College Students' Sports and Psychological Integrated Management Platform Based on Multi-Source Health Monitoring

Bing Yi[1], Xueting Zhang[1*], Pu Liu[1], Guibo Liu[1]

Abstract
The perfect multi-source health monitoring system is essential to effective enhancement of the physical and mental health of college students. This paper innovatively applies the analysis of multi-source health monitoring data to the integrated management of college students' sports and psychology and designs the functions of the integrated management platform. Firstly, a college students' sports and psychological integrated management platform was established, along with its functional modules and platform framework. Next, the sensing approach for multi-source health data was introduced, and data sensing was carried out to obtain the data about the physiological and mental health of college students. On this basis, the integrated health of college students was predicted, the common diseases of college students were diagnosed based on electronic medical records (EMRs), and the emotions of college students were regulated through personalized recommendation of sports. Finally, an outlier detection method was given for multi-source health monitoring data. To verify its performance, the proposed approach was applied to derive the psychological states of subjects through empirical modal decomposition of multisource health data and recommend personalized sports to subjects. The results confirm the effectiveness of multi-source health monitoring and analysis in the integrated management of college students' sports and psychology.

Keywords: multi-source health monitoring; sports; mental health; integrated management; college students

1. Introduction

College students are a pool of talents in each country. It is important for them to maintain a strong body and a healthy mind (Chu & Yin, 2021; Toscos et al., 2018; Zhang, 2021). The perfect multi-source health monitoring system is essential to effective enhancement of the physical and mental health of college students (Firdausi et al., 2021; K. Li & Yu, 2021; Muthusamy & Durairaj, 2019; Rahaman et al., 2019; Yan et al., 2019). In recent years, many high and new technologies have been introduced to the field of multi-source health monitoring, thanks to the boom in big data analysis, cloud computing, artificial intelligence (AI), and wireless sensing network (WSN). Many research institutes, and colleges have stepped up the investment into the research and development (R&D) of medical aid and monitoring systems, trying to find a more convenient way to solve the various physiological and psychological problems of college students (Bouida et al., 2020; Ha et al., 2018; Nouioua et al., 2020; Ornek et al., 2020; Sucena et al., 2017). G. Li (2021) adopted a symptom self-assessment scale to test the mental health of college students, and analyzed the correlation between their mental health and interest in sports, revealing that the relationship between the two factors is a U-shaped curve instead of a rising straight line. Jingjing et al. (2020) established a sports health model according to the health state of female college students, which analyzes weight from caloric intake and caloric consumption. The model lays the basis for the design of subsequent health models for college students. Chen et al. (2020) proposed a visual analysis system to enhance the relevance between different classes of questionnaires and reduce the uncertainty of mental health analysis; a Circos view was designed to visualize the individual answers and associations between multi-source questionnaires, such that the users can easily identify the students with uncertain results. Hu et al. (2018) surveyed the contribution of sports intervention on college students, and observed and compared the health check indices of college students before and after the intervention. Yarong (2017) analyzed and summarized the status quo of sports fitness and the physical education (PE) reform in China, and obtained the statistics on the PE teaching reform under sports fitness development. Gang (2017) probed into the factors affecting the physical health of college students, and highlighted the importance of development of college students' physical health, providing the certain theoretical basis and scientific support for the development of college students' physical health. Considering the decline of physical health among Chinese college students, Yang (2017) explored how to improve the individual physical health of college students from the angle of PE reform, and proposed two options for PE and a strategy to individualize sports therapies by analyzing the theories and methods of

[1] School of Physical Education, Beihua University, Jilin 132000, China.
E-Mail: zxtbb11@126.com

personalized PE teaching.

The existing research has shown the room for development in the sharing and exchange of college students' physical and psychological monitoring data, and the mining of such data with machine learning. Therefore, it is promising to integrate big data analysis into the physical and psychological monitoring of college students. In most cases, the multi-source health monitoring data are applied to track and test the physiological state of humans. There is no application to the physical and mental health management of college students, including mental state monitoring, emotional fluctuation tacking, and sports intervention. Some scholars have constructed platforms for national fitness monitoring data standardization and service application. Referring to the architecture and construction mode of the existing platforms, this paper designs a comprehensive management platform for college students' sports and psychology based on multi-source health monitoring and discusses reasonable application scenarios of the platform. The main contents are as follows: (1) Setting up a college students' sports and psychological integrated management platform, along with its functional modules and platform framework; (2) Introducing the sensing approach for multi-source health data, and sensing the data about the physiological and mental health of college students; (3) Predicting the integrated health of college students, diagnosing the common diseases of college students based on electronic medical records (EMRs), and regulating the emotions of college students through personalized recommendation of sports; (4) Detecting outliers in multi-source health monitoring data. The proposed platform was proved feasible through experiments.

2. Platform Construction

The proposed integrated management platform of college students' sports and psychology mainly detects the health of college students in different age groups and disciplines and provides them with suggestions on physiological and mental health. Every user on the platform has his/her own account and can easily learn the variation in his/her health state by checking the monitoring results on his/her health information in various aspects. The platform can offer the sports training scheme to each student, according to the evaluation results on individual psychological and physiological health. Based on the scheme, each college student can actively participate in physical exercises, enhance his/her physical fitness, and reduce the risk of common diseases, thereby easing the psychological problems to different degrees. Figure 1 shows the functional modules of the integrated management platform of college students' sports and psychology.

The functional design of the platform considers three types of people on two levels, namely, friends, emergency contacts, and parents, as well as college students using the platform. Emergency contacts might overlap with the other types of people. Figure 2 shows the relationship between college students using the platform.

