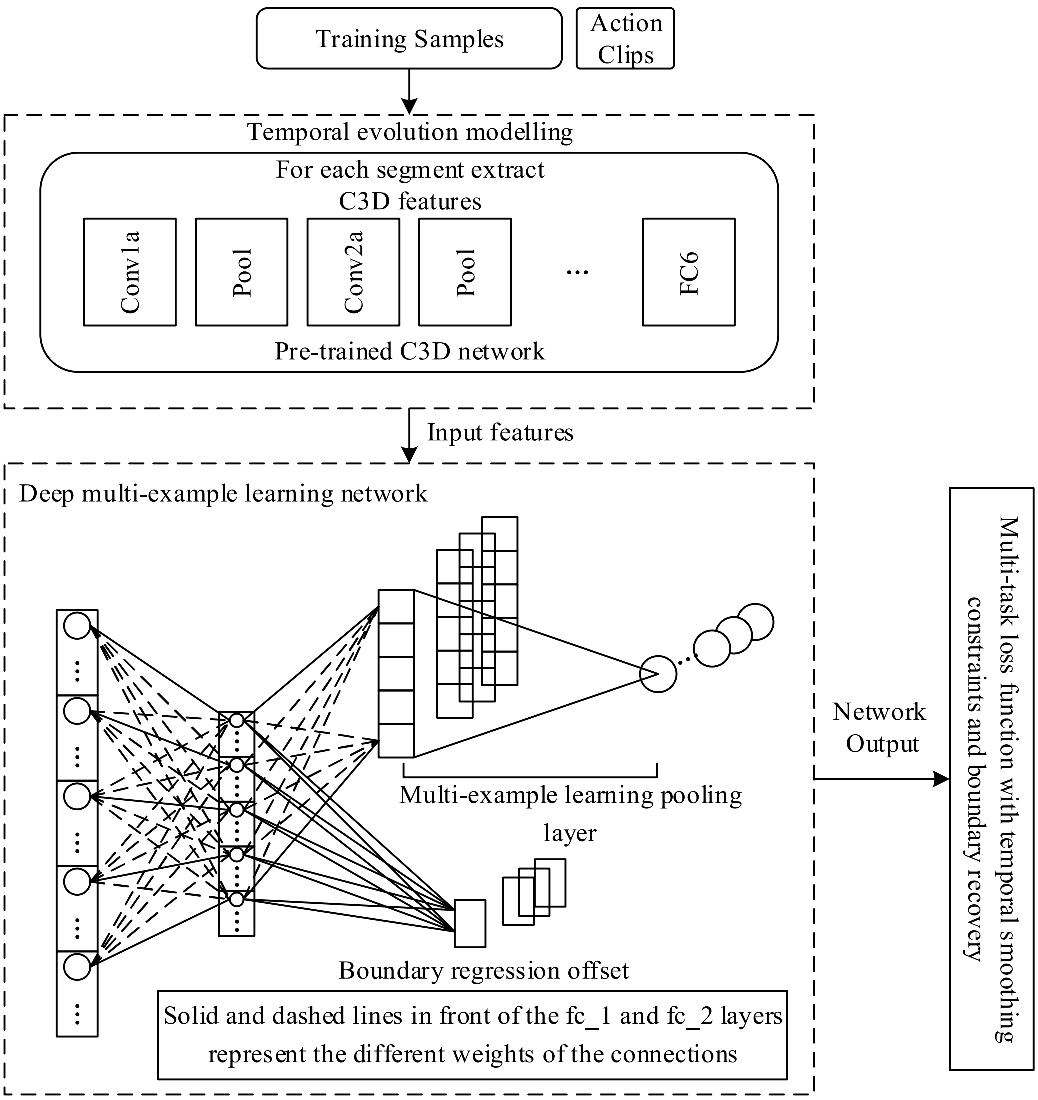

Figure 1: Training action detection model framework structure

(1) C3D feature [23] extraction

The C3D network has a total of eight convolution operations and five pooling operations, and its network structure is shown in Fig. 2. where all the convolution kernels are of size 3 × 3 × 3 with step size (1,1,1). In order not to shrink prematurely in timing, the size of the pooling kernel in the first layer is 1 × 2 × 2 and the size of the pooling kernel in the other pooling layers is 2 × 2 × 2. Finally the network obtains the final output after going through the fully-connected layer and the softmax layer twice. The input dimension of the network is (3,16,112,112) where 3 is the RGB three channels, 16 is the number of frames in the input image and 112 is the length and width of the image frame. Since this paper uses C3D network to extract features, the output of the fc6 layer is used as the feature value of the video clip.

