Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Research on the Construction and Application of Predictive Models for Physical Activity Behaviors in Children and Adolescents

Peng Liao
Wuhan Sports University
Liu Mingyi
mingyi84@aliyun.com

Wuhan Sports University
Kong Nianxin
Wuhan Sports University
Gong Haipei
Wuhan Sports University
Cai Jingjie
Wuhan Sports University
Yin Zikang
Wuhan Sports University
Du Zhiwei
Wuhan Sports University
Lu Donglei
Wuhan Sports University
Zhang Kaiyi
Wuhan Sports University

Research Article

Keywords: Machine Learning, Urinary Retention□Children And Adolescents□Physical Activity□Physical Fitness

Posted Date: June 9th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-9538306/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Research on the Construction and Application of Predictive Models for Physical Activity Behaviors in Children and Adolescents

Peng Liao [1,2]，Liu Mingyi [1,2,*] ，Kong Nianxin[1,2]，Gong Haipei[1,2]，Cai Jingjie[1,2]，Yin Zikang[1,2]，Du Zhiwei[1,2]，Lu Donglei[1,2]，Zhang Kaiyi[1,2]

1. China Youth Sports and the Integration of Sports and Education Public Policy Research Center, Wuhan, China； 2. The Research Center for High-Quality Development of Characteristic Competitive Sports, Wuhan Sports University, Wuhan, China

Corresponding author details： Mingyi Liu

No. 461, Luoyu Road, Hongshan District, Wuhan 430079, Hubei, China

027-87191823

mingyi84@aliyun.com

Abstract

Objective Adequate adherence is a necessary condition for success with any intervention. Physical activity is crucial to the physical and mental health as well as the holistic development of children and adolescents, exerting beneficial effects on the enhancement of physical fitness, psychological growth, and social competence. Improving adherence to physical activity training among children and adolescents constitutes a vital prerequisite for ensuring the effectiveness of exercise interventions, promoting their comprehensive physical and mental development, and advancing the implementation of the Healthy China Strategy. However, developing adherence support systems capable of just-in-time adaptive reminders requires understanding the factors that predict adherence, particularly an imminent adherence lapse.

Methods To this end, a random sampling approach was adopted in this study, where children and adolescents participating in physical activity training were divided into two groups. The dataset was then used to predict participants' adherence across different tiers. Subsequently, five machine learning algorithms—logistic regression, k-nearest neighbor, support vector machine, naive Bayes, and decision tree—were employed to construct predictive models for children and adolescents' physical activity behavior. These models were evaluated and compared on the validation set, with their predictive performance assessed using the area under the receiver operating characteristic curve on the test set.

Results A final predictive model was thus developed, which leveraged a range of baseline measures (demographic characteristics, attitudes, and experiences) to predict adherence to physical activity participation among children and adolescents.

Conclusion Ultimately, the logistic regression model outperformed the other four algorithms across all metrics, including accuracy, precision, recall, AUC, and F1-score, in both the training and validation sets. The final logistic regression model achieved an AUC of 0.836 (>0.8) on the test set, indicating a good model fit.

Keywords：Machine Learning；Urinary Retention；Children And Adolescents；Physical Activity；Physical Fitness

Introduction

Overall physical activity levels are low worldwide. The *Global Physical Activity Status Report*(2022), issued by the World Health Organization (WHO) in 2022, reveals that over 80% of children and adolescents globally are encountering a health predicament of insufficient physical activity [1].Physical inactivity is widely recognized as one of the leading causes of mortality and morbidity due to its impact on several noncommunicable diseases [2].However, Physical inactivity has been found to further increase as children enter adolescence, with physical activity levels reducing at an average rate of 7% per year between the ages of 10 years and 19 years [3].Studies have shown that physical inactivity is also associated with obesity, and has been linked to the accelerating global prevalence of childhood obesity Heart disease, stroke, diabetes and multiple cancers since the 1980s [4].More importantly, physical inactivity can have serious consequences on the physical and mental health of young children that can extend into adulthood [5]. Therefore, addressing the gradual decline in physical activity levels, especially in children and adolescents, is an extremely important goal, which has a significant impact on the healthy growth and independence of children and adolescents, but also has a larger social impact, which is related to increasing or decreasing the professional resources for physical education required by individual adolescents.

WHO defines adherence as “the extent to which a person’s behavior agrees with the recommendations of a health care provider.” It is generally recognized that adherence is influenced simultaneously by a host of individual difference factors such

as socioeconomic status, health, cognition, functional ability, and attitudes, as well as properties of the intervention itself. Adherence research has primarily focused on medication, diet, behavioral therapies [6-12], and far fewer studies have examined adherence to physical activities [13].

Although studies of adherence in other domains provide some basis for the understanding of factors that might predict cognitive training adherence, there may also be unique factors important for determining adherence to this type of intervention specifically. Specifically, cognitive training has different requirements and challenges compared to other health interventions.

On the one hand, the allocation of sports resources is uneven in different regions, and the sports environment and atmosphere available to children and adolescents are significantly different [14]. **On the other hand,** the intervention forms of physical activity have different effects on different children and adolescents, which may bring other effects to children and adolescents who participate in physical activity, especially for children and adolescents in the sensitive period of growth and development [15]. Thus, it is unclear whether there are similar or unique factors that predict physical activity adherence in children and adolescents. A detailed understanding of these factors would make it possible to predict adherence and adherence lapses and consequently maximize intervention effectiveness.

Machine learning has emerged as a critical tool in behavioral science, excelling in managing complex nonlinear relationships among variables. It leverages data from diverse sources to identify predictive factors and generate accurate forecasts, thereby

constructing more robust predictive models capable of reliably classifying unknown data into known categories [16]. However, traditional statistical methods used for analyzing variable relationships and building predictive models often fail to effectively validate the original data. For example, some studies have integrated factors such as adolescents' interest in physical exercise and autonomy, employing statistical techniques to establish predictive models linking children and adolescents' physical fitness test indicators with academic and sports performance. Based on these predictions, they further identified key factors influencing academic and sports outcomes to predict levels of physical activity adherence in children and adolescents. Nevertheless, the reliability and accuracy of the constructed predictive models remain relatively limited [17,18]. To address this issue, this study investigates the development of predictive models using five machine learning algorithms: Logistic Regression, k-Nearest Neighbors, Support Vector Machine, Naive Bayes, and Decision Tree. The performance of these models is rigorously evaluated and compared to identify the optimal predictive model. This model aims to assist local physical activity organizations, youth sports professionals, and children and adolescents themselves in systematically identifying and analyzing factors influencing physical activity behaviors. Furthermore, it provides a foundation for designing more targeted and effective physical activity intervention measures or promotion strategies.

Research Objectives

Existing relevant research adopted a two-stage physical activity training intervention, consisting of a 12-week structured training stage with fixed and

standardized exercise schedules, as well as a subsequent 12-week unstructured stage that allowed participants to arrange training independently [19]. The core intervention adopted in the prior research was 12 weeks of continuous aerobic running training, with the primary research aim of verifying whether targeted motivational message interventions could enhance participants' compliance with the aerobic running training scheme. However, the motivational message manipulation adopted in the study failed to produce significant promotional effects. This outcome indicated that individual differential characteristics might serve as the key factors influencing participants' training adherence. In terms of analytical methods, the previous study adopted generalized linear mixed-effects models (GLMM) to conduct predictive analysis on training adherence.

Different from prior research, the present study adopted data from child and adolescent participants of the 2023 national thematic fitness initiative "Run for the Youth", which covered 31 provinces, autonomous regions, and municipalities across China. This research mainly focuses on exploring the predictive performance of machine learning and deep learning algorithms in forecasting adherence to cognitive training games. The dataset utilized in this study integrates multiple dimensions of variables, including participants' demographic profiles, objective and autonomous behavioral indicators of physical activity, after-school exercise experience, and sports interest levels. Additionally, the dataset also covers evaluation indicators of whether participants met the physical activity standards specified by the International Physical Activity Guidelines. Distinct from the traditional single-dimensional adherence

evaluation standard, the newly developed adherence indicators in this research take both the completion quantity of training sessions and the compliance of single-session training duration with experimental requirements into comprehensive consideration.

