RESEARCH ARTICLE

Dynamic early warning of physical health risk among college students based on XGBoost

Haitao Lu[1], Fengmin Ge[1], Faxiang Fan[2,*]

[1]Xiamen Ocean Vocational College, Xiamen, Fujian, China. [2]Qiannan Normal University for Nationalities, Qiannan, Guizhou, China.

Received: May 14, 2026; **accepted:** September 30, 2026.

The physical health of college students has become an increasingly prominent public health problem worldwide. With the popularization of information lifestyle and increasing academic pressure, the physical activity level of college students has decreased significantly, and the physical health indicators have shown a structural weakening trend. The traditional evaluation method based on "National Students' Physical Health Standard" mainly relies on static grading and lacks the ability of dynamic monitoring and forward-looking early warning of risk evolution. Machine learning (ML) technology, especially gradient lifting algorithm, shows great potential in health risk prediction, but its systematic application in the stratification of college students' physical health risks is still relatively scarce. This study established a dynamic early warning framework for college students' health risks by combining a comprehensive evaluation system, which covered body shape, body function, and physical quality based on the "National Standards for Students' Physical Health", with machine learning classification. K-means clustering method was used for risk stratification, and the traditional grade evaluation was transformed into data-driven risk identification. A multiclassification risk early warning model based on eXtreme Gradient Boosting (XGBoost) was then constructed and compared with six benchmark models. The results revealed that the percentage of outstanding grades decreased annually, while the number of fail and high-risk groups increased sharply in senior grades, while the proportion of moderate-risk groups remained the largest. College students' physical health (CSPH) declined progressively with advancing grades. Comparative model results showed that XGBoost achieved the best prediction performance among male, female, and students of all grades with an accuracy of 91.4% for males, 89.1% for females, and the highest area under the curve (AUC) of 0.962, which outperformed other comparison models overall and demonstrated strong nonlinear modeling and generalization capabilities. This study established a dynamic early warning (DEW) system that integrated ML prediction with hierarchical physical health evaluation. The system had significant practical implications for promoting intelligent management of physical health in colleges and universities and provided data support and a technical pathway for CSPH management and precise intervention.

Keywords: machine learning; XGBoost; K-means; physical health; risk warning.

[*]**Corresponding author:** Faxiang Fan, Qiannan Normal University for Nationalities, Qiannan, Guizhou 558000, China. Email: 15980761151@163.com.

Introduction

In recent years, with the popularization of information lifestyle and increasing academic pressure, the physical activity level of college students has dropped significantly, and the physical health status has shown a structural weakening trend [1]. Long-term sedentary

behavior and excessive use of digital devices not only lead to the decline of physical function, but also may lead to mental fatigue, metabolic disorders, and other issues, which have become an important topic of concern in the global public health field [2]. The university stage is a critical period for the development of young people's physique and an important stage for the formation of healthy behaviors and lifestyles. Therefore, it is of great practical significance to build a scientific, accurate, and operable physical health risk identification and early warning mechanism for improving the efficiency of college sports management and realizing hierarchical and classified intervention.

At present, the physical health evaluation of colleges and universities in China is mainly based on the National Student Physical Health Standard to carry out comprehensive scoring and grading [3]. With the development of health monitoring technology, the national evaluation system has gradually changed to a comprehensive evaluation model that considers physical function, health behavior, and long-term risk management [4], which requires deep mining and intelligent analysis of physical health data. In recent years, machine learning (ML) has been gradually applied in health risk prediction, especially the gradient lifting algorithm shows strong classification ability and generalization performance in mental health risk stratification [5]. eXtreme Gradient Boosting (XGBoost) and other tree-based models have strong predictive ability and generalization performance when dealing with high-dimensional, nonlinear, and multivariable data. Ding *et al*. applied ML to predict the mental health risks of adolescents in China under stressful life events, and both internal and external validations showed excellent performance [6]. Huang *et al*. combined logistic regression with XGBoost to analyze the risk factors of insomnia using the U.S. National Health and Nutrition Examination Survey (NHANES) dataset, achieving effective risk prediction [7]. Wu *et al*. used XGBoost to predict the risk of osteoporosis, demonstrating that the model was capable of identifying key risk factors without directly measuring bone mineral density [8]. By integrating XGBoost with random forest, Md *et al*. enhanced the accuracy of stroke prediction [9]. By merging routine medical monitoring data with electronic clinical records and employing XGBoost and Bi-LSTM (Long Short-Term Memory) to extract important features for heart disease prediction, Alzakari *et al*. further illustrated the potential of ML in the integration of multisource health data [10]. Furthermore, XGBoost has shown promise for application in college students' physical health (CSPH). For high-precision identification of physical health status, Liang *et al*. employed the differential evolution method to tune the global parameters of XGBoost and used a discrete coding approach to evaluate college students' physical fitness test data [11]. Qin *et al*. combined association rules with XGBoost to explore the key influencing factors of students' mental health, thereby providing an effective tool for early risk warning [12]. To assess the significance of environmental influences on mental health, Liu *et al*. integrated SHAPley additive explanations (SHAP) with XGBoost and highlighted the necessity of model interpretability in precise intervention [13]. However, the current research on college physical health evaluation system lacks the ability of hierarchical prediction and dynamic early warning for different grades and gender groups. The evaluation of physical health in colleges and universities is still mainly based on static grading, lacking dynamic monitoring and forward-looking early warning of the risk evolution process, and systematic research in the field of physical health risk stratification and dynamic early warning in colleges and universities is still relatively scarce [14, 15]. Traditional statistical methods have limited ability to describe complex nonlinear relationships when dealing with multi-index and high-dimensional data [16, 17]. Although ML methods begin to enter the field of health prediction in recent years, most of the research is still limited to static cross-sectional analysis, lacking hierarchical prediction for different grades and gender groups.