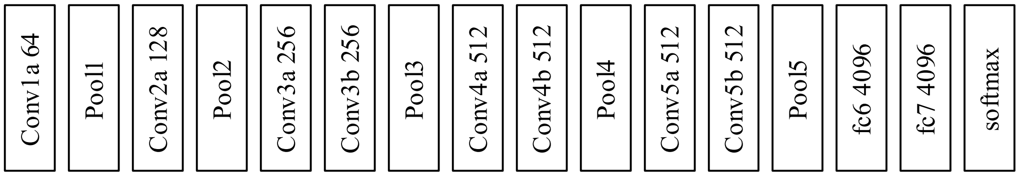

Figure 2: C3D network framework

(2) Temporal evolution modeling

In order to utilize the inherent temporal structure of actions, a simple and effective method is proposed to model action segments based on the temporal evolution of actions. Given an action segment X, it is uniformly divided into three parts, X_s, X_m, and X_e, which represent the beginning, middle, and end of the action, respectively. The features of these three parts are extracted as intrinsic features. In addition to the intrinsic features, the timing context of X is taken into account. X_i and X_r denote the part before and the part after the action segment X, respectively, which have a fixed length, and the features of these two parts are extracted as context features. Finally, concatenate the contextual and intrinsic features as the final feature representation of the action fragment f_X:

$$f_X = F(X_i) \| F(X_s) \| F(X_m) \| F(X_e) \| F(X_r) \tag{6}$$

Where □ denotes the concatenation of vectors and F denotes the feature extractor.

(3) Deep Multi-example Learning Network [24]

This paper proposes a deep multi-example learning network in which a multi-example learning pooling layer is used to output video clip action category probabilities based on the action probabilities of the examples in the video clips.

The proposed network consists of three fully connected layers (fc_1, fc_2, fc_3) and a multi-example learning pooling layer. The fc_1 layer is an intermediate layer, the fc_2 layer produces a temporal example score for each action category, and the fc_3 layer outputs the boundary offsets of the segments. Where the number of nodes in the fc_1 layer is 1000, the number of nodes in the fc_2 layer is $C{*}m$ (C represents the total number of categories, and m stands for the number of exemplars in each category), and the number of nodes in the fc_3 layer is $C{*}2$ (C also represents the total number of categories, and *2 is because there are offsets before and after the segments). Here, because the fc_2 layer is expected to produce temporal example scores, the temporal order of the input space needs to be maintained in the passing of the network, i.e., the features in each part of the fragment are encouraged to be more correlated with the corresponding part in the next network layer. Therefore, additional predefined weights were added before the fc_1 layer and before the fc_2 layer.

A global pooling function called the Noisy-and pooling function was used to generate action category scores for the segments. It assumes that a package is a positive sample when and only when the number of positive examples contained in the package exceeds a certain threshold:

$$P_c = g_c(\{p_j^c\}) = \frac{\sigma(a(p_{\bar{j}}^c - b_c)) - \sigma(-ab_c)}{\sigma(a(1-b_c)) - \sigma(-ab_c)} \tag{7}$$

where $p_{\bar{j}}^c = \frac{1}{|j|}\sum_j p_j^c$ and σ is the sigmoid function. The Noisy-and pooling function can activate the probability of the packet hierarchy P_C when the score $p_{\bar{j}}^c$ of the example hierarchy exceeds a certain threshold, a behavior that mimics the logics and functions in the field of probability. The parameters a and b_c control the shape of the activation function. b_c is a set of parameters learned in training to represent the appropriate threshold for each category c. a is also a parameter learned in training that controls the slope of the activation function. The terms $\sigma(a(1-b_c))$ and $\sigma(-ab_c)$ are used to normalize P_c.

(4) Loss function

The multitask loss function L for simultaneous training of classification and action boundary regression is:

$$L = L_{cls} + \lambda_1 L_{ts} + \lambda_2 L_{reg} \tag{8}$$

where L_{cls} is a standard multicategorical cross-entropy loss, L_{ts} is the temporal smoothing constraint, which adds a temporal smoothing constraint to the example scores by minimizing the scores of the two neighboring examples, and L_{reg} is the loss of the piecewise boundary regression. λ_1 and λ_2 are hyperparameters that are set empirically.

