

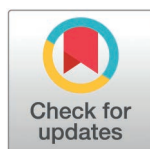
RESEARCH ARTICLE

SHAP-enhanced machine learning identifies modifiable obesity predictors across adolescent weight groups: A 2021 YRBSS analysis

Yuhai Peng¹, Zehan Xu², Songjian Du³, Tianyuan Hou³, Jin Yan^{3*}

1 School of Physical Education, Henan University of Economics and Law, Henan, China, **2** Faculty of Science, University of Sydney, Sydney, New South Wales, Australia, **3** School of Physical Education and Sports Science, Soochow University, Suzhou, China

* jinyan1126@suda.edu.cn



OPEN ACCESS

Citation: Peng Y, Xu Z, Du S, Hou T, Yan J (2025) SHAP-enhanced machine learning identifies modifiable obesity predictors across adolescent weight groups: A 2021 YRBSS analysis. PLoS One 20(10): e0334502. <https://doi.org/10.1371/journal.pone.0334502>

Editor: Mohamed Ahmed Said, ISSEP Kef: Universite de Jendouba Institut Supérieur du Sport et de l'Education Physique du Kef, TUNISIA

Received: June 24, 2025

Accepted: September 29, 2025

Published: October 10, 2025

Peer Review History: PLOS recognizes the benefits of transparency in the peer review process; therefore, we enable the publication of all of the content of peer review and author responses alongside final, published articles. The editorial history of this article is available here: <https://doi.org/10.1371/journal.pone.0334502>

Copyright: © 2025 Peng et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/),

Abstract

Background

The growing prevalence of obesity in adolescents around the world poses a major threat to public health. This research uses machine learning models to examine the main causes of obesity, in contrast to standard information that typically rely on a single chance. The important fat-related steps were identified and ranked in this assessment to provide information on the effectiveness of the expected solutions.

Methods

Data from the 2021 Youth Risk Behavior Surveillance System (YRBSS) were used in a cross-sectional analysis of adolescents aged 12–18 years. Random Forest and XGBoost models were implemented to investigate behavioral, dietary, sleep, and substance use factors. Model interpretability was enhanced using SHapley Additive exPlanations (SHAP).

Results

Breakfast frequency, moderate-to-vigorous physical activity (MVPA) days, sleep duration, fruit intake, and screen time emerged as the most important predictors of obesity, with vaping also contributing to risk. Random Forest achieved an accuracy of 66.4% and XGBoost 66.3%, both with modest discriminative ability (AUC ~0.58). Fewer MVPA days, lower breakfast frequency, shorter sleep duration, lower fruit intake, and longer screen time were associated with increased obesity risk. SHAP analysis confirmed breakfast frequency and MVPA days as the top-ranked factors.

Conclusion

Machine learning models identified key predictors of adolescent obesity, providing insights into the complex interplay of behavioral and lifestyle factors. Public health

which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data availability statement: All relevant data are within the paper and its [Supporting Information](#) files.

Funding: The author(s) received no specific funding for this work.

Competing interests: NO authors have competing interests.

strategies should prioritize daily breakfast and fruit consumption, regular physical activity, sufficient sleep, reduced screen time, and vaping prevention to mitigate rising obesity rates among adolescents.

1 Introduction

The number of people with obesity has increased significantly around the world, affecting people of all ages [1]. Reports show that obesity around the world has nearly tripled since 1975. This increase is caused by biological, life, and social variables [2]. In 2016, over 650 million people were considered overweight, and the levels of obesity in children and adolescents were likewise increasing rapidly [3]. In the U.S., the number of overweight adolescents risen from 10.5% between 1988 and 1994 to 20.6% in 2017 and 2018, according to various national health surveys [4]. A study in 41 countries found that 8.9% of hungry adolescents were also overweight. This shows that in some regions, there are both thin and obese people at the same time [2].

In some areas of Asia, the increase in BMI is particularly significant in children and adolescents. In some rich countries, the rapid increase in BMI has slowed down and stopped growing as fast as before [2,4–6]. During adolescence, it is important to foster good behavior, but the growing rates of obesity during this time can lead to severe health problems later on. Research has shown that excessive weight gain during adolescence is associated to health problems in parents, like heart disease, type 2 diabetes, and some types of cancer [7–9]. Furthermore, being extremely heavy can harm mental health and the general quality of life. This makes adolescents especially sensitive to feelings like being judged by others and feeling sad or stressed in social situations [10]. Therefore, the objective of this study is to identify modifiable predictors of adolescent obesity using Shapley Additive exPlanations (SHAP)-enhanced machine learning models applied to the 2021 YRBSS dataset, in order to provide interpretable evidence for targeted public health interventions.

In the United States, many adolescents are suffering from obesity, and the latest statistics show that more than 20% of them are considered overweight. This matches information from national health surveys showing a significant rise in obesity rates among adolescents [11]. This design is part of a more extensive worldwide matter, with the number of obese children and adolescents rising nearly ten times over the past four decades [2,12]. The rising trend may be explained by lifestyle changes, such as less physical activity, unhealthy eating habits, more moments spent in front of screens, and quick access to bad foods that are high in calories but low in nutrients [13,14]. Moreover, family incomes and educational levels are significantly impacts whether individuals can get good food and opportunities to practice. This helps explain why some groups have higher rates of obesity than others [3]. YRBSS conducted biennially by the CDC since 1991, provides nationally representative data on health behaviors among U.S. high school students. The 2021 YRBSS dataset used in this study covers a wide range of demographic and behavioral factors relevant to adolescent obesity.

