Chen et al. BMC Public Health (2026) 26:1535
https://doi.org/10.1186/s12889-026-27050-8

BMC Public Health

RESEARCH

Open Access

Why they do not move: An explainable machine learning analysis of physical activity barriers in obese adolescents and tool translation

Cheng Chen[1], Kai Chen[2†], Wenqian Du[1†], Shengtao Li[1†], Wenling Gou[3], Jing Yang[4], Pedro Forte[5,6,7], Xiaoran Zhang[1], Yuwen Shangguan[8], Yongyu Huang[9], Hao Zhang[10], Xiaofei Zhang[11,12], Zhiyi Lin[13], Xiaolin Yao[14] and Huan Li[15*]

Abstract

Background Despite the well-documented benefits of physical activity, insufficient activity remains highly prevalent among adolescents with obesity. This study is the first to apply interpretable machine learning methods to identify the barriers, facilitators, and U-shaped determinants of physical activity in this population.

Methods We analyzed data from 1,041 adolescents with obesity from the China Education Panel Survey. A range of personal, family, and school-level variables were incorporated to construct six machine learning models for predicting physical activity attainment. The Shapley Additive Explanations (SHAP) method, a game-theoretic approach for explainable artificial intelligence, was used to identify key predictive factors and quantify their relative contributions.

Results The Random Forest model demonstrated the best performance, achieving an accuracy of 85.30% and an AUC of 0.720 on the test set. SHAP analysis revealed several key factors associated with physical activity. Positive facilitators included parents with an education level beyond high school (≥ 3.27 for mothers, ≥ 3.51 for fathers), higher school rankings (≥ 3.77), adequate school sports facilities (≥ 1.44), and a personal interest in sports. Negative barriers included excessive screen time (≥ 5.51 h) and school location in central urban areas (≥ 4.23). Notably, U-shaped relationships were identified for academic workload, sleep problems, and self-perceived appearance. Specifically, moderate levels of these factors were associated with lower physical activity, whereas both low and high extremes promoted activity.

Conclusion This study demonstrates that physical activity among adolescents with obesity is shaped by a complex interplay of individual, family, and school-level factors, with parental education emerging as the strongest predictor. The identification of specific risk thresholds (e.g., screen time ≥ 5.51 h) and U-shaped relationships offers precise, actionable targets for intervention. These findings underscore the need for multi-level strategies: families should

†Kai Chen, Wenqian Du and Shengtao Li contributed equally to this work and share first authorship.

*Correspondence:
Huan Li
doctorlihuan@163.com

Full list of author information is available at the end of the article

prioritize fostering cultural capital, schools should ensure facility accessibility beyond regular hours, and policymakers must address environmental constraints in urban settings. To facilitate the translation of these insights into practice, a web-based tool has been deployed to help physical education teachers identify at-risk students and design targeted interventions.

Keywords Adolescent obesity, Physical activity, Machine learning, Weight management, Academic workload

Introduction

Adolescent obesity has emerged as a critical public health challenge in China [1]. The prevalence has surged dramatically, with rates among Chinese boys increasing from 1.3% in 1990 to 15.2% in 2022, and among girls from 0.6% to 7.7% over the same period [2]. This condition poses significant health risks, contributing not only to cardiometabolic disorders such as cardiovascular disease, diabetes, and hypertension [3] but also to psychological and social difficulties, including low self-esteem and social exclusion [4]. Crucially, obesity during adolescence is a strong predictor of premature mortality in adulthood [5] and is notoriously difficult to reverse; an estimated 80% of adolescents with obesity will remain obese as adults [6, 7]. This persistence underscores the urgent need for effective interventions during this developmental stage.

Physical activity represents a low-cost, non-pharmacological intervention with substantial potential for both preventing and managing adolescent obesity [8]. Regular engagement in physical activity not only facilitates weight control but also enhances self-efficacy and mental well-being [9]. Despite these well-established benefits, a significant proportion of adolescents with obesity fail to meet recommended activity levels [10], a paradox that has been consistently documented [11]. Previous research has explored the underlying reasons for this disconnect. Some studies have emphasized psychological barriers, such as a lack of interest or low self-efficacy [12], while others have highlighted the overriding influence of external social and environmental factors [13, 14].

The socio-ecological model offers a valuable theoretical framework for understanding the multi-level determinants of physical activity. This model posits that health behaviors are shaped by the complex interplay of factors across multiple levels: intrapersonal (e.g., beliefs, attitudes), interpersonal (e.g., family support), organizational (e.g., school environment), community, and policy domains [15, 16]. Applied to adolescent physical activity, this framework suggests that interventions targeting only individual-level factors are unlikely to succeed if the surrounding environmental and social contexts are not concurrently addressed [17].

Consistent with this perspective, a substantial body of literature has identified three primary levels of influence on physical activity among adolescents with obesity. At the individual level, key barriers include lack of interest, perceived laziness, low self-efficacy, poor physical capacity, and body image dissatisfaction [18–22]. At the family level, factors such as parental support, parental education level, parental engagement in physical activity, and family socioeconomic status play crucial roles [23–26]. At the school level, the availability of sports facilities, experiences of peer bullying, and the structure of physical education programs have been shown to significantly impact participation [27–29]. Collectively, these findings indicate that physical activity behavior in this population is a product of multiple, interacting forces spanning individual traits, family environment, and school context.

Despite these valuable insights, significant gaps remain in our understanding. First, most prior studies have examined these factors in isolation using traditional statistical approaches, which are often ill-equipped to handle the complexity and interactions among numerous variables across different ecological levels [30]. These methods typically assume linear relationships and treat predictors independently, thereby failing to capture potential nonlinearities and interdependencies. Second, while a wide array of influencing factors has been identified, their relative importance—and how they might interact with one another to shape behavior—remains unclear. This knowledge gap hinders the development of comprehensive, multi-level intervention strategies that target the most critical determinants.

To address these limitations, the present study leverages data from the China Education Panel Survey (CEPS), comprising 1,041 adolescents with obesity in junior high school. By integrating individual, family, and school-level variables, we construct six machine learning models to predict physical activity levels. We then apply Shapley Additive Explanations (SHAP), an interpretable machine learning method grounded in cooperative game theory [31], to identify key factors and their specific thresholds of influence. SHAP quantifies the marginal contribution of each feature to the model's prediction for every individual, allowing us to precisely determine the impact of variables such as screen time, parental education, or academic workload. Unlike traditional methods, SHAP can capture complex non-linear patterns and interaction effects, enabling us to identify not only the most important factors but also the direction of their influence (positive or negative) and any U-shaped relationships they may have with physical activity among adolescents with obesity.

The specific objectives of this study are to:

1. Develop and validate machine learning models using personal, family, and school-level variables to predict whether adolescents with obesity meet recommended daily physical activity standards.
2. Employ explainable machine learning methods (SHAP) to identify the key determinants of physical activity participation and determine their specific influence thresholds.

Data and methods

Data sources

This study utilized data from the China Education Panel Survey (CEPS), a large-scale, nationally representative survey (available at: http://ceps.ruc.edu.cn) [32]. We analyzed data from both the baseline survey (2013–2014 academic year) and the first follow-up survey (2014–2015 academic year). All questionnaire items analyzed in this study were derived directly from the publicly available CEPS student, parent, and school administrator questionnaires, as has been done in previous studies using similar questionnaire items [33]. The baseline survey employed a stratified, multistage sampling design with probability proportional to size, randomly selecting 28 county-level administrative units across China (covering 22 provinces, 5 autonomous regions, and 4 municipalities) and 112 schools. It surveyed two cohorts of students in the 7th and 9th grades, encompassing 19,487 students in total.

The follow-up survey tracked the cohort of students who were in the 7th grade at baseline (2013–2014). The primary analysis for the current study was conducted using data from this follow-up survey; therefore, our analytical sample consists of students who were in the 8th grade at the time of outcome measurement. This dataset provided information on students' social behavior and development. After excluding cases with missing data on key variables, the initial sample comprised 7,397 students.

To identify adolescents with obesity, we applied the BMI classification standards from the Chinese National Student Physical Fitness and Health Survey [34]. These age- and sex-specific standards are widely used for school-based health monitoring in China and are officially endorsed for identifying weight status in children and adolescents. Based on these criteria for 8th-grade students (typically aged 14–15 years):

- For male students, overweight was defined as $22.6 \leq \text{BMI} < 25.2$, and obesity as $\text{BMI} \geq 25.3$.
- For female students, overweight was defined as $22.3 \leq \text{BMI} < 24.8$, and obesity as $\text{BMI} \geq 24.9$.

After screening for and excluding outliers (e.g., cases of extreme obesity or underweight), the final analytical sample consisted of 1,041 adolescents with obesity. In the Chinese education system, children typically start primary school at age 6, and junior high school (middle school) covers grades 7 to 9. Thus, the 8th-grade students in our sample were generally 14 to 15 years old.

This study adhered to the ethical principles outlined in the Declaration of Helsinki (2013 revision). All data were de-identified and publicly available, with no issues concerning the disclosure of personal privacy.