This study categorized training adherence into minimum and full compliance standards at both overall and weekly levels. On this basis, multiple machine learning and deep learning prediction models were constructed to forecast participants' overall adherence during the 12-week structured training period, as well as their weekly minimum and full adherence status. Two core research questions (RQs) were proposed in this study:

RQ1: Can machine learning models, based on baseline data including demographic characteristics, sports attitudes, and cognitive ability indicators collected at the initial stage of the trial, effectively predict participants' overall adherence to technology-assisted physical activity training programs? Furthermore, which variables exert the most prominent predictive effects?

RQ2: Can deep learning models, utilizing weekly real-time interactive data generated during daily training, accurately predict participants' adherence status in the subsequent week for technology-assisted physical activity training programs? What are the key predictive variables for weekly adherence?

The key theoretical and practical contributions of the present study to the research landscape of physical activity adherence are summarized below:

1. Prior studies have extensively deployed machine learning approaches to forecast behavioral adherence in medication management, clinical rehabilitation, and

dietary intervention research [20,21]. However, relevant applications targeting physical activity compliance among children and adolescents remain scarce, which constitutes an underexplored research niche in the current literature.

2. This study performs a comparative analysis of diverse machine learning models based on multi-dimensional adherence assessment standards. By adopting differentiated adherence definitions for model training and validation, this work achieves comprehensive and systematic prediction of individuals' compliance with physical exercise training regimens.

3. This research successfully extracts and validates critical influencing factors that can effectively predict users' long-term overall training adherence and short-term weekly adherence at varying compliance levels. The identified feature variables offer solid empirical support for follow-up studies on physical activity behavior prediction.

4. The empirical results derived from this research can guide the design of dynamic, real-time intervention strategies. Such targeted intervention schemes are conducive to improving children and adolescents' persistence in physical training and further enhancing the overall effectiveness of public fitness promotion programs.

The remainder of this paper is structured in six sections. Section 2 reviews and synthesizes recent literature focusing on physical activity training adherence. Section 3 elaborates the study framework, data processing workflows, model development pipelines, and model interpretability methods. Section 4 reports and analyzes the experimental outcomes of all predictive models. Section 5 interprets the key findings,

discusses the practical implications for boosting training adherence, and proposes potential avenues for future research. Finally, Section 6 concludes the entire study.

Related Work and Background

In recent year, sports for Health Promotion has been undergoing a transformation towards integrating digital technologies including mobile applications offering dietary advice, physical activity plans, and screen time management, alongside digital technologies such as health monitoring devices, this approach applies behavioral change theory to enhance adherence to physical activity among children and adolescents through goal setting, self-monitoring, and feedback.However, only a few studies have evaluated the effectiveness of different mobile application components in enhancing physical activity adherence and improving health outcomes among children and adolescents.

conducted a review of studies that have examined app components associated with increased engagement. Measures of engagement can be categorized into quantitative measures which are test-based, and qualitative measures which are based on subjectivity (i.e., participant opinions). Across digital tools for different health conditions, three major types of components were found to be associated with increased engagement and usage：(1) diaries and feedback, (2) coaching and education, (3) reminders (e.g., app notifications, text reminders). In particular, text messages or push notifications have been shown to improve the use of apps for mental health [22], and weight loss [23,24]. The review did not identify any studies that have explored strategies and best practices for sending the reminders. Our study aims to facilitate the

development of effective strategies for sending timely reminders to promote adherence in mobile app interventions.

Recently, systematically evaluated factors influencing children and adolescents' willingness and adherence to physical activity participation, with a specific focus on determinants associated with their engagement in physical activity interventions [25]. The study revealed that physical activity behaviors among youth are positively associated with male gender, older age, certain racial backgrounds, higher literacy levels, supportive environments—including family, peers, and teachers—greater self-efficacy, and higher satisfaction with school-based physical education. In contrast, these behaviors are negatively associated with elevated BMI, prolonged sedentary behavior and screen time, tobacco and alcohol use, and the presence of negative emotional states. These findings highlight a range of individual, social, and environmental factors across multiple levels, offering evidence-based guidance for the development of targeted and effective intervention strategies.

From a theoretical research perspective, current physical activity prediction models for children and adolescents suffer from notable limitations. First, the dimensionality of predictors is overly simplistic, with most confined to sociodemographic characteristics or self-reported physical activity data, while neglecting the interactions and dynamic changes among family environment, psychological cognition, and community context [26]. Second, the generalizability of models developed in most studies is inadequate, with significant performance variations across different age groups, genders, and ethnic/racial populations.

Furthermore, these models lack validation of applicability in real-world settings: most remain at the theoretical construction stage and have not been tested for feasibility and effectiveness through clinical or public health intervention practices, thereby failing to meet the practical demands of precision public health [27].

Materials and Methods

The Physical Activity Training Intervention and the Dataset

A randomized controlled trial was conducted among 227 overweight adolescents, who were randomly allocated into three groups: the virtual reality (VR) exercise group, the conventional exercise group, and the control group. The intervention duration was 8 weeks. Participants in the VR exercise group were provided with personalized immersive exercise guidance by an artificial intelligence (AI)-enabled virtual coach, using equipment such as the Oculus Quest 2 head-mounted display, combined with custom exercise accessories and heart rate monitoring sensors. The results showed that adolescents in the VR exercise group achieved an average body fat reduction of 4.28 kg, with no statistically significant difference in fat loss effect compared with the conventional exercise group. Moreover, 6 months after the intervention, the improvements in physical fitness and exercise willingness in the VR group remained sustained. Meanwhile, the injury rate in the VR exercise group was 7.69%, which was significantly lower than the 13.48% in the conventional exercise group. This study demonstrates that technology can enhance adolescents' exercise adherence and the long-term effectiveness of interventions by improving the enjoyment and safety of physical activity [28].

A prospective interventional study was carried out among 100 adolescents with simple obesity, who were divided into two groups: the conventional management group and the peer-supported group. In addition to conventional health management, the peer-supported group was provided with motivational interviewing delivered by certified psychologists, along with full-course peer accompaniment during physical exercise. Following a six-month intervention period, the findings indicated that the full attendance rate for weekly multidisciplinary team consultations in the peer-supported group was 89%, which was significantly higher than the 57% observed in the conventional management group. Moreover, the peer-supported group demonstrated better adherence to moderate-to-vigorous physical activity, achieving a health management effectiveness rate of 83%—almost double the 43% in the conventional management group. This study demonstrates that motivational interviewing can elicit adolescents' intrinsic motivation for exercise, and peer accompaniment can effectively reduce attrition rates. The integration of these two approaches significantly enhances exercise adherence and optimizes intervention efficacy [29].

Overall Design

First, prior to constructing the prediction model, a dedicated Adolescent Physical Activity Adherence Questionnaire was independently developed and specifically designed for the present study, rather than adopted from previously published studies. The study population comprised participants of the 2023 "Run for Health · Youth Initiative" Children and Adolescents Themed Fitness Activity Participant Questionnaire and Satisfaction Evaluation（Supplementary Material 1）, with primary

data covering 31 provinces, autonomous regions, and municipalities directly under the Central Government of China. A stratified random cluster sampling method was employed to recruit 2000 children and adolescents from this population as study participants. To ensure the reliability and accuracy of questionnaire data, a pilot survey was conducted concurrently, which included centralized training for all participants.

Second, the 2000 participants were randomly divided into a training set (n=1600) and a test set (n=400) at an 8:2 ratio. Multivariate Logistic regression analysis and Spearman's correlation coefficient analysis were applied to systematically identify key factors influencing physical activity adherence among children and adolescents.

Finally, the identified key factors were incorporated into the modeling framework. Five machine learning algorithms—Logistic regression, K-nearest neighbor (K-NN), support vector machine (SVM), naive Bayes, and decision tree—were used to construct predictive models for children and adolescents' physical activity behavior. Model performance was evaluated and compared in the test set. The predictive efficacy of each model was quantified by calculating the area under the receiver operating characteristic curve in the test set, and the optimal predictive model was ultimately determined. Figure 1 illustrates the overall design of this study.