To handle adolescent obesity, we need to know how genes, activities, environment, and social variables all work together. This situation is made more complicated by cultural factors, city living, and new technologies that have caused people to be less active [15]. Obesity has many causes, so we need a careful way to prevent and treat it. This means concentrating on early prevention or primary prevention, creating specific health programs, and building a friendly environment that encourages healthy behavior [16,17].

Moreover, machine learning improves how effectively we can create and understand projections. Strategies like SHAP help us understand how much each data part affects the estimates. This makes understanding and applying machine learning designs for public health activities easier. For instance, SHAP has been successfully used to identify various health results and to know how different factors work together in intricate models [18,19]. This ability to focus on essential aspects allows for certain activities that are stronger at lowering the risk of obesity [20,21].

Even though it could be beneficial, applying machine learning in studies about obesity in adolescents is still not very popular. Most studies primarily focus on individual risk factors, rather than considering the broader context [22]. Structure understanding is still not used frequently in adolescent obesity information despite its potential benefits. Instead of focusing on the broader context, the majority of opinions are on special risk aspects. To capture potential differences across subpopulations, adolescents were also stratified by weight categories, which may reveal unique behavioral and environmental predictors of obesity. Utilizing machine learning models to identify and categorize the main factors of adolescent obesity, the aim of this study was to advance the field. In this study, influential predictors and their relative importance are revealed through the use of machine learning techniques [23]. This review provides a nuanced knowing that supports more precisely targeted public health interventions. In order to improve the transparency of the models, techniques like SHAP are employed; this allows for a more precise understanding of each factor's influence on the obesity prediction [18,24].

To determine the most important factors of adolescent obesity, this study analyzed machine learning models. Our method places emphasis on identifying potentiating aspects that could be addressed by comprehensible algorithms. The findings from this study are expected to enrich the current body of knowledge by delineating the major predictors of adolescent obesity, which can be leveraged to design targeted public health strategies [25]. By identifying the key indicators of adolescent obesity, public health practitioners may be better equipped to design and implement interventions aimed at reducing obesity rates among adolescents, thereby contributing to improved health outcomes and lower long-term health-care costs [26,27].

2 Methods

2.1 Study design and participants

We used data from the 2021 Youth Risk Behavior Surveillance System (YRBSS), a nationally representative dataset that provides comprehensive information on adolescents' health behaviors in the United States for the cross-sectional design of this study. The YRBSS dataset was chosen because of its thorough analysis of important demographic, behavioral, and environmental factors [28,29]. The sample was representative of American children in grades 9–12 when the data were gathered using a stage-based cluster trial design [30]. The cross-sectional nature of the study made it possible to examine the connections between various elements and obesity status (accessed on 10 December 2024) [31].

In the review sample were young YRBSS study participants who were between the ages of 12 and 18 in 2021. A two-step imputation technique was applied to handle missing values and retain all available data. Specifically, demographic variables (age, sex) were used as auxiliary variables in SPSS to impute missing values, ensuring completeness for subsequent analyses. The initial results were based on the age (98 cases, 0.6%), sex, and median age (both at 15 years old) and SPSS mode, respectively. Stata used the ages and sexes as sign parameters to find any missing data by making a number of assumptions. This method made sure the data was trustworthy and accessible later for analysis [32,33].

Based on the YRBSS's de-identified, publicly available information, the best administrative review panel gave approval to the research. Ethical approvals for the study were secured at the national or regional level, with each country obtaining approval from an ethics review board or an equivalent regulatory body specific to the government. The current study was approved by the Ethics Committee at Soochow University (SUDA20240626H06). In conducting this study, we adhered to the guidelines outlined in the Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) [34].

2.2 Measurement of variables

Weight status was defined using the CDC sex-specific BMI-for-age percentiles (BMIPCT variable in YRBSS). Participants were classified as underweight (< 5th percentile), normal weight (5th – 84th percentile), overweight (85th – 94th percentile), or obese ($\geq$ 95th percentile) [35]. The YRBSS evaluation, which had a variety of options for getting socioeconomic, cognitive, and health-related data, was used to analyze the factors in this review. Interviewees' ages ranged from "12 years old or younger" to "18 years older or older" when asked how old they were. The gender was chosen, regardless of whether it was "Male" or "Female". The two-step study asked respondents to be "White", "Black or African American", "Asian", "Hispanic/Latino" and others. For analysis, these responses were consolidated into four main categories: White, Black or African American, Hispanic/Latino, and All Other Races.

Moderate-to-vigorous physical activity (MVPA) days was measured by asking how many days in the past week participants were physically active for at least 60 minutes, with responses ranging from "0 days" to "7 days." Screen time was assessed based on the number of hours spent in front of screens on an average school day, excluding schoolwork, with options ranging from "less than 1 hour per day" to "5 or more hours per day." Sleep duration was determined by asking how many hours participants typically slept on school nights, with options from "4 or less hours" to "10 or more hours."