Variable definitions

All variables in this study were constructed based on items from the CEPS questionnaires, which are publicly available and have been widely used in previous research [32, 33].

Outcome variable

The primary outcome was physical activity attainment status. Data were derived from the CEPS student questionnaire, which asked about the number of days per week students engaged in physical activity and the typical daily duration of such activity.

According to the World Health Organization (WHO) guidelines, adolescents should engage in at least 60 min of moderate-to-vigorous physical activity (MVPA) daily. However, recognizing the unique physiological capacities and potential barriers faced by adolescents with obesity, we made an adaptive adjustment to this standard. Specifically, participants were classified as "meeting the standard" (coded as 1) if their self-reported average daily MVPA duration was ≥ 30 min, and as "not meeting the standard" (coded as 0) otherwise.

Predictor variables

Predictor variables were selected based on the socio-ecological model and organized into three levels: individual, school, and family.

Individual-level variables

- Gender: Self-reported (0 = female, 1 = male).
- Birth weight: Parent-reported, categorized as 0 = low (<2500g), 1 = normal (2500–4000g), 2 = high (>4000g).
- Household registration (hukou): Self-reported (0 = agricultural, 1 = non-agricultural).
- Self-perceived appearance: Self-reported on a 5-point Likert scale (1 = very ugly, 5 = very beautiful).
- Interest in sports: Derived from a multiple-response question on hobbies; coded as 1 if "sports activities" was selected, 0 otherwise.

- Dietary habits: Self-reported frequency of unhealthy food consumption (e.g., fast food, sugary snacks) on a 5-point scale (1 = never, 5 = always).
- Sleep problems: A composite score (range: 0–9) created by summing responses to nine yes/no questions about sleep issues (e.g., difficulty falling asleep, waking during the night, daytime fatigue). Each problem was coded as 1 if present.
- Screen time: A continuous variable (range: 2–12 hours) representing the total self-reported weekend hours spent on television viewing and internet/gaming.
- Academic workload: A continuous variable (range: 3–18 hours) representing the total self-reported weekend hours spent on homework, tutoring, and self-study.

School-level variables

- School type: Categorized as 1 = public, 2 = private, 3 = private school for migrant workers' children.
- School ranking: Subjective ranking of the school within its county (1 = worst, 5 = best), reported by the school administrator.
- School location: Coded as 1 = rural, 2 = township, 3 = urban-rural fringe, 4 = peripheral urban area, 5 = central urban area.
- School sports facilities: A composite score (range: 0–3) created by summing the presence (1) or absence (0) of a sports field, gymnasium, and swimming pool, as reported by the school administrator.

Family-level variables

- Parental education: Parent-reported educational attainment, coded on a scale from 1 = no education to 7 = graduate degree or higher. Separate variables were created for mother's education (M_Edu) and father's education (F_Edu).
- Family economic status: Parent-reported on a 5-point Likert scale (1 = very difficult, 5 = very affluent).
- Family structure: Parent-reported indicator of whether the child is an only child (0 = no, 1 = yes).

Research methods

Feature selection algorithm

To identify the most relevant predictors while minimizing overfitting, we employed a feature selection approach combining random forest importance and AUC performance. First, we assessed the contribution of each variable using a random forest model and ranked them by their predictive importance. Subsequently, we sequentially added features in order of importance and calculated the cumulative AUC after each addition. As illustrated in Fig. 1, the AUC increased with each added feature and plateaued after reaching its maximum, indicating that features beyond this point did not meaningfully enhance model performance. The final set of features was selected based on this optimal point, ensuring both predictive power and model parsimony [35].

Machine learning methods

Using the scikit-learn library in Python, we developed six machine learning models based on the 12 features selected through our feature selection algorithm: logistic regression, support vector machine (SVM), multilayer perceptron (MLP), XGBoost, LightGBM, and random forest [36]. The total sample (N = 1,041) was randomly split into a training set (n = 729) and a test set (n = 312) using a 70:30 ratio. The training set was used for model development and hyperparameter tuning, while the test set was reserved exclusively for final model evaluation and SHAP analysis.

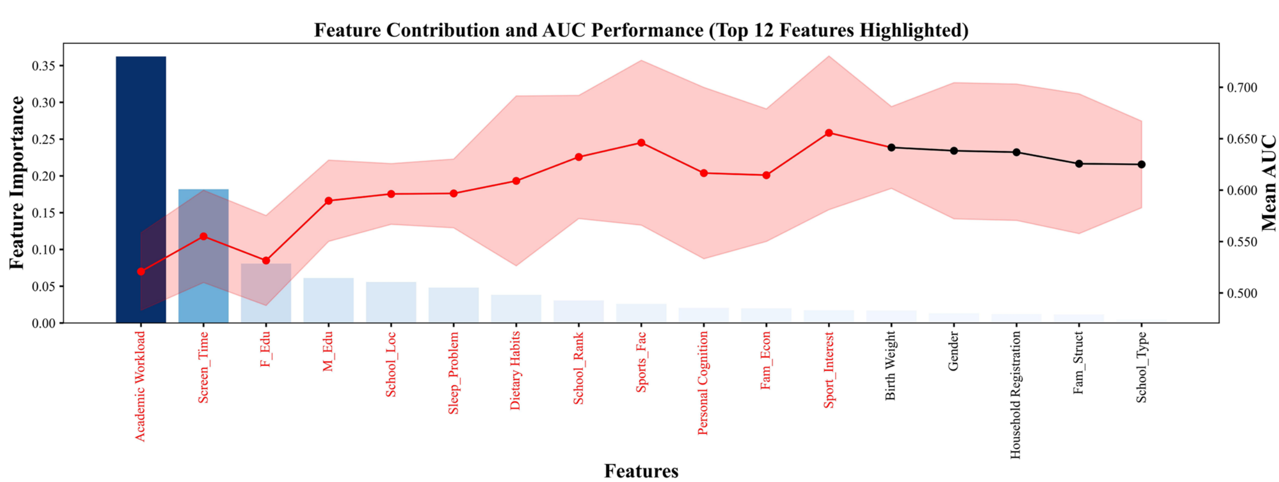

Fig. 1 Feature screening

The outcome variable exhibited substantial class imbalance, with only 16.42% of participants meeting the physical activity standard. Rather than employing resampling techniques, we addressed this imbalance by adjusting the models' built-in class weight parameters. Model performance was primarily assessed using the area under the receiver operating characteristic curve (ROC-AUC), a threshold-independent metric that is robust to class imbalance.

To identify the optimal hyperparameter configuration for each model, we performed grid search across a predefined range of values, following established best practices for machine learning applications in physical activity research [37, 38]. The complete hyperparameter search spaces and the final optimized parameters for each model are provided in Table S1 [39].

SHAP for model interpretability

To enhance the interpretability of our machine learning models, we employed Shapley Additive Explanations (SHAP), a game theory-based approach that quantifies the contribution of each feature to individual predictions [40]. SHAP values enable a granular analysis of each variable's impact on the model output: positive SHAP values indicate that a variable increases the predicted probability of meeting physical activity standards, whereas negative values reflect a suppressive effect. This method allows us to move beyond simple feature importance rankings to understand both the direction and magnitude of each factor's influence.

Statistical methods

Descriptive statistics were computed for all variables. Categorical variables were presented as frequencies and percentages [n (%)], with group comparisons conducted using chi-square tests or Fisher's exact tests, as appropriate based on expected cell counts. All statistical tests were two-sided, with statistical significance set at $P < 0.05$.

For continuous variables, normality was assessed prior to analysis. Normally distributed variables were expressed as mean ± standard deviation, with group comparisons performed using independent samples *t*-tests. Non-normally distributed variables were reported as median with interquartile range (Q_1, Q_3), and group differences were evaluated using the Mann-Whitney U test.

Ethics declaration

Clinical trial number

Not applicable.

Ethics approval and consent to participate

This study conducted a secondary analysis of de-identified, publicly available data from the China Education Panel Survey (CEPS). The original CEPS data collection was approved by the Research Ethics Committee of Renmin University of China. Informed consent was obtained from all participants and their guardians by the CEPS research team prior to data collection. As this analysis involved only anonymized, publicly accessible data, no additional ethical approval or consent was required.

Consent for publication

Not applicable.

Results

Participant characteristics

A total of 1,041 s-year junior high school students (8th grade) with obesity were included in the final analysis. Of these, only 171 students (16.42%) met the adapted physical activity standard (≥ 30 min of daily MVPA), while 870 students (83.58%) did not.

Table 1 presents the complete descriptive statistics and group comparisons. Key differences between students who met the physical activity standard and those who did not are summarized below.

At the individual level, a significant association was found between interest in sports and physical activity attainment (χ^2 = 30.38, **p** < 0.001). Among students reporting an interest in sports, 86 (25.6% of that subgroup) met the standard, compared to 85 (12.1% of that subgroup) among those without such interest. Sleep problems were also significantly associated with activity status. In the "meeting standards" group, 60.82% reported no sleep problems, compared to 52.41% in the "not meeting standards" group (**p** = 0.022). For screen time, while the median was 5 h for both groups, the interquartile range was slightly narrower in the meeting standards group, indicating a statistically significant difference (Z = -2.10, **p** = 0.035).