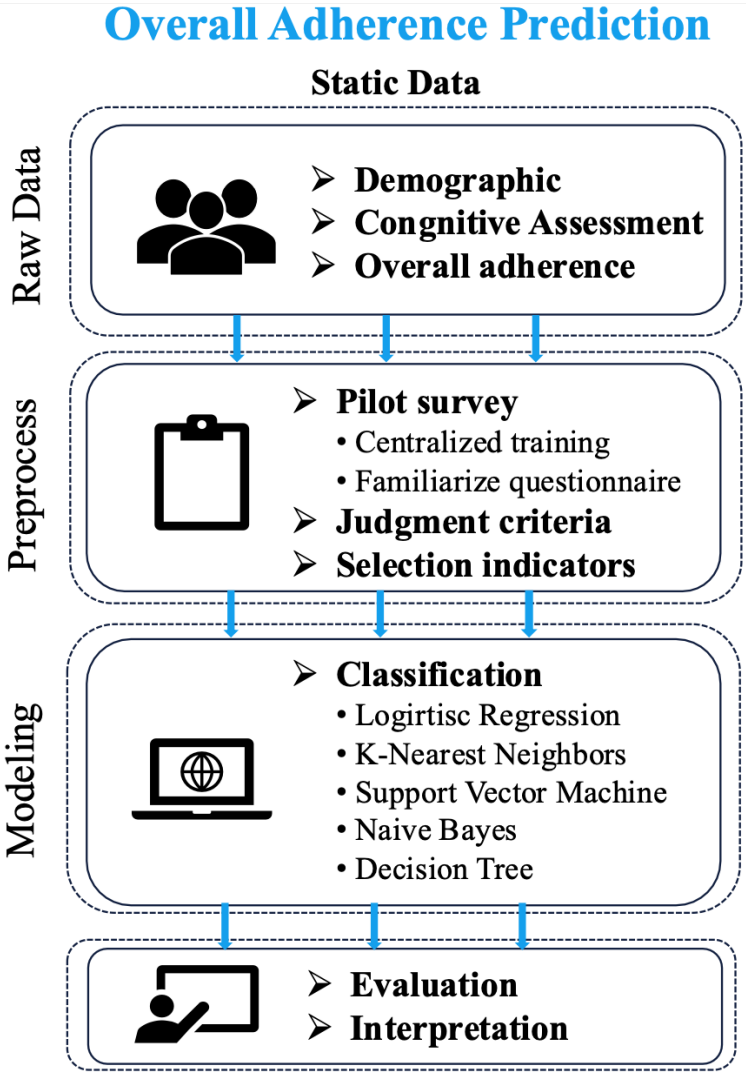


Figure 1 Illustration of the overall design of the study.

Measuring Adherence

First, this study was based on the participant questionnaire from the 2023 "Run for the Youth" National Children and Adolescents Thematic Fitness Campaign. The questionnaire consisted of three components: sociodemographic information, physical activity behavior, and campaign satisfaction. Sociodemographic information included core indicators such as gender, age, region of residence, and grade level. Children and adolescents' physical activity behavior was assessed using the revised Physical Activity Rating Scale (PARS-3) [30], with results presented as scores corresponding to three

activity levels (high, moderate, and low). This scale comprises three dimensions: exercise intensity per session, exercise duration per session, and exercise frequency per week, and employs a 3-point Likert scale (scores ranging from 1 to 3). The weekly physical activity score was calculated as follows: weekly physical activity score = exercise intensity per session score×exercise duration per session score × exercise frequency per week score. Higher scores indicate greater physical activity volume. Specific scoring criteria were defined as follows: exercise intensity per session was categorized into three levels—"negligible intensity" (1 point), "low physical activity intensity" (2 points), and "moderate-to-high or higher physical activity intensity" (3 points); exercise duration per session was divided into three grades—"0–30 minutes" (1 point), "30–60 minutes" (2 points), and "more than 60 minutes" (3 points); exercise frequency per week was classified into three categories—"1–2 times per week" (1 point), "3–4 times per week" (2 points), and "6 or more times per week" (3 points).The criteria for physical activity volume classification were: total score≤9 points (low volume); 10–18 points (moderate volume); and ≥19 points (high volume). The maximum possible score of the scale was 27 points, corresponding to the highest activity level: moderate-to-high or higher intensity per session, duration exceeding 60 minutes per session, and 6 or more sessions per week. Additionally, the test-retest reliability of the scale was 0.82, which met the methodological requirements of this study.

Stringent quality control measures were implemented throughout the study. Prior to the formal survey, a pilot study was conducted to determine the questionnaire

completion time and assess its applicability. All participants received centralized training, with key survey precautions emphasized to ensure full familiarity with the questionnaire content. The research purpose and significance were clearly stated in the questionnaire, supplemented by standardized introductory instructions, and the completion time was restricted to within 20 minutes. During data collation, invalid questionnaires with obvious logical inconsistencies or a missing data rate exceeding 15% were excluded. Overall, the questionnaire was designed with good representativeness, providing a reliable reference for assessing physical activity behavior among children and adolescents.

Second, the primary outcome measure—physical activity compliance—was defined as adherence to the international PA guideline. This definition was based on the core recommendation from the WHO: "Children and adolescents should accumulate at least 60 minutes of moderate-to-vigorous physical activity per day." The variable was dichotomously coded： 1 = adherence to the international physical activity guidelines, 0 = non-adherence to the international physical activity guidelines.

Finally, data entry, collation, and statistical analyses were performed using SPSS 24.0 software and Python 3.6.9. First, descriptive statistical analyses were conducted for participants' baseline characteristics, exercise intensity per session, exercise duration per session, and exercise frequency per week: data following a normal distribution were presented as mean ± standard deviation ($\bar{x}\pm s$). Second, univariate analyses were performed with "adherence to international physical activity guidelines" as the dependent variable to identify statistically significant factors influencing physical

activity behavior among children and adolescents (significance level α=0.05, i.e., P05 was considered statistically significant).Five machine learning algorithms—Logistic regression, K-NN, SVM, Naive Bayes, and decision tree—were implemented on the Python platform to construct predictive models for children and adolescents' physical activity behavior. Model performance was evaluated using four metrics: precision, recall, F1-score (harmonic mean of precision and recall), and the AUC of the ROC. Data were split into a training set (n=1600, 80%) and a test set (n=400, 20%) at an 8:2 ratio. Ten-by-ten cross-validation was applied to the training set to prevent model overfitting and enhance generalizability. Based on the performance evaluation results, the optimal predictive model for children and adolescents' physical activity behavior was ultimately established.

Predicting Overall Adherence

To investigate the predictive value of individual difference indicators for overall physical activity adherence, we first performed univariate analyses on factors associated with physical activity behavior among children and adolescents, selecting variables with statistical significance for inclusion in machine learning model construction. Specifically, based on the training set data and univariate analysis results, eight objective or autonomous indicators were ultimately identified as independent variables: age, gender, urban-rural distribution, weekly exercise frequency, exercise duration per session, exercise intensity per session, extracurricular training experience, and exercise interest. Results showed that five indicators—age, gender, urban-rural distribution, extracurricular training experience, and exercise interest—exhibited

statistically significant differences between the group adhering to international physical activity guidelines and the non-adherent group (all P<0.05) (see Table 1 for demographics).

Table 1 Univariate analysis of factors related to physical activity levels in children and adolescents

Variables	Total [Number (Proportion/%)] （n=2000）	Met the Standard [Number (Proportion/%)] （n=153）	Failed to Meet the Standard [Number (Proportion/%)] （n=1847）	X^2 值	P 值
Age				33.81	<0.001
7-9	527（26.35）	32（20.92）	495（26.80）		
10-12	695（34.75）	46（30.07）	649（35.14）		
13-15	517（25.85）	41（26.80）	476（25.77）		
16-18	261（13.05）	34（22.22）	227（12.29）		
Gender				87.58	<0.001
Male	956（47.80）	92（60.13）	864（46.78）		
Female	1044（52.20）	61（39.87）	983（53.22）		
Region of Residence				98.35	0.103
Northeast China	44（2.20）	1（0.65）	43（2.33）		
Northwest China	62（3.10）	7（4.58）	55（2.98）		
North China	926（46.30）	80（52.29）	846（45.80）		
East China	410（20.50）	28（18.30）	382（20.68）		
Central China	118（5.90）	12（7.84）	106（5.74）		
Southwest China	440（22.00）	25（16.34）	415（22.47）		
Urban-Rural Distribution				35.34	<0.001
Urban	1153（57.65）	112（73.20）	1041（56.36）		
Rural-Township	847（42.35）	41（26.80）	806（43.64）		

