Research Article

Research on Data Mining of College Students' Physical Health for Physical Education Reform

Lei Zhu¹ and Ling Zhang²

¹Wuhu Institute of Technology, Physical Education Department, Anhui 241000, China

²Wuhu Institute of Technology, Institute of Materials Engineering, Anhui 241000, China

Correspondence should be addressed to Lei Zhu; z_l@whit.edu.cn

Received 10 January 2022; Revised 11 February 2022; Accepted 23 February 2022; Published 7 March 2022

Academic Editor: Zhiguo Qu

Copyright © 2022 Lei Zhu and Ling Zhang. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

The traditional data mining method of students' physical health has some problems, such as low recall rate of data mining, long mining time, and poor mining accuracy. Therefore, this paper proposes a data mining method of college students' physical health for physical education reform. Using association rules to construct the correspondence between the fitness test data, the fitness test data can be classified and the data training model can be built. The decision tree of data attribute was built, and the physical health data was segmented by the segmentation technology. The information entropy of health data was calculated by the decision tree, and the information gain of health data sample set was obtained. The C4.5 algorithm was used to improve the ID3 algorithm. The improved decision tree was used to obtain the physique data splitting attribute, and the information gain rate was obtained by the ID3 algorithm correction. The *k*-means algorithm is used to divide the data into clusters, according to which the physical health data mining of college students is realized. Experimental results show that the recall rate of the physical health data mining method proposed in this paper is as high as 96%, the data mining time is only 3 s, and the accuracy of data mining is as high as 98%, indicating that the method proposed in this paper can improve the physical health data mining effect.

1. Introduction

With the advent of the era of big data, the massive data and powerful analysis ability of big data will provide more valuable possibilities for students' physical health promotion in concept and how to reveal hidden, previously unknown, and potentially valuable information from the massive fuzzy and noisy historical data in the practical application database [1–5]. In order to comprehensively analyze students' physical health data by using advanced technology, explore the characteristics of students' physical quality projects with different levels of physical health and provide more accurate scientific data support for students' physical health promotion needs by using the relevance, depth, and forward-looking thinking mode of big data, it provides a reference basis for the placement and implementation of college students' physical exercise service policy [6, 7].

Clarifying the concept of physical quality and physique and the relationship between them is the premise and basis

for understanding students' physical condition and formulating personalized physical health promotion plan. Physique refers to the quality of human body [8]. It is a relatively stable comprehensive feature of human body in terms of physical quality, morphological structure, sports ability, physiological function, and psychological development based on congenital heredity and acquired acquired. Its scope covers the development level of physique and physical fitness, physiological function, adaptability, and mental state [9–12].

Physical quality refers to the comprehensive ability displayed by various organ systems in sports, labor, and daily life under the regulation of the central nervous system. It is mainly reflected in the body abilities such as endurance, strength, speed, sensitivity, and flexibility. The ability depends on the physiological and anatomical characteristics of human muscles, the energy supply during muscle work, and the regulation of internal organs and nervous system [13, 14]. It can be seen that physique includes physical

quality. However, the physique promotion mentioned in this study actually refers to the promotion of physical quality, especially for the three basic qualities of human endurance, strength, and flexibility [15]. In other words, the term “physique” in this study refers to the combination of endurance, strength, and flexibility. It should be pointed out that both physique and physical quality are not only related to heredity but also closely related to the acquired environment, nutrition, physical exercise, and sanitary conditions. In particular, scientific and appropriate physical exercise can improve physical quality in all aspects and promote physical development [16].

At present, some progress has been made in the research on the physical health of college students at home and abroad, but there are also some difficulties. For example, in China, the promotion of students’ physical health has always been an aspect of the school’s close attention and development. In recent years, all schools have strictly implemented the students’ physical health test in accordance with the National Students’ Physical Health Standard. However, the massive data of physical health tests in most colleges and universities are only “stored” in the form of database, and few “stored” databases are mined and fully utilized [17]. According to the survey, most colleges and universities conduct simple statistical analysis and report when processing students’ physical health data, resulting in the “non benign cycle” of insufficient data mining, limited depth of data analysis, and relevant research unable to be used for health promotion [18–21]. For example, Japan has the most complete research data on the physique of college students. Its understanding is roughly the same as that of China, including morphological structure, psychological factors, physical quality, and sports ability, but it is only different in form and formulation. The Japanese are a combination of physical factors, physical factors, and mental factors. Physical factors mainly refer to the body’s physique, body shape, physical ability, and the ability to respond to and adapt to external environmental stimuli, while mental speed refers to some psychological factors, such as will, temperament, intelligence, and judgment. The United States began to pay attention to the system as early as the 1980s, and the content and indicators of physical fitness measurement in the United States continue to develop and progress with the development of social productivity, people’s living standards, and quality of life. The measurement of college students’ physical health data in the United States is mainly the measurement of physical system. With the continuous development of physical research, people’s cognition of physique also changes. Therefore, the contents and indicators of college students’ physical health measurement are also changing, mainly including cardiopulmonary function, muscle strength, endurance, physical flexibility, and body composition. However, foreign research on physical health data mining is insufficient. This led to limited follow-up research.