At the family level, parental educational attainment emerged as a strong differentiator. In the meeting standards group, 59.65% of mothers had attained a high school education or higher, substantially exceeding the 40.92% in the non-meeting group (χ^2 = 26.27, **p** < 0.001). Similarly, 68.42% of fathers in the meeting standards group had at least a high school education, compared to 48.16% in the non-meeting group (χ^2 = 29.60, **p** < 0.001).

At the school level, the availability of sports facilities was significantly associated with physical activity attainment (χ^2 = 11.43, **p** = 0.010). Notably, 25.00% of students in schools with three sports facilities met the standard (15 out of 60), compared to only 6.67% in schools with no such facilities (3 out of 45).

In summary, significant bivariate associations with physical activity attainment were observed for interest in sports, sleep problems, screen time, parental education (both mother and father), and the number of school

Table 1 Basic characteristics and differences of participants

Variables	Total (n=1041)	Physical Activity Not Meeting Standards (n=870)	Physical Activity Meeting Standards (n=171)	Statistic	P
Individual level					
Sport Interest, n(%)				χ^2=30.38	< 0.001
No interest	705 (67.72)	620 (71.26)	85 (49.71)		
Interest in sports	336 (32.28)	250 (28.74)	86 (50.29)		
Personal Cognition, n(%)				χ^2=1.71	0.789
Very ugly	33 (3.17)	27 (3.10)	6 (3.51)		
Somewhat ugly	83 (7.97)	72 (8.28)	11 (6.43)		
Average	807 (77.52)	675 (77.59)	132 (77.19)		
Fairly beautiful	62 (5.96)	52 (5.98)	10 (5.85)		
Very beautiful	56 (5.38)	44 (5.06)	12 (7.02)		
Dietary Habits, n(%)				χ^2=2.83	0.587
Never	54 (5.19)	45 (5.17)	9 (5.26)		
Rarely	349 (33.53)	287 (32.99)	62 (36.26)		
Sometimes	510 (48.99)	432 (49.66)	78 (45.61)		
Often	119 (11.43)	100 (11.49)	19 (11.11)		
Always	9 (0.86)	6 (0.69)	3 (1.75)		
Sleep Problem, n(%)				-	0.022*
0	560 (53.79)	456 (52.41)	104 (60.82)		
1	237 (22.77)	201 (23.10)	36 (21.05)		
2	132 (12.68)	116 (13.33)	16 (9.36)		
3	65 (6.24)	58 (6.67)	7 (4.09)		
4	25 (2.40)	23 (2.64)	2 (1.17)		
5	15 (1.44)	13 (1.49)	2 (1.17)		
6	4 (0.38)	2 (0.23)	2 (1.17)		
7	2 (0.19)	0 (0.00)	2 (1.17)		
8	1 (0.10)	1 (0.11)	0 (0.00)		
Screen Time, M (Q_1, Q_3)	5.00 (4.00, 7.00)	5.00 (4.00, 7.00)	5.00 (4.00, 6.00)	Z=-2.10	0.035
Academic Workload, M (Q_1, Q_3)	6.00 (5.00, 8.00)	6.00 (5.00, 8.00)	6.00 (5.00, 8.00)	Z=-0.76	0.448
Family Level					
Fam Econ, n(%)				χ^2=6.50	0.164
Very Difficult	21 (2.02)	20 (2.30)	1 (0.58)		
Somewhat difficult	86 (8.26)	72 (8.28)	14 (8.19)		
Moderate	780 (74.93)	658 (75.63)	122 (71.35)		
Somewhat affluent	144 (13.83)	113 (12.99)	31 (18.13)		
Very affluent	10 (0.96)	7 (0.80)	3 (1.75)		
M Edu, n(%)				χ^2=26.27	< 0.001
No education	0	0	0		
Elementary school	162 (15.56)	150 (17.24)	12 (7.02)		
Junior high school	421 (40.44)	364 (41.84)	57 (33.33)		
Vocational high school/technical school/general high school	270 (25.94)	208 (23.91)	62 (36.26)		
Associate Degree	84 (8.07)	69 (7.93)	15 (8.77)		
Bachelor's Degree	95 (9.13)	71 (8.16)	24 (14.04)		
Graduate and above	9 (0.86)	8 (0.92)	1 (0.58)		
F Edu, n(%)				χ^2=29.60	< 0.001
No education	0	0	0		
Elementary school	118 (11.34)	108 (12.41)	10 (5.85)		
Junior high school	387 (37.18)	343 (39.43)	44 (25.73)		
Vocational high school/technical school/general high school	305 (29.30)	245 (28.16)	60 (35.09)		
Associate Degree	99 (9.51)	70 (8.05)	29 (16.96)		

Table 1 (continued)

Variables	Total ($n=1041$)	Physical Activity Not Meeting Standards ($n=870$)	Physical Activity Meeting Standards ($n=171$)	Statistic	P
Bachelor's Degree	113 (10.85)	88 (10.11)	25 (14.62)		
Graduate and above	19 (1.83)	16 (1.84)	3 (1.75)		
School Level					
School Rank, n(%)				χ^2=2.73	0.604
Lowest	64 (6.15)	58 (6.67)	6 (3.51)		
Below Average	4 (0.38)	3 (0.34)	1 (0.58)		
Middle	170 (16.33)	142 (16.32)	28 (16.37)		
Upper-middle	537 (51.59)	445 (51.15)	92 (53.80)		
Best	266 (25.55)	222 (25.52)	44 (25.73)		
School Loc, n(%)				χ^2=5.87	0.209
Rural	150 (14.41)	128 (14.71)	22 (12.87)		
Township	122 (11.72)	94 (10.80)	28 (16.37)		
Urban-rural fringe	157 (15.08)	129 (14.83)	28 (16.37)		
Peripheral urban areas	74 (7.11)	60 (6.90)	14 (8.19)		
Central urban areas	538 (51.68)	459 (52.76)	79 (46.20)		
Sports Fac, n(%)				χ^2=11.43	0.010
0	45 (4.32)	42 (4.83)	3 (1.75)		
1	687 (65.99)	586 (67.36)	101 (59.06)		
2	249 (23.92)	197 (22.64)	52 (30.41)		
3	60 (5.76)	45 (5.17)	15 (8.77)		

Z Mann-Whitney test, χ^2 Chi-square test, -: Fisher exact, *: Simulated *p*-value

M Median, Q_1 1st Quartile, Q_3 3st Quartile

sports facilities. These findings provide preliminary evidence for the multivariable modeling and mechanism identification that follow.

Model comparison

Figure 2 presents the performance of six machine learning models in classifying physical activity attainment among adolescents with obesity, evaluated on both the training and test sets. The models compared were: Support Vector Machine (SVM), Extreme Gradient Boosting (XGBoost), Light Gradient Boosting Machine (LightGBM), Logistic Regression (LR), Random Forest (RF), and Multi-Layer Perceptron (MLP).

In terms of accuracy (Fig. 2A), the Random Forest model achieved the highest performance on the training set (90.52%), followed by XGBoost (86.13%) and SVM (85.03%). On the test set, RF again demonstrated the highest accuracy (85.30%), with XGBoost and MLP tied for second (84.35%). This pattern suggests that the RF model strikes an effective balance between learning from the training data and generalizing to unseen data.

For discriminative ability, as assessed by the area under the receiver operating characteristic curve (AUC) (Fig. 2B), the RF model achieved the highest training AUC (0.980), followed by XGBoost (0.885) and LightGBM (0.830). On the test set, RF maintained its leading position (AUC = 0.720), with MLP (0.714) and XGBoost (0.709) showing comparable but slightly lower performance.

To select the optimal model for subsequent interpretation, we considered multiple performance metrics in accordance with established guidelines for machine learning applications in health behavior research [41]. Our primary criterion was test-set AUC, as it provides a threshold-independent measure of discriminative ability and is particularly robust to class imbalance [42, 43]. Secondary criteria included test accuracy, the discrepancy between training and test performance (as an indicator of overfitting), and balanced accuracy.

Based on these criteria, the Random Forest model was selected as the most suitable for further analysis. As shown in Fig. 2, RF achieved the highest test AUC (0.720) and test accuracy (85.30%) among all models evaluated. Although the training-test performance gap (training AUC = 0.980; test AUC = 0.720) suggests some degree of overfitting, this level of performance degradation is acceptable for complex social science data, where some noise and irreducible error are expected [44]. In comparison, the next best-performing models—MLP (AUC = 0.714) and XGBoost (AUC = 0.709)—demonstrated slightly lower discriminative ability.