The temporal smoothing constraint is specified as:

$$L_{ts} = \frac{1}{N}\sum_{i=1}^{N}\sum_{c=1}^{C} l_i^c \left[\sum_{j}^{m-1}(p_{j,i}^c - p_{j+1,i}^c)^2\right] \tag{9}$$

where N is the batch size, C represents the total number of action categories, and l_i^c is the label. When the category corresponding to the i th training sample is c, then $l_i^c = 1$, otherwise, $l_i^c = 0$. m is the number of examples in each category, and p indicates the score of the examples.

The piecewise boundary regression loss function is specified as:

$$L_{reg} = \frac{1}{N}\sum_{i=1}^{N}\sum_{c=1}^{C} l_i^c [R(\hat{o}_{s,i}^c - o_{s,i}^c) + R(\hat{o}_{e,i}^c - o_{e,i}^c)] \tag{10}$$

where N, C, and l_i^c have the same meaning as above. $\hat{o}$ is the boundary regression estimated offset, o is the true labeled offset, R denotes the Manhattan distance, and the subscripts s and e denote the beginning and end of the segment, respectively.

IV. Data mining of physical fitness of high school students

Data mining method is adopted to analyze the multivariate statistical analysis of college students' physical fitness data.

(1) Data collection

The data used in this paper comes from a sports college, and the data includes: data collected from students' physical fitness test data. The dataset has a total of 11,392 entries. The dataset includes four grades of 2020, 2021, 2022 and 2023. Students' physical fitness data, the data of 2022 grade is taken as the sample of research data, the sample data is taken SQL Server to store and manage in this college, and synchronized with the students' personal information in the teaching system. After deleting the relevant personal privacy information and irrelevant fields, the final sample information includes the following fields: gender, height, weight, lung capacity, 50-meter run, 800 meters for girls, 1000 meters for boys, standing long jump, seated forward bending, 1-minute sit-ups for girls, pull-ups for boys, naked eye vision in the left eye, naked eye vision in the right eye, total scores and total scores of the grade.

(2) Data pre-processing

The data preprocessing process includes four steps: data selection, data cleaning, data integration and data normalization. Generally speaking, there is a large amount of incomplete, fuzzy and noisy data in the original data, and data preprocessing is to select and clean these data. Some data that do not have analytical value are cleaned up, and more appropriate data are selected for experimental analysis, so as to improve the efficiency and accuracy of data mining. In addition to this, data integration, data normalization and data statute can reduce the redundancy of data, maintain data integrity, so as to improve the level of data mining.

1) Data Selection

There are a large number of redundant and worthless fields in the original data, and synchronizing the information of these fields for data analysis will consume a lot of computing cycles and computing resources, and will produce errors in the experimental results. Therefore, data selection is the first step of data preprocessing. By observing the data information in the data collection and combining the objectives of the mining project, data selection is carried out to obtain more streamlined data information. So that the efficiency of data mining is improved, but also to discover the hidden internal connections and laws between data information and attributes.

2) Data Cleaning

As there is a large amount of missing, redundant, erroneous and noisy data in the data, it is necessary to perform data cleaning on the data. In practice, after getting the data, to understand its basic situation, to determine what data is unreasonable and then common data cleaning methods for cleaning. The actual operation is mainly to delete duplicate data, supplement incomplete data, correct erroneous data, etc., making the cleaned data standard, clean and continuous, in line with the requirements of subsequent calculations.

3) Data normalization

Data normalization is a basic operation of data mining, usually in the process of data mining processing, different features in the data set are inconsistent, the difference between the values is too large, which in turn affects the results of data processing. Therefore, it is necessary to normalize the data to facilitate data analysis.

4) Data integration

Data integration is an important step in the formation of a data warehouse, in the actual use of the process is often due to the application system to take different types of databases, it is necessary to extract between the various application systems, integration, unified transformation format, de-emphasis, merger and other operations, and finally imported into a unified database.

V. Construction of a dataset of key points for physical training exercise postures

Due to the specificity of the application scenario of the algorithm in this paper, which mainly involves three kinds of exercises, namely, pull-up, 50m and sit-up, the private human movement key point dataset is constructed by collecting the picture samples by itself and manually labeling them. At the same time, in order to improve the amount of data for training, the data are enhanced using image processing methods.