This study developed a dynamic early warning framework for college students' health risks by combining comprehensive evaluation system with ML classification to realize the transformation from traditional static grade evaluation to data-driven dynamic risk identification. The study developed a thorough system for evaluating the physical health of college students using test data, determined the overall physical fitness score using standardized processing and weighted scoring, and stratified students' risks using the K-means clustering method, thereby realizing the shift from "grade evaluation" to "risk identification". XGBoost algorithm was applied to build a physical Health Risk Early Warning (HREW) model that was then compared with many models to analyze the prediction ability and stability of the model under the condition of multi-dimensional physical characteristics. The weighted comprehensive scoring system of National Students' Physical Health Standard was innovatively combined with the K-means clustering risk stratification method. Through the four-category risk early warning model of college students' physical health based on XGBoost algorithm, the changing law of physical structure and risk evolution trend were revealed from the two dimensions of gender and grade. An analysis framework integrating physical assessment, risk stratification, and intelligent early warning was constructed, which provided a new technical path for precise intervention and dynamic management of college students' physical health and had important practical significance for promoting intelligent management of college students' physical health.

Materials and methods

Datasets collection

The data used in this study mainly collected from the physical health questionnaire survey of college students and the annual physical fitness test results from Xiamen Ocean Polytechnic College (Xiamen, Fujian, China). The questionnaire was conducted online using the Wjx platform (https://v.wjx.cn/vm/PgBJd7G.aspx#), involving a total of 300 undergraduate students in 2024 academic year. The data sample consisted of 156 males (52.0%) and 144 females (48.0%) across four grades with 85 freshman (28.3%), 80 sophomore (26.7%), 72 junior (24.0%), and 63 senior (21.0%). The age of participants was ranged from 18 to 22 years old with the average of 20.1 ± 1.3 years. During data collection, the original physical fitness test data was arranged in compliance with the “National Standards for Students' Physical Fitness and Health”, which included core indicators of body mass index (BMI), vital capacity, 50-meter running, standing long jump, sitting forward, pull-ups (male) and sit-ups (female), and 1,000-meter running (male) and 800-meter running (female). The students' basic information data including gender, grade, and test year information were used to construct hierarchical analysis and time series structure. All the procedures of this research were approved by Ethics Review Committee of Xiamen Ocean Polytechnic College (Xiamen, Fujian, China) with the approval number of HYXY-IRB-2024-017.

Index system and scoring structure

According to "National Standards for Students' Physical Health", this study selected the core indicators of college students' physical evaluation including body shape, body function, and physical quality three categories [18]. Index weights were allocated according to national standards to ensure the scientific contribution of various indicators to the comprehensive score [19]. The specific weight distribution included body shape index with the BMI being weighted 15%, physical function index with the weights of vital capacity, 50-meter running, and standing long jump as 15%, 20%, and 10%, the physical fitness index with the forward flexion of sitting position, pull-ups for male/sit-ups for female, and 1,000 m running for male/800 m running for female being weighted as 10%, 10%, and 20%. BMI was used to evaluate the status of weight control, which was an important structural variable of physical health risk and was calculated as below [20].

$$BMI = \frac{W}{H} \quad (1)$$

Where *W* was weight (kg). *H* was height (m). To eliminate the influence of dimension, all indexes were standardized. For positive indicators of vital capacity and long jump performance, range standardization was adopted and calculated as follows.

$$x_i' = \frac{x_i - x_{min}}{x_{max} - x_{min}} \quad (2)$$

where x_i was the original observation value. x_{min} and x_{max} were the minimum and maximum values of the sample. This process made the variable interval unified to [0, 1]. For negative indicators of 1,000/800-meter running time and 50-meter running time, reverse standardization was adopted and calculated as below.

$$x_i' = \frac{x_{max} - x_i}{x_{max} - x_{min}} \quad (3)$$

To avoid the influence of extreme values, Z-score standardization was also carried out for comparison as below.

$$Z_i = \frac{x_i - \mu}{\sigma} \quad (4)$$

where μ was the mean. σ was the standard deviation. This approach was more stable when the data distribution was near normal. After the data standardization was completed, the comprehensive score was calculated according to the national standard weight as follows.

$$S = \sum_{j=1}^{n} w_j x_j' \quad (5)$$

where w_j was the weight of the j-th index. x_j' was the standardized value. n was the total number of indicators. The comprehensive score S was used to describe the physical health level of individuals. The constitution grade could be expressed as a piecewise function as below.

$$G(S) = \begin{cases} A, S \geq 90 \\ B, 90 > S \geq 80 \\ C, 80 > S \geq 60 \\ D, S < 60 \end{cases} \quad (6)$$

where A was excellent (90.0 or above). B was good (80.0 to 90.0). C was pass (60.0 to 80.0). D was fail (below 60.0).

Constructing risk stratification based on K-means cluster analysis

Risk stratification adopted K-means clustering in unsupervised learning method [21]. The comprehensive score and individual indicators were combined into a characteristic matrix and normalized to eliminate the dimensional differences of different indicators. The cluster center was optimized by minimizing Within-Cluster Sum of Squares (WCSS) as below [22].

$$min \sum_{k=1}^{K} \sum_{x_i \in C_k} \| x_i - \mu_k \|^2 \quad (7)$$

where C_k was the K-th cluster. μ_k was the cluster center. Different cluster numbers were evaluated by contour coefficient to ensure that the selected cluster number was the most reasonable in population structure and distribution as follows.

$$s(i) = \frac{b(i) - a(i)}{max(a(i), b(i))} \quad (8)$$

where $a(i)$ was the average distance from the sample to the cluster. $b(i)$ was the average distance from the sample to the nearest cluster. Four clusters including A (health), B (mild risk), C (moderate risk), and D (high risk) were defined corresponding to the ideal number of clusters in this study. The risk classification method used in this study was that, according to the weighted scoring system, the comprehensive score was calculated for each student. The characteristic matrix was composed of the comprehensive score and each single standardized index. K-means cluster analysis was carried out. According to the relative position of the clustering center in the comprehensive score dimension, the clustering results were mapped into four risk

levels. The lower the score, the higher the risk level. This method transformed the traditional four-level evaluation (excellent, good, passing, failing) into four-level risk classification (health, mild risk, moderate risk, high risk) and realized data-driven risk identification.