Therefore, the Random Forest model was selected for subsequent SHAP analysis to identify the key factors influencing physical activity among adolescents with

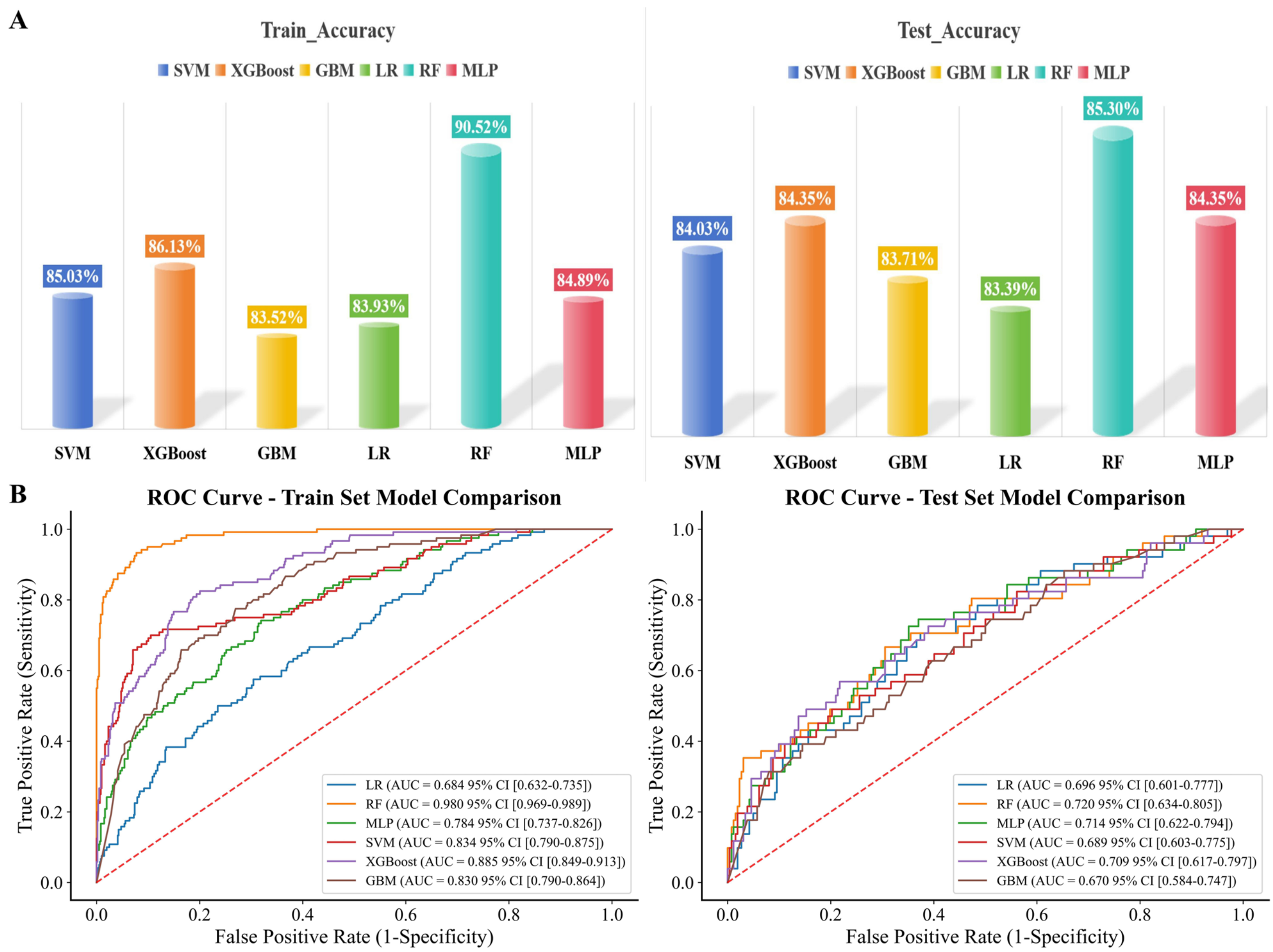

Fig. 2 Comparison of prediction performance of different machine learning models in training set and test set. **A** Accuracy; **B** Area under the curve (AUC)

obesity and to elucidate the direction and magnitude of their effects.

Key determinants of physical activity in adolescents with obesity

To identify the key determinants of physical activity attainment among adolescents with obesity, we applied SHAP (Shapley Additive Explanations) to the best-performing Random Forest model. Figure 3 presents the feature importance ranking (Fig. 3A) and the direction of each variable's influence (Fig. 3B), while Fig. 4 visualizes the nonlinear relationships and identifies specific threshold values.

Feature importance ranking: multi-level determinants

As shown in Fig. 3A, maternal education (M_Edu) had the highest mean absolute SHAP value, indicating it was the strongest predictor of physical activity attainment among adolescents with obesity. This was closely followed by interest in sports (Sport_Interest), school location (School_Loc), paternal education (F_Edu), and academic workload (Academic_Workload). The presence of five top-ranked variables spanning family (parental education), school (location), and individual (sports interest, academic workload) levels underscores the multi-dimensional nature of physical activity determinants in this population.

Direction of influence and threshold effects

Figure 3B illustrates the direction of each variable's effect on the model's prediction, while Fig. 4 provides a detailed visualization of the relationship between individual variables and their SHAP values, enabling identification of specific thresholds. Based on these analyses, we classified the key predictors into three categories: positive facilitators, negative barriers, and variables exhibiting U-shaped relationships.

Positive facilitators

Increases in these variables were associated with a higher predicted probability of meeting physical activity standards:

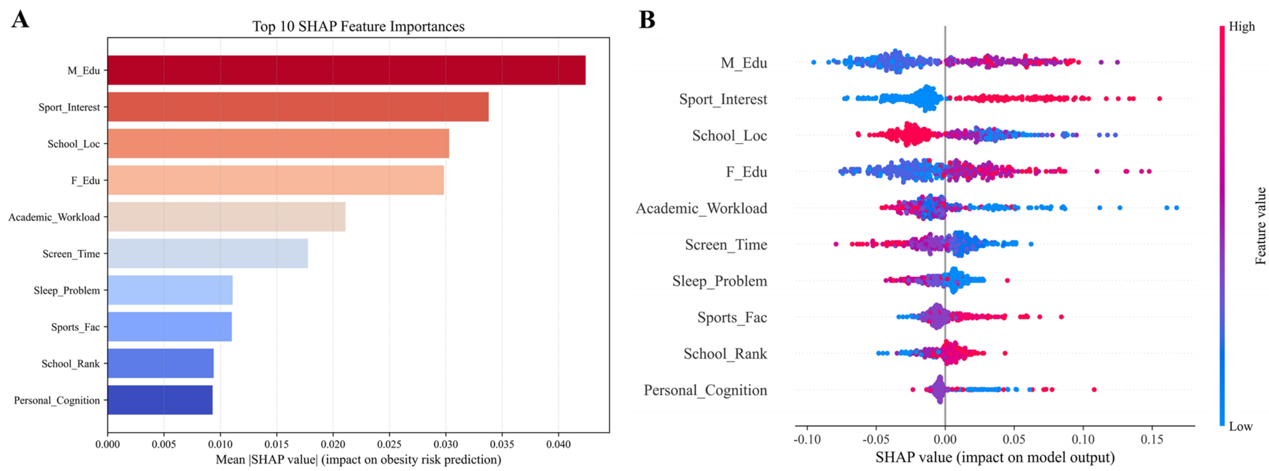

Fig. 3 Feature importance ranking and feature direction. **A** Mean absolute SHAP values (feature importance ranking); **B** SHAP values (direction of influence)

- Parental education: When maternal education reached ≥3.27 (equivalent to high school level or above) and paternal education reached ≥3.51 (high school level or above), the likelihood of physical activity attainment increased.
- Interest in sports: Having a personal interest in sports was a strong positive facilitator.
- School sports facilities: When the sports facility score reached ≥1.44 (approximately equivalent to having at least one to two facilities), physical activity participation was more likely.
- School ranking: Schools ranked at ≥3.77 (upper-middle level or above) were associated with higher physical activity attainment among students with obesity.

Negative barriers

Increases in these variables were associated with a reduced probability of meeting physical activity standards:

- Screen time: When weekend screen time reached ≥5.51 hours, the likelihood of physical activity attainment decreased substantially.
- School location: Schools located in "core urban areas" (location code ≥4.23) were associated with lower physical activity participation among adolescents with obesity.

U-shaped relationships

These variables exhibited a pattern where moderate levels were associated with reduced physical activity, while both low and high extremes were associated with increased activity:

- Academic workload: Workloads in the range of 4.8 to 12 hours showed a negative association with physical activity. However, both lighter workloads (<4.8 hours) and heavier workloads (>12 hours) were associated with increased likelihood of meeting activity standards.
- Sleep problems: The presence of sleep problems (from >0.40 to approximately 5 issues) was negatively associated with physical activity. Notably, when the number of sleep problems exceeded 5, the model predicted an increased likelihood of meeting physical activity standards.
- Self-perceived appearance: Adolescents who perceived their appearance as "average" or "fairly attractive" (scores ranging from 2.89 to 4 on the 5-point scale) were less likely to engage in physical activity. In contrast, those at both extremes—perceiving themselves as "very ugly" or "fairly ugly" (scores <2.89) or as "very attractive" (score = 5)—showed a higher predicted probability of physical activity attainment.