In recent years, a considerable number of scholars have studied the evaluation system of students’ physical health test data, such as Bartel’s research and development of college students’ physical health evaluation system and exercise prescription, Wang’s thinking based on regression and cor-

relation analysis, and the relationship between the weight and contribution rate of physical health evaluation indicators. This kind of research pays more attention to the evaluation methods and related analysis models of students’ physical health results, but there are still some limitations, such as a large amount of information contained in the database is buried, and the practical problems that can be solved are quite limited.

The current research does not conduct in-depth analysis on the relevance of students’ physical health data, which leads to problems such as low recall rate of data mining, long mining time, and poor mining accuracy. Therefore, this paper takes solving the problems existing in traditional methods as the research goal. This paper puts forward a data mining method of college students’ physical health for physical education teaching reform, which is conducive to database mining and improving data utilization, especially the research of scientific evaluation and analysis of data.

2. Construction of Association Rules for Health Data of Physical Fitness Test

When constructing the association rules of fitness test health data, take the initial fitness data collected by the hardware structure as the processing object and set $I = \{I_1, I_2, \dots, I_m\}$ as the data set of m test items [22]. At this time, the fitness test items are I_k , that is, the item set with length k . Suppose there is a health data subset T in the data set, at this time, the mapping of subsets on the test set can be defined as relevance, and the quantitative relationship of relevance can be expressed as

$$\sup(k \Rightarrow T) = \frac{\sigma(I_k \cup k)}{|D|}, \quad (1)$$

where $|D|$ represents the total number of transactions, and $\sigma(I_k \cup k)$ represents the number of health data [23–26]. Under the above quantitative relationship, the probability of health data in physical fitness test data is constructed, and the numerical relationship can be expressed as:

$$\text{conf}(k \Rightarrow T) = \frac{\text{sup_count}(I_k \cup k)}{\text{sup_count}(k)}, \quad (2)$$

where $\text{sup_count}(I_k \cup k)$ represents the event containing the data length, and $\text{sup_count}(k)$ indicates the number of physical test items containing health data. According to the above calculation formula, the probability value can reflect the proportion of college students’ physical fitness test health data in the total data set. When the probability data value is large, it can be used as the minimum confidence of mutual rules between health data. Take the minimum confidence as the screening criterion, repeatedly screen the data items in the data, and then test health data [27]. Continuously integrate health data into a health data set, set physical fitness test indicators for judging physical fitness, and gradually extract the rules that meet the requirements as the screening rules of health data, as shown in Figure 1.

FIGURE 1: Schematic diagram of corresponding relationship between physical fitness test and health data.

In Figure 1, after constructing association rules, classify the association results of physical fitness test data output to form a health data test group, which is divided into the training data group and verification data group [28–31]. Under the control of data proportion, build a data training model to measure the prediction data of the project, in order to control the accuracy of the classification function formed by the evaluation, the binary model is used to divide the algorithm model and the calculated data into different types of data. Under the condition of ensuring that they are all real data, the actual accuracy of the data training model can meet the requirements.

3. Data Mining of College Students' Physical Health for Physical Education Reform

For the course of physical education, schools have always paid little attention to the “tradition”. It is not surprising that some schools and education management departments lack the necessary attention to the specific implementation of physical health test [32]. For example, some schools even lie or hide the results of physical health test, and it seriously limits the guiding role of sports health testing data in strengthening physical education teaching reform of the school itself.