Construction of XGBoost physical HREW model

XGBoost was used to construct a four-class model of physical health risk. As an ensemble learning algorithm based on the gradient boosting tree, it continuously added decision trees *via* the additive model and optimized prediction accuracy by leveraging the first- and second-order information of the loss function. Compared with the traditional Gradient Boosting Decision Tree (GBDT), XGBoost controlled model complexity by incorporating a regularization term to prevent over-fitting and supported parallel computing to improve training efficiency. Moreover, its built-in automatic handling of missing values and feature importance evaluation capabilities improved model robustness when processing complex physical fitness data, while also enhancing the interpretability of classification results. The iterative training process of XGBoost algorithm included the start of serial iterative training after the original input data to constitute a sample set [23]. The basic weak learner was initialized, and the residual and negative gradient were calculated. In each subsequent round, the negative gradient (residual) obtained in the last round was used as the learning target to fit the Classification and Regression Tree (CART) classification regression tree. The overall model was updated according to the newly trained decision tree. The new residual and negative gradient were calculated again and sent to the next round of circulation. The iterative process was repeated until the termination condition was met, and the prediction outputs of all decision trees were finally accumulated to get the final prediction result of the model. XGBoost adopted the form of addition model and gradually superposed multiple regression trees to form the final prediction function. The predicted value of the i-th sample was expressed as below.

$$\hat{y}_i = \sum_{t=1}^{T} f_t(m_i), f_t \in \mathcal{F} \quad (9)$$

where $\hat{y}_i$ was the prediction score of the i-th sample. T was the total number of trees. $f_t(m_i)$ was the t-th regression tree. m_i was the feature vector of the i-th sample. $\mathcal{F}$ was the function space of all possible tree structures. The basic idea behind XGBoost was to construct an objective function with a loss function and regularization term to balance the model's complexity and fitting ability. Iteration t 's optimization goal was stated as below.

$$\mathcal{L}^{(t)} = \sum_{i=1}^{n} l(y_i, \hat{y}_i^{(t-1)} + f_t(m_i)) + \Omega(f_t) \quad (10)$$

where n was the number of samples. $\mathcal{L}^{(t)}$ was the objective function value of t-round iteration. $l(\cdot)$ was the multi-classification cross entropy loss function. y_i was the real category label. $\hat{y}_i^{(t-1)}$ was the cumulative prediction value of the previous $t-1$ round. $\Omega(f_t)$ was the regularization term. $f_t(m_i)$ was the predicted output of the currently added tree. By reducing prediction error and model complexity, the objective function struck a compromise between accuracy and generalization capacity. To prevent the model from overfitting, XGBoost added a structural regularization component to the objective function, which limited the complexity of the tree as follows.

$$\Omega(f_t) = \gamma q + \frac{1}{2}\lambda \sum_{j=1}^{J} w_j^2 \quad (11)$$

where q was the number of leaf nodes. J was the total number of leaf nodes. w_j was the weight of the j-th leaf node. γ was the penalty coefficient of node splitting. λ was the regular parameter of the weight. The first item controlled the complexity of the tree structure, and the second item limited the weight of leaf nodes. In the process of tree generation, it was necessary to judge whether to split nodes. XGBoost measured the decline of loss function after splitting by calculating splitting gain as below.

$$Gain = \frac{1}{2}\left[\frac{G_L^2}{H_L+\lambda} + \frac{G_R^2}{H_R+\lambda} - \frac{G^2}{H+\lambda}\right] - \gamma \quad (12)$$

where G_L and H_L were the first- and second-order gradations of the left child node. G_R and H_R were the gradient sums of the right child node. G and H were the gradient sums of the parent node. When $Gain > 0$, it showed that splitting could effectively reduce the loss function, and nodes would be divided, otherwise they would stop growing. Because this study was a four-category risk prediction, XGBoost used Softmax function to transform scores of multiple categories into probability distribution in the output stage as follows.

$$P(y = c|x) = \frac{e^{\hat{y}_c}}{\sum_{k=1}^{K} e^{\hat{y}_k}} \quad (13)$$

where K was the number of risk levels. $\hat{y}_c$ was the forecast score of category c. The final prediction category was the category with the greatest probability. Softmax function ensured that the sum of probability of each category was 1, so that the model could complete the multi-classification risk level discrimination task.