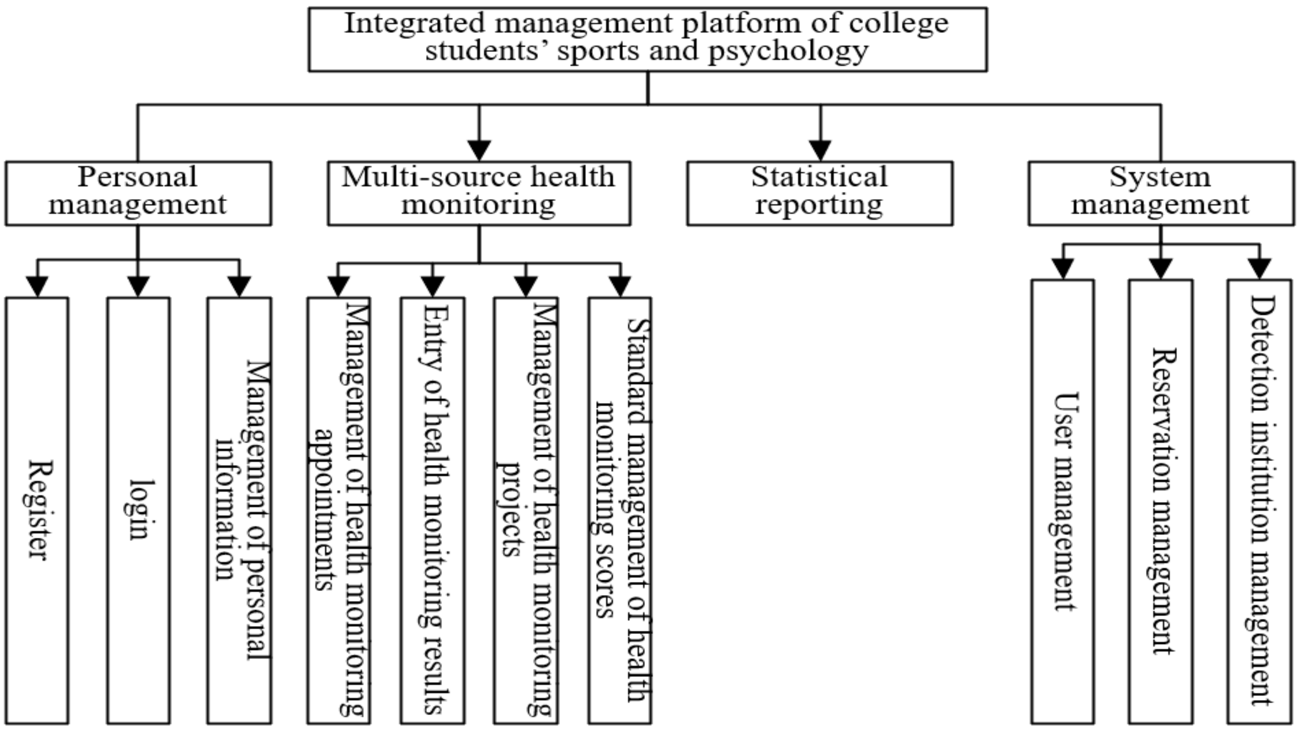

Figure 1. Functional modules of the integrated management platform of college students' sports and psychology

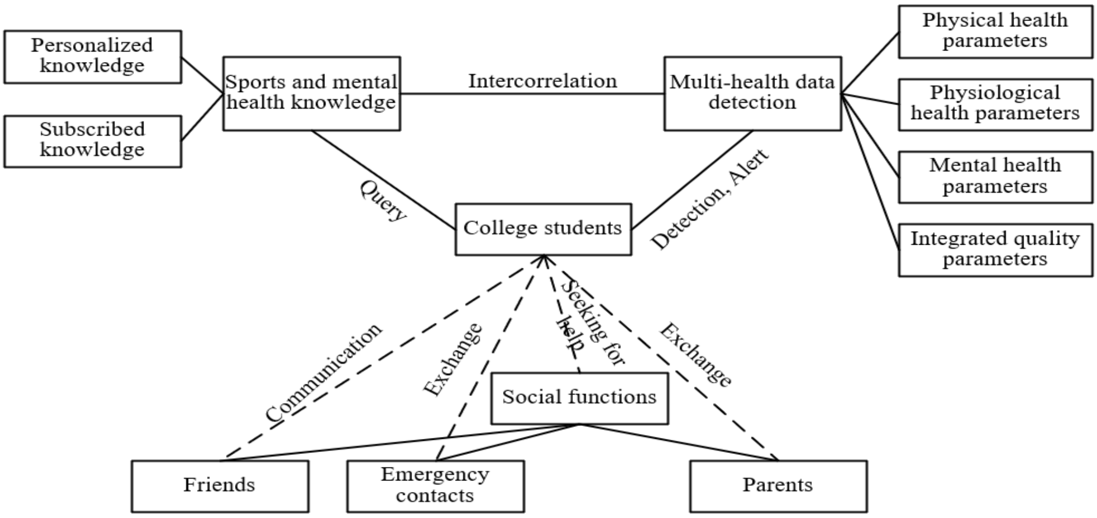

Figure 2. Relationship between college students using our platform

The Android operating system has many advantages: easy to development, rich in hardware, and free and open-source codes. Therefore, our integrated management platform adopts Android-based intelligent terminals, which connect multi-source health information acquisition devices mainly via the Internet and Bluetooth. The basic architecture of Android system is shown in Figure 3.