3.1. Information Gain Based on Decision Tree. The decision tree algorithm can obtain a decision tree with less instability through segmentation technology. Let S set species have s sample data and s category attributes with different values: $C_i (i = 1, 2, \dots, m)$ and then assume that the number of samples in class C_i is S_j . For this sample, the total information entropy is

$$I(s_1, s_2, \dots, s_m) = - \sum_{i=1}^m a_i P_i \log_2(P_i). \quad (3)$$

The probability that the sample belongs to C_i is expressed as P_i or s_i/s . Using the decision tree classification, assuming that the test attribute is A and A contains k different values $\{a_1, a_2, \dots, a_k\}$, the set S can be divided into k subsets $\{S_1, S_2, \dots, S_k\}$, all the samples with a value of a_j belong to the subset S_j , and S_j is the subset of the generated new leaf nodes [33]. Assuming that the number of samples with category G in subset S_j is S_{ij} , the information entropy of the divided samples is

$$E(A) = \sum_{j=1}^k a_j \frac{s_{1j} + s_{2j} + \dots + s_{mj}}{s} I(S_{1j}, S_{2j}, \dots, S_{mj}), \quad (4)$$

where $I(S_1, S_2, \dots, S_m)$ and P_{ij} represent the probability of samples with category attribute C in subset S_j . Finally, the information gain (Gain) of sample set S is

$$\text{Gain}(A) = I(S_1, S_2, \dots, S_m) - E(A). \quad (5)$$

According to the information entropy formula (5) of sample set S , the information gain $\text{Gain}(A)$ increases with the information entropy $E(A)$. The uncertainty of classifying the set with A as the test attribute will be reduced. There are k different attributes A in the set, corresponding to k subsets in the set S . The above steps are repeatedly called recursively, so as to generate other attributes as the child nodes of the node, and finally build a complete decision tree.

3.2. Based on C4.5 Calculation of Information Gain Rate of Algorithm. ID3 algorithm and C4 The main difference between 5 algorithms is that the judgment criteria selected for classification attributes are different. ID3 algorithm gains information, while C4.5 algorithm through information gain rate.

Split information for attribute A :

$$\text{Split}(A) = - \sum_{j=1}^m \frac{|S_j|}{|S|} \times \log_2 \frac{|S_j|}{|S|}. \quad (6)$$

Information gain rate of sample set is as follows:

$$\text{GainRatio}(A) = \frac{\text{Gain}(A)}{\text{Split}(A)}. \quad (7)$$

Pass C4.5 when the algorithm constructs the decision tree, and the splitting attribute of the current node is determined by the maximum information gain rate [34]. If the calculated attribute information gain rate becomes smaller, the later attribute with larger information gain rate is regarded as the split attribute.

C4.5: the improvement of the ID3 algorithm is as follows:

Step 1. The C4.5 algorithm discretizes the continuous features, so that it has the ability to deal with the continuous and discrete attribute categories.

In the second step, the information gain variable is introduced to correct the problem that the information gain in ID3 algorithm tends to be biased to the features with more values.

The third step is to deal with the missing samples [35].

In the fourth step, the regularization coefficient is introduced to prune the decision tree.

3.3. Physical Health Data Mining Based on the K-Means Algorithm. There are many methods to calculate the distance between physical health data objects. Euclidean distance is usually used to calculate the distance between each data object. For a given sample $D = (x_1, x_2, \dots, x_m)$, the K -means algorithm minimizes the square error of the cluster partition $C = \{C_1, C_2, \dots, C_k\}$ according to the class [36].

$$E = \sum_{i=1}^k \sum_{x \in C_i} \|x - u_i\|^2, \quad (8)$$

$$u_i = \frac{1}{|C_i|} \sum_{x \in C_i} x, \quad (9)$$

where x is the mean vector of cluster C_i . Equation (8) describes the compactness of the samples in the cluster around the cluster mean vector to a certain extent. The smaller the E value, the higher the similarity of the samples in the cluster. The algorithm steps are as follows:

Input: Sample set $D = (x_1, x_2, \dots, x_m)$; Number of clusters k ;

Process: Repeat.

Make C_i an empty set, $1 \leq i \leq k$.