Experimental environment and parameters setting

Windows 10 operating system served as the foundation for the study's experimental setup, which included an Intel Core i7 processor and 16 GB of RAM. Python 3.10 (https://www.python.org/downloads/release/python-3100/) was used for both model creation and data processing, primarily utilizing Pandas (https://pandas.pydata.org/) for data management, Scikit-learn (https://scikit-learn.org/stable/index.html) for conventional ML model construction, and the XGBoost (https://docs.xgboost.cn/en/stable/) for gradient lifting model training. The dataset was split into 70% training and 30% test sets. To improve the model's stability and capacity for generalization, 10% cross-validation was used during the training phase. Six popular ML models including Logistic regression (LR) (https://scikit-learn.cn/stable/modules/generated/sklearn.linear_model.LogisticRegression.html), Support vector machine (SVM) (https://scikit-learn.cn/stable/modules/svm.html), Random forest (RF) (https://scikit-learn.cn/stable/modules/ensemble.html#random-forests-and-other-randomized-tree-ensembles), K-nearest neighbors (KNN) (https://scikit-learn.cn/stable/modules/neighbors.html#nearest-neighbor-algorithms), Decision tree (DT) (https://scikit-learn.cn/stable/modules/tree.html), and Gradient boosting decision tree (GBDT) (https://scikit-learn.cn/stable/modules/ensemble.html#gradient-boosted-trees) were chosen as comparison models to confirm the prediction advantages of XGBoost model. All models were optimized by grid search to ensure that different models could be compared fairly under the optimal parameters. The XGBoost model's fundamental parameters included a maximum tree depth of 6, a learning rate of 0.1, a sub-sample sampling ratio of 0.8, a feature sampling ratio of 0.8, a minimum sub-node weight of 1, an L2 regularization coefficient of 1, a splitting penalty coefficient of 0.1, and 300 iterative trees. The model's performance evaluation metrics were accuracy, precision, recall, F1 value, and area under the curve (AUC) value. The confusion matrix was used to examine the categorization effect of different risk levels to carefully evaluate the model's practical application value and prediction performance.

Statistical analysis

SPSS 26.0 software (IBM, Armonk, New York, USA) was employed for statistical analysis of this study. Continuous variables were expressed as mean ± standard deviation (SD). The comparison of physical health indicators between different genders and grades was performed using one-way analysis of variance (ANOVA). When significant differences were detected, pairwise comparisons were conducted using Tukey's post hoc test. The distribution comparison of physical grade and risk grade among different gender and grade groups was performed using the chi-square test. P value less than 0.05 was defined as the statistical significance.

Results and discussion

Physical health indicators by genders and grades

Table 1. Descriptive statistics of male students' physical health indicators (mean ± SD).

Grade	BMI (kg/m²)	Vital capacity (ml)	50-m sprint (s)	Standing long jump (cm)	Sit-and-reach (cm)	Pull-ups (times)	1,000-m run (s)
Freshman	22.11 ± 2.84	4012 ± 756	7.62 ± 0.71	231.4 ± 28.6	12.5 ± 6.1	16.8 ± 5.7	245.3 ± 32.4
Sophomore	22.35 ± 2.91	3956 ± 801	7.68 ± 0.74	228.9 ± 29.7	12.1 ± 6.3	15.7 ± 5.4	249.7 ± 35.2
Junior	22.54 ± 3.02	3898 ± 845	7.75 ± 0.79	224.6 ± 31.2	11.6 ± 6.8	14.6 ± 5.1	256.4 ± 37.1
Senior	22.73 ± 3.11	3815 ± 902	7.88 ± 0.83	219.8 ± 32.5	10.9 ± 7.2	13.4 ± 4.9	263.2 ± 39.6

The results of stratified statistics according to gender (male/female) and grade (freshman to senior) demonstrated that there were significant differences in various physical health indicators among different gender groups ($P < 0.05$). At the level of physical morphology indicators, the BMI of male students gradually increased from 22.11 kg/m² in the first year to 22.73 kg/m² in the senior year with a cumulative increase of 2.8% (Table 1). The BMI of female students gradually increased from 20.73 kg/m² in the first year to 21.22 kg/m² in the senior year with a cumulative increase of 2.4% (Table 2). Although the increase in BMI with grade was relatively small, it was still statistically significant ($P < 0.05$), suggesting that the pressure of weight management gradually increased with grade. There was a significant difference in the baseline values of BMI between male and female students with male students' BMI in each grade being higher than that of female students, which was in line with the physiological differences between genders. At the level of physical function indicators, the lung capacity of male students gradually decreased from 4,012 mL in the first year to 3,815 mL in the senior year with a cumulative decrease of 4.9% (Table 1). The lung capacity of female students gradually decreased from 2,984 mL in the first year to 2,814 mL in the senior year with a cumulative decrease of 5.7% (Table 2). There was a significant difference in the baseline values of lung capacity between male and female students with male students' lung capacity in each grade being significantly higher than that of female students ($P < 0.05$), reflecting the physiological advantage of males in the heart and lung functions. The relative decrease rate of lung capacity for female students was slightly higher than that for male students, suggesting that the decline rate of heart and lung functions for female students might be faster. The 50-meter running time of male students increased from 7.62 seconds in the first year to 7.88 seconds in the senior year with a cumulative increase of 3.4% (Table 1), while that of female students increased from 8.52 seconds in the first year to 8.85 seconds in the senior year with a cumulative increase of 3.9% (Table 2). Both male and female students showed a significant decline in speed qualities, and the decline rate for female students was slightly higher than that for male students. The standing long jump score of male students decreased from 231.4 cm in the first year to 219.8 cm in the senior year with a cumulative decrease of 5.0% ($P < 0.05$) (Table 1), while that of female students decreased from 176.4 cm in the first year to 165.7 cm in the senior year with a cumulative decrease of 4.9% ($P < 0.05$) (Table 2). There was a significant difference in the baseline values of standing long jump between male and female students with male students' scores in each grade being significantly higher than those of female students, but the decline rate of speed for male students was slightly higher than that for female students. At the level of physical fitness indicators, the sit-and-reach score of male students decreased from 12.5 cm in the first year to 10.9 cm in the senior year with a cumulative decrease of 12.8% ($P < 0.05$) (Table 1), while that of female students decreased from 15.2 cm in the first year to 13.9 cm in the senior year with a cumulative decrease of 8.5% ($P < 0.05$) (Table 2). The flexibility scores of female students in each grade were significantly higher than those of male students, and the decline rate was also smaller than that of male students, indicating that the flexibility level of female students was

Table 2. Descriptive statistics of female students' physical health indicators (mean ± SD).