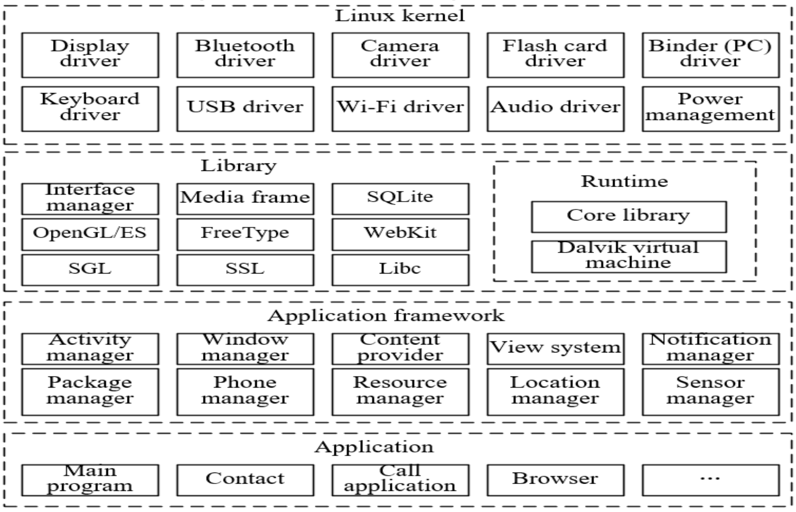

Figure 3. Basic architecture of Android system

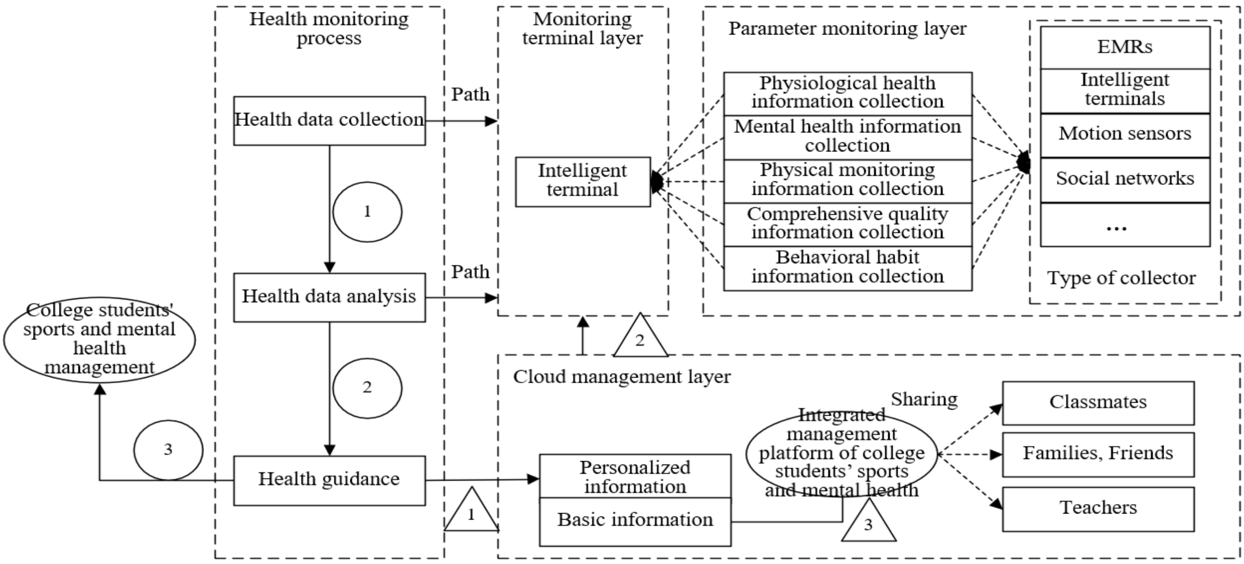

Figure 4. Framework of our integrated management platform.

The above analysis mainly discusses the working mechanism and selects the software basis of our integrated management system. Firstly, the multi-component platform needs to realize real-time monitoring of college students, with the aid of acquisition devices of physical fitness, medical examination, and psychological information. For this purpose, a basic parameter monitoring layer was designed for the platform. The latest information monitored by this layer was transmitted to the cloud management layer via intelligent terminals for further fusion, analysis, storage, and output of the data. Next, the personalized information and basic information of college students were combined to judge if the monitored information is abnormal, and guidance of physical exercise was provided proactively to college students, in order to realize sports intervention of college students' physiological and mental health. Figure 4 shows the framework of our integrated management platform. It can be observed that the entire platform consists of three layers, namely, parameter monitoring layer, health monitoring terminal layer, and cloud management layer, which are interconnected via the Internet.

3. Multi-Source Health Data Analysis

This paper first introduces the ways to sense multi-source health data, and establishes data perception from such aspects as college students' physiological health data and their mental health data. The obtained data provide the basis for subsequent integrated management of college students' physical health, and the analysis on their mental health state.

3.1 Integrated prediction of college students' health

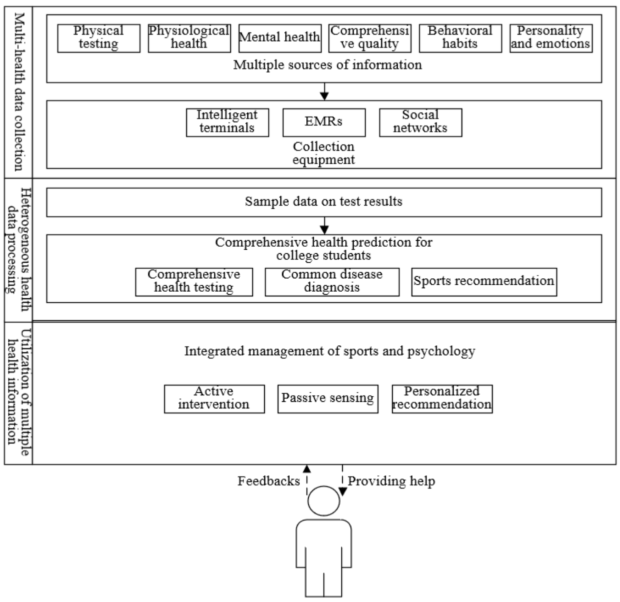

Figure 5. Framework of college students' comprehensive health prediction system

For the college students' comprehensive health prediction system, the prediction function is mainly completed through three steps: collection, processing, and utilization. Figure 5 presents the framework of the system. Specifically, the collection of multi-source health data aims to acquire the physical fitness, physical examination results, psychological test results of college students. The processing of multi-source health data mines the personal health information from the collected original data. Based on the valuable mined information, the utilization of multi-source health data tries to provide information query or recommendation services. The three steps form an organic whole.

This paper predicts the comprehensive health of college students with a model based on the decision tree (DT) algorithm. A tree structure was set up by the algorithm, according to the collected samples about the physical fitness, physical examination results, psychological test results of college students, where A1-A4 represent physical health feature, physiological health feature, mental health feature, and comprehensive quality feature, respectively. Each internal node of the DT stands for the test on a feature, and each leaf node on the tree represents a class. The DT was constructed by classifying the collected samples by feature. The DT can be described by purity entropy CD:

$$CD(E) = -\sum_{i=1}^{M} T(i) * log_2 T(i) \tag{1}$$

where, M=2 for the college students' health test samples need to be allocated into two classes (0 means unhealthy; 1 means healthy). Let D_i be the number of samples in the test set E, which belong to a class; D be the total number of samples in E. Then, $T(i)$ can be calculated by:

$$T(i) = \frac{D_i}{D}, i = 0,1 \tag{2}$$

Substituting formula (2) to formula (1), the purity entropy EN of sample set E can be obtained. The EN value is positively correlated with the difficulty in classifying E. Since the DT is binary, the value domain $U(G)$ of feature $A=\{A1, A2, A3, A4\}$ normally contains two subdomains. For the two sample sets divided by the separation value, the number of samples is denoted as $D(u_j)$, the total number of samples in E as D, and the purity entropy of each set obtained by formula (1) as $CD(E_{uj})$. Then, the information gain $ZY(E, G)$ can be defined as the purity entropy EN difference of feature G relative to sample set E:

$$ZY(E,G) = CD(E) - \sum_{u\in U(G)} \frac{D(v_j)}{D} * CD\left(E_{u,j}\right) \tag{3}$$