Experience of Off-Campus Training				69.67	<0.001
Yes	808（40.40）	111（72.55）	697（37.74）		
No	1192（59.60）	42（27.45）	1150（62.26）		
Sports Interest				146.97	<0.001
Yes	1951（97.55）	152（99.35）	1799（97.40）		
No	49（2.45）	1（0.65）	48（2.60）		

Experimental Design and Evaluation

Model training and optimization were conducted using PyCharm software. First, the Logistic regression, K-NN, SVM, naive Bayes, and decision tree modules from the Scikit-learn library, as well as the XG Boost classifier module from the XG Boost library, were imported. Subsequently, the training and validation sets were loaded: five features selected via univariate Logistic regression analysis from the training set served as input variables, and "whether children and adolescents meet the international physical activity guidelines" was defined as the outcome variable for initial model training. To obtain the optimal hyperparameter combinations and enhance model performance and training efficiency, optuna was employed for hyperparameter tuning of all models. Through a parallel optimization strategy, various attributes of hyperparameters were scanned to select the optimal subset within the minimum number of experiments, thereby determining the best hyperparameter configurations for each model. Following model training and hyperparameter optimization, the validation set was used for unified model validation. The performance of each model was comprehensively evaluated by calculating five metrics: accuracy, precision, recall, AUC, and F1-score [31].

Logistic regression

Logistic regression is a classical classification algorithm belonging to the family of log-linear models. Its core principle is to construct a regression equation for the classification boundary based on existing data, thereby enabling the classification of target data [32,33]. Building on traditional linear regression, this model incorporates the Sigmoid activation function to compress the output of the linear model into the interval [0,1], endowing the result with probabilistic significance. Specifically, the Sigmoid function maps any input to the range [0,1], realizing the conversion from a numerical value to a probability [34]. The mathematical expression of the Logistic regression model is as follows：

$$P(x)=Sigmoid(XW)=\frac{1}{1+e^{(w_1x_1+w_2x_2+....w_kx_k+b)}} \quad (1)$$

Herein, the variable X denotes the feature vector [x_1, x_2, …, x_k, 1], where the trailing "1" serves as a bias term (replacing the intercept b) to integrate the constant term and feature variables in the linear model.

The k-nearest neighbor

The K-NN algorithm is a classical statistical method for classification and regression. Its core principle adheres to the "neighbor similarity" criterion: based on the feature information and corresponding class labels of training samples, the distances between the test sample and all training samples are calculated to identify the k nearest neighbor samples of the test sample. Subsequently, the final class of the test sample is determined by a voting mechanism according to the class distribution characteristics of these k nearest neighbors [35]. Its mathematical expression is as follows：

$$Sim(\frac{\sum_{k=1}^{M} w_{ik} \times w_{jk}}{\sqrt{(\sum_{k=1}^{M} w_{ik})(\sum_{k=j}^{M} w_{jk^2})}}) \quad (2)$$

$$\mu_j(x) = \sum_{i=1}^{K} \mu_j(x_i) sim(X, X_i) \quad (3)$$

Herein, the indicator variable $\mu_j(x) \in \{0,1\}$: $\mu_j(x)=1$ indicates that sample X_i belongs to class c_j, while $\mu_j(x)=0$ indicates that X_i does not belong to class c_j. $sim(X, X_i)$ denotes the similarity between the test sample X and the training sample X_i. The specific decision rule of the algorithm is as follows：

$$\mu_1 = max\mu_j(X), Decision\ X \in c_1 \quad (4)$$

The support vector machine

The SVM is a classical classification algorithm based on the "structural risk minimization" criterion. Its core objective is to find the optimal hyperplane in the feature space that effectively separates samples of different classes, while maximizing the distance (typically Euclidean distance) between the hyperplane and the nearest sample points of each class—a characteristic that endows it with significant advantages in small-sample data scenarios.

$$y(x) = w^T \emptyset(x) + b \quad (5)$$

Herein, ω denotes the weight vector, ∅(x)is the kernel function that maps samples to a high-dimensional feature space, b represents the bias term, and y(x)is the predicted value of sample.

The training process of the SVM essentially involves solving the following convex optimization problem to determine the optimal hyperplane parameters:

$$min 1_{w,b,\xi} \frac{1}{2} \boldsymbol{w}^T \boldsymbol{w} + \boldsymbol{C} \sum_{i=1}^{m} \boldsymbol{\xi}_i \quad (6)$$

$$\{y_i[w^T \phi(x_i + b)] \geq 1 - \xi i, \xi i \geq 0, i = 1,2 \cdots, m\} \quad (7)$$

Where C is a hyperparameter that balances model complexity and classification error; ξi are slack variables allowing a certain degree of classification tolerance, enabling the model to adapt to data scenarios that are not perfectly linearly separable.

Naive Bayes

Naive Bayes is a classical classification algorithm based on Bayes' theorem and the "conditional independence assumption of features" [36]. Its core logic involves using the prior probabilities of features to derive their posterior probabilities via Bayes' theorem, and classifying samples by selecting the class with the maximum posterior probability.Within the framework of probability statistics, let event A (representing the class) and event B (representing the feature set) be two events in the same sample space. Bayes' theorem can be derived from the conditional probability formula:

$$P(A|B)=\frac{P(B|A)P(A)}{P(B)} \quad (8)$$

Herein，P(A) denotes the prior probability, representing the prior distribution probability of class A; P(B)=0(a necessary premise for the probability formula, as the denominator cannot be zero), P(B|A) is the conditional probability (also referred to as the likelihood probability), indicating the probability of the feature set B occurring given the class A; and P(A|B) is the posterior probability, i.e., the posterior inferred probability of class A observed based on the feature set B.

Decision tree

The decision tree is a fundamental supervised learning algorithm widely applied to data classification tasks [37]. Its core objective is to construct a highly interpretable

predictive model that learns implicit classification rules from the training dataset to accurately predict the class labels of test samples.

The structure of a decision tree consists of two types of nodes: leaf nodes and internal nodes. Leaf nodes correspond to the final classification results, with their class labels determined by a majority voting mechanism among the training samples reaching the node; internal nodes represent feature-based decision judgments, generating branches based on different values of the selected feature to direct samples to corresponding child nodes.The core of the decision tree model construction (taking the impurity-based splitting criterion such as Gini coefficient as an example) can be expressed as:

$$H(D)=-\sum_{k=1}^{k} p_i log(p_i)=-\sum_{k=1}^{k} \frac{|C_k|}{|D|} log \frac{C_k}{D} \quad (9)$$

Herein, D denotes the total number of samples in the dataset at the current node, K is the total number of classes of the target variable, and $|C_k|$ represents the number of samples belonging to the k-th class in the current dataset. This formula quantifies the impurity of the node (e.g., Gini coefficient) to guide the optimal feature splitting.

Results

Adherence Distribution

In this study, the surveyed children and adolescents were assigned to two groups: the first group underwent physical activity training4–7times per week, whereas the second group participated in physical activity only 1–2 times weekly. By analyzing the distribution characteristics of physical activity adherence across the two groups, the study sought to explore the association between exercise frequency and adherence, and

to verify whether higher weekly physical activity frequency correlates with superior adherence levels. Figure 2 depicts the overall adherence distribution of the two cohorts. The results reveal that both distributions exhibit a bimodal profile, a hallmark indicative of substantial heterogeneity in physical activity adherence among the pediatric and adolescent population. Moreover, the findings further confirm that children and adolescents engaging in physical activity 4–7 times weekly demonstrated significantly higher adherence compared with their low-frequency counterparts.

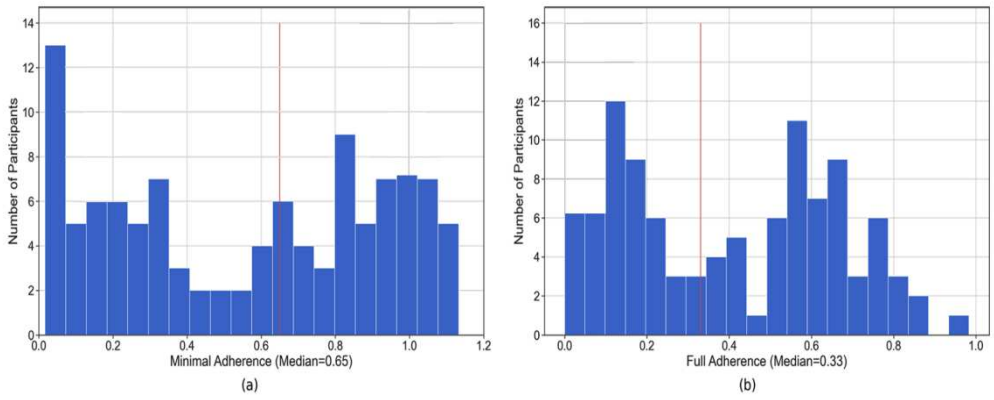


Figure 2 Distribution of overall adherence at (a) minimal level; and (b) full level.