Discussion

Individual-level influencing factors

Our findings reveal several important individual-level factors that shape physical activity among adolescents with obesity, including evidence of a potential vicious cycle, a nuanced "time displacement" effect, and novel U-shaped relationships.

First, our results suggest that adolescents with obesity may become trapped in a self-reinforcing vicious cycle: lack of interest in exercise → reduced physical activity → increased obesity → greater fatigue during exercise → further diminished interest. We observed a positive association between interest in physical activity and actual activity time, consistent with prior research [21]. This

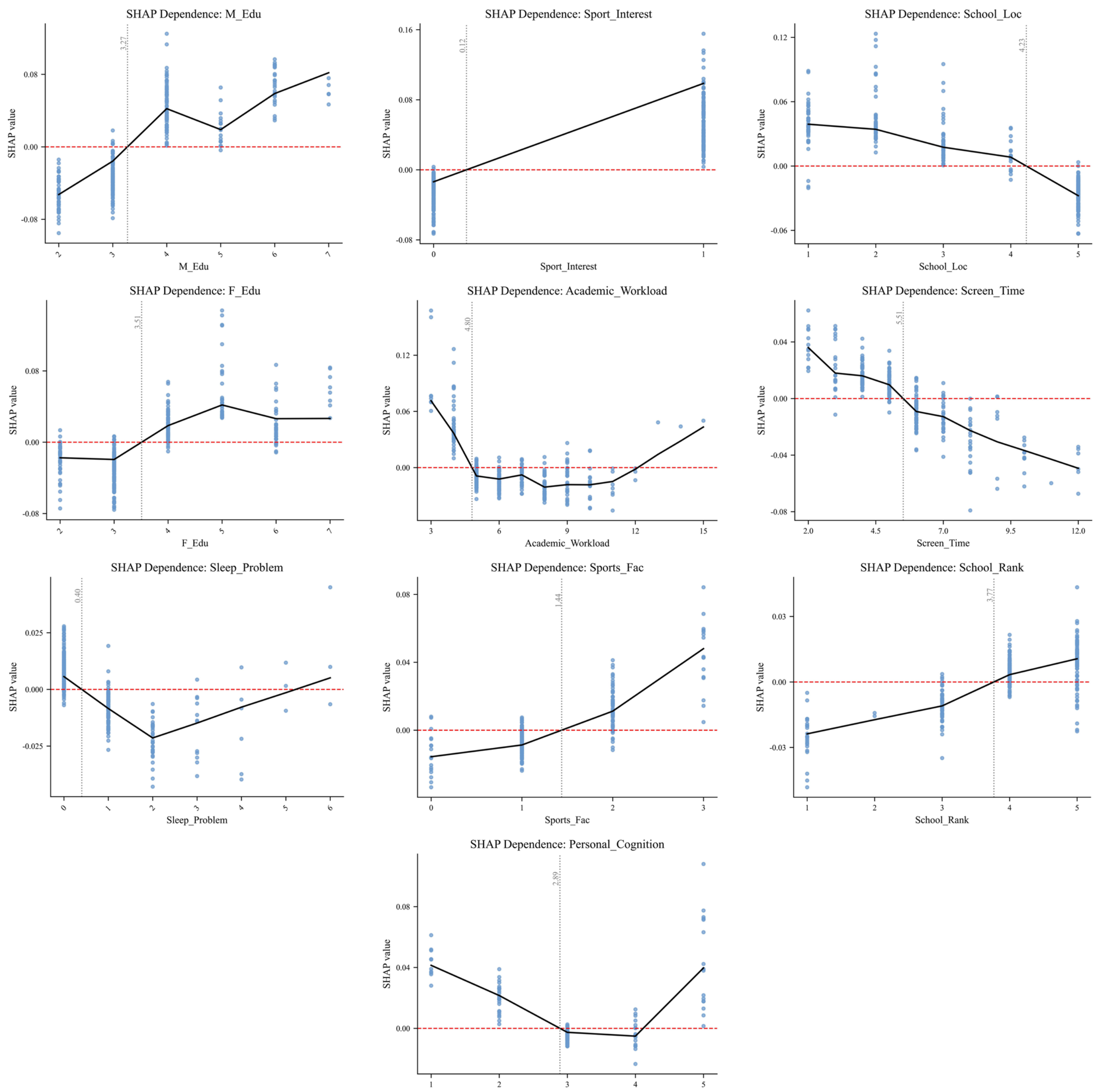

Fig. 4 Feature dependency graph

relationship can be understood through the proposed "vicious cycle of obesity and physical activity." Whether obesity is present from early childhood or develops during adolescence, reduced physical function often leads to lower mood and heightened perceptions of fatigue during and after exertion [45]. These negative experiences, in turn, diminish interest in physical activity [46]. As interest wanes, motivation to engage in exercise declines, resulting in sustained inactivity and further exacerbation of obesity [47]. This cyclical pattern underscores the importance of early intervention to disrupt the cycle before it becomes entrenched.

Second, our findings provide partial support for the "time displacement" effect of screen-based media use. Consistent with previous studies [48], we identified prolonged screen time (≥ 5.51 h on weekends) as a significant barrier to physical activity. This reflects a classic "time crowding out" phenomenon: time spent on electronic screens displaces time available for physical activity. However, our results offer a novel perspective on this effect when considering academic workload. Logically, academic demands might be expected to exert a similar crowding-out effect, a pattern supported by some prior research [49]. Yet, we found that among adolescents with obesity, heavy academic workload (≥ 12 h on weekends) was actually associated with increased likelihood of meeting physical activity standards. This counterintuitive finding points to a potentially deeper explanatory factor: self-discipline. We hypothesize that highly self-disciplined students are better able to manage their time effectively, allowing them to meet academic demands while still prioritizing physical activity. For these students, the time crowding-out effect of academic workload may be attenuated or even reversed, resulting in a positive association between study time and extracurricular activity. This interpretation, while speculative, highlights the need for future research examining the role of self-regulation and time management skills in moderating the relationship between competing demands and physical activity.

Finally, this study identified U-shaped relationships between two individual-level factors—self-perceived appearance and sleep problems—and physical activity among adolescents with obesity. These findings represent one of the key innovations of our study.

Regarding self-perceived appearance, adolescents who rated themselves as "very ugly" or "very beautiful" were more likely to engage in physical activity, while those perceiving themselves as "average" were least likely to do so. Traditional perspectives emphasize a direct link between negative body image and exercise avoidance among adolescents with obesity [50]. However, our results reveal a more complex pattern: extreme body perceptions—whether extremely negative or extremely positive—may actually motivate physical activity engagement. This duality can be explained through complementary mechanisms. Adolescents who perceive themselves as "very ugly" may be motivated to improve their body shape through exercise, viewing physical activity as a means of enhancing self-image and achieving weight-related goals [9]. Conversely, those who perceive themselves as "very beautiful" may already have established exercise habits that contribute to and reinforce their positive self-perception [51]. This finding suggests that interventions should not only address the inhibitory effects of negative body image but also recognize that certain negative self-perceptions, under the right conditions, can serve as catalysts for behavior change.

Regarding sleep problems, we observed a similarly nuanced pattern. Contrary to previous conclusions that sleep problems are uniformly negatively correlated with physical activity [52], our study found that both very few (≤ 0.4 reported problems) and very many (≥ 5 reported problems) sleep issues were associated with increased physical activity. The former finding is intuitive: good sleep quality enhances physical vitality and mental alertness, which in turn increases motivation and capacity for physical activity. The latter finding, however, is counterintuitive and warrants careful interpretation. While we might expect multiple sleep problems to hinder exercise participation, our results suggest that when adolescents experience more than five sleep-related difficulties, physical activity levels actually increase. This phenomenon may reflect an active coping mechanism: adolescents facing significant sleep disturbances may intentionally increase their physical activity as a strategy to improve sleep quality [53]. Alternatively, this could reflect reverse causality, whereby higher physical activity levels (perhaps from school sports teams or training) lead to greater physical fatigue and, paradoxically, more reported sleep problems. Given the cross-sectional nature of our data, we cannot determine causal direction, and this relationship warrants investigation in future longitudinal studies.

Family-level influencing factors

This study found that higher parental educational attainment was associated with increased physical activity among adolescents with obesity. Specifically, when maternal education reached ≥ 3.27 (equivalent to high school level or above) and paternal education reached ≥ 3.51 (high school level or above), the likelihood of meeting physical activity standards increased. This finding aligns with existing literature, suggesting that parents with higher education levels provide more resources and support for physical activity, thereby increasing their children's participation [54]. A key contribution of this study is the identification of specific educational attainment thresholds. This also implies that in Chinese families where both parents have a high school education or lower, adolescents with obesity may face a relative lack of resources and support for physical activity.