Dietary habits were captured through questions regarding the frequency of eating breakfast, consuming fruit juice, fruit, vegetables (green salad, potatoes, carrots, and other vegetables), and drinking soda and milk in the past week, with responses indicating frequencies from "0 days" or "I did not consume" to "7 days" or "4 or more times per day." Substance use was measured for alcohol, cigarettes, electronic vapor products, marijuana, and prescription pain medicine misuse. Participants were asked how many days in the past 30 days they used alcohol, cigarettes, or electronic vapor products, with response options ranging from "0 days" to "All 30 days." For marijuana and pain medicine misuse, participants were asked how many times in their lifetime they had used these substances, with answers ranging from "0 times" to "100 or more times."

2.3 Statistical analysis

In this study, we applied both conventional statistical methods and machine learning models to examine factors associated with adolescent obesity. The 2021 YRBSS, which contained a variety of health behavior data, was used as the analytic dataset. To decide which of the 20 potential indicators should be used, Chi-square tests were applied to screen candidate variables. Variables showing statistical significance ($p < 0.05$) in Chi-square tests were retained, resulting in 16 features that included demographic, behavioral, and environmental factors relevant to adolescent obesity [36]. For these statistical analyses, R software (version 4.4.0) was used [28].

To complement the regression analyses, we implemented two ML algorithms—Random Forest and XGBoost—to provide additional insights into variable importance in complex datasets [37]. Training (70%) and testing (30%) sets were included in the dataset. A fixed random seed (set.seed = 3554) was applied to ensure reproducibility. Hyperparameter tuning was performed using grid search with cross-validation, focusing on key parameters of each model. To address class imbalance, Synthetic Minority Over-sampling Technique (SMOTE) was applied as the primary method, supplemented by class weight adjustment. Model performance was evaluated using repeated 5-fold cross-validation on the training set and then assessed on the held-out testing set (30%) using accuracy, sensitivity, specificity, AUC of the ROC curve, and weighted F1-score [38]. Classification thresholds were set according to the default behavior of the algorithms, where each observation was assigned to the class with the highest predicted probability (argmax rule). In addition, imbalance-sensitive

metrics were reported, including the mean one-vs-rest AUC, which are particularly relevant given the unequal distribution of weight categories in the YRBSS dataset. These evaluations were used primarily to characterise model behavior rather than to optimise predictive accuracy.

The primary correlates of overweight and obesity were examined using multivariable logistic regression. Odds ratios (ORs) and 95% confidence intervals (CIs) were used to calculate statistical significance when the p -value was less than 0.05. For this analysis, SPSS (version 27.0) was used.

To further interpret model performance, we examined the relationships between predictors and outcomes. Feature importance was assessed using the Mean Decrease Gini index in the Random Forest model, whereas Gain quantified each feature's contribution to node splits in the XGBoost model [39]. SHAP were implemented, following Han and Wang (2023) [39]. For each model, SHAP values were computed for individual predictions and then aggregated across all participants using mean absolute SHAP values to determine overall feature importance. The distribution of SHAP values for each predictor was visualised using beeswarm plots, and the top 15 predictors of adolescent obesity were highlighted. Analyses and visualisations were conducted using R (version 4.4.0) and Origin (version 2021). Compared with the initial submission, the revised analyses incorporated hyperparameter tuning, imbalance handling, and cross-validation to strengthen robustness.

3 Results

Table 1 presents the distribution of demographic and behavioral characteristics across weight categories. Significant associations were observed for sex, age, and race/ethnicity. Obesity prevalence was higher among males (60.2%) than females (39.8%) ($p=0.022$), and older adolescents (16–18 years) had a higher prevalence of obesity ($p<0.001$). Obesity prevalence was highest among Hispanic/Latino participants (24.1%); the proportion of White participants within the obese group was 45.7% ($p<0.001$). More detailed information is provided in the supplementary materials.

Table 2 provides the multivariable logistic regression results. Male participants were more likely to be overweight or obese compared to females (OR = 1.30, 95% CI: 1.22–1.40, $p<0.001$). Race/ethnicity also played a substantial role: Black

Table 1. Demographic characteristics across different weight categories.

Variables	Overall (n = 17232) n (%)	Non-overweight (n = 11459) n (%)	Overweight (n = 2749) n (%)	Obese (n = 3024) n (%)	χ^2	p
Sex						
Female	8152(47.3)	5564(48.6)	1384(50.3)	1204(39.8)	23.684	0.022
Male	9080(52.7)	5895(51.4)	1365(49.7)	1820(60.2)		
Age						
12 years old	39(0.2)	18(0.2)	10(0.4)	11(0.4)	85.446	< 0.001
13 years old	62(0.4)	38(0.3)	13(0.5)	11(0.4)		
14 years old	3403(19.7)	2227(19.4)	601(21.9)	575(19.0)		
15 years old	4525(26.3)	2997(26.2)	728(26.5)	800(26.5)		
16 years old	4276(24.8)	2857(24.9)	635(23.1)	784(25.9)		
17 years old	3904(22.7)	2638(23.0)	607(22.1)	659(21.8)		
18 years old	1023(5.9)	684(6.0)	155(5.6)	184(6.1)		
Race						
White	9375(54.4)	6641(58.0)	1351(49.1)	1383(45.7)	257.976	< 0.001
Black or African American	2384(13.8)	1401(12.2)	434(15.8)	549(18.2)		
Hispanic/Latino	3331(19.3)	1943(17.0)	659(24.0)	729(24.1)		
All Other Races	2142(12.4)	1474(12.9)	305(11.1)	363(12.0)		

<https://doi.org/10.1371/journal.pone.0334502.t001>

Table 2. Multivariable logistic regression analysis of factors associated with overweight and obesity.