Grade	BMI (kg/m²)	Vital capacity (ml)	50-m sprint (s)	Standing long jump (cm)	Sit-and-reach (cm)	Sit-ups (times)	800-m run (s)
Freshman	20.73 ± 2.45	2984 ± 521	8.52 ± 0.63	176.4 ± 22.8	15.2 ± 6.7	34.6 ± 6.4	232.6 ± 28.5
Sophomore	20.91 ± 2.57	2927 ± 556	8.61 ± 0.66	173.8 ± 23.6	14.6 ± 7.1	33.2 ± 6.1	236.8 ± 29.9
Junior	21.04 ± 2.61	2875 ± 590	8.73 ± 0.69	169.5 ± 24.7	13.8 ± 7.4	31.9 ± 5.8	241.3 ± 31.7
Senior	21.22 ± 2.69	2814 ± 624	8.85 ± 0.72	165.7 ± 26.1	13.9 ± 7.8	30.4 ± 5.6	248.9 ± 34.2

better. The number of pull-ups of male students decreased from 16.8 times in the first year to 13.4 times in the senior year with a cumulative decrease of 20.2% (Table 1). The number of sit-ups of female students decreased from 34.6 times in the first year to 30.4 times in the senior year with a cumulative decrease of 12.1% ($P < 0.05$) (Table 2). The absolute and relative decrease amounts of strength indicators for male students were greater than those for female students, suggesting that the strength quality of male students had deteriorated more severely. The 1,000-meter running time of male students increased from 245.3 seconds in the first year to 263.2 seconds in the senior year with a cumulative increase of 7.3% (Table 1). The 800-meter running time of female students increased from 232.6 seconds in the first year to 248.9 seconds in the senior year with a cumulative increase of 7.0% (Table 2). Both endurance indicators showed a significant grade-increasing trend and were statistically significant ($P < 0.05$), indicating that endurance quality was the most sensitive indicator of physical changes. There were also significant differences among different grade groups ($P < 0.05$). From freshman to senior, the physical health indicators of males and females all demonstrated a downward trend with the increase of grade. For males, BMI was increasing year by year, while vital capacity was decreasing year by year, and the time of 50-meter running continued to increase. The standing long jump distance and the performance of sitting forward were decreasing year by year, while the number of pull-ups was decreasing year by year, and the time of 1,000-meter running was increasing year by year, which indicated that the body shape, cardiopulmonary endurance, speed, explosiveness, flexibility, upper limb strength, and endurance quality were all deteriorating. Females also showed the trends of increasing BMI, decreasing vital capacity, increasing the time of 50-meter and 800-meter running, and decreasing the standing long jump distance and the number of sit-ups, but the decline of sitting forward was slow and slightly fluctuating from junior to senior. The indexes of senior students were significantly worse than those of freshmen, and the changes between grades had obvious stage characteristics. The standard deviation increased with the grade, suggesting that the individual differences within the senior group further increased. The findings of this research were consistent with the previous studies on the downward trend of college students' physique. The observed deterioration of grade-related cardiopulmonary endurance and muscle strength was consistent with previous reports that academic pressure and reduced participation in sports activities led to a continuous decline in physical fitness in college. The high-risk rate of female students, especially senior students, might be attributed to gender-specific physiological differences and lifestyle patterns including lower baseline muscle strength and endurance levels.

Physical fitness grades by genders and grades

The proportion of students with excellent grades in physical health declined with advancing grades with males accounting for 18.4% in freshman year but falling to 12.6% in senior year and females showing 15.1% in freshman year and 10.4% in senior year (Table 3). The results indicated that, with increasing grade level, the proportion of students with high-level physical fitness gradually decreased, which might be related to increased academic pressure and

Table 3. Distribution of students' physical fitness grades by gender and grade.

Gender	Grade	Excellent (A, ≥90)	Good (B, 80–89.9)	Pass (C, 60–79.9)	Fail (D, <60)
Male	Freshman	18.40%	34.60%	36.20%	10.80%
	Sophomore	16.70%	31.90%	39.40%	12%
	Junior	14.30%	29.80%	42.70%	13.20%
	Senior	12.60%	27.40%	44.80%	15.20%
Female	Freshman	15.10%	32.70%	41.30%	10.90%
	Sophomore	13.80%	30.60%	43.90%	11.70%
	Junior	11.90%	28.40%	45.70%	14%
	Senior	10.40%	26.10%	47.20%	16.30%

reduced time for exercise. The trends for males and females were similar, although the proportion of males with excellent performance was slightly higher than that of females. The proportion of students with good grades also declined with grade level, dropping from 34.6% among freshmen to 27.4% among seniors for males and from 32.7% to 26.1% for females. These changes suggested that the ability of mid- to high-level students to maintain their physical fitness was slightly weakened in the later years of college. Overall, the magnitude of change in the proportion of good grades was not as pronounced as that of excellent and passing grades, but it still exhibited a trend of declining physical fitness with grade level. The proportion of students with passing grades increased gradually with grade level with the rate for males rising from 36.2% in freshmen to 44.8% in seniors, while that for females rising from 41.3% to 47.2%. Although most students' physical fitness levels remained essentially up to par, their physical fitness levels became increasingly concentrated in the passing range as grades rise, suggesting that overall physical health among senior students had deteriorated. The proportion of students with failing grades also rose with grade level, reaching the highest increase among the four categories with 15.2% for males and 16.3% for females in the senior year. The failure rate for female students was somewhat greater than that for male students, indicating that the deterioration among female senior students was more noticeable in strength and endurance performance. The physical health level of college students showed an obvious downward trend in grades, and the proportion of excellent grades decreased with grades, while the proportion of "passing" and "failing" increased year by year with the physical decline of senior students being the most significant. This trend reflected that, after college students entered the senior year, due to the increase of professional curriculum pressure and the increase of time occupied by internship and employment preparation, the physical exercise time was seriously squeezed. Meanwhile, the living habits of senior students tended to be solidified, the sedentary time was prolonged, and the quality of sleep was reduced. These factors jointly accelerated the degradation process of physical health. The failure rate of female students was always slightly higher than that of male students, which might be related to their low basic strength level and weak willingness to participate in high-intensity sports activities. It suggested that colleges and universities should fully consider gender differences when formulating physical health intervention strategies and design more attractive sports and incentive mechanisms for female students. In addition, the phenomenon that the passing grade accounted for the highest proportion and continued to rise demonstrated that "maintaining basic standards" had become the physical condition of most senior students. Although this group had not reached the failing standard, it was on the edge of sub-health and would be the focus of future physical intervention. If effective measures were not taken in time, these students were likely to face greater health risks when they entered the early stage of society after graduation. Therefore,