For $j=1, 2, \dots, m$ do

Calculate the distance between the sample x_j and each mean vector u_i ;

The distance calculation formula is as follows:

$$d_{ij} = \|x_j - u_i\|^2. \quad (10)$$

The cluster marker of x_j is determined according to the nearest mean vector. The nearest mean vector function is expressed as

$$\lambda = \arg \min_{i \in \{1, 2, \dots, k\}} d_{ij}. \quad (11)$$

Classify the sample x_j into the corresponding cluster $C_{\lambda j}$:

$$C_{\lambda j} = C_{\lambda j} \cup \{x_j\}. \quad (12)$$

End for

For $j=1, 2, \dots, m$ do

Calculate the new mean vector u'_i ,

$$u'_i = \frac{x}{|C_i|}. \quad (13)$$

If $u'_i \neq x$, then

Update the current mean vector u_i to u'_i ;

Else

Keep the current mean vector unchanged;

End if

End for

Until the current mean vector is not updated;

Output: Partition cluster $C = \{C_1, C_2, \dots, C_k\}$;

According to the above classification process, the physical health data mining analysis of college students for physical education reform is realized.

4. Experiment

4.1. Data Preprocessing. Generally speaking, the original data contains too much noise data and incomplete data, which needs to go through the processes of data cleaning, data screening, data processing, and so on. Data cleaning is a process of re examination and verification of data. The purpose is to delete duplicate information, correct existing errors, and provide data consistency. For example, if the vital capacity in a record is "9999", which is obviously unconventional, screening conditions can be set to delete such data. Data filtering is to delete unnecessary fields according to actual needs. What this study needs is the relationship between the total score of physical health test and the impact degree of each test item. Filter the redundant fields such as year, grade number, class number, class name, ethnic code, and date of birth, leaving only useful fields. Data processing is to process data into a form that meets the requirements of the model according to the requirements of the specific data mining model.

According to the statistical results in Tables 1 and 2, Modeler 18.0 of SPSS is used, and 0 software is divided into male and female students. It analyzes the association rules of 7 items, four levels, and a total of 28 variables in the physical test results, including morphology, function, and quality. It excavates the deep relationship with more than 60% support and confidence among indicators and levels and presents them in the form of charts according to the relationship between the front and rear items.

4.2. Data Mining Results of College Students' Physical Health

4.2.1. Accuracy of Data Mining. In order to verify the accuracy of this method for college students' physical health data mining, reference [5] method, reference [6] method, reference [7] method, and this method are used to calculate the accuracy of physical health data mining. The results are shown in Figure 2.

Analysis of Figure 2 shows that there are significant differences in the accuracy of health data mining under different methods. When the amount of physique data mined in this paper is 20 GB, the data mining accuracy of the reference [5] method is 68%, the data mining accuracy of the reference [6] method is 63%, the data mining accuracy of the reference [7] method is 74%, and the data mining accuracy of this method is 99%. When the amount of physique data mined in this paper is 60 GB, the data mining accuracy of the reference [5] method is 33%, that of the reference [6]

TABLE 1: Score statistics of some items (male).

Project	Average value	Minimum value	Maximum	Standard deviation	Count
Total score	71.272	11	115	8.892	35 503
Vital capacity (mL)	4 357.37	1 004	6 574	861.278	35 503
50 m run score	77.781	0	100	11.704	35 503
1000 m run score	63.42	0	100	15.348	35 503
Standing long jump score	63.917	0	100	19.21	35 503
Sitting flexion score	69.484	0	100	15.795	35 503
Pull up score	37.039	0	100	31.350	35 503

TABLE 2: Statistics of scores of some items (female).

Project	Average value	Minimum value	Maximum	Standard deviation	Count
Total score	73.792	21.8	103.6	7.679	30 032
Vital capacity (mL)	2 903.63	729	9 272	660.29	30 032
50 m run score	69.329	0	100	12.639	30 032
1000 m run score	68.403	0	100	15.75	30 032
Standing long jump score	69.221	0	100	18.298	30 032
Sitting flexion score	70.416	0	100	16.449	30 032
Pull up score	63.462	0	100	13.966	30 032

FIGURE 2: Accuracy curve of physical health data mining.

method is 48%, that of the reference [7] method is 40%, and that of this method is 98%. With the increase of data volume, the data mining accuracy of the reference [5] method, reference [6] method, and reference [7] method has decreased significantly, while this method remains at a high level, which shows that the data mining accuracy of this method is better.

4.2.2. Time Consuming Data Mining. In order to verify the efficiency of this method on college students' physical health data mining, reference [5] method, reference [6] method, reference [7] method, and this method are used to calculate the time-consuming of physical health data mining. The results are shown in Figure 3.