Variables	Odds Ratio	95%CI	p
Age (ref= 12 years old)			
13 years old	0.76	0.33-1.76	0.522
14 years old	0.70	0.36-1.35	0.292
15 years old	0.67	0.35-1.29	0.228
16 years old	0.64	0.33-1.24	0.187
17 years old	0.61	0.32-1.17	0.139
18 years old	0.63	0.32-1.22	0.172
Sex (ref=Female)			
Male	1.30	1.22-1.40	< 0.001
Race (ref= White)			
Black or African American	1.51	1.37-1.67	< 0.001
Hispanic/Latino	1.58	1.45-1.72	< 0.001
All Other Races	1.02	0.92-1.13	0.683
MVPA days (ref=0 days)			
1 day	0.95	0.83-1.10	0.512
2 days	0.95	0.84-1.08	0.452
3 days	0.88	0.78-0.99	0.049
4 days	0.77	0.68-0.88	< 0.001
5 days	0.84	0.75-0.95	0.007
6 days	0.55	0.47-0.65	< 0.001
7 days	0.63	0.56-0.70	< 0.001
Screen time (ref=Less than 1 hour per day)			
1 hour per day	1.00	0.82-1.22	0.994
2 hours per day	1.05	0.90-1.24	0.527
3 hours per day	0.94	0.80-1.09	0.416
4 hours per day	1.01	0.87-1.18	0.861
5 hours per day	1.12	0.97-1.29	0.112
Sleep duration (ref=4 or less hours)			
5 hours	0.98	0.86-1.12	0.794
6 hours	0.98	0.87-1.10	0.706
7 hours	0.89	0.79-1.00	0.056
8 hours	0.98	0.86-1.12	0.785
9 hours	0.99	0.81-1.20	0.886
10 or more hours	1.04	0.80-1.35	0.785
Breakfast days (ref=0 days)			
1 day	0.87	0.77-0.98	0.023
2 days	0.94	0.84-1.05	0.270
3 days	0.73	0.64-0.82	< 0.001
4 days	0.74	0.64-0.85	< 0.001
5 days	0.76	0.66-0.87	< 0.001
6 days	0.65	0.55-0.77	< 0.001
7 days	0.56	0.51-0.63	< 0.001
Fruit (ref=Did not eat)			
1 to 3 times	1.06	0.95-1.18	0.283
4 to 6 times	1.15	1.02-1.30	0.021
1 time per day	1.00	0.87-1.16	0.977

(Continued)

Table 2. (Continued)

Variables	Odds Ratio	95%CI	p
2 times per day	1.05	0.91-1.21	0.529
3 times per day	1.18	0.98-1.42	0.075
4 or more times per day	1.18	0.98-1.41	0.076
Other vegetables (ref=Did not eat)			
1 to 3 times	1.01	0.92-1.11	0.856
4 to 6 times	1.12	1.01-1.24	0.040
1 time per day	1.01	0.88-1.16	0.838
2 times per day	1.06	0.90-1.25	0.475
3 times per day	1.32	1.06-1.65	0.015
4 or more times per day	1.07	0.84-1.37	0.582
Soda (ref=Did not drink)			
1 to 3 times	1.01	0.93-1.10	0.754
4 to 6 times	1.04	0.94-1.16	0.459
1 time per day	1.01	0.87-1.16	0.925
2 times per day	1.05	0.89-1.23	0.591
3 times per day	1.18	0.93-1.49	0.181
4 or more times per day	1.18	0.96-1.44	0.121
Milk (ref=Did not drink)			
1 to 3 times	1.06	0.97-1.15	0.177
4 to 6 times	1.14	1.02-1.28	0.021
1 time per day	1.11	0.99-1.24	0.075
2 times per day	0.99	0.86-1.14	0.886
3 times per day	1.10	0.89-1.35	0.376
4 or more times per day	1.03	0.85-1.26	0.755
Alcohol (ref=0 days)			
1 or 2 days	0.88	0.79-0.99	0.026
3 to 5 days	0.85	0.73-1.00	0.051
6 to 9 days	0.73	0.58-0.91	0.005
10 to 19 days	0.93	0.70-1.23	0.599
20 to 29 days	1.24	0.72-2.15	0.435
All 30 days	1.12	0.69-1.82	0.656
Cigarettes (ref=0 days)			
1 or 2 days	1.53	1.20-1.95	< 0.001
3 to 5 days	0.96	0.63-1.44	0.834
6 to 9 days	1.17	0.64-2.15	0.604
10 to 19 days	1.49	0.92-2.41	0.102
20 to 29 days	1.11	0.49-2.52	0.803
All 30 days	0.95	0.60-1.49	0.817
Electronic vapor (ref=0 days)			
1 or 2 days	0.94	0.79-1.11	0.454
3 to 5 days	1.24	1.01-1.53	0.045
6 to 9 days	1.07	0.83-1.38	0.612
10 to 19 days	0.95	0.76-1.19	0.682
20 to 29 days	1.01	0.79-1.29	0.936
All 30 days	1.07	0.89-1.27	0.488