Comparison of Five Machine Learning Predictive Models

We implemented the analysis via PyCharm, importing the logistic regression, k-nearest neighbors, support vector machine, naive Bayes, and decision tree modules from the Scikit-learn library, alongside the XGBoost classifier module from the XGBoost library. Using the training dataset, we selected 5 features (filtered via univariate logistic regression) as input variables, with "whether children and adolescents met the physical activity recommendations of the International Physical Activity Guidelines" as the outcome variable, to train the models. Optuna was

employed for hyperparameter optimization to determine the optimal parameter configurations for each predictive model. Following model training, all models were validated uniformly using the hold-out validation dataset, and metrics including accuracy, precision, recall, AUC, and F1-score were computed for each model.

The results demonstrated that all five machine learning algorithms could predict the adherence of children and adolescents to physical activity recommendations to some extent. Among these, the logistic regression model exhibited the strongest performance: it achieved a precision of 83.5% and a recall of 83.9% in identifying compliant physical activity behavior, with an F1-score of 84.9% (indicating an excellent balance between precision and recall). Its AUC value reached 83.6%, further emphasizing its superior ability to discriminate between compliant and non-compliant physical activity levels in this population. The naive Bayes model performed second-best: while its AUC value (82.3%) exceeded 80%, its accuracy, precision, recall, and F1-score were all lower than those of the logistic regression model. The performance differences across machine learning predictive models are summarized in Table 2.

Table 2 Comparative analysis of differences between different machine learning prediction models（n=1600）

Evaluation Metrics（%） / Algorithm	Accuracy	Precision	Recall	AUC	F1-Score
Logistic Regression	86.7%	83.5%	83.9%	83.6%	84.9%
K-Nearest Neighbors (K-NN)	83.1%	82.6%	83.9%	79.6%	81.9%

Support Vector Machine (SVM)	84.3%	86.6%	85.1%	76.3%	79.4%
Naive Bayes	76.5%	74.3%	75.1%	82.3%	79.6%
Decision Tree	72.6%	74.4%	73.1%	79.4%	77.7%

Final Construction of the Machine Learning Predictive Model

A stepwise regression approach was adopted to incorporate five predictive variables with statistically significant differences (P<0.05) identified in the univariate analysis into the logistic regression model, namely age, gender, urban-rural distribution, and sports interest. The definitions and coding criteria of these variables are summarized in Table 3, while the detailed results of the logistic regression analysis (including regression coefficients, odds ratios [ORs], and 95% confidence intervals [CIs]) are presented in Table 4.

Table 3 Variables and Assignments

Variable Name	**Coding Criteria**
Compliance with International Physical Activity Guidelines (Y)	No=0，Yes=1
Age (X_1) (Reference group: 7–9 years old)	
10–12 years old (X_{1a})	No=0，Yes=1
13–15 years old (X_{1b})	No=0，Yes=1
Gender（X_2）	Female=0; Male=1
Urban-Rural Distribution（X_3）	Rural=0; Urban=1
Sports Interest（X_4）(Reference group: Rarely exercise)	
Occasionally exercise （X_{4a}）	No=0，Yes=1
Frequently exercise（X_{4b}）	No=0，Yes=1
Extracurricular Training Experience（X_5）	No=0，Yes=1

As indicated by the logistic regression results in Table 4, children and adolescents aged ≥10 years, male, residing in urban areas, engaging in regular physical activity, and with extracurricular physical activity training experience were significantly more likely to meet the International Physical Activity Guidelines and exhibited a stronger propensity for physical activity participation. Based on these statistically significant predictive factors, the predictive model for physical activity

behavior in children and adolescents established in this study is as follows：P（Y=1）=1/{1+exp[-(1.256+0.507X1a+0.803X1b-0.814X2+1.742X3+2.438X4a+3.524X4b+1.807X5)]}。

Table 4 Logistic regression analysis results

Variable	Regression Coefficient	Standard Error (SE)	Wald χ^2 Value	P Value	Odds Ratio (OR)	95% Confidence Interval (95%CI)
Age (Reference group: 7–9 years old)			11.672	＜0.010		
10–12 years old	0.507	0.150	4.800	0.013	1.725	(1.013,2.437)
13–15 years old	0.803	0.161	10.785	＜0.010	2.348	(1.364,3.332)
Gender (Reference group: Male)						
Female	-0.814	0.056	29.502	＜0.010	1.412	(0.707,2.117)
Urban-Rural Distribution (Rural)						
Urban	1.742	0．644	2.511	＜0.010	2.980	(1.772,6.452)
Sports Interest (Reference group: Rarely exercise)			30.702	0.015		
Occasionally exercise	2.438	0.145	45.765	＜0.010	1.465	(0.815,2.280)
Frequently exercise	3.524	0.213	32.132	＜0.010	2.698	(1.973,2.698)
Extracurricular Training Experience (Reference group: No)						
Yes	1.807	0.161	68.278	＜0.010	2.523	(1.767,2.523)
Constant	1.256	0.208	15.578	＜0.010	2.365	

Evaluation of the Performance of the Final Predictive Model

corresponding variable values of children and adolescents in the test set were substituted into the aforementioned model to calculate the predictive probabilities of their physical activity behaviors. Based on the actual measured results, the ROC curve was plotted, as shown in Figure 3. The AUC of the predictive model for physical

activity behaviors in children and adolescents was 0.836>0.80, indicating a good fitting effect of the model. The P-value of the HL goodness-of-fit test was 0.815>0.05, which demonstrated that the model passed the HL test with a high degree of consistency between the predicted and observed values. In conclusion, the predictive model for physical activity behaviors in children and adolescents established in this study exhibits satisfactory fitting performance.

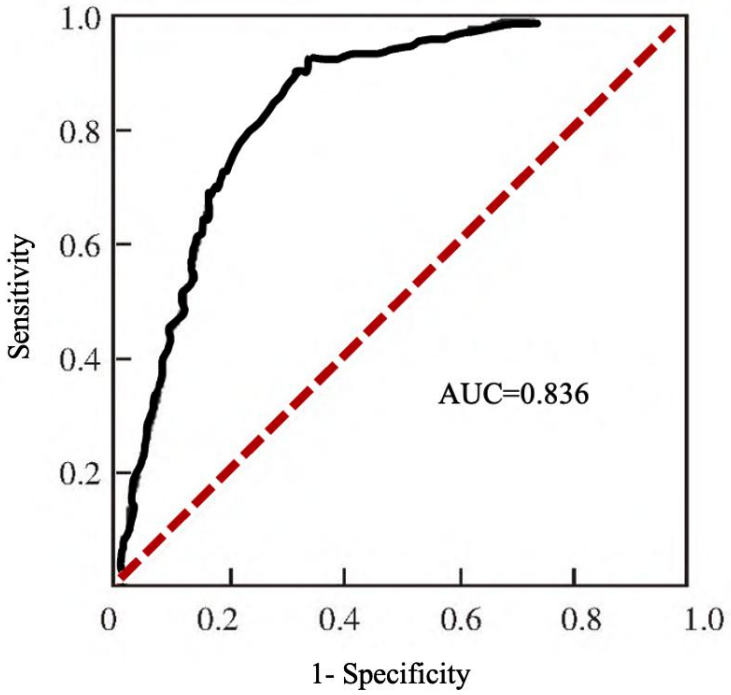


Figure 3 ROC curves for the predictive model of physical activity behavior of children and adolescents in the test set

Discussion

Main Findings

To the best of our knowledge, this is the first study that uses machine learning to predict adherence to cognitive training programs. Adherence constitutes a core determinant of the effectiveness of physical activity interventions, and its quantitative assessment directly validates the real-world efficacy of such programs [38].This study

makes several original contributions to the experimental design of adherence-focused research on physical activity among children and adolescents. First, we defined distinct adherence levels (i.e., minimal adherence and full adherence), systematically analyzed the associations between demographic characteristics, subjective attitudes, motor skills, and relevant experiences in physically active children and adolescents, and quantitatively evaluated the overall adherence of this population to physical activity programs across different tiers. The findings from this factor analysis provide empirical insights for interpreting the distribution characteristics of physical activity adherence across age groups of children and adolescents, as well as for formulating targeted strategies to enhance adherence. Second, we employed machine learning algorithms to develop predictive models for children and adolescents' adherence to physical activity training protocols. Notably, these models can accurately identify the critical junctures at which adherence declines during the intervention process. The output data serve as an evidence base for the timely design and implementation of adherence-promoting strategies, thereby facilitating the optimization and scaling-up of physical activity interventions.