Another family-level variable, family economic status, did not rank among the top 10 features in terms of predictive importance. However, previous studies have identified economic conditions as a critical factor influencing physical activity participation among adolescents with obesity [55]. This discrepancy can be explained by examining how family economic status was measured and distributed in our sample. As detailed in Sect. 2.2.2, this variable was derived from a subjective, self-reported, five-point Likert scale. Table 1 shows that nearly 75% of adolescents in this study reported their family's economic status as "moderate" (category 3 on the 5-point scale), while only 2.02% reported "very difficult" and 0.96% reported "very affluent." This high concentration of responses in the middle category indicates limited statistical variability in perceived economic standing within this cohort.

This homogeneity in perceived economic status statistically weakens its power to predict differences in physical activity behavior. In contrast, parental education levels (M_Edu and F_Edu) showed much greater variability across seven educational categories, ranging from "elementary school" to "graduate degree and above." This broader distribution allows education level to emerge as a stronger differentiator in predictive models.

Therefore, in the context of this Chinese sample, the influence of subjectively perceived family economic status on adolescents' physical activity may be overshadowed by the more differentiating factor of cultural capital, as represented by parental education. This does not imply that economic factors are unimportant in general, but rather that within the relatively homogeneous perceived economic conditions of our sample, cultural capital emerges as a stronger predictor. Future research employing more objective and varied measures of family wealth may reveal different patterns.

School-level influencing factors

School location, ranking, and the availability of sports facilities all significantly influence physical activity levels among adolescents with obesity.

First, although urban schools generally have better access to sports resources on paper, this study found that adolescents with obesity attending schools in central urban areas (school location ≥ 4.23) actually face greater barriers to physical activity. This paradox highlights a critical distinction: the mere existence of sports resources does not equate to their accessibility. In many urban schools in China, sports facilities are routinely closed after school hours, limiting students' opportunities to use these resources beyond scheduled physical education classes. Furthermore, Chinese middle schools are typically assigned based on proximity to school districts, meaning that students attending urban schools generally reside in those same high-density urban areas. In such environments, public recreational space per capita is often severely limited, with parks, playgrounds, and sports fields in short supply [56]. Together, institutional access barriers (restricted facility hours) and physical space constraints (limited public recreation infrastructure) create compounded obstacles to physical activity among adolescents with obesity in urban settings.

Second, school ranking also plays an important role. This study found that schools ranked in the upper-middle and top tiers (rank ≥ 3.77) are more likely to promote physical activity participation among students with obesity. Several mechanisms may explain this association. Higher-ranked schools in China generally command greater educational resources, including better-funded physical education programs and superior sports facilities [57]. Students attending these schools also tend to exhibit greater self-discipline and time management skills, enabling them to balance academic demands with extracurricular engagement. Additionally, higher-ranked schools often set higher expectations for students' holistic development, including physical fitness, and may therefore be more proactive in encouraging physical activity. The finding that sports facility availability (Sports_Fac ≥ 1.44) promotes physical activity among adolescents with obesity is consistent with this interpretation, suggesting that material resources and institutional culture work in tandem to create an environment conducive to physical activity.

Limitations and future prospects

Despite the strengths of this study, including its application of interpretable machine learning methods to identify key factors influencing physical activity among adolescents with obesity, several limitations should be acknowledged.

First, physical activity duration was measured using self-report questionnaires, which may be subject to recall bias and social desirability effects. Participants may overestimate their activity levels or misremember actual duration, potentially affecting the accuracy of our outcome variable. Future studies could address this limitation by employing more objective measurement methods, such as wearable devices or activity trackers, to capture continuous, real-time physical activity data.

Second, while this study identified multiple factors associated with physical activity, the cross-sectional nature of the data precludes establishing causal relationships between these factors and activity outcomes. The associations we observed may reflect bidirectional influences or unmeasured confounding variables. For instance, the U-shaped relationship between sleep problems and physical activity could indicate that increased activity improves sleep, that poor sleep motivates activity as a coping strategy, or that a third factor (e.g., stress) influences both. Future longitudinal studies are needed to examine how these factors interact over time and to clarify the direction of causality.

Third, our sample was drawn exclusively from Chinese adolescents with obesity, which may limit the generalizability of findings to other cultural contexts or weight categories. Replication studies in diverse populations would help determine whether the identified thresholds and U-shaped relationships are culturally specific or universally applicable.

Finally, although we included a comprehensive set of individual, family, and school-level variables, other potentially important factors—such as peer influences, neighborhood built environment, or genetic predispositions—were not available in the CEPS dataset. Future research incorporating a broader range of ecological levels could provide a more complete understanding of physical activity determinants.

Practical implications

The findings of this study offer actionable implications for families, schools, policymakers, and physical education teachers seeking to promote physical activity among adolescents with obesity.

For families

The finding that parental education—particularly mother's education—emerged as the strongest predictor of adolescent physical activity suggests that interventions should focus on enhancing parents' knowledge and skills related to physical activity promotion, rather than solely providing economic support. Parents with lower educational attainment may benefit from targeted educational programs that offer practical guidance on how to encourage and support their children's physical activity, model active lifestyles, and create a home environment conducive to regular exercise. Additionally, the U-shaped relationship between self-perceived appearance and physical activity suggests that parents should foster a healthy body image that avoids both extreme self-deprecation and overconfidence, while consistently encouraging activity participation regardless of body image concerns.

For schools

The identification of screen time ≥ 5.51 h as a critical threshold suggests that school-based interventions should incorporate media literacy education and screen time reduction strategies into health curricula. Schools should also consider extending access to sports facilities beyond regular school hours, particularly in urban areas where public recreational space is limited. The finding that higher-ranked schools promote physical activity demonstrates that academic excellence and physical activity are not mutually exclusive; schools can and should integrate physical activity promotion into their broader mission of student development. Furthermore, the positive association between sports facility availability (≥ 1.44 facilities) and physical activity attainment underscores the importance of adequate infrastructure, though facilities must be accompanied by accessible hours to be effective.

For policymakers

The paradox that schools in central urban areas—despite having better facilities on paper—are associated with lower physical activity participation highlights the need for policies that ensure facility accessibility, not just availability. This may include mandating after-school facility access in public schools or investing in public recreational spaces in high-density urban environments where per capita green space is limited. The U-shaped relationship between academic workload and physical activity suggests that education policies should aim for a balanced approach to academic demands, avoiding both excessive leniency that may lead to sedentary screen time and excessive pressure that could crowd out physical activity.

For physical education teachers

Based on these findings, we have developed and deployed a web-based screening tool (available at: https://decoding-the-pa-paradox-in-obese-adolescents-rf-gaa3orm8yyhzmx.streamlit.app/). This tool allows teachers to input student information across the identified risk factors and receive immediate feedback on the student's predicted probability of meeting physical activity standards. By identifying at-risk students early, teachers can implement targeted interventions, such as personalized activity plans, counseling, or referral to school-based physical activity programs, potentially disrupting the vicious cycle of inactivity before it becomes entrenched.

Conclusions

This study yields several key conclusions. First, both screen time and academic workload exert a "time crowding" effect on physical activity among adolescents with obesity, but students with high self-discipline appear better able to resist this effect, maintaining activity engagement despite competing demands. Second, in Chinese families, cultural capital—as reflected in parental education—has a substantially greater influence on adolescents' physical activity participation than economic capital, at least within the relatively homogeneous perceived economic conditions of our sample. Third, higher school rankings and increased availability of sports facilities promote physical activity among adolescents with obesity, though the urban paradox we identified reveals significant heterogeneity among schools in central urban areas, where facility access barriers may undermine otherwise favorable conditions.

The primary innovation of this study lies in its application of interpretable machine learning methods, which address the limitations of traditional statistical approaches in handling high-dimensional data and capturing nonlinear relationships. For the first time, this study identified U-shaped relationships between sleep problems, self-perceived appearance, academic workload, and physical activity among adolescents with obesity, providing specific risk ranges and thresholds that offer precise targets for intervention. These findings provide a new perspective for understanding the complex, multi-determined nature of exercise behavior in this population.

To translate these insights into practice, parents should encourage positive body image and active lifestyles regardless of weight status; schools should extend access to sports facilities and promote time management skills; and policymakers should prioritize reducing screen time and increasing public recreational spaces, particularly in dense urban environments. A web-based screening tool has been developed and deployed to assist physical education teachers in identifying students at risk of physical inactivity, thereby enabling the formulation of targeted, evidence-based intervention strategies.

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12889-026-27050-8.

Supplementary Material 1.

Supplementary Material 2.

Supplementary Material 3.

Supplementary Material 4.

Supplementary Material 5.

Supplementary Material 6.

Supplementary Material 7.

Supplementary Material 8.

Supplementary Material 9.

Supplementary Material 10: Table S1. Hyperparameter tuning settings and optimal parameters for each machine learning model.