(1) Sample Collection

In this paper, the model is applied to the type detection of three kinds of sports: pull-up, 50m and sit-up, and the picture sample set needs to include the scenes of the above three kinds of sports, in order to converge to get the sports type detection model with good generalization performance and recognition effect. The picture samples are obtained by frame extraction and decoding from the motion video, and the time interval of frame extraction is set according to the degree of change of the actual video image to obtain the picture samples with high differentiation. In order to prevent the occurrence of overfitting problem, it is necessary to record videos under many different backgrounds. A picture de-duplication algorithm based on hash fingerprinting algorithm and Hamming distance is also implemented to set the similarity threshold according to the similarity of the actual pictures, and filter the images extracted from the videos with the same background to remove the pictures with higher similarity.

(2) Sample labeling

Sample labeling is carried out for all the picture samples obtained in the previous step, and this paper involves the detection of three types of sports, setting the pull-up label as 0, the 50m label as 1, and the sit-up label as 2. Since the picture samples of the three types of sports and the human body's key point data have been organized in three folders when the samples are collected, sample labeling in this paper directly adds the labels to the three folders of sample data uniformly. The labeling in this paper can be added uniformly to the sample data of the three folders.

(3) Data Enhancement

In order to solve the problem that the model is prone to recognition errors and poor generalization performance due to the small number of picture datasets, this paper adopts the data enhancement method to increase the amount and diversity of data, and at the same time balances the number of samples of the three types of motion to prevent overfitting phenomenon. Data enhancement is realized through image processing technology, and since this paper is feature extraction of human key point data, and the processing methods such as scaling and panning do not change the relative position of the key points, small-angle rotation, flipping, offset affine and perspective transformation methods are chosen. The model only classifies and detects three kinds of motions at present, and a total of 7743 image samples are finally constructed, which are divided into training set, validation set and test set in the ratio of 6:2:2.

VI. Example of principal component-based multivariate statistical analysis and optimization of physical training

VI. A. Multivariate statistical analysis of physical fitness data

In this paper, 60 sophomore students of a university, divided into class 1 and class 2, were used as research subjects to conduct multivariate statistical analysis of the physical fitness quality of pre-exercise students by means of statistical analysis methods (descriptive statistics, independent samples t-test, and principal component analysis, etc.). The specific experimental number of people was the sample size of 30 people (15 males and 15 females) in class 1 and 30 people (15 males and 15 females) in class 2. The study subjects who participated in the test were voluntary and in good health without any abnormality. The study mainly screened the height, weight, lung capacity, forward body flexion, 1000m, 800m, 50m, sit-ups and pull-ups from the most recent physical fitness test data of the students of both classes.

Table 1 shows the results of the comparison of the physical fitness data of the students in the two classes before the experiment. As can be seen from the data in Table 1, the data gap between the two classes in the physical fitness tests of height, weight, lung capacity, and forward body flexion was small and there was no significant difference. For example, the average height of class 1 and class 2 is 163.25±15.17 and 162.79±14.49cm respectively, $P=0.794>0.05$. The data of 1000m running of class 1 is 4.57±0.44min, and the data of class 2 is 4.49±0.41min, and the P is also greater than 0.05. In order to better analyze the comprehensive qualities of the two classes of students' physical fitness test, the following section adopts the method of principal component analysis for the multivariate analysis of physical fitness data. component analysis method for multivariate statistical analysis of physical fitness data.

Table 1: The results of the previous two students' physical data comparison results

Index	Variable	N	Mean	Sd	P
Height (cm)	Class 1	30	163.25	15.17	0.794
	Class 2	30	162.79	14.49	
Weight (kg)	Class 1	30	60.81	10.81	0.828
	Class 2	30	60.13	10.25	
Lung capacity (ml)	Class 1	30	3087	1171.43	0.705
	Class 2	30	3125	1098.46	
Predisposition (cm)	Class 1	30	13.62	9.42	0.753
	Class 2	30	13.29	8.87	
1000m (min)	Class 1	15	4.57	0.44	0.911
	Class 2	15	4.49	0.41	
800m (min)	Class 1	15	4.11	0.36	0.743
	Class 2	15	4.09	0.34	
50m (s)	Class 1	30	7.02	0.88	0.853
	Class 2	30	6.97	0.92	
Sit-ups	Class 1	15	33.00	12.13	0.882
	Class 2	15	33.33	13.46	
Lead up	Class 1	15	9.33	6.12	0.749
	Class 2	15	9.20	5.79	