(Continued)

Table 2. (Continued)

Variables	Odds Ratio	95%CI	p
Marijuana (<i>ref=0 times</i>)			
1 or 2 times	1.09	0.94-1.28	0.253
3 to 9 times	1.12	0.92-1.36	0.251
10 to 19 times	1.12	0.89-1.40	0.352
20 to 39 times	0.85	0.66-1.10	0.225
40 or more times	0.86	0.70-1.06	0.149
Pain medicine (<i>ref=0 times</i>)			
1 or 2 times	1.11	0.92-1.33	0.280
3 to 9 times	1.01	0.78-1.30	0.929
10 to 19 times	1.10	0.75-1.62	0.631
20 to 39 times	1.05	0.54-2.04	0.877
40 or more times	1.19	0.75-1.89	0.460

<https://doi.org/10.1371/journal.pone.0334502.t002>

or African American (OR = 1.51, 95% CI: 1.37–1.67, $p < 0.001$) and Hispanic/Latino adolescents (OR = 1.58, 95% CI: 1.45–1.72, $p < 0.001$) had higher odds of obesity than White participants. MVPA on 6 days per week (OR = 0.55, 95% CI: 0.47–0.65, $p < 0.001$) and daily breakfast consumption (OR = 0.56, 95% CI: 0.51–0.63, $p < 0.001$) showed the strongest protective effects. For screen time, adolescents reporting ≥ 5 hours per day had higher odds of obesity (OR = 1.12, 95% CI: 0.97–1.29, $p = 0.112$) compared with those reporting less than 1 hour per day.

Table 3 compares the performance of the Random Forest and XGBoost models. The Random Forest model achieved a slightly higher accuracy (66.4%) than XGBoost (66.3%), with very similar AUC values (0.577 vs. 0.576). Both models demonstrated extremely low specificity (3.9% vs. 4.8%), indicating limited ability to correctly identify negative cases. Random Forest showed a higher negative predictive value (0.60 vs. 0.47), whereas XGBoost performed marginally better on positive predictive value (0.677 vs. 0.672). The F1 scores were comparable (0.800 vs. 0.797), suggesting balanced performance on precision and recall for the positive class.

Fig 1 shows the feature importance in both models based on the Mean Decrease Gini index (Random Forest) and Gain metrics (XGBoost). “Breakfast days” and “MVPA days” consistently emerged as the top predictors across both models, followed by “Sleep duration” and “Fruit” intake. The highest Mean Decrease Gini index was observed for Breakfast days, highlighting its crucial role in predicting obesity.

Fig 2 displays the beeswarm plot of SHAP values, highlighting the top 15 predictors of adolescent obesity in the XGBoost model. Higher MVPA days, longer sleep duration, and more frequent breakfast consumption were associated

Table 3. Comparison of evaluation performance between random forest and XGBoost.

	Random Forest	XGBoost
Accuracy	0.6636	0.6625
95% CI	(0.6506, 0.6765)	(0.6494, 0.6754)
McNemar’s Test P-Value	< 0.001	< 0.001
Specificity	0.0389	0.0477
Pos Pred Value	0.6720	0.6765
Neg Pred Value	0.6000	0.4655
F1	0.7995	0.7972
AUC	0.5769	0.5763

<https://doi.org/10.1371/journal.pone.0334502.t003>

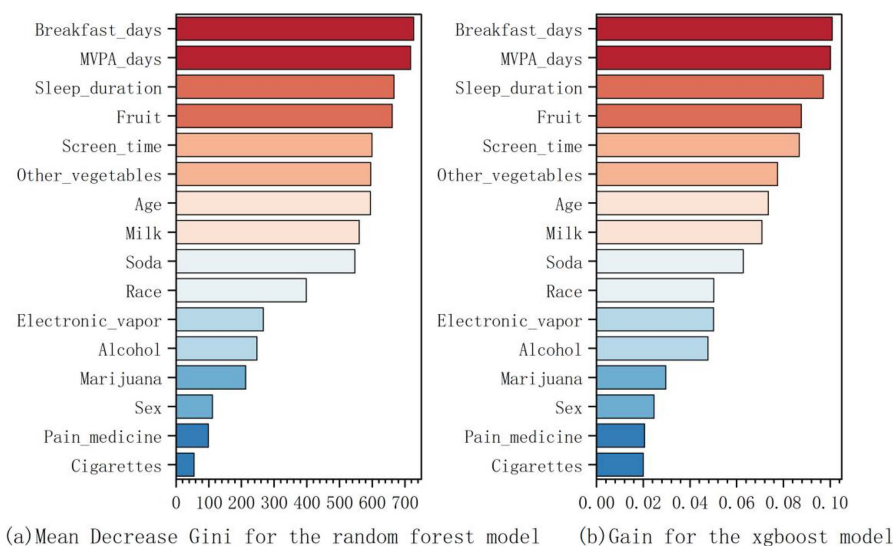


Fig 1. Importance ranking of factors influencing youth obesity with Random Forest and XGBoost models.