Formula (3) makes it possible to obtain the information gain of feature G under a specific value domain. During the construction of the DT, it is necessary to select the split feature for each node. By comparing the information gain between features, the feature with the highest information gain can be selected as the split feature:

$$TZ = \underset{G}{argmax}\{ZY(E, G)\} \quad (4)$$

Based on the split feature TZ obtained by formula (4), the DT algorithm can recursively complete the DT for the physiological health prediction of college students. Given the data on the features of college students (e.g., physical fitness, mental health, and comprehensive quality), the DT can predict the comprehensive health of college students, providing a reference for the integrated management and intervention of their PE and psychology.

3.2 EMR-based diagnosis of common diseases

Table 1.

Data size changes in the processing of EMR data

Steps		Data size changes
Original records		Number of EMRs: 14213; Inspection feature records: 841107
Irrelevant data processing		Number of EMRs: 8578; Inspection feature records: 792252
Irrelevant feature processing	Data integration	Number of EMRs: 8578; Inspection feature records: 351
	Removal of invalid features	Number of EMRs: 8578; Inspection feature records: 38
	Removal of repetitive data	Number of EMRs: 8540; Inspection feature records: 37
Data labeling		Medical records of common diseases: 512; Medical records of normal people: 9847
Data transformation		Number of EMRs: 8540; Expanding the number of features per record to 55

The diagnosis of common diseases among college students was modeled based on the EMRs of college students in a Grade A Class 3 municipal hospital, which used to cooperate with our research team. Table 1 shows the data size changes as the EMR data were processed through data integration, feature processing, data transformation, etc. In the end, 8,540 valid EMRs were obtained, each of which records 55 different features. Among them, 512 EMRs are about common diseases, and 9,847 are about normal people.

Based on the statistical results of college students' physiology and psychology monitored by an authentic hospital, machine learning was introduced to predict and prevent common diseases among college students. Three common disease diagnosis models were established for college students, based on DT, support vector machine (SVM), and artificial neural network (ANN). The diagnosis results of the three models were integrated by ensemble method to optimize the prediction effect.

The Gini index (GI)-based DT algorithm was adopted for modeling. The purity of college students' EMR dataset C can be measured by GI (C):

$$GI(C) = \sum_{l=1}^{|b|} \sum_{l' \neq l} T_l T_{l'} = 1 - \sum_{l=1}^{|b|} T_l^2 \quad (5)$$

Formula (5) shows, when two random samples are selected from C for classification, the probability that the two samples belong to different classes is $GI(C)$. The smaller the $GI(C)$, the simpler the classification, and the purer the dataset.

For a single feature g, the GI can be defined as:

$$g^* = \underset{g \in G}{argmin}\ GII(C, g) \quad (6)$$

In the entire candidate feature set G, the classification by feature g with the smallest $GII(C, g)$ can be selected as the optimal classification:

$$GII(C, g) = \sum_{u=1}^{U} \frac{|C^u|}{|C|} GI(C^u) \quad (7)$$

The depth of the DT model was set to 3-15 layers. The testing and training accuracies of DTs with different depths were compared, revealing that our DT model achieved the highest test accuracy at the depth of 8. Therefore, the tree structure with the depth of 8 was chosen as the diagnosis model for common diseases of college students.

Next, the ANN was selected to build up the common disease diagnosis model for college students, with rectified linear unit (ReLU) as the activation function:

$$b = max(0, \theta^T a + r) \quad (8)$$

The trained model is easy to face overfitting, if there are too many model parameters and too few training samples. Test results show that our model has a Dropout rate of 0.5. To reduce the overfitting risk, L2 regularization was adopted to process the network. The loss function of the network can be given by:

$$Loss(\omega) = \frac{1}{2n}\left[\sum_{i=1}^{n}(f_\omega(a_i) - b_i)^2 + \mu\left(\sum_{j=1}^{m}\omega_j^2\right)\right] \quad (9)$$

The limited memory - Broyden–Fletcher–Goldfarb–Shanno (L-BFGS) algorithm was employed to optimize the loss function, that is, the approximation matrix F_l was updated based on the latest n curvatures:

$$\begin{aligned} F_l &= (U_{l-1}^T \dots U_{l-n}^T)F_0^l(U_{l-n} \dots U_{l-1}) \\ &+\sigma_{l-n}(U_{l-1}^T \dots U_{l-n+1}^T)E_{l-n}E_{l-n}^T(U_{l-n+1} \dots U_{l-1}) \\ &+\sigma_{l-n+1}(U_{l-1}^T \dots U_{l-n+2}^T)E_{l-n+1}E_{l-n+1}^T(U_{l-n+2} \dots U_{l-1}) \\ &+\dots+\sigma_{l-1}E_{l-1}E_{l-1}^T \end{aligned} \quad (10)$$

As shown in formula (10), the first step is to solve the a_l in the l-th iteration, and save the curvature as $\{e_i, b_i\}$. Then, the initial matrix $F^0{}_l$ is corrected by the previous n curvatures to obtain F_l, which is usually set as a proportional coefficient o_l. In this way, the actual matrix size can be estimated based on the curvature of the previous iteration.

Our SVM model was established based on a linear kernel function. Finally, ensemble method was adopted to integrate the three models through weighted averaging:

$$F(a) = \sum_{i=1}^{\psi}\theta_i f_i(a) \quad (11)$$

where, θ_i is the weight of a single model f_i, which satisfies:

$$\theta_i \geq 0, \sum_{i=1}^{\psi}\theta_i = 1 \quad (12)$$

Compared with the three single models, the integrated model can accurately diagnose the common diseases among college students

3.3 Emotional regulation of college students through personalized recommendation of psychical exercise

Studies have shown that college students' psychological diseases are increasing year by year, while optimistic and stable emotions play a very important role in the control of such diseases. At the same time, stable emotions help to balance the inner physiological state of college students. Our integrated management platform can regulate college students' psychology through PE intervention, namely, pushing suitable physical exercise and relevant contents (e.g., tutorial videos) to those in need, and thereby promote the physiological state of college students with common diseases.