The results of the total variance of the principal component analysis were analyzed by performing principal component analysis on each of the physical fitness quality indicators, as shown in Figure 3. By calculating the cumulative variance contribution rate, it was found that when the principal component score was 2, the cumulative variance contribution rate was 85.4%, which was higher than the threshold value of 85% generally used to determine the principal component score. Therefore, when the master achievement score is 2, it synthesizes most of the information from the results of the physical fitness test.

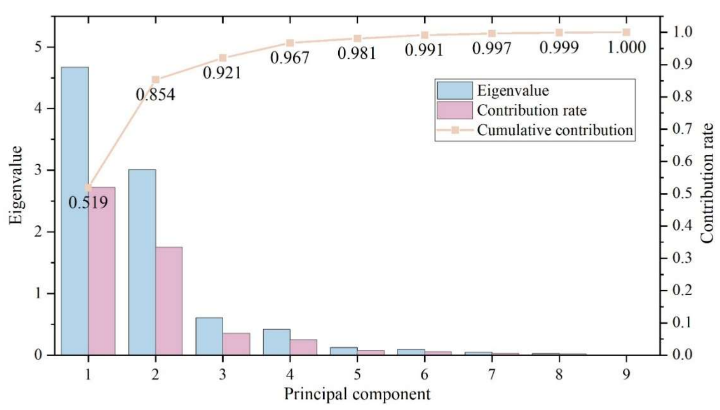

Figure 3: The total variable results of the main component analysis

In order to analyze the relationship of the main indexes in different principal components, the first principal component PC1 and the second principal component PC2 were subjected to load matrix rotation respectively, and the results of the load matrices of each component are shown in Table 2.

From the table, it can be seen that the information of lung capacity, body forward flexion, 1000m, 800m, 50m, sit-ups and pull-ups number are mainly synthesized in the first principal component, and PC1 can be named as the active quality factor. When PC1 is bigger the parameter of active quality factor is bigger. And the second principal component PC2 mainly synthesizes the information of height, weight and lung capacity, which can be named as inert quality factor, and when PC2 is bigger, the inert quality factor parameter is bigger. The contribution rate of the two principal components was used as the weight, respectively, to construct the principal component comprehensive evaluation model: PC = 0.712PC1 + 0.288PC2. where PC is the composite score of the students' physical fitness data, combined with the test results, we obtained a composite score of 1.59 for class 1 and a composite score of 1.57 for class 2, which indicates that the speed, strength, endurance, coordination, agility, flexibility and other physical fitnesses of the students of the two classes do not differ much. physical quality are not much different.

Table 2: The load matrix results of each group

Index	PC1	PC2
Height	0.145	0.452
Weight	0.162	0.493
Lung capacity	0.411	0.185
Predisposition	0.329	0.053
1000m	0.475	0.048
800m	0.387	0.046
50m	0.423	0.064
Sit-ups	0.459	0.054
Lead up	0.391	0.031

VI. B. Training movement localization and detection accuracy measurement

In order to validate the application performance of the physical training movement detection model based on time evolution modeling and multiple examples learning designed in this paper, the training movement localization and detection accuracy were measured for three exercises, namely, pull-ups, 50m and sit-ups. The relevant parameters for the experiment are as follows: the processor is Intel Core i5, the CPU is @2.20GHz, the operating system is Windows 10 64-bit, RAM/GB is 4, and the simulation software is Matlab.

Assisted virtual action localization refers to the virtual training image as the basis for localizing the training action in the image, which lays the foundation for the subsequent action detection. The results of the training action localization accuracy of the three sports are shown in Figure 4. From the results in the figure, it can be seen that the action localization accuracy of the three sports is relatively low compared to that of 50m running. This may be due to the fact that the training photo data of 50m running was acquired in an outdoor venue, which resulted in lower data quality due to uneven lighting and other reasons. However, the physical training action localization accuracy of the model in this paper always stays above 85%, and the highest can reach 97.3%. This fully demonstrates that the model designed in this paper is able to accurately localize the training actions in the physical training images.