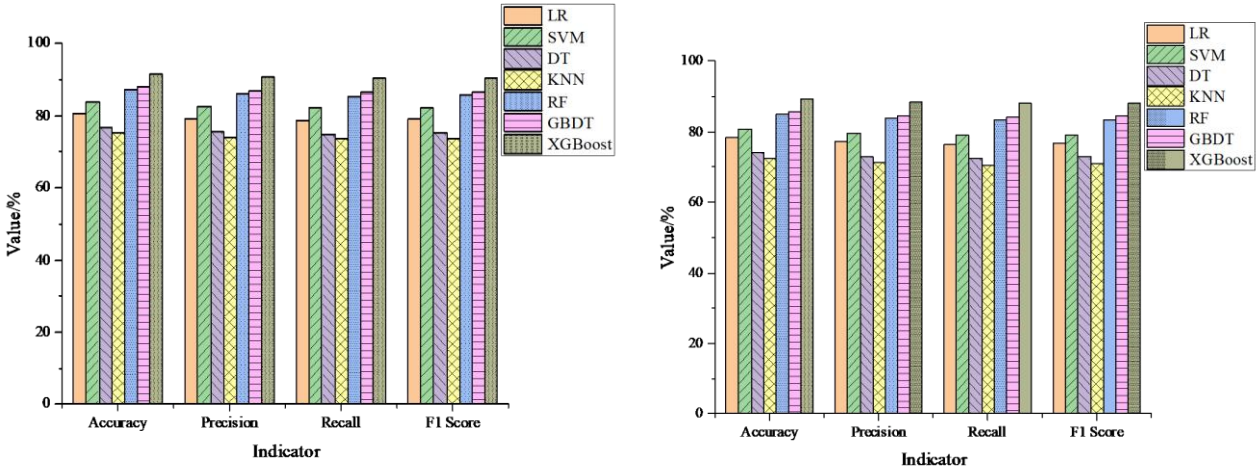
Table 4. Distribution of physical health risk grades of different genders and grades.

Gender	Grade	Healthy (A)	Mild Risk (B)	Moderate Risk (C)	High Risk (D)
Male	Freshman	20.30%	33.50%	34.70%	11.50%
	Sophomore	18.10%	31.40%	38.60%	11.90%
	Junior	15.20%	29.70%	41.80%	13.30%
	Senior	13.70%	27.80%	43.60%	14.90%
Female	Freshman	17.60%	35.20%	36.10%	11.10%
	Sophomore	15.40%	33.80%	39.50%	11.30%
	Junior	13.20%	30.60%	42.70%	13.50%
	Senior	11.50%	28.90%	44.20%	15.40%

colleges and universities should establish a dynamic monitoring mechanism of physical health from freshman to sophomore and block the vicious circle of physical decline through early warning and layered intervention.

Physical health risks of different genders and grades

The proportion of healthy students decreased with advancing grades, declining from 20.3% in freshman year to 13.7% in senior year for males and from 17.6% to 11.5% for females, which indicated that the population of physically healthy students among senior grades was shrinking gradually while health risks were increasing. The proportion of students classified as having mild risk showed a slight decrease with grade level from 33.5% to 27.8% for males and from 35.2% to 28.9% for females. Although the proportion had declined, it still accounted for a relatively large segment of the student population, suggesting that a considerable number of students exhibited mild physical risks and needed to be monitored closely. The proportion of students in the moderate-risk category was the highest, rising from 34.7% to 43.6% for males and from 36.1% to 44.2% for females. The proportion of moderate–risk students increased significantly in senior grades, indicating that most students' physical fitness levels had shifted from healthy or mild-risk status to moderate-risk, and that the need for physical intervention was now urgent. The proportion of high-risk students had risen year by year from 11.5% to 14.9% for males and from 11.1% to 15.4% for females (Table 4). High-risk groups were most concentrated in the senior year, suggesting that students' physical fitness declined significantly in the later years of college, making them primary monitoring and intervention targets. The high-risk rate for females was slightly higher than that for males, which might be related to a more rapid decline in endurance and strength performance. From the perspective of the dynamic evolution of risk levels, college students' physical health showed the characteristics of risk migration with the proportion of health level and mild risk level decreasing synchronously with the increase of grade. While the moderate risk level, as a reservoir, continued to absorb people who had been degraded from the upstream and eventually transported them to the high-risk level. This migration chain showed that the physical health risk was not a sudden event, but a gradual and cumulative development process. The relative decline of female students' health grade was slightly larger than that of male students, suggesting that the physical health foundation of female students might be more fragile. The proportion of moderate risk groups had exceeded 40% in the senior level, which meant that nearly half of senior students were in a state of sub-health. If not intervened in time, this group of people would face higher risk of chronic diseases when they entered society after graduation. Although the proportion of high-risk group was relatively low, its absolute number could not be ignored. This group often deteriorated with many physical indexes at the same time, which was the key intervention object of sports health work in colleges and universities.