The emotional testing and regulation of college students often consume lots of manpower, materials, and money. Therefore, this paper only labels the emotions of some data samples on college students' behaviors and verifies all the labels through transfer learning. During the labeling phase, the source domain data input a_E come from a few college students, including basic information and mental health data; the target data a_C for estimating the probability $\upsilon(d|a_C)$ of college students' emotions come from the other college students. Since a_E and a_C might exist in different feature spaces, this paper sets up a translator $\gamma(h_A, h_C) \propto \upsilon(h_A, h_C)$ to connect the two feature spaces, where unlabeled features h_A and h_C are independent under a_E, which belongs to set A_E. $\upsilon(h_A, h_C)$ can be calculated by:

$$\upsilon(h_C, h_E) = \int_{A_E}\upsilon(g_C, a_E)\upsilon(g_A|a_E)a_E \quad (13)$$

Let T and W be the probable distributions of a_E and h_A, respectively. Then, h_A and h_C can be connected by the Jensen-Shannon (JS) divergence:

$$JS(T||W) = \frac{1}{2}\left(C_{KL}(T||W) + C_{KL}(T||W)\right) \quad (14)$$

where, N=1/2(T+W) and C_{KL}(.) are both Kullback–Leibler (KL) divergence.

In the verification phase, the emotional label space LS might differ from the labeled emotion of college students. Therefore, the similarity $N=\{n_1, \dots, n_m\}$ between them must be compared under the framework of transfer learning:

$$S(d,n) = PC^2[LS_d, LS_n] \quad (15)$$

where, $LS_d=\{b_i|i=1, \dots, m_d\}$ and $LS_n=\{c_i|i=1, \dots, m_n\}$ are a set of documents representing the emotions in D and D, respectively; $PC^2[LS_d, LS_n]$ is the maximum mean difference used to compute the similarity between D and N. The term frequency–inverse document frequency (TF-IDF) vectors can be represented by b_i and c_i. If $S(d, n)$ is greater than the preset threshold, the most similar label in D will be verified. After a period of emotional labeling and verification, the similarity comparison through transfer learning could save some of the high processing cost and long processing time, and produce a suitable training set. Further, the emotional samples can be divided into 7 emotions by the hidden Markov model (HMM). The trained model can accurately test the real-time emotions of college students.

After detecting the real-time emotions of college students, it is necessary to identify the attributes of the sports teaching video contents, normally with the aid of the inherent labels of the video.

Next, the matching degree was computed between the emotional attributes of college students and the contents of sports teaching video, i.e., the similarity between their emotional attributes before and after watching the video was calculated. Let $GR^i{}_v$ and $GR^i{}_u$ be the scores of the emotional attributes of college students before and after watching the video, respectively. Then, we have:

$$\xi = \frac{1}{m}\sum_{i=1}^{m}(E_v^i - E_u^i) \tag{16}$$

Formula (16) shows, the smaller the ξ value, the greater the matching degree between the emotional attributes of college students and the recommended physical exercise. If ξ is below the preset threshold, the two are matched; If ξ is above that threshold, the two are not matched.

The emotional attributes of college students and the attributes of the recommended physical exercise can be measured by viewing history and the inherent labels of the video, respectively. If the two are unmatched, the physical exercise that interests college students can be obtained by looking up their search history of physical exercises or the viewing history of relevant videos and processing the historical data through singular value decomposition (SVD) and potential semantic indexing. The suitable physical exercise can be recommended as follows:

Suppose there are a total of n college students $\{v_i|i=1, 2, \ldots, n\}$ and e sports teaching videos, which belong to m different physical exercises $\{\tau_i|i=1, 2, \ldots, m\}$. A college student has watched g_{ij} videos on physical exercise τ_j. If he/she has not watched any video on that physical exercise, $g_{ij}=0$. Then, the matrix G about the watching of sports teaching videos can be expressed as:

$$G = \begin{bmatrix} g_{11} & g_{12} & \cdots & g_{1m} \\ g_{21} & g_{22} & \cdots & g_{2m} \\ \vdots & \vdots & & \vdots \\ g_{n1} & g_{n2} & \cdots & g_{nm} \end{bmatrix} \tag{17}$$

Let $rank(G)=o$ be the rank of matrix G. Through SVD of the matrix, we have $G \approx \alpha V^T$, where V is an $n \times o$ matrix with orthogonal left singular vectors. For matrix α of the size $o \times o$, the nonzero diagonal elements are known as singular values, and satisfy $\varepsilon_1 \geq \varepsilon_2 \geq \ldots \geq \varepsilon_s > 0$. Another orthogonal matrix U is of the size $o \times m$, in which the vectors are right singular vectors. G can be decomposed as the product between three matrices: V, α and U:

$$V = \begin{bmatrix} v_{11} & \cdots & v_{1s} \\ \vdots & \vdots & \ddots \\ v_{n1} & \cdots & v_{ns} \end{bmatrix} \Sigma = \begin{bmatrix} \varepsilon_1 & & \\ & \ddots & \\ & & \varepsilon_s \end{bmatrix} U = \begin{bmatrix} u_{11} & \cdots & u_{1m} \\ \vdots & \vdots & \ddots \\ u_{o1} & \cdots & u_{om} \end{bmatrix} \tag{18}$$

The physical exercise τ_i that interests college student v_i be given by potential semantic indexing. The emotional attributes of college students were all represented as row vectors in V, while the attribute features of physical exercises as column vectors in U. The correlation between each row in V and each column in U can be described by a singular value matrix. The first two columns in V and the first two rows in U can be respectively denoted as v and u, respectively:

$$v = \begin{bmatrix} v_{11} & v_{12} \\ \vdots & \vdots \\ v_{n1} & v_{n2} \end{bmatrix} u = \begin{bmatrix} u_{11} & \cdots & u_{1m} \\ u_{21} & \cdots & u_{2m} \end{bmatrix} \tag{19}$$

To obtain the comprehensive relationship between college students' interests and physical exercises, $\{(v_{11}, v_{12}), \ldots, (v_{n1}, v_{n2})$ and $(u_{11}, u_{12}), \ldots, (u_{n1}, u_{n2})\}$ were projected to a two-dimensional (2D) plane. Then, clustering was carried out to obtain the sports teaching videos that interest the college students but not watched yet. In this way, it is possible to realize personalized recommendation of videos.