According to the findings, general self-efficacy, objective cognitive and memory measures with delayed recall, and technology self-efficacy are most predictive of a participant's overall adherence. Various studies have also shown a strong correlation between self-efficacy, treatment adherence and intervention outcomes [39].

Regarding weekly adherence prediction, the models predicting adherence at both minimal level and full level achieved high performance. We found that children and

adolescents who engaged in physical activity training more frequently per week exhibited significantly higher levels of adherence. Furthermore, we found that the logistic regression model could better capture the complex linear relationships among variables and identify the predictive factors for physical activity behavior in children and adolescents, thus emerging as the optimal predictive model for this population's physical activity behavior. Additionally, the employment of multiple machine learning algorithms allowed us to identify the top-performing algorithm in the context of the present dataset, which facilitates the selection of machine learning predictive models with optimal or relatively superior performance.

Theoretical and Practical Implications for Adherence Promotion

Using the model to predict children and adolescents' physical activity behaviors and provide high-quality and balanced youth sports resources

Our findings indicate that regional disparities constitute a crucial factor influencing physical activity participation among children and adolescents, with urban participants exhibiting significantly higher levels of engagement than their rural counterparts—a result consistent with those of numerous previous studies. Research has demonstrated that external resources from both governmental and non-governmental organizations are relatively concentrated in urban areas compared to rural regions, directly enabling urban children and adolescents to access more abundant sports resources [40]. Additionally, other studies have shown that substantial regional variations in China’s economic and social development directly contribute to the

uneven distribution of specialized sports resources across urban and rural areas, different regions, and individual schools. Rural areas face deficiencies in sports facility allocation, the composition of physical education faculty, educational philosophies, and the implementation of physical education programs, indicating that the quality and standard of rural physical education still require further improvement [41,42].

In this study, we established a predictive model for physical activity behavior among children and adolescents, using a sample of 2,000 participants from the 2023 "Run for the Youth" National Fitness Campaign for Children and Adolescents. Based on the comparison of AUC values during model construction, and given that the dependent variable in this study was binary, the logistic regression model was selected for analysis. Compared with other algorithms, this analytical approach offers high computational efficiency, along with strong model interpretability and applicability [43].The application of the predictive model can effectively identify disparities in the allocation of sports resources for children and adolescents across different regions, and support targeted policy-making and timely adjustments. This not only provides a guarantee for the high-quality and balanced allocation of sports resources in various regions but also offers scientific references for improving the physical health of children and adolescents. Specific suggestions are as follows：

First, optimize the allocation of physical education (PE) teachers. All regions are encouraged to explore the establishment of a regional shared mechanism for PE teacher resources. Through models such as grouped school-running, regular exchanges and seminars, and the construction of urban-rural education communities, PE teachers are

promoted to teach across schools and educational stages, facilitating the balanced distribution of PE teacher resources among schools, between urban and rural areas, and across regions to meet the basic needs of PE teaching. Meanwhile, all regions should formulate hierarchical and classified training programs for PE teachers, implement periodic and regular training, and focus on enhancing the professional knowledge reserve and teaching competence of rural PE teachers.

Second, improve the supply of sports venues and equipment. All regions are encouraged to explore the establishment of a co-construction and sharing mechanism for venues among schools, sports schools, youth sports clubs, and other institutions. Combined with the actual situation of local youth sports development, social sports venue resources should be reasonably activated to improve the utilization rate of sports facilities, providing solid support for the development of youth sports activities.

Third, strengthen financial guarantee. All regions are encouraged to construct an urban-rural integrated financial input mechanism for youth sports that emphasizes balance. Through differentiated per-student funding subsidies, comprehensive rewards and subsidies, and regional resource coordination, necessary financial support is provided for the development of youth sports activities, PE teaching, and training in economically underdeveloped areas.

Using the model to predict children and adolescents' physical activity behaviors and design adaptive youth physical activity programs

Our preliminary findings revealed that demographic factors including age and gender are also pivotal determinants of physical activity participation among children

and adolescents. Specifically, the proportion of individuals meeting the recommended levels of physical activity specified in international guidelines increases progressively with age. In China, the detection rate of moderate-to-vigorous physical activity (MVPA) among children and adolescents has exhibited an overall upward trend. This pattern can be attributed to two primary factors: first, preschool children are characterized by a severe lack of MVPA, stemming from the fact that early childhood educators rarely organize MVPA-focused school-based physical activities due to safety concerns. Second, adolescents aged 13–15 years are in a critical period of puberty and represent an active cohort for physical activity engagement, demonstrating a strong willingness to participate in diverse forms of physical activities. Furthermore, the implementation of the physical education assessment system incorporated into the senior high school entrance examination has further prompted schools and parents to elevate their emphasis on, and enhance support for, physical activity among children and adolescents [44].

The prediction model for children and adolescents' physical activity behavior established in this study can effectively identify whether an individual's physical activity level meets the recommended standards specified in international physical activity guidelines, and is capable of formulating personalized adaptive physical activity programs for children and adolescents across different age groups and genders. The specific application pathways are as follows：

First, targeted intervention measures should be formulated and implemented based on the laws of growth and development and the characteristics of motor ability

development in children and adolescents. It is recommended that schools at all levels, youth sports clubs, and other relevant sports organizations design and implement adaptive physical activity intervention programs according to the developmental characteristics of children and adolescents in different academic stages and based on their actual demands for sports participation. Meanwhile, a dynamic monitoring support mechanism for children and adolescents' health should be established to help families, schools, and adolescents themselves grasp health status in real time, promote home-school collaboration, and provide scientific guidance for children and adolescents to engage in effective physical activities. In addition, adolescents can be accurately classified into subgroups based on differences in physical activity levels, and stratified intervention programs can be scientifically designed to improve the physical health of children and adolescents in a targeted manner.

Second, the integrated design of physical education curricula spanning primary, junior high, and senior high schools should be strengthened to cultivate children and adolescents' capacity and habits for lifelong physical exercise. Schools serve as the core venues for children and adolescents to engage in physical activities. Therefore, it is advisable for all types of schools to follow the laws of physical and mental development of children and adolescents, actively construct a coherent physical education curriculum system covering primary, junior high, and senior high schools, and adopt a teaching model combining structured knowledge imparting and skill training to provide professional physical education guidance for students in different academic stages. This approach can effectively avoid the drawbacks of fragmented curriculum design such as

superficial and repetitive teaching, address the practical problem that students fail to master a single specialized sports skill despite receiving over a decade of physical education, and truly realize the continuity, in-depth engagement, and practical applicability of children and adolescents' sports participation, thereby helping them develop good habits of lifelong physical exercise.

Using the model to predict children and adolescents' physical activity behaviors and implement standardized youth physical activity strategies

Our exploratory analysis revealed that sports interest and extracurricular physical activity training experience are two independent factors influencing physical activity behaviors among children and adolescents. Specifically, those with strong sports interest and a history of extracurricular physical activity training generally achieve the recommended physical activity levels specified in international guidelines. Relevant studies have explored the moderating effects of motivation and interest on the relationship between the social environment and adolescents' physical activity. By constructing structural equation models, these studies confirmed that adolescents' engagement in physical activities is directly correlated with their positive sports interest [45]. Lindley conducted a 3-year empirical study to explore the correlation between three dimensions of psychological factors—perceived competence, autonomy, and intrinsic motivation—and physical activity behaviors among children and adolescents. Their findings indicated that the level of need satisfaction was significantly positively correlated with intrinsic motivation, physical activity intention, and motor competence. These results suggest that enhancing need satisfaction and fostering awareness of

autonomous physical activity can effectively promote children and adolescents' engagement in physical activities and improve their physical fitness [46].