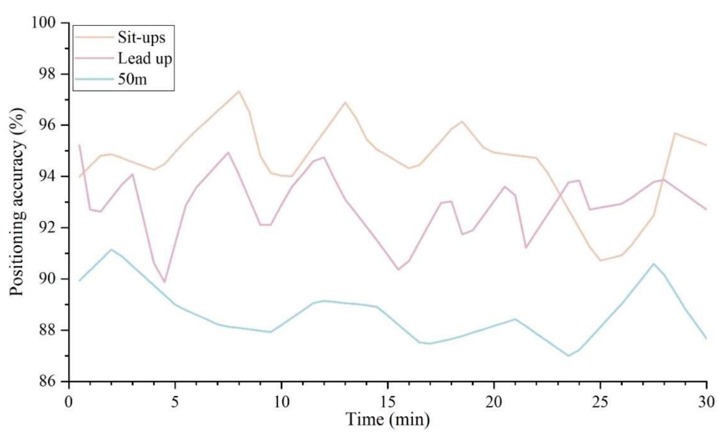

Figure 4: The comparison of the positioning accuracy of the three movements

The results of the comparison of the detection accuracy of the training movements can be more intuitively analyzed to the effectiveness of the model of this paper on the detection of training movements. The comparison results of the detection accuracy of three kinds of sports training actions are shown in Figure 5. Analysis of the detection accuracy results in the figure shows that: with the continuous increase in the number of virtual training images, the model of this paper for three kinds of physical training action detection accuracy are maintained at more than 90%, which can meet the needs of students' physical action standardization. The experimental results show that the action detection model designed in this paper has a better effect on the localization and detection of physical training actions. It shows that the technology has strong practical applicability and is applied to students' daily physical training.

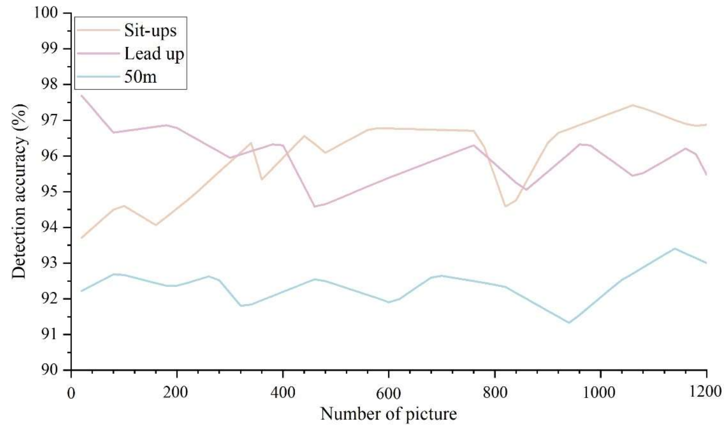

Figure 5: The comparison of the detection accuracy of the three movements

VI. C. Examination of the effect of optimization and adjustment of physical training

As can be seen from the above analysis, there is no significant difference between the comprehensive physical fitness quality of students in class 1 and class 2 of a physical education institute. 60 students from the two classes were taken as the research object to carry out the teaching experiment of physical fitness optimization training in physical education, eliminating the influence of gender differences on the experimental results. Class 1 was the experimental group, and the physical training movement detection model designed in this paper was integrated into the physical training to standardize the students' physical training movements. class 2 was the control group, and the traditional physical training method was used. After 2 months of physical training, the overall physical fitness quality of the two classes was compared to test the effect of the movement detection model on the optimization of students' physical fitness. The physical fitness test data were the same as above, including: height, weight, lung capacity, forward body flexion, 1000m, 800m, 50m, sit-up and pull-up scores.

Table 3: The results of the results of the students' physical data in the two classes

Index	Variable	N	Mean	Sd	P
Height (cm)	Class 1	30	163.91	14.96	0.827
	Class 2	30	163.13	14.04	
Weight (kg)	Class 1	30	57.44	11.23	0.683
	Class 2	30	59.87	10.62	
Lung capacity (ml)	Class 1	30	4383	661.29	0.000
	Class 2	30	3317	1307.35	
Predisposition (cm)	Class 1	30	17.19	7.55	0.017
	Class 2	30	14.23	9.21	
1000m (min)	Class 1	15	4.04	0.27	0.006
	Class 2	15	4.32	0.47	
800m (min)	Class 1	15	3.54	0.24	0.003
	Class 2	15	3.79	0.42	
50m (s)	Class 1	30	6.79	0.65	0.002
	Class 2	30	6.93	0.91	
Sit-ups	Class 1	15	45	9.07	0.002
	Class 2	15	36	14.28	
Lead up	Class 1	15	16	4.26	0.009
	Class 2	15	11	7.63	