The research of the relationship between the fluctuations and influencing factors of college students' real-time emotions must consider the different sensitivities of college students to different influencing factors. To realize accurate personalized recommendation, the universal model should be replaced with a personalized prediction model that reflects the diversity and difference between college students.

Suppose there are n college students $\{v_i|i=1, 2, \ldots, n\}$, and e sports teaching videos $\{u_i|i=1, 2, \ldots, e\}$. Each student has k features and each video has l features. The feature matrices for college students and videos are recorded as V and U, respectively. Based on the evaluations of each college student for his/her watched videos, a matrix O can be built up. Because each student only evaluates the videos that interest or disgust him/her, the evaluation matrix contains rather sparse elements. Then, the many unknown values in O can be predicted with the aid of the matching degree O between college students' emotional attributes and the recommended physical exercise attributes. The baseline prediction can be described as $r_{vu}=\lambda+r_v+r_u$. The user preference for a potential influencing factor can be expressed as t_v. The video factors representing the score of a video under the influence of these factors can be denoted by w_u. The mean evaluation of all college students in r_{vu} can be depicted as λ. The bias in the different requirements among students on physical exercise, which arise from personal differences, can be expressed as r_v. The bias of videos reflecting the degree of preference for a physical exercise can be described as r_u. Then, the established model can be described as:

$$\dot{o}_{vu} = r_{vu} + t_v^T w_u \tag{20}$$

Let μ be the regularization constant. Then, the assisted prediction problem can be converted into an optimization problem:

$$\min_{r_*, w_*, t_*} \sum_{(v,u)}(o_{mu} - \lambda - r_v - r_u - t_v^T w_u)^2 + \mu_1 r_v^2 + \mu_2 r_u^2 + \mu_3 \|t_v^2\| + \mu_4 \|w_u^2\| \tag{21}$$

The optimization problem was solved through stochastic gradient descent. To prevent overfitting, it is assumed that $s_{vu}= o_{vu}-\dot{o}_{vu}$ and the learning rate is δ:

$$\begin{cases} r_v \leftarrow r_v + \delta_1(s_{vu} - \mu_1 r_v) \\ r_u \leftarrow r_u + \delta_2(s_{vu} - \mu_2 r_u) \\ t_v \leftarrow t_v + \delta_3(s_{vu} w_u - \mu_3 t_v) \\ w_u \leftarrow t_v + \delta_4(s_{vu} t_v - \mu_4 w_u) \end{cases} \quad (22)$$

By formula (22), the evaluation of each college student for an unwatched sports teaching video can be obtained iteratively.

4. Outlier Detection of Multi-Source Health Monitoring Data

Considering the features and size of multi-source health monitoring data, this paper decides to detect the outliers in the data with the local abnormal factor algorithm, which can quickly operate on medium to high-dimensional datasets. The algorithm is a density-based unsupervised outlier detection method, capable of flexibly adjusting proximity on demand. To a certain extent, it can regulate the outlier detection limit.

Let $\varphi_K(p)$ be the distance between the midpoint P and the k-th nearest point of the multi-source health monitoring data; $\varphi(p, p')$ be the distance between sample points p and p'. On this basis, the k-nearest neighbors of point p, which contain all the points no further than $\varphi_K(p)$ from point p, can be established as:

$$M_K(p) = \{p'|p' \in E, \phi(p.p') \leq \phi_K(p)\} \quad (23)$$

If $\varphi(p, p')$ is greater than $\varphi_l(p)$, the reachable distance from p to p' can be defined as $\varphi(p, p')$; if $\varphi(p, p')<\varphi_K(p)$, the reachable distance is $\varphi_K(p)$:

$$\phi_{RD}(p,p') = max\{\phi_K(p), \phi(p,p')\} \quad (24)$$

The minimum proximity threshold for the sample points can be determined by parameter K, which depends on the statistical results of college students' multi-source health monitoring data or the actual needs. The relative density of sample points can be calculated based on all the points within the minimum proximity range. The local reachable density can be calculated by:

$$LRD_K(p) = \|M_K(p)\| / \textstyle\sum_{p' \in M_K(p)} \phi_{RD}(p,p') \quad (25)$$

The mean ratio of minimum proximity density of point p to the reachable range K can be defined as the local outlier factor $LOF_K(p)$:

$$LOF_K = \sum_{p' \in M_K(p)} \frac{LRD_K(p')}{LRD_K(p)} / \|M_K(p)\| \quad (26)$$

If the $LRD_K(p)$ of point p is large and the relative density of the point is low within its minimum proximity range, then point p is within the outlier range, and very likely to be an abnormal value.

5. Experiments and Results Analysis

The main reason for the collected local multi-source health data to fluctuate significantly over time is the presence of suddenly changing data. To control the effect of abnormal values in the collected data on the platform functions, the stable local trend of multi-source health data was obtained, the high-frequency modal functions were eliminated, and the low-frequency ones were reserved, thereby filtering the original data sequence of the collected data. Figure 6 shows the modal reconstruction states of multi-source health data. It can be observed that the produced data sequence agrees well with the original sequence of multi-source health data. The modal reconstruction successfully derives the stable and smooth trend of multi-source monitoring data on health.