FIGURE 3: Time consuming curve of data mining.

According to the analysis of Figure 3, when the amount of physical fitness data is 10 GB, the data mining time of the reference [5] method is 7 s, the data mining time of the reference [6] method is 17 s, the data mining time of the reference [7] method is 3 s, and the data mining time of this method is 0.5 s. When the amount of physical fitness data is 70 GB, the data mining time of the reference [5] method is 43 s, that of the reference [6] method is 42 s, that of the reference [7] method is 40 s, and that of this method is 3 s. This method takes less time to mine students' physical health data, which shows that the mining efficiency of this method is better than other methods.

FIGURE 4: Recall rate curve of health data mining.

FIGURE 5: Mean square error curve of physical health data mining.

4.2.3. Recall Rate of Data Mining. In order to verify the recall rate of this method for physical health data mining, reference [5] method, reference [6] method, reference [7] method, and this method are used to compare the recall rate of physical health data mining. The results are shown in Figure 4.

As can be seen from Figure 4, when students' physical health data mining is carried out, the trend of mining recall rate curve generally decreases with the increase of data volume. When the amount of data is 50 GB, the data mining recall rate of the reference [5] method is 33%, that of the reference [6] method is 35%, that of the reference [7] method is 50%, and that of the method in this paper is 96%. The data mining recall rate of the method in this paper is much higher

than that of traditional methods. This is because the method in this paper uses association rules to build the correspondence between the health data of the physical fitness test, classifies the association results of the physical fitness test data output, and forms the health data test group, which is divided into training data group and verification data group. Under the control of data proportion, the data training model is constructed to improve the recall rate of data mining and the effect of physical data mining.

4.2.4. Mean Square Error of Data Mining. In order to verify the accuracy of the method in this paper for physical health data mining of college students, reference [5] method and reference [6] method were selected from the above methods

FIGURE 6: Security of physical health data mining under different network attack intensities.

as the control group and methods in this paper as the experimental group. The mean square error analysis of data mining of physical health data mining under different methods was analyzed, and the results were shown in Figure 5.

Figure 5 shows that the errors of all physical health data mining methods are different under different methods. In the first experiment, the mean square error of the data mining method in reference [5] is 12, the mean square error of the data mining method in reference [6] is 8, and the mean square error of the data mining method in this paper is 1. In the tenth experiment, the mean method error of the data mining method in reference [5] is 13, the mean square error of the data mining method in reference [6] is 18, and the mean square error of the data mining method in this paper is 2. The data mining average method error of the proposed method is significantly smaller than that of the methods in reference [5] and reference [6], which indicates that the proposed method can achieve more accurate physical health data mining.

4.2.5. Data Mining Security. Based on the above experimental results, the safety of physical health data mining is verified by using the reference [5] method, reference [6] method, reference [7] method, and this paper. The results are shown in Figure 6.

As shown in Figure 6, the security of physical health data mining decreases with the increase of network attack intensity. When the network attack intensity is 20 dps, the health data mining security of the reference [5] method is 86%, that of the reference [6] method is 63%, that of the reference [7] method is 55%, and that of this method is 98%. When the network attack intensity is 80 dps, the health data mining security of the reference [5] method is 42%, the health data mining security of the reference [6] method is 38%, the health data mining security of the reference [7] method is 60%, and the health data mining security of the method in this paper is 96%. This shows that the method in this paper

can obtain high mining effect under high network attack and data mining security is high.

5. Conclusion

This paper proposes a data mining method of college students' physical health for physical education reform. Association rules were used to construct the correspondence between health data of physical fitness test, and the data of physical fitness test were classified. The attribute decision tree of physical fitness data was constructed, and the information gain was obtained by calculating the information entropy value of health data. The k -means algorithm is used to divide the data into clusters, according to which the physical health data mining of college students is realized. The following conclusions are drawn through the experiment:

- (1) When the amount of data mined in this paper is 60 GB, the data mining accuracy of the method in this paper is 98%, which always maintains a high level, indicating that the data mining accuracy of the method in this paper is high
- (2) When the data volume is 70 GB, the data mining time of the method in this paper is only 3 s, indicating that the mining time of students' physical health data is low and the mining efficiency is high
- (3) When the amount of data is 50 GB, the data mining recall rate of this method is 96%, which is much higher than that of traditional methods, indicating that this method can improve the effect of physical data mining
- (4) The error of all data mining methods in this paper is less than 2, which indicates that the method in this paper can achieve more accurate physical health data mining
- (5) When the intensity of network attack is 80 dps, the security of health data mining method in this paper is 96%. The method in this paper can obtain better mining effect under high network attack, and the data mining security is high

Data Availability

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

Conflicts of Interest

The authors declared that they have no conflicts of interest regarding this work."