<https://doi.org/10.1371/journal.pone.0334502.g001>

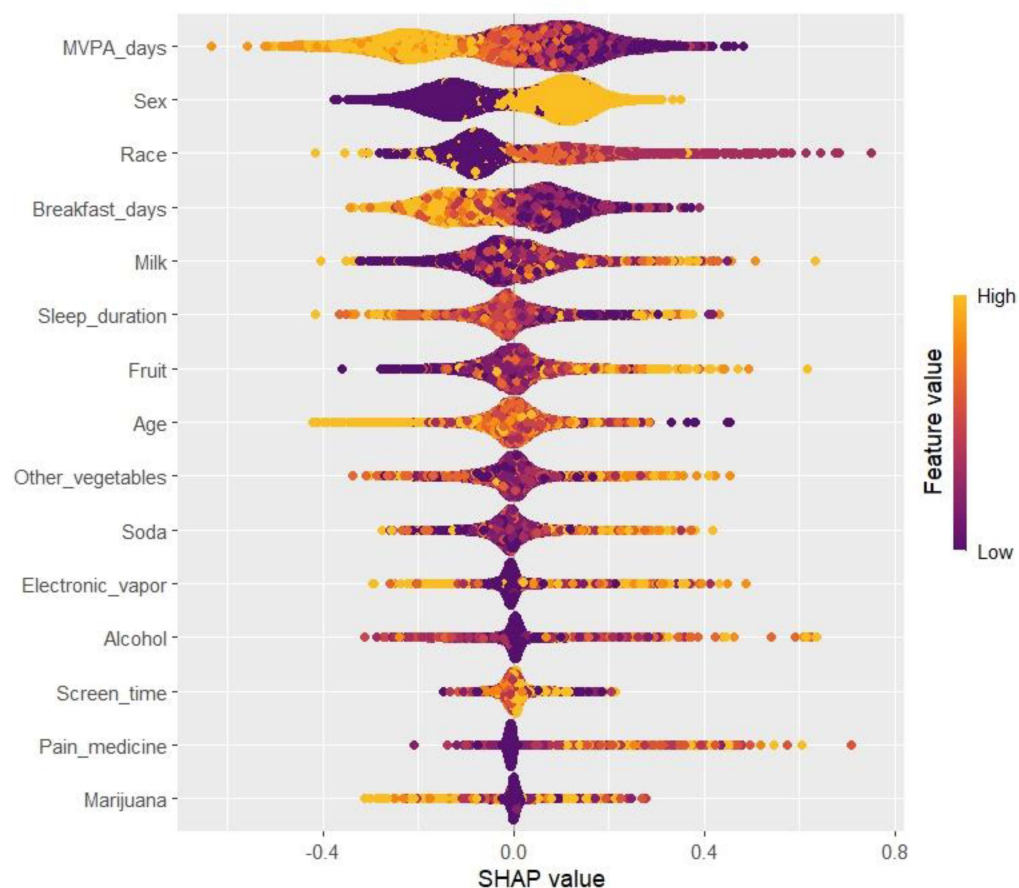


Fig 2. Beeswarm plot of SHAP values for the top 15 predictors of adolescent obesity in the XGBoost model. (Note: Each point represents an individual participant. The x-axis shows the SHAP value (impact on log-odds of obesity), where positive values indicate increased risk and negative

values indicate reduced risk. The y-axis lists predictors ranked by overall importance, and point colour reflects the feature value (yellow = high, purple = low). The least influential variable (Cigarettes) was excluded).

<https://doi.org/10.1371/journal.pone.0334502.g002>

with reduced obesity risk, whereas higher screen time, greater soda intake, and electronic vapor use contributed to increased risk. The top features were ranked according to their mean absolute SHAP values across all participants, with point colours indicating feature values (yellow = high, purple = low).

Fig 3 presents a nomogram constructed based on the Random Forest and XGBoost models, illustrating the relative contributions of key predictors to obesity risk. The nomogram assigns point values to predictors such as age, race, MVPA days, and dietary habits. Higher total scores, particularly those associated with fewer MVPA days, shorter sleep duration, lower breakfast frequency, and increased screen time, indicated a higher likelihood of obesity.