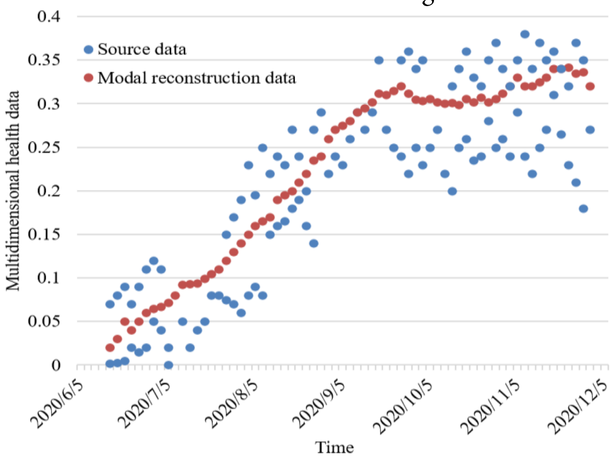

Figure 6. Modal reconstruction states of multi-source health data

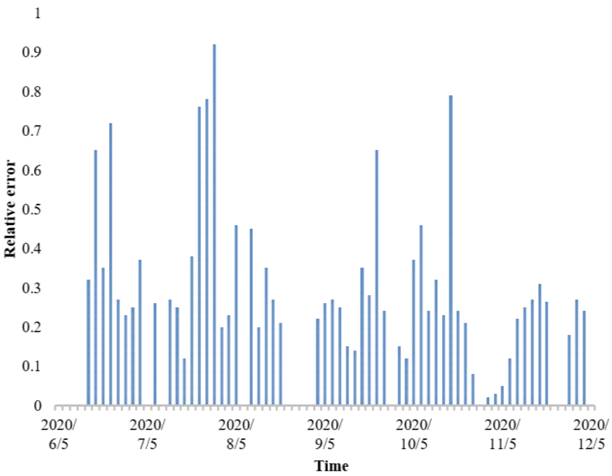

Figure 7. Computed relative error

To further improve the degree of contrast based on the data of empirical modal decomposition and multi-source health data, the relative error between the original sequence of multi-source health data and the modal reconstruction data was calculated (Figure 7). Referring to expert experience, a threshold was configured preliminarily to control the relative error: If the relative error is greater than 0.3, the data are considered abnormal; if the relative error is smaller than 0.3, the data are considered normal.

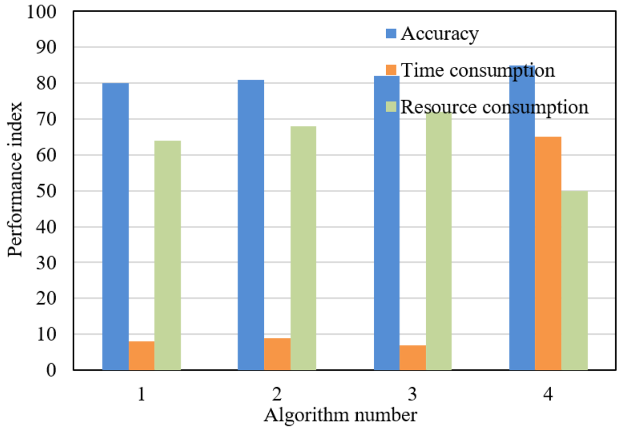

Figure 8. Core indices of different models

Figure 8 compares the diagnosis accuracy, time consumption, and resource consumption of different EMR-based diagnosis algorithms for common diseases among college students. It can be inferred that DT1, SVM2, ANN3, and ensemble model 4 all achieved quite good diagnosis effects, as their accuracies were all above 80%: 80.11% for DT model; 82.11% for SVM model; 83.01% for ANN model; 84.12% for ensemble model. As for time consumption, DT1, SVM2, and ANN3 consumed a short time, while the ensemble model took over 60s, i.e., had the lowest operating efficiency. As for memory occupation, DT1, SVM2, and ANN3 differed slightly in memory cost, while the ensemble model occupied the smallest memory. Overall, the ensemble model achieved the best effect. Figure 9 compares the accuracies of different algorithms on training set and test set of common disease diagnosis. It can be seen that the ensemble model outperformed each single model in the diagnosis of test samples.

Figure 10 shows the variation of the response time of our integrated management platform for personalized recommendation of physical exercises. It can be observed that the mean response time per request increased with the number of concurrent requests for personalized recommendation. As the number of concurrent requests increased gradually from 10 to 800, the maximum, minimum, and moderate time delays exhibited fixed change laws. When 500 concurrent requests occurred, the maximum, minimum, and moderate time delays were 480ms, 89ms, and 57ms, respectively. When the number reached 700 and above, the time delay of the integrated management platform fluctuated significantly due to high load, reaching at least 1,500ms. Therefore, the concurrent load was set to 500ms for our platform. Considering the time delay induced by the up- and down-link data, our platform can effectively meet user requirement on performance, when the concurrent requests are below 500ms.

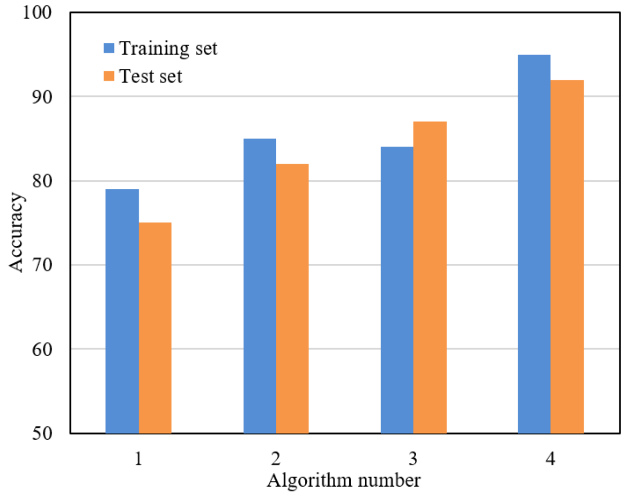

Figure 9. Accuracies of different algorithms on training set and test set of common disease diagnosis

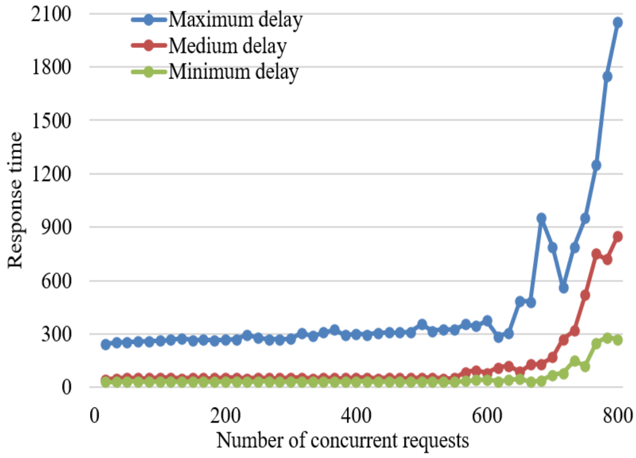

Figure 10. Test results on personalized recommendation performance

6. Conclusions

This paper applies the analysis of multi-source health monitoring data to the integrated management of college students' sports and psychology, and designs the functions of the integrated management platform. Specifically, a college students' sports and psychological integrated management platform was established, along with its functional modules and platform framework. In addition, the sensing approach for multi-source health data was introduced, and data sensing was carried out to obtain the data about the physiological and mental health of college students. Furthermore, the integrated health of college students was predicted, the common diseases of college students were diagnosed based on EMRs, and the emotions of college students were regulated through personalized recommendation of sports. To verify the effectiveness of our outlier detection method for multi-source health monitoring data, experiments were carried out to obtain the modal reconstruction states of multi-source health data, and the relative errors. It can be observed that the produced data sequence agrees well with the original sequence of multi-source health data. The comparison of core indices, as well as the comparison of accuracies of training set and test set of common disease diagnosis, demonstrate that our integrated model achieved better diagnosis effect on the test samples than the existing models. Finally, a performance test was carried out on the personalized recommendation of physical exercises. The test confirms that our platform can effectively meet user requirement on performance, when the concurrent requests are below 500ms.