Authors' contributions

Cheng Chen: Conceptualization, Methodology, Formal analysis, Writing – original draft, Writing – review & editing, Visualization. Kai Chen: Data curation, Investigation, Validation, Writing – review & editing. Wenqian Du: Software, Formal analysis, Data curation, Writing – review & editing. Shengtao Li: Methodology, Investigation, Writing – review & editing. Wenling Gou: Conceptualization, Methodology, Writing – review & editing. Jing Yang: Investigation, Resources, Writing – review & editing. Pedro Forte: Validation, Writing – review & editing. Xiaoran Zhang: Translation, Validation, Writing – review & editing. Yuwen Shangguan: Physiological interpretation, Writing – review & editing. Yongyu Huang: Statistical analysis, Visualization, Writing – review & editing. Hao Zhang: Literature review, Theoretical framework, Writing – review & editing. Xiaofei Zhang: Methodology, Data collection supervision, Writing – review & editing. Zhiyi Lin: Clinical perspective integration, Writing – review & editing. Xiaolin Yao: Resources, Supervision, Writing – review & editing. Huan Li: Supervision, Project administration, Writing – review & editing, Corresponding author. All authors reviewed and approved the final version of the manuscript and agree to be accountable for all aspects of the work.

Funding

This study was supported by the Applied Basic Research Project of Changzhou (CJ20252030).

Data availability

The data that support the findings of this study are publicly available from the China Education Panel Survey (CEPS) database. Researchers can apply for access through the official CEPS website (http://ceps.ruc.edu.cn).

Declarations

Ethics approval and consent to participate

This study was a secondary analysis of de-identified, publicly available data from the China Education Panel Survey (CEPS). The original CEPS data collection was approved by the Renmin University of China Research Ethics Committee. Informed consent was obtained from all participants and their guardians by the CEPS team. Therefore, no additional ethical approval or consent was required for this analysis.

Consent for publication

Not applicable.

Competing interests

The authors declare no competing interests.

Author details

[1]Graduate School, Harbin Sport University, Harbin, China
[2]Yantai Health and Health Vocational College, Yantai, China
[3]SanTESiH Laboratory, Faculty of Sports Sciences and Techniques, University of Montpellier, Montpellier, France
[4]School of Psychology and Sociology, Mianyang Normal University, Mianyang, China
[5]Department of Sports, Higher Institute of Educational Sciences of the Douro, Penafiel, Portugal
[6]Departament of Sports Sciences, Instituto Politécnico de Bragança, Bragança, Portugal
[7]Research Center for Active Living and Wellbeing (Livewell), Instituto Politécnico de Bragança, Bragança, Portugal
[8]Department of Exercise Physiology, Kunsan National University, Gunsan, Jeollabuk-do 54150, South Korea
[9]School of Physical Education and Sports Science, South China Normal University, Guangzhou, China
[10]Department of Physical Education and Research, Central South University, Changsha, China
[11]School of Health and Social Development, Faculty of Health, Deakin University, Melbourne, Australia
[12]Centre for Healthcare and Community Transformation, Coventry University, Coventry, UK
[13]School of Physical Education and Sport Science, Fujian Normal University, Fujian, China
[14]Institute of Sports Humanities and Society, Harbin Sport University, Harbin, China
[15]Department of Articular Orthopedics, The First People's Hospital of Changzhou, The Third Affiliated Hospital of Soochow University, Changzhou, China

Received: 29 December 2025 / Accepted: 13 March 2026
Published online: 01 April 2026

References

1. Mingazov RN, H MP, Gureev SA, A ГС, Zotov VV, B ЗВ, et al. Global risks of obesity in adolescent and teenage populations. Probl Social Hygiene Public Health History Med. 2022;30:1067–71. https://doi.org/10.32687/0869-866X-2022-30-s1-1067-1071.
2. Lister NB, Baur LA, Felix JF, Hill AJ, Marcus C, Reinehr T, et al. Child and adolescent obesity. Nat Rev Dis Primers. 2023;9:24. https://doi.org/10.1038/s41572-023-00435-4.
3. Must A, Strauss RS. Risks and consequences of childhood and adolescent obesity. Int J Obes. 1999;23:S2–11. https://doi.org/10.1038/sj.ijo.0800852.
4. Reinehr T. Long-term effects of adolescent obesity: time to act. Nat Rev Endocrinol. 2018;14:183–8. https://doi.org/10.1038/nrendo.2017.147.
5. Liu H, Wu Y-C, Chau PH, Chung TWH, Fong DYT. Prediction of adolescent weight status by machine learning: a population-based study. BMC Public Health. 2024;24:1351. https://doi.org/10.1186/s12889-024-18830-1.
6. Simmonds M, Llewellyn A, Owen CG, Woolacott N. Predicting adult obesity from childhood obesity: a systematic review and meta-analysis. Obes Rev. 2016;17:95–107. https://doi.org/10.1111/obr.12334.
7. The NS, Suchindran C, North KE, Popkin BM, Gordon-Larsen P. Association of adolescent obesity with risk of severe obesity in adulthood. JAMA.