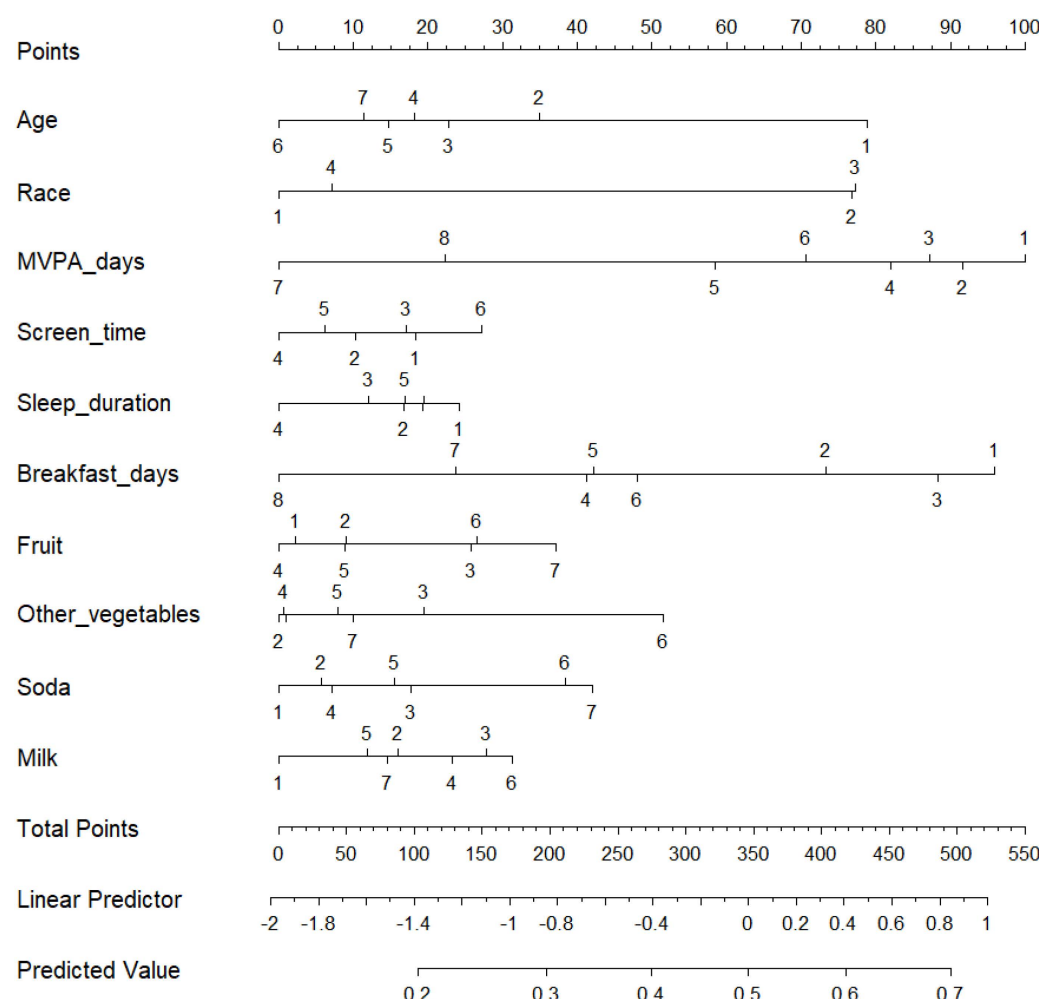


Fig 3. Nomogram of predictors of adolescent obesity. (Note: Each predictor is shown as a horizontal scale, with values mapped to points according to their relative contribution. The “Points” axis at the top indicates the score assigned to each predictor level. Summing all scores yields a “Total Points” value, which is then projected onto the “Linear Predictor” and “Predicted Value” axes at the bottom to estimate the probability of obesity. Higher total points correspond to greater predicted obesity risk).

<https://doi.org/10.1371/journal.pone.0334502.g003>

4 Discussion

This study applied machine learning models (Random Forest and XGBoost) combined with SHAP to identify key predictors of adolescent obesity. The use of SHAP improved model interpretability by quantifying the contribution of each predictor to obesity risk, thereby bridging the gap between complex modelling and clinically relevant insights [40,41]. Both machine learning and traditional regression analyses consistently highlighted breakfast frequency, MVPA days, sleep duration, fruit intake, and screen time as the most important modifiable predictors of obesity [42–44]. These findings strengthen the evidence base for prioritising behavioral interventions targeting these factors in adolescent obesity prevention.

MVPA

In this study, MVPA days were identified as one of the strongest predictors of an adolescent's BMI status, ranking just after breakfast frequency. According to research that links physical activity to maintaining normal weight, adolescents who regularly participated in MVPA were significantly less likely to be overweight or obese [45,46]. Increased muscle mass, improved insulin sensitivity, and increasing energy expenditure, all of which contribute to lowering natural BMI increases, may be a result of MVPA's efforts to lower fat regulations. Since then, Wang et. al emphasize the value of regular physical activity, who found that adolescents who adhered to the daily MVPA guidelines were significantly less likely to become obese [47]. Similarly, Lister et al. (2023) demonstrated that maintaining physical activity during adolescence is essential to preventing obesity later in life [48]. In addition to the WHO guidelines recommending at least 60 minutes of MVPA per day, these findings support the necessity of school-based activities that use normal physical activity for all children. In both Random Forest and XGBoost models, MVPA days consistently ranked among the top predictors, second only to breakfast frequency. This finding matches the data Klein et al. provided, which demonstrated that school initiatives that promote regular physical activity lower obesity prevalence and improve long-term health outcomes [49]. Given the significant effect of physical activity on BMI, increasing options for MVPA should be a crucial focus of public health approaches aimed at adolescent obesity prevention.