The prediction model for children and adolescents' physical activity behavior established in this study can provide scientific references for the organization of physical activities by youth sports institutions such as schools and youth sports clubs, thereby giving full play to the core role of physical activities in promoting the healthy growth of children and adolescents. The specific implementation suggestions are as follows：First, adopt an interest-oriented approach and continuously optimize activity implementation strategies. It is recommended that youth sports organizations proactively assume their social responsibility for promoting adolescents' physical fitness, take cultivating sports interest and developing exercise habits as the core orientation, and combine the laws of physical and mental development of children and adolescents with the rules of motor skill formation. Taking life skill training, motor skill practice, sports competition organization, and characteristic project development as the core carriers, efforts should be made to continuously enrich and improve the physical activity content system, and design and carry out innovative, interesting, and diversified physical activities that are compatible with the developmental characteristics of children and adolescents at different age groups. Meanwhile, physical activity guidelines adapted to different developmental stages of children and adolescents should be formulated, and activity content should be scientifically designed, and activity intensity and volume should be regulated in accordance with guideline standards, with the establishment of a regular feedback and evaluation mechanism. Second, take

standardization as the fundamental principle and enhance the effectiveness of activity implementation supervision. Local governments need to further refine the management norms for youth sports social organizations such as schools and youth sports clubs, focusing on strengthening supervision over the formulation of activity plans, activity safety control, qualification review of practitioners, compliance of activity venues, and suitability of activity content. This will ensure the safe and orderly conduct of various physical activities and promote the healthy and sustainable development of youth sports organizations.

Limitation and Future Work

Several limitations should be noted in this study. The models we developed in this study are specific to our dataset. First, we will further conduct work to demonstrate the generalizability of the models to other domains. Nonetheless, we believe that the framework of using machine learning to predict adherence at a certain interval can be generalized to other mobile interventions. Second, this study has not yet conducted cluster analysis based on the adherence data of participating children and adolescents, and thus failed to identify subgroups with similar adherence patterns. In subsequent research, we will further analyze the correlations between participants' demographic characteristics, subjective attitudes, activity levels, and adherence patterns, and develop more targeted machine learning-based predictive models for specific subgroups of participants with distinct adherence patterns.

Conclusions

In this study, a random sampling method was adopted to collect physical activity data of children and adolescents covering 31 provinces (autonomous regions and municipalities directly under the Central Government) nationwide, aiming to evaluate the feasibility of using machine learning and deep learning algorithms for predicting adherence to physical activity training. The results indicated that children and adolescents who engaged in physical activity training more frequently per week exhibited significantly higher levels of training adherence. Five machine learning algorithms—logistic regression, K-NN, SVM, Naive Bayes, and decision tree—were employed to construct predictive models for physical activity behavior among children and adolescents. Comprehensive evaluation and comparative analysis revealed that both the logistic regression and naive Bayes models achieved an AUC of over 75% on the test set, exhibiting favorable predictive efficacy for children and adolescents' physical activity behavior. This finding indicates that the two models can effectively incorporate multi-dimensional influencing factors and resolve the complex interactions among these factors, thus demonstrating practical applicability in predicting physical activity behavior in this population [47]. Moreover, they provide novel technical tools and implementation approaches for the early identification and intervention of physical activity participation among children and adolescents. Further results of this study showed that, compared with the naive Bayes algorithm, the logistic regression model exhibited superior performance and generalization ability in constructing predictive models for children and adolescents' physical activity behavior, with a predictive accuracy of up to 83.6%. In conclusion, in terms of predictive objectives and

application outcomes, the logistic regression model is more suitable for identifying the influencing factors of physical activity behavior in children and adolescents and accurately predicting their future physical activity behavior.

Acknowledgements

We would like to thank the participants for their involvement in this study.

Authors' contributions

PLA was in charge of questionnaire distribution, data analysis, and text writing. LMY provided overall control of the study. KNX, GHP, CJJ, YZK, DZW, LDD and ZKY revised the manuscript. All authors have read and agreed to the published version of the manuscript.

Funding

Not applicable.

Data and materials Availability

Data will be made available on request, if necessary, please contact the authors.

Declarations

Ethics approval and consent to participate

Firstly, this study strictly follows the principles mentioned in the Declaration of Helsinki, and all methods were carried out in accordance with the relevant guidelines and regulations. Secondly, this study has been endorsed by the Ethics Committee of

Wuhan Sports University. Finally, informed consent was obtained from all the children and adolescents for this study. For all participants under the age of 16, written informed consent was additionally obtained from their parents or legal guardians.

Consent for publication

Not applicable.

Competing interests

The authors declare no competing interests.

References

[1] Van, S., Ekelund, U., Crochemore-Silva, I.,2021. Physical activity behaviours in adolescence: current evidence and opportunities for intervention. Lancet.31,398(10298),429-442.

[2] Kohl, H.W., Craig, C.L., Lambert, E.V., Inoue, S., Alkandari, J.R., Leetongin, G., Kahlmeier, S.,2012. Lancet physical activity series working G. The pandemic of physical inactivity:global action for public health. Lancet.380(9838):294–305.

[3] Dumith, S.C., Gigante, D.P., Domingues. M.R., Kohl, H.W.,2011. Physical activity change during adolescence: a systematic review and a pooled analysis. Int J Epidemiol.40: 685–98.

[4] Han, J.C, Lawlor, D.A, Kimm, S.Y. Childhood obesity., 2010. Lancet.375: 1737–48.

[5] Beauchamp, M.R., Puterman, E., Lubans, D.R.,2018. Physical inactivity and mental health in late adolescence. JAMA Psychiatry.75:543–44.

[6] García-Casares, N., Gallego, F.P., Barbancho, M., López-Gigosos, R., García-Rodríguez, A., Gutiérrez-Bedmar, M.,2021. Alzheimer's Disease, Mild Cognitive Impairment and Mediterranean Diet. A Systematic Review and Dose-Response Meta-Analysis. Journal of Clinical Medicine.10(20), 4642. 10.3390/jcm10204642.

[7] Koutsonida, M., Kanellopoulou, A., Markozannes, G., Gousia, S., Doumas, M.T., Sigounas, D.E., Tzovaras, V.T., Vakalis, K., Tzoulaki, I., Evangelou, E., Rizos, E.C., Ntzani, E., Aretouli, E., Tsilidis, K.K.,2021. Adherence to Mediterranean Diet and Cognitive Abilities in the Greek Cohort of Epirus Health Study. Nutrients.13(10), 3363. 10.3390/nu13103363.

[8] Kuo, W.Y., Chen, MC., Lin, Y.C., Yan, S.F., Shyu,Y.I., 2021. Trajectory of adherence to home rehabilitation among older adults with hip fractures and cognitive impairment. Geriatric Nursing.42(6), 1569–1576.

[9] Koesmahargyo, V., Abbas, A., Zhang, L., Guan, L., Feng, S., Yadav, V., Galatzer-Levy, I.R., 2020. Accuracy of machine learning-based prediction of medication adherence in clinical research. Psychiatry Research. 294, 113558.

[10] Lo-Ciganic, W.H., Donohue, J.M., Thorpe, J.M., Perera, S., Thorpe, C.T., Marcum, Z.A, Gellad, W.F.,2015. Using Machine Learning to Examine Medication Adherence Thresholds and Risk of Hospitalization. Medical Care.53(8), 720–728.

[11] Kim, N., McCarthy, D.E., Loh, W.Y, Cook, J.W., Piper, M.E., Schlam, T.R., Baker, T.B.,2019. Predictors of adherence to nicotine replacement therapy: Machine learning evidence that perceived need predicts medication use. Drug and Alcohol Dependence.205, 107668.

[12] Scioscia, G., Tondo, P., Foschino, M.P, Sabato, R., Gallo, C., Maci, F., Lacedonia, D.,2021. Machine learning-based prediction of adherence to continuous positive airway pressure (CPAP) in obstructive sleep apnea (OSA). Informatics for Health and Social Care.0(0), 1–9.