Table 3 shows the results of the comparison of the physical fitness data of the two classes of students after the experiment. After the end of teaching, the physical training performance of the experimental group rose rapidly, and except for height and weight, which were not significantly different from the control group ($P > 0.05$), the rest of the indexes showed significant differences from the control group ($P < 0.05$). The average scores of lung capacity, forward body flexion, 1000m, 800m, 50m, sit-ups and pull-ups in the experimental group were 4383±661.29ml, 17.19±7.55cm, 4.04±0.27min, 3.54±0.24min, 6.79±0.65s, 45±9.07, and 16±4.26, and those in the control group were respectively were 3317±1307.35ml, 14.23±9.21cm, 4.32±0.47min, 3.79±0.42min, 6.93±0.91s, 36±14.28 and 11±7.63. Through the principal component comprehensive evaluation model constructed above: $PC = 0.712PC1 + 0.288PC2$, after teaching, the comprehensive physical fitness quality of the experimental group and the control group was 2.37 and 1.75, respectively, which showed a significant difference. As a result, the application of the physical training movement detection model designed in this paper helps to improve the comprehensive physical fitness quality of students.

VII. Conclusion

This study realizes the precision and science of students' physical fitness training in college physical education through multivariate statistical analysis and optimization adjustment model. The main conclusions are as follows:

The multivariate statistical method based on principal component analysis effectively extracted the main features of physical fitness data. It was found that when the number of principal components was 2, the cumulative variance contribution rate reached 85.4%, which exceeded the threshold value of 85%, proving that the method was able to synthesize most of the information of the results of the physical fitness test. Through the loading matrix analysis, the first principal component mainly represents the active quality factor and the second principal component represents the inert quality factor, which provides a scientific basis for physical fitness training.

The physical training action detection model based on time evolution modeling and multiple examples learning shows high reliability. The experimental results show that the model maintains a movement localization accuracy of more than 85%, up to 97.3%, for the three exercises of pull-up, 50-meter run and sit-up; the movement detection accuracy maintains more than 90%, which is able to effectively identify and correct irregular training movements.

The experiment proved the significant effect of this research method in improving students' physical fitness level. After two months of training, students in the experimental group outperformed the control group in lung capacity, forward bending, 1000-meter run, sit-ups and other indexes, and the comprehensive physical fitness quality score increased to 2.37, which was significantly higher than that of the control group (1.75), which indicated that the physical fitness training method based on multivariate statistical analysis and movement detection model could comprehensively improve the physical fitness level of students.

The method proposed in this study provides a new idea for the reform of physical education teaching in colleges and universities, which can not only accurately assess the physical condition of students, but also effectively standardize the training movements and improve the training effect, which is of great significance in promoting the quality of physical education teaching in colleges and universities.

References

[1] Demir, K. A. (2021). Smart education framework. Smart Learning Environments, 8(1), 29.

[2] Ma, J. (2024). Design and deconstruction of the intelligent system of college physical education in the era of 5G+ artificial intelligence. Journal of Artificial Intelligence Practice, 7(2), 86-96.

[3] Dudley, D., Mackenzie, E., Van Bergen, P., Cairney, J., & Barnett, L. (2022). What drives quality physical education? A systematic review and meta-analysis of learning and development effects from physical education-based interventions. Frontiers in Psychology, 13, 799330.

[4] Nazirjonovich, K. Z. (2021). The use of modern educational technologies in the organization of physical education is a guarantee to increase the effectiveness of education. ACADEMICIA: An International Multidisciplinary Research Journal, 11(10), 477-480.

[5] Xia, W., Huang, C. H., Guo, Y., Guo, M. G., Hu, M., Dai, J., & Deng, C. H. (2021). The physical fitness level of college students before and after web-based physical education during the COVID-19 pandemic. Frontiers in pediatrics, 9, 726712.

[6] Yang, Y., & Liu, W. (2021). THE INFLUENCE OF PUBLIC PHYSICAL EDUCATION CURRICULUM ON COLLEGE STUDENTS'PHYSICAL HEALTH. Revista Brasileira de Medicina do Esporte, 27, 83-86.