(a) Male students

(b) Female students

Figure 1. Comparison of prediction performance of students of different genders in different models.

The phenomenon that the high-risk ratio of female students was always slightly higher than that of male students might be related to the low basal metabolic rate, relatively small muscle mass, and weak willingness to participate in high-intensity resistance training. It suggested that gender differences should be fully considered when formulating risk intervention programs, and more targeted exercise prescriptions and health education strategies should be designed for female students. In addition, from the vertical evolution trend, sophomore to junior was the accelerated period of risk level migration, which was at the turning point of the sharp increase of professional curriculum pressure and the decrease of community activities. It was suggested that colleges and universities strengthened the frequency of physical health monitoring and implemented early warning and layered intervention in this critical window period to block the path of continuous risk deterioration.

The performance comparison of different models

The results showed that XGBoost was significantly superior to other models in both male and female samples with the accuracy rate in male of 91.4%, the precision of 90.7%, the recall of 90.3%, and F1 of 90.5%, which were higher than that in random forest and GBDT. The accuracy rate of XGBoost in female samples was 89.1%, the precision was 88.3%, the recall was 87.9%, and F1 was 88.1%, which were also better than other models (Figure 1). The results suggested that integrated methods like RF, GBDT, XGBoost were obviously superior to linear or distance-based methods of LR, SVM, KNN, DT. XGBoost was the most robust in dealing with nonlinear and multidimensional physical characteristics. The prediction index of male was often two to three percentage points higher than that of female, which suggested that the physical traits of men were easier to recognize or more stable in the model. XGBoost model demonstrated the best prediction performance in gender grouping, which was due to its unique algorithm design. Compared with the Bagging strategy of random forest, the gradient lifting framework adopted by XGBoost could continuously correct the residual of the antecedent model through iterative optimization, and the regularization term introduced effectively controlled the complexity of the model and avoided the risk of over-fitting. When dealing with physical health data, XGBoost also played an important role in the automatic processing mechanism of missing values because individual indicators were often missing in physical test data. The results showed that the

integrated learning method was better than the linear model. There was not a simple linear superposition relationship between physical health risk and multiple physical indicators, but a complex nonlinear interaction. The change of a single indicator might produce differentiated risk effects under different combination conditions. The phenomenon that the prediction performance of male samples was better than that of female samples might be related to the greater dispersion of male physical fitness indicators and higher feature discrimination, or it might reflect that the influencing factors of female physical health risks were more diverse, and besides physical indicators, they were also disturbed by variables such as physiological cycle and eating habits that were not included in the model.

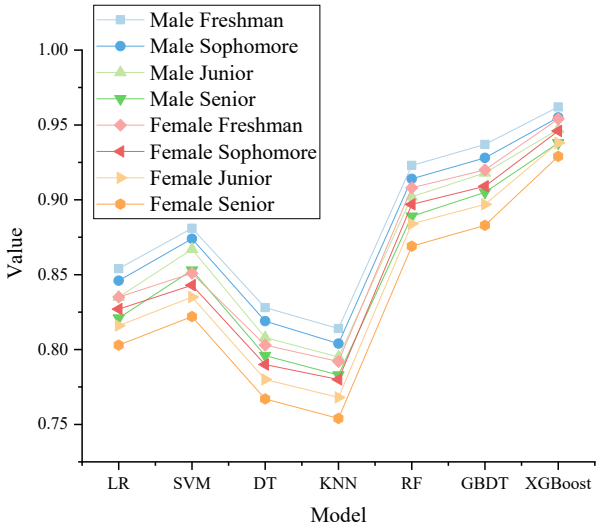


Figure 2. Comparison of AUC values of male and female models in different grades.

Comparison of AUC values of male and female models in different grades

The changing trend of AUC of each model in different grades and genders demonstrated that XGBoost was the highest in all groups with the AUC value of freshman male as 0.962 and that of senior male as 0.938. The AUC value of the freshman female was 0.954 and that of the senior female was 0.929. However, with the increase of grade years, AUC decreased slightly, indicating that the distribution of physical data of senior samples was more complicated and it was more difficult to distinguish. In contrast, although the AUC values of RF and GBDT were close, they were still 0.02 to 0.03 behind, while the AUC values of linear models or simple models were mostly in the range of 0.80 to 0.88, and the gap was obvious (Figure 2). XGBoost had a leading ability in the EW task of physical health risks in colleges and universities, and it still maintained a high stability under the stratification of grade and gender. However, the prediction difficulty of senior samples needed to be combined with more behavioral or time series characteristics to further improve the performance. The AUC value of each model showed a downward trend with the increase of grade, which revealed the core difficulty of physical health risk prediction. The distribution of physical fitness indexes of junior students was relatively concentrated, and the boundary between health and risk groups was clear, so the model was easy to establish effective discriminant rules. However, due to the long-term exposure of senior students to academic pressure, sedentary behavior, and other risk factors, the distribution of physical fitness indicators tended to be scattered, and the overlapping areas between different risk levels expanded, which led to the decline of model discrimination ability. Although the AUC of XGBoost decreased in the senior samples, the magnitude of this decline was significantly smaller than that of other models, demonstrating robust performance, which was attributed to the model's ability to adaptively partition the high-dimensional feature space and maintain relatively stable decision boundaries amidst complex data distributions. As representatives of integrated learning, RF and GBDT were close to but always slightly inferior to XGBoost. The main reason was that RF's random feature selection strategy might lose some information in the scene of strong correlation between physical fitness data, while GBDT lacked the second derivative optimization and explicit regularization mechanism of XGBoost. The performance of linear model was particularly

poor in senior samples, which further confirmed the nonlinear nature of physical health risk, and simple weight addition could not capture the interaction effect and threshold effect between indicators. In future research, longitudinal tracking data should be considered, and the dynamic characteristics such as the change rate and fluctuation range of physical fitness indexes can be incorporated into the model to improve the prediction accuracy of complex samples in senior grades.