Acknowledgments

This work supported by Project type: Provincial Massive Open Online Course (MOOC) Demonstration Project, Project Name: Volleyball, Project Number: 2019MOOC402.

References

- [1] J. Lin, L. Dakun, P. Wu, and W. Z. Xu, "Modeling and analysis of mental health early warning based on social data mining," *Electronic technology and software engineering*, vol. 21, no. 8, pp. 2–16, 2020.
- [2] L. Wu and Z. Liangliang, "Bridge structure health detection based on data mining technology," *Journal of Jilin University: Engineering Edition*, vol. 50, no. 2, pp. 7–21, 2020.
- [3] Z. Chonglin, S. Wang, H. Wang, and H. Dadao, "Physical health knowledge discovery of college students in physical exercise behavior stage based on association rule data mining," *Journal of Jinggangshan University: Natural Science Edition*, vol. 41, no. 3, pp. 5–18, 2020.
- [4] S. Zhaoming, Y. Jing, and R. Dong, "Data mining: C5 application of 0 decision tree algorithm in physical analysis of students in police colleges," *Modern Electronic Technology*, vol. 32, no. 1, pp. 52–55, 2021.
- [5] G. Shuling, "Analysis on the influence of pattern rope skipping on physical health of college students majoring in physical education," *Journal of Jiujiang University: Natural Science Edition*, vol. 34, no. 4, pp. 3–15, 2019.
- [6] P. Yan, "Thoughts and Enlightenment on the application of data mining in students' physical health test in China," *Science and Technology Information*, vol. 15, no. 22, pp. 199–200, 2017.
- [7] X. Luo, L. Yaqi, and Z. Zhongxiang, "Comparative study on physical fitness test of college students based on attribute correlation analysis," *Journal of Longdong University*, vol. 24, no. 5, pp. 106–109, 2021.
- [8] Z. Chonglin, L. Yu, and W. Wu, "Application of association rule data mining technology in physical fitness test analysis," *Journal of Shanghai Institute of Physical Education*, vol. 36, no. 2, pp. 42–44, 2020.
- [9] J. Tingting, L. Xiaozuo, and F. Linjie, "Design and implementation of physical health management system," *Electronic Design Engineering*, vol. 23, no. 4, pp. 44–46–44–50, 2015.
- [10] B. Zhu, "Application analysis of middle school students' physical health data in physical education," *Sports Fashion*, vol. 18, no. 10, pp. 101–103, 2019.
- [11] H. M. Ni, Y. M. He, and X. M. Yang, "Optimization model research on major underlying factors in the subhealth condition evaluation in 1 City and 7 provinces in China," *World Journal of Traditional Chinese Medicine*, vol. 5, no. 2, pp. 17–26, 2019.
- [12] F. Deng and W. Shaoyi, "Research on teaching countermeasures based on the analysis of college students' physical health test data," *Contemporary Sports Science and Technology*, vol. 10, no. 2, pp. 129–131, 2020.
- [13] P. Yongjie and H. Xia, "Analysis on the changes of physical health indexes of female college students in Shaanxi Province – Taking 2000–2010 as an example," *Sports Research and Education*, vol. 28, no. 2, pp. 125–128, 2021.
- [14] H. Zhiyu, "Discussion on the influence of tennis elective course on college students' physical health," *Heilongjiang Science*, vol. 11, no. 17, pp. 3–18, 2020.
- [15] L. Genping, "Influence of college sports intervention measures on college students' physical health," *Journal of Taiyuan Urban Vocational College*, vol. 17, no. 12, pp. 98–100, 2021.
- [16] M. Ma, L. Hongxia, and Y. Yang, "Comparative analysis of physical health test of Uyghur, Kazakh, Mongolian and Han college students in Xinjiang," *Science and Technology Information*, vol. 18, no. 34, pp. 203–205, 2011.
- [17] L. Xinfeng, "Analysis on influencing factors of physical health of college students – taking Henan Province as an example," *Journal of Nanyang Normal University*, vol. 11, no. 3, pp. 99–102, 2012.
- [18] L. Bo and G. Yang, "Research on the development strategy of college students' physical health based on big data," *Bonding*, vol. 41, no. 3, pp. 82–85, 2020.