[7] Calestine, J., Bopp, M., Bopp, C. M., & Papalia, Z. (2017). College student work habits are related to physical activity and fitness. International journal of exercise science, 10(7), 1009.

[8] Yang, W. Z. (2024). Teamwork triumphs: fostering student development through cooperative learning in university physical education. International Journal of Current Educational Practice, 12(2), 114-120.

[9] Liu, D. (2024). Research on the Path of Implementing the Fundamental Task of “Cultivating Virtue and Building Character” in Public Physical Education in Zhejiang Province’s Universities. Contemporary Education Frontiers, 2(2), 54-60.

[10] Ho, W., Rafael, K., Ahmed, D., & Xie, J. Y. Y. (2024). Women in Physical Education and their Career Development: Current Trends and Issues: Women in Physical Education and their Career Development. International Sports Studies, 46(2), 50-63.

[11] Hsieh, Y. C., & Chen, T. H. (2019, May). Digital Archive Use in Physical Education and Sports Culture. In Archiving Conference (Vol. 16, pp. 120-124). Society for Imaging Science and Technology.

[12] Wang, R. (2017). Research on physical education problems and management reform of chinese colleges and universities. Eurasia Journal of Mathematics, Science and Technology Education, 13(10), 7057-7069.

[13] Zhang, S., Song, Z., Li, X., & Li, L. (2023). Physical Education and Health Teaching Behavior Realistic Dilemma and Optimization Path Research. Pacific International Journal, 6(4), 09-13.

[14] Liu, Y. (2020, September). College students' physical fitness test data analysis, visualization and prediction using data mining techniques. In Journal of Physics: Conference Series (Vol. 1631, No. 1, p. 012121). IOP Publishing.

[15] Andres, A. S. (2017). Physical education of students, considering their physical fitness level. Physical education of students, 21(3), 103-107.

[16] Barendrecht, M., Barten, C. C., Smits‐Engelsman, B. C., van Mechelen, W., & Verhagen, E. A. M. (2021). A retrospective analysis of injury risk in physical education teacher education students between 2000‐2014. Translational Sports Medicine, 4(5), 597-605.

[17] Chobotko, I. I., Skrypchenko, I. T., & Rozhechenko, V. M. (2023). Methods of organizing classes in special physical training in the system of physical education and sports. Publishing House "Baltija Publishing".

[18] Cantos-Gomez, P. (2019). Multivariate statistics commonly used in Multi-Dimensional Analysis. Multi-dimensional analysis: Research methods and current issues, 97.

[19] Cai, L., & Hansen, M. (2018). Improving educational assessment: multivariate statistical methods. Policy Insights from the Behavioral and Brain Sciences, 5(1), 19-24.

[20] Nkpordee, L., & Ibinabo, M. O. (2022). On the multivariate analysis of students' academic performance in WASSCE in public senior secondary schools in Rivers state (2018-2020). Journal of Mathematical Sciences & Computational Mathematics, 3(4), 441-491.

[21] Na Wang,Feng Li & Xiaodong Fan. (2025). Quality assessment of Bupleurum chinense by coupling Belousov-Zhabotinsky oscillation with principal component analysis. International Journal of Electrochemical Science,20(7),101033-101033.

[22] Jean Claude Uwamahoro,John Bosco Habarulema,Dalia Buresova,Nigussie Mezgebe Giday,Valence Habyarimana,Kateryna Aksonova... & Ange Cynthia Umuhire. (2025). Simultaneous evaluation of solar activity proxies during geomagnetic storms using principal component analysis: Case study of the African low and mid-latitude regions. Journal of Atmospheric and Solar-Terrestrial Physics,270,106477-106477.

[23] Fei Ren,Chao Ren & Tianyi Lyu. (2025). IoT-based 3D pose estimation and motion optimization for athletes: Application of C3D and OpenPose. Alexandria Engineering Journal,115,210-221.

[24] John Ziegler,Jaclyn F. Hechtman,Satshil Rana,Ryan N. Ptashkin,Gowtham Jayakumaran,Sumit Middha... & Ahmet Zehir. (2025). A deep multiple instance learning framework improves microsatellite instability detection from tumor next generation sequencing. Nature Communications,16(1),136-136.