Conclusion

A dynamic risk early warning framework combining hierarchical evaluation and intelligent forecasting was constructed in this study. The results showed that the physical health level of college students was obviously decreasing with grades. XGBoost achieved the best prediction performance in all subgroups. Moderate risk group was the biggest and most urgent intervention target, and the high-risk ratio of female senior students' needs targeted intervention strategies. However, there were still some limitations in this study. The cross-sectional design could not infer the time dynamics of physical decline. Single-institution samples might limit the generalization of the results to a wider group of college students. Standardized physical fitness tests failed to capture health behaviors such as eating habits and sleep quality that might affect the risk level. Future research should include longitudinal design, multi-agency sampling, and comprehensive health behavior assessment to enhance the practical application value of early warning framework.

Acknowledgements

This research was supported by China Student Sports Federation (Grant No. L202504012) and Xiamen Ocean Vocational and Technical College's high-level talent research and entrepreneurship funding program (Grant No. KYG202551).

References

1. Wang JL. 2019. The association between physical fitness and physical activity among Chinese college students. J Am Coll Health. 67(6):602-609.
2. Kumareswaran S. 2023. Detrimental impact of sedentary behaviour on health. Eur J Med Health Sci. 5(1):18-22.
3. Sun J, Sun Z, Kong J, Tian X, Wang L, Wang Q, *et al*. 2024. Regular meta-analysis of the impact of sports activities intervention on some items of the national student physical health standards for adolescents. Front Physiol. 15:1419441.
4. Gresham G, Christopher CN, Hamaya R, Hazra A, Manson JE, Sesso HD. 2025. Use of wearable technology for physical activity and health monitoring in individuals diangnosed with cancer. Cancer Res. 85(8_Supplement_1):7365-7365.
5. Park YH, Suh JH, Kim YW, Kang DR, Shin J, Yang SN. 2022. Machine learning based risk prediction for Parkinson's disease with nationwide health screening data. Sci Rep. 12(1):19499.
6. Ding H, Li N, Li L, Xu Z, Xia W. 2025. Machine learning-enabled mental health risk prediction for youths with stressful life events: A modelling study. J Affect Disord. 368:537-546.
7. Huang AA, Huang SY. 2023. Use of machine learning to identify risk factors for insomnia. PLoS One. 18(4):e0282622.
8. Wu X, Park S. 2023. A prediction model for osteoporosis risk using a machine-learning approach and its validation in a large cohort. J Korean Med Sci. 38(21):e162.
9. Md SA, Kiran M, Lakkampelly A, Ms RS. 2025. Fusion of XGBOOST and random forest for accurate stroke prediction. Int J Sci Res Eng Manag. 9(6):1-7.
10. Alzakari SA, Menaem AA, Omer N, Abozeid A, Hussein LF, Abass IAM, *et al*. 2024. Enhanced heart disease prediction in remote healthcare monitoring using IoT-enabled cloud-based XGBoost and Bi-LSTM. Alex Eng J. 105:280-291.
11. Liang B, Qin W, Liao Z. 2025. A differential evolutionary-based XGBoost for solving classification of physical fitness test data of college students. Mathematics. 13(9):1405.
12. Qin M, Li Y, Cao J. 2025. Application of intelligent algorithms in student mental health management and risk early warning model research. J Comb Math Comb Comput. 127:5379-5397.
13. Liu K, Liao C. 2024. Examining the importance of neighborhood natural, and built environment factors in predicting older adults' mental well-being: An XGBoost-SHAP approach. Environ Res. 262:119929.
14. Cao J, Chen H. 2025. Research on sports injury risk assessment and rehabilitation strategy based on big data analysis. Mol Cell Biomech. 22(3):1081.
15. Iannario M, Romano R, Vistocco D. 2023. Dyadic analysis for multi-block data in sport surveys analytics. Ann Oper Res. 325(1):701-714.
16. Abhadiomhen SE, Nzeakor EO, Oyibo K. 2024. Health risk assessment using machine learning: systematic review. Electronics. 13(22):4405.
17. Mondal S, Nag A, Barman AK, Karmakar M. 2023. Machine Learning-based maternal health risk prediction model for IoMT framework. Int J Exp Res Rev. 32(1):145-159.
18. Zhu Z, Yang Y, Kong Z, Zhang Y, Zhuang J. 2017. Prevalence of physical fitness in Chinese school-aged children: Findings from

the 2016 Physical Activity and Fitness in China—The Youth Study. J Sport Health Sci. 6(4):395-403.
19. Wang Y, Liu J, Liu Y. 2024. Recommended amount of physical activity for Chinese students according to national student physical fitness standards: 321. Med Sci Sports Exerc. 56(10S):103-103.
20. Mohajan D, Mohajan HK. 2023. Body mass index (BMI) is a popular anthropometric tool to measure obesity among adults. J Innov Med Res. 2(4):25-33.
21. Shin S, Yoon WS, Byeon SH. 2022. Trends in occupational infectious diseases in South Korea and classification of industries according to the risk of biological hazards using K-means clustering. Int J Environ Res Public Health. 19(19):11922.
22. Nasser AA, Al-Samawi YA, Alghawli ASA, Essayed AA. 2025. Unveiling health security patterns in the European union through a hybrid entropy-CoCoSo and K-Means clustering framework. F1000Res. 14:600.
23. Kumar D, Sood SK, Rawat KS. 2023. Early health prediction framework using XGBoost ensemble algorithm in intelligent environment. Artif Intell Rev. 56(Suppl 1):1591-1615.