Acknowledgement

This paper was supported by Horizontal Topic Project, Beihua University (Project name: Design and Application of Integrated Management Platform of College Students' Physical and Mental Health).

References

Bouida, A., Beladgham, M., Bassou, A., Benyahia, I., Ahmed-Taleb, A., Haouam, I., & Kamline, M. (2020). Evaluation of Textural Degradation in Compressed Medical and Biometric Images by Analyzing Image Texture Features and Edges. *Traitement du Signal, 37*(5), 753-762. https://doi.org/10.18280/ts.370507

Chen, X., Tong, M., Shi, C., Zhang, Y., Zhang, J., Chen, X., & Zhou, Z. (2020). Visual analysis of multi-source college students' mental health questionnaire data. *Journal of Computer-Aided Design & Computer Graphics, 32*(2), 181-193. https://doi.org/10.3724/SP.J.1089.2020.17929

Chu, Y., & Yin, X. (2021). Data Analysis of College Students' Mental Health Based on Clustering Analysis Algorithm. *Complexity, 2021*. https://doi.org/10.1155/2021/9996146

Firdausi, N., Lestari, F., & Ismiyati, A. (2021). Disaster Preparedness Analysis of Public Health Centers in DKI Jakarta Province in 2020. *International Journal of Safety and Security Engineering, 11*(1), 91-99. https://doi.org/10.18280/ijsse.110110

Gang, L. (2017). The Promotion of Personalized Physical Health of College Students Based on Physical Education Reform. *AGRO FOOD INDUSTRY HI-TECH, 28*(3), 1852-1855.

Ha, M., Lim, S., & Ko, H. (2018). Wearable and flexible sensors for user-interactive health-monitoring devices. *Journal of Materials Chemistry B, 6*(24), 4043-4064. https://doi.org/10.1039/C8TB01063C

Hu, Q., Chen, L., Lv, P., & Yang, M. (2018). Promotion of College Students' Physical Health: A Case Study on Physical Education Intervention. *2018 9th International Conference on Information Technology in Medicine and Education (ITME)* (pp. 540-544). IEEE. https://doi.org/10.1109/ITME.2018.00125

Jingjing, S., Hangqi, J., & Jingjing, Z. (2020). Establishment of sports health model for college students based big data. *Journal of Physics: Conference Series. 1607*(1) (pp. 012105). IOP Publishing. https://doi.org/10.1088/1742-6596/1607/1/012105

Li, G. (2021). The Relationship between the Degree of Interest in Physical Education and the Mental Health of College Students. *2021 2nd Asia-Pacific Conference on Image Processing, Electronics and Computers* (pp. 846-849). Association for Computing Machinery. https://doi.org/10.1145/3452446.3452648

Li, K., & Yu, W. (2021). A Mental Health Assessment Model of College Students Using Intelligent Technology. *Wireless Communications and Mobile Computing, 2021*. https://doi.org/10.1155/2021/7485796

Muthusamy, P., & Durairaj, P. V. (2019). An Overview of Microwave UWB Antenna for Structural Health Monitoring of Wind Turbine Blades: Optimal Design and Analysis. *Instrumentation, Mesures, Métrologies, 18*(1), 75-81. https://doi.org/10.18280/i2m.180112

Nouioua, N., Seddiki, A., & Ghaz, A. (2020). Blind Digital Watermarking Framework Based on DTCWT and NSCT for Telemedicine Application. *Traitement du Signal, 37*(6), 955-964. https://doi.org/10.18280/ts.370608

Ornek, A. H., Ervural, S., Ceylan, M., Konak, M., Soylu, H., & Savasci, D. (2020). Classification of Medical Thermograms Belonging Neonates by Using Segmentation, Feature Engineering and Machine Learning Algorithms. *Traitement du Signal, 37*(4), 611-617. https://doi.org/10.18280/ts.370409

Rahaman, A., Islam, M. M., Islam, M. R., Sadi, M. S., & Nooruddin, S. (2019). Developing IoT Based Smart Health Monitoring Systems: A Review. *Rev. d'Intelligence Artif., 33*(6), 435-440. https://doi.org/10.18280/ria.330605

Sucena, A., Carneiro, J. F., Almeida, F. G., & Viana, F. L. (2017). Screening reading habilities: a comparison between engineering education, health and psychology in portuguese college students. *International Journal of Engineering Education, 33*(1A), 151-161. http://hdl.handle.net/1822/50064

Toscos, T., Carpenter, M., Drouin, M., Roebuck, A., Kerrigan, C., & Mirro, M. (2018). College students' experiences with, and willingness to use, different types of telemental health resources: do gender, depression/anxiety, or stress levels matter? *Telemedicine and e-Health, 24*(12), 998-1005. https://doi.org/10.1089/tmj.2017.0243

Yan, Y., Wu, Z., Wu, X., Zhou, X., & Weng, C. (2019). A Linux-Based Integrated Structural Health Monitoring System for Bridges in Remote Regions. *Instrumentation, Mesures, Métrologies, 18*(6), 527-534. https://doi.org/10.18280/i2m.180603

Yang, D.-M. (2017). Research on college students' individual physical health promotion-Based on the perspective of physical education reform. *AGRO FOOD INDUSTRY HI-TECH, 28*(1), 937-941.

Yarong, L. (2017). The Promotion Mechanism of Physical Health of College Students from the Perspective of PE Teaching Reform. *AGRO FOOD INDUSTRY HI-TECH, 28*(3), 517-521.

Zhang, J. (2021). A study on mental health assessments of college students based on triangular fuzzy function and entropy weight method. *Mathematical Problems in Engineering, 2021*. https://doi.org/10.1155/2021/6659990