- [19] C. Chuan, Y. Yang, and Y. Xinguo, "Effect of Taijiquan combined with traditional Chinese medicine tea on physical health level of Obese College Students," *Chinese Journal of Gerontology*, vol. 39, no. 17, pp. 4235–4238, 2019.
- [20] L. Mengyuan and X. Liqing, "Simulation Research on intelligent distribution of College Students' dormitory based on improved genetic algorithm," *Computer Simulation*, vol. 37, no. 7, pp. 224–228, 2020.
- [21] L. Zhu, H. Yuling, and J. Lanxin, "Investigation on sports injury in physical health test of college students," *Sports Science and Technology Literature Bulletin*, vol. 28, no. 4, pp. 62–63, 2020.
- [22] P. Yongjie and H. Xia, "Analysis of physical health indexes of college students with abnormal height and weight in Shaanxi Province," *Contemporary Sports Technology*, vol. 15, no. 26, pp. 2–13, 2021.
- [23] H. Yang, Y. Luo, X. Ren et al., "Risk prediction of diabetes: big data mining with fusion of multifarious physical examination indicators," *Information Fusion*, vol. 75, no. 21, pp. 140–149, 2021.
- [24] L. J. Deborah, S. C. Rajkumar, and P. Vijayakumar, "Medical decision support system using data mining," *Science Direct*, vol. 17, no. 32, pp. 55–66, 2021.
- [25] B. Robson and S. Boray, "Studies in the use of data mining, prediction algorithms, and a universal exchange and inference language in the analysis of socioeconomic health data," *Computers in Biology and Medicine*, vol. 112, no. 21, pp. 103369–103375, 2019.
- [26] T. Zhang, "Data mining can play a critical role in COVID-19 linked mental health studies," *Asian Journal of Psychiatry*, vol. 54, no. 2, pp. 102399–102407, 2020.
- [27] C. Eseab, D. Co, E. Mhm, and G. LLF, "Data mining to transform clinical and translational research findings into precision health-science direct," *Artificial Intelligence in Precision Health*, vol. 19, no. 23, pp. 149–173, 2020.
- [28] Y. Liu, E. Benjamin, E. Pau et al., "Epi graph DB: a database and data mining platform for health data science," *Bioinformatics*, vol. 16, no. 17, pp. 23–38, 2020.
- [29] C. Dai, "School health program: impacting physical activity behaviors among disadvantaged students," *Journal of School Health*, vol. 89, no. 6, pp. 468–475, 2019.
- [30] C. L. Zhong and Y. L. Li, "Internet of things sensors assisted physical activity recognition and health monitoring of college students," *Measurement*, vol. 159, no. 21, pp. 107774–107787, 2020.
- [31] C. Wang, "Analysis method of college student physical education quality based on big data analysis," in *The International Conference on Cyber Security Intelligence and Analytics*, pp. 543–557, Springer, Cham, 2020.
- [32] S. S. M. Ajibade, N. Bahiah, B. Ahmad, and S. M. Shamsuddin, "An investigation on educational data mining to analyze and predict the student's academic performance using

- visualization,” *Information Systems Design and Intelligent Applications*, vol. 173, no. 21, pp. 69–78, 2019.
- [33] J. Corral, S. T. Konstantinidis, and P. D. Bamidis, “Learning analytics, education data mining, and personalization in health professions education,” *Digital Innovations in Healthcare Education and Training*, vol. 19, no. 25, pp. 54–65, 2021.
- [34] L. Liping, C. Shanping, X. Pan, and C. Yinglan, “Undergraduates’ attitudes toward National Student Physical Health Standard,” *Shandong Sports Science & Technology*, vol. 41, no. 2, pp. 26–30, 2019.
- [35] G. Ramaswami, T. Susnjak, A. Mathrani, J. Lim, and P. Garcia, “Using educational data mining techniques to increase the prediction accuracy of student academic performance,” *Information and Learning Sciences*, vol. 120, pp. 451–467, 2019.
- [36] L. Z. Javier, J. Torralbo, and R. Cristobal, “Early prediction of student learning performance through data mining: a systematic review,” *Psicothema*, vol. 33, no. 3, pp. 456–465, 2021.