2010;304(18):2042–7. https://doi.org/10.1001/jama.2010.1635. PMID: 21063014; PMCID: PMC3076068.
8. Nelson TF, Stovitz SD, Thomas M, LaVoi NM, Bauer KW, Neumark-Sztainer D. Do Youth Sports Prevent Pediatric Obesity? A Systematic Review and Commentary. Curr Sports Med Rep. 2011;10:360. https://doi.org/10.1249/JSR.0b013e318237bf74.
9. Gualdi-Russo E, Rinaldo N, Zaccagni L. Physical Activity and Body Image Perception in Adolescents: A Systematic Review. Int J Environ Res Public Health. 2022;19:13190. https://doi.org/10.3390/ijerph192013190.
10. Ng K, Cooper J, McHale F, Clifford J, Woods C. Barriers and facilitators to changes in adolescent physical activity during COVID-19. BMJ Open Sport Exerc Med. 2020;6. https://doi.org/10.1136/bmjsem-2020-000919.
11. GAVIN J, MCBREARTY M, MALO K, ABRAVANEL M. Adolescents' Perception of the Psychosocial Factors affecting Sustained Engagement in Sports and Physical Activity. Int J Exerc Sci. 2016;9:384–411. https://doi.org/10.70252/YDGA7242.
12. Deforche BI, De Bourdeaudhuij IM, Tanghe AP. Attitude toward physical activity in normal-weight, overweight and obese adolescents. J Adolesc Health. 2006;38:560–8. https://doi.org/10.1016/j.jadohealth.2005.01.015.
13. Goldweber A, Waasdorp TE, Bradshaw CP. Examining Associations Between Race, Urbanicity, and Patterns of Bullying Involvement. J Youth Adolescence. 2013;42:206–19. https://doi.org/10.1007/s10964-012-9843-y.
14. Oreskovic NM, Goodman E, Park ER, Robinson AI, Winickoff JP. Design and implementation of a physical activity intervention to enhance children's use of the built environment (the CUBE study). Contemp Clin Trials. 2015;40:172–9. https://doi.org/10.1016/j.cct.2014.12.009.
15. McLeroy KR, Bibeau D, Steckler A, Glanz K. An Ecological Perspective on Health Promotion Programs. Health Educ Q. 1988;15:351–77. https://doi.org/10.1177/109019818801500401.
16. Bronfenbrenner U. The ecology of human development: experiments by nature and design. Harvard University Press; 1979. https://doi.org/10.2307/j.ctv26071r6.
17. Sallis JF, Owen N, Fisher EB. Ecological models of health behavior. In K. Glanz, B. K. Rimer, & K. Viswanath (Eds.). Health behavior and health education: Theory, research, and practice (4th ed., pp. 465–485). Jossey-Bass/Wiley; 2008.
18. Daley AJ, Copeland RJ, Wright NP, Wales JKH. `I Can Actually Exercise If I Want To; It Isn't As Hard As I Thought'. J Health Psychol. 2008. https://doi.org/10.1177/1359105308093865.
19. Power TG, Ullrich-French SC, Steele MM, Daratha KB, Bindler RC. Obesity, cardiovascular fitness, and physically active adolescents' motivations for activity: A self-determination theory approach. Psychol Sport Exerc. 2011;12:593–8. https://doi.org/10.1016/j.psychsport.2011.07.002.
20. Scotto di Luzio S, Martinent G, Popa-Roch M, Ballereau M, Chahdi S, Escudero L, et al. Obesity in Childhood and Adolescence: The Role of Motivation for Physical Activity, Self-Esteem, Implicit and Explicit Attitudes toward Obesity and Physical Activity. Children. 2023;10:1177. https://doi.org/10.3390/children10071177.
21. Stankov I, Olds T, Cargo M. Overweight and obese adolescents: what turns them off physical activity? Int J Behav Nutr Phys Act. 2012;9:53. https://doi.org/10.1186/1479-5868-9-53.
22. Valerio G, Gallarato V, D'Amico O, Sticco M, Tortorelli P, Zito E, et al. Perceived Difficulty with Physical Tasks, Lifestyle, and Physical Performance in Obese Children. Biomed Res Int. 2014;2014:735764. https://doi.org/10.1155/2014/735764.
23. Bridger Staatz C, Kelly Y, Lacey RE, Blodgett JM, George A, Arnot M, et al. Socioeconomic position and body composition in childhood in high- and middle-income countries: a systematic review and narrative synthesis. Int J Obes. 2021;45:2316–34. https://doi.org/10.1038/s41366-021-00899-y.
24. Øen G, Kvilhaugsvik B, Eldal K, Halding A-G. Adolescents' perspectives on everyday life with obesity: a qualitative study. International Journal of Qualitative Studies on Health and Well-being; 2018.
25. Seo D-C, Lee CG. Association of School Nutrition Policy and Parental Control With Childhood Overweight. J Sch Health. 2012;82:285–93. https://doi.org/10.1111/j.1746-1561.2012.00699.x.
26. Wheeler S. The significance of family culture for sports participation. Int Rev Sociol Sport. 2012;47:235–52. https://doi.org/10.1177/1012690211403196.
27. Lampard AM, MacLehose RF, Eisenberg ME, Neumark-Sztainer D, Davison KK. Weight-Related Teasing in the School Environment: Associations with Psychosocial Health and Weight Control Practices Among Adolescent Boys and Girls. J Youth Adolescence. 2014;43:1770–80. https://doi.org/10.1007/s10964-013-0086-3.
28. Meaney KS, Hart MA, Griffin LK. Do You Hear What I Hear? Overweight Children's Perceptions of Different Physical Activity Settings. 2011. https://doi.org/10.1123/jtpe.30.4.393
29. Thomas HM, Irwin JD. What Is a Healthy Body Weight? Perspectives of Overweight Youth. Can J Diet Pract Res. 2009;70:110–6. https://doi.org/10.3148/70.3.2009.110.
30. Tang T, Li Z, Lu X, Du J. Development and validation of a risk prediction model for anxiety or depression among patients with chronic obstructive pulmonary disease between 2018 and 2020. Annals of Medicine; 2022.
31. Li X, Zhao Y, Zhang D, Kuang L, Huang H, Chen W, et al. Development of an interpretable machine learning model associated with heavy metals' exposure to identify coronary heart disease among US adults via SHAP: findings of the US NHANES from 2003 to 2018. Chemosphere. 2023;311:137039. https://doi.org/10.1016/j.chemosphere.2022.137039.
32. Overview-China Education Tracking Survey. http://ceps.ruc.edu.cn/English/Overview/Overview.htm. Accessed 9 Mar 2026.
33. Zhang Y, Du K. The impact of parental migration on left-behind children's tooth health in China. BMC Public Health. 2025;25:72. https://doi.org/10.1186/s12889-024-21193-2.
34. Notice of the Ministry of Education on Issuing the National Student Physical Health Standard. (2014 Revision) -Official Website of the Ministry of Education of the People's Republic of China. http://www.moe.gov.cn/s78/A17/tyws_left/moe_938/moe_792/s3273/201407/t20140708_171692.html. Accessed 19 Feb 2026.
35. You J, Guo Y, Kang J-J, Wang H-F, Yang M, Feng J-F, et al. Development of machine learning-based models to predict 10-year risk of cardiovascular disease: a prospective cohort study. Stroke Vasc Neurol. 2023;8. https://doi.org/10.1136/svn-2023-002332.
36. Ye S, Sun X, Kang B, Wu F, Zheng Z, Xiang L, et al. The kinetic profile and clinical implication of SCC-Ag in squamous cervical cancer patients undergoing radical hysterectomy using the Simoa assay: a prospective observational study. BMC Cancer. 2020;20:138. https://doi.org/10.1186/s12885-020-6630-0.
37. Han J, Gondro C, Reid K, Steibel JP. Heuristic hyperparameter optimization of deep learning models for genomic prediction.
38. Han J, Gondro C, Reid K, Steibel JP. Heuristic hyperparameter optimization of deep learning models for genomic prediction. G3 Genes|genomes|genet. 2021;11:jkab032. https://doi.org/10.1093/g3journal/jkab032.
39. Choe J-P, Lee S, Kang M. Machine learning modeling for predicting adherence to physical activity guideline. Sci Rep. 2025;15:5650. https://doi.org/10.1038/s41598-025-90077-1.
40. Lundberg SM, Erion G, Chen H, DeGrave A, Prutkin JM, Nair B, et al. From local explanations to global understanding with explainable AI for trees. Nat Mach Intell. 2020;2:56–67. https://doi.org/10.1038/s42256-019-0138-9.
41. Machine learning in. health promotion and behavioral change: scoping review. J Med Internet Res. 2022;24. https://doi.org/10.2196/35831.
42. Bradley AP. The use of the area under the ROC curve in the evaluation of machine learning algorithms. Pattern Recognit. 1997;30:1145–59. https://doi.org/10.1016/S0031-3203(96)00142-2.
43. Mandrekar JN. Receiver operating characteristic curve in diagnostic test assessment. J Thorac Oncol: Off Publ Int Assoc Study Lung Cancer. 2010;5:1315–6. https://doi.org/10.1097/JTO.0b013e3181ec173d.
44. Montesinos López OA, Montesinos López A, Crossa J, Overfitting. Model Tuning, and Evaluation of Prediction Performance. In: Montesinos López OA, Montesinos López A, Crossa J, editors. Multivariate Statistical Machine Learning Methods for Genomic Prediction. Cham: Springer International Publishing; 2022. pp. 109–39. https://doi.org/10.1007/978-3-030-89010-0_4.
45. Øen G, Kvilhaugsvik B, Eldal K, Halding A-G. Adolescents' perspectives on everyday life with obesity: a qualitative study. Int J Qual Stud Health Well-being. 2018;13:1479581. https://doi.org/10.1080/17482631.2018.1479581.
46. Deforche BI, Bourdeaudhuij IMD, Tanghe AP. Attitude toward physical activity in normal-weight, overweight and obese adolescents. J Adolesc Health. 2006;38:560–8. https://doi.org/10.1016/j.jadohealth.2005.01.015.
47. Lee KX, Quek KF, Ramadas A. Dietary and lifestyle risk factors of obesity among young adults: a scoping review of observational studies. Curr Nutr Rep. 2023;12:733–43. https://doi.org/10.1007/s13668-023-00513-9.
48. Hu FB, Leitzmann MF, Stampfer MJ, Colditz GA, Willett WC, Rimm EB. Physical activity and television watching in relation to risk for type 2 diabetes mellitus in men. Arch Intern Med. 2001;161:1542–8. https://doi.org/10.1001/archinte.161.12.1542.
49. Frömel K, Šafář M, Jakubec L, Groffik D, Žatka R. Academic stress and physical activity in adolescents. Biomed Res Int. 2020;2020:4696592. https://doi.org/10.1155/2020/4696592.

50. Chen L-J, Fox KR, Haase AM. Body image and physical activity among overweight and obese girls in taiwan. Women's Stud Int Forum. 2010;33:234–43. https://doi.org/10.1016/j.wsif.2010.01.003.
51. Hausenblas HA, Fallon EA. Exercise and body image: a meta-analysis. Psychol Health. 2006;21:33–47. https://doi.org/10.1080/14768320500105270.
52. Vanderlinden J, Boen F, van Uffelen JGZ. Effects of physical activity programs on sleep outcomes in older adults: a systematic review. Int J Behav Nutr Phys Act. 2020;17:11. https://doi.org/10.1186/s12966-020-0913-3.
53. Sherrill DL, Kotchou K, Quan SF. Association of physical activity and human sleep disorders. Arch Intern Med. 1998;158:1894–8. https://doi.org/10.1001/archinte.158.17.1894.
54. Kantomaa MT, Tammelin TH, Näyhä S, Taanila AM. Adolescents' physical activity in relation to family income and parents' education. Prev Med. 2007;44:410–5. https://doi.org/10.1016/j.ypmed.2007.01.008.
55. Nielsen G, Grønfeldt V, Toftegaard-Støckel J, Andersen LB. Predisposed to participate? The influence of family socio-economic background on children's sports participation and daily amount of physical activity. Sport in Society. 2012;15(1):1–27. https://doi.org/10.1080/03031853.2011.625271.
56. Sridhar GR, Gumpeny L. Built environment and childhood obesity. World J Clin Pediatr. 2024;13:93729. https://doi.org/10.5409/wjcp.v13.i3.93729.
57. Trilk JL, Ward DS, Dowda M, Pfeiffer KA, Porter DE, Hibbert J, et al. Do physical activity facilities near schools affect physical activity in high school girls? Health Place. 2011;17:651–7. https://doi.org/10.1016/j.healthplace.2011.01.005.
58. Zhao X, Wang Y, Li J, Liu W, Yang Y, Qiao Y, et al. A machine-learning-derived online prediction model for depression risk in COPD patients: A retrospective cohort study from CHARLS. J Affect Disord. 2025;377:284–93. https://doi.org/10.1016/j.jad.2025.02.063.

Publisher's Note