[13] Turunen, M., Hokkanen, L., Bäckman, L., Stigsdotter-Neely, A., Hänninen, T., Paajanen, T., Soininen, H., Kivipelto, M., Ngandu, T.,2019. Computer-based cognitive training for older adults: Determinants of adherence. PLOS ONE.14(7), e0219541.

[14] Peng, L., Yang, Z., Chang, D., Liu, C.,2025. Disparities in sports participation behavior among Chinese residents. Sci Rep.Mar 14;15(1),8788.

[15] Hashemi Javaheri, F.S., Mousavi, Z., Mohammadi, S., Amjadi, A., Mobarakeh, K.A., Hayedeh Mousavi Shalmani, S., Torkaman, M., Mirshafaei, M.A., Khoshdooz, S., Saeedirad, Z., Kalantari, N., Mirzaei, P., Doaei, S., Gholamalizadeh, M., 2025. The effect of a comprehensive intervention on anthropometric indices, dietary intake, and physical activity of adolescent boys with overweight. BMC Pediatr.25(1), 67.

[16] Dong, J., Feng, T., Thapa-Chhetry, B., Cho, B.G., Shum, T., Inwald, D.P., Newth, C.J.L., Vaidya, V.U., 2021. Machine learning model for early prediction of acute kidney injury (AKI) in pediatric critical care. Crit Care.25(1), 288.

[17] Hu, X., 2023. Research on intelligent prescription of college students' physical health based on machine learning [D]. Xihua University.

[18] Guo, X., 2023. How family structure influences middle school students' participation in physical exercise and their academic performance. TRAMES-J humanit soc.10(1), 1–9.

[19] Deborah, M.T., Rebecca, M.M., Silmara, G., 2025. Moving beyond moderate-to-vigorous physical activity: A longitudinal study on adherence to 24-hour Movement Guidelines in adolescents. Journal of Science and Medicine in Sport.28(2), 147-153.

[20] Jun, W.D., Min, Y.C., Yi, R.X., 2025. A predictive healthcare model using machine learning and psychological factors for medication adherence. Healthcare Analytics.7, 100397.

[21] Wellington, K., Absalom, E.E., Tsitsi, M., Jean, V., 2025. Machine learning-based classification of medication adherence among patients with noncommunicable diseases. Informatics in Medicine Unlocked.52, 101611.

[22] Bidargaddi, N., Almirall, D., Murphy, S., Nahum-Shani, I., Kovalcik, M., Pituch, T., Maaieh, H., Strecher, V., 2018. To Prompt or Not to Prompt? A Microrandomized Trial of Time-Varying Push Notifications to Increase Proximal Engagement With a Mobile Health App. JMIR MHealth and UHealth.6(11), e10123.

[23] Dolan, P.T., Afaneh, C., Dakin, G., Pomp, A., Yeo, H.L., 2019. Lessons Learned From Developing a Mobile App to Assist in Patient Recovery After Weight Loss Surgery. Journal of Surgical Research.244, 402–408.

[24] Morrison, L.G., Hargood, C., Lin, S.X., Dennison, L., Joseph, J., Hughes, S., Michaelides, D.T., Johnston, D., Johnston, M., Michie, S., Little, P., Smith, P.W., Weal, M.J., Yardley, L., 2014. Understanding Usage of a Hybrid Website and Smartphone App for Weight Management: A Mixed-Methods Study. Journal of Medical Internet Research.16(10), e201.

[25] Shao, T., Zhou, X., 2023. Correlates of physical activity habits in adolescents: A systematic review. Frontiers in Physiology.14, 1131195.

[26] Zhao, C., Yuan, J., Huang, W.W., 2025. Multilevel determinants of physical activity in children and adolescents: a meta-analysis guided by social ecological model. BMC Sports Science, Medicine and Rehabilitation, 17(1), 1-12.

[27] Jones, A.B., Smith, C.D., Lee, E.F., 2024. Generalization and validation of youth physical activity prediction models: A systematic review. Preventive Medicine Reports. 19, 102896.

[28] Wang, J., Qin, Y., Li, H., 2025. An adaptive AI-based virtual reality sports system for adolescents with excess body weight: a randomized controlled trial. Nature Medicine.31(6), 2255-2268.

[29] Li, M., Zhang, Y., Wang, L., Chen, J., 2023. Effects of motivational interviewing combined with peer support on exercise adherence and health management efficacy in adolescents with simple obesity: A randomized controlled trial. Chinese J Contemp Pediatr.25(4), 412-418.

[30] Liang, D.Q., 1994. Stress levels of college students and their relationship with physical exercise. Chinese Mental Health Journal.8(1), 5-6.

[31] Fang, Z., Yu, S., 2022. Prediction of potential civil aviation passengers based on IGA-Optuna-LightGBM. Foreign Electronic Measurement Technology.41(10), 142-147.

[32] Ramsdale, E., Kunduru, M., Smith, L., 2023. Supervised learning applied to classifying fallers versus non-fallers among older adults with cancer. Journal of Geriatr Oncol.14(4), 101498.

[33] Zhu, Z.X., Sun, J.W., Li, X.S., 2022. A construction method of scorecard using machine learning and logical regression. Procedia Computer Sci.214, 1541–1545.

[34] Wang, F.B., Wang, M., 2018. Motivation and interest: Mediating factors between social environment and urban adolescents' physical activity. China Youth Studies.5, 76-83+67.

[35] Liu, C., Yan, S.Y., Zha, H.B., 2012. A study on KNN classification of EEG signals for multiclass motor imagery tasks. Chinese Journal of Scientific Instrument.33(8), 1714-1720.

[36] Li, Z.F., Li, Y., Li, A., 2020. Public opinion analysis and application based on Hadoop and machine learning. Application Research of Computers.37(S1), 43-46.

[37] Safavian, S., Landgrebe, D.A., 1991. Survey of decision tree classifier methodology. IEEE Transactions on Systems, Man, and Cybernetics.21(3), 660-674.

[38] André, N., Grousset, M., Audiffren, M., 2024. A Behavioral Perspective for Improving Exercise Adherence. Sports Med - Open.10, 56.

[39] Okuboyejo, S., Mbarika, V., Omoregbe, N., 2018. The effect of self-efficacy and outcome expectation on medication adherence behavior. Journal of Public Health in Africa.9(3), 826.

[40] Huo, L.T., Wang, G.X., 2019. Spatial Equilibrium Changes of Preschool Education Resources in China: On the Implementation Effectiveness of Two Phases of the "Three-Year Action Plan for Preschool Education". Journal of Modern Education Management.10, 14-20.

[41] Peng, X.W., Wang, B., Mao, Z.M., 2024. School Physical Education Modernization for "a Vast Population Scale": Connotations, Advantages, Challenges and Promotion Strategies. Journal of Wuhan Institute of Physical Education.58(11), 11-17.

[42] Zhao, P.T., Guo, L.Y., Chen, K., Huang, L., 2024. Practical Dilemmas and Development Pathways of Rural School Physical Education. Sports Culture Guide.6, 1-5.

[43] Li, Y.L., Chen, J., Xu, M.Y., Ren, Z.Y., Hou, P.F., He, W.Y., Xie, S.S., 2023. Establishment and Validation of a Risk Prediction Model for Unsuccessful Treatment in Patients with Pulmonary Tuberculosis. Chinese Journal of Preventive Medicine.24(9), 912-916.

[44] Fan, X., Cao, Z., 2017. Physical activity among Chinese school-aged children: National prevalence estimates from the 2016 Physical Activity and Fitness in China—The Youth Study. Journal of Sport and Health Science.6, 388-394.

[45] Wang, J., Tao, Q., 2008. Machine Learning: The State of the Art. IEEE Intelligent Systems.23(6), 49–55.

[46] Lindley, M., Anne, E.C., Meghan, H.M., 2014. Need fulfillment and motivation in physical education predict trajectories of change in leisure-time physical activity in early adolescence. Psychology of Sport and Exercise.15, 471-480.

[47] Grendas, L., Chiappella, L., Rodante, D., 2022. Comparison of traditional model-based statistical methods with machine learning for the prediction of suicide behavior. J Psychi Res.145, 85-91.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- Questionnaire2023RunforHealthYouthInitiativeChildrenandAdolescentsThemedFitnessActivityParticipantQuestionnaireandSatisfactionEvaluation.docx