Screen time

Although screen time emerged as a relevant predictor in ML models, its association with obesity was not statistically significant in regression analyses ($p=0.112$). This discrepancy highlights the complexity of behavioral influences and the value of complementary analytic approaches. Adolescents reporting ≥ 5 hours of daily screen time had slightly higher odds of obesity compared with those reporting <1 hour, although this association did not reach statistical significance. This objective is in line with the findings of Christofaro et al. (2016) and Al-Hazzaa (2018), which established a clear link between excessive screen time and poor eating habits [50,51]. One possible pathway linking prolonged screen exposure to higher BMI is through displacement of physical activity and promotion of obesogenic behaviors such as snacking. Buchanan et al. (2016) made a case for the benefits of sedentary behavior and how screen media exposure can lead to weight gain [52]. Public health interventions should therefore aim to reduce excessive screen exposure and encourage more active alternatives to limit its potential impact on BMI.

Dietary predictors

Breakfast frequency was identified as the strongest dietary factors associated with a lower risk of high BMI. According to previous research, adolescents who had breakfast daily were significantly less overweight or obese [53,54]. According to Szajewska & Ruszczyński (2010), better appetite regulation and a healthier overall diet are directly related to one's BMI [55]. Additionally, Gordon-Larsen et al. (2006) found who have breakfast are more likely to maintain a balanced diet throughout the day to prevent overeating or consuming unhealthy snacks [56]. Adolescents who frequently skip breakfast

are more likely to have a higher BMI as a result of changes in their daily eating patterns [57]. Our regression and machine learning results consistently demonstrated that daily breakfast consumption was strongly protective against obesity, aligning with Al-Hazzaa et al.(2012), which emphasizes the importance of breakfast in adolescent health [58].

Milk intake also appeared as a secondary dietary factor associated with adolescent BMI. Regression findings were inconsistent across intake levels (including an elevated odds ratio at 4–6 times/week), suggesting milk's association with BMI may depend on type and context of consumption. Milk provides essential nutrients such as calcium, protein, and vitamin D, which are important for metabolic health and bone development [59]. However, the relationship with BMI may vary by type (whole vs. low-fat) and frequency of consumption. While our study only measured overall milk frequency, future research should differentiate by milk type to refine dietary recommendations [60–62]. Fruit intake emerged as an important predictor in the ML/SHAP analysis; however, regression results were mixed across consumption categories, with one intermediate level associated with higher odds of obesity. This discrepancy may reflect measurement and residual confounding. Adolescents reporting higher fruit consumption were less likely to be obese, consistent with evidence that fruit contributes to satiety and healthier dietary patterns [63,64]. In contrast, vegetable consumption showed less consistent associations, possibly due to differences in measurement and preparation methods [65]. These findings suggest that promoting daily fruit intake, alongside other healthy dietary practices, may play a meaningful role in reducing obesity risk.

Substance Use

Electronic vapor use was associated with higher BMI in the machine learning models, whereas cigarette use did not consistently emerge as a significant predictor. This relationship may be driven by behavioral patterns such as reduced physical activity and poor dietary choices, which often cluster with substance use [66,67]. Addressing vaping in adolescence may therefore have additional benefits for obesity prevention alongside its well-established health risks [68]. The associations between alcohol and marijuana use with BMI were less clear, as these substances did not consistently predict obesity risk across analyses [69,70]. The unexpected inverse association with alcohol in some models likely reflects residual confounding related to adolescent social behaviors rather than a direct effect of alcohol consumption. Future studies should explore these relationships in greater depth, considering other behavioral and environmental influences that may contribute to the observed trends [71].

Implications

The need for qualified public health interventions focused on these crucial areas is underscored by the more substantial effects of physical activity, breakfast frequency, sleep duration, and fruit intake on BMI, with screen time playing a more modest role. Schools should establish rules to reduce excessive screen exposure, encourage healthier eating habits, and increase opportunities for daily physical activity. To avoid negative impacts on BMI, it is also crucial to address adolescent substance use, particularly vaping/e-cigarette use. Public health professionals can use the insights from machine learning models to inform interventions, while recognising the modest predictive performance of these models [72,73].

Study limitations

Although this research contains important data, there are some limitations. First, the cross-sectional design prevents causal inference, and the identified associations should be interpreted as correlational rather than causal. Second, all measures relied on self-reported data, which introduces the potential for recall and reporting bias [74], particularly in physical activity, nutrition, and substance use. Third, we did not include longitudinal validation, so it remains uncertain whether the identified predictors can consistently forecast obesity risk over time. Additionally, model performance was modest (accuracy ~66%, with very low specificity), reflecting reliance on self-reported behavioral data. Despite tuning and imbalance handling, predictive power remained limited; thus, the ML analyses were intended to highlight correlates rather

than provide precise prediction. Future research should therefore adopt longitudinal designs, incorporate wearable-based measures, and expand the scope of machine learning models to include genetic and environmental factors, which could further improve predictive power and utility in public health applications [75,76].

5 Conclusion

This study provides valuable insight into the most critical predictors of adolescent obesity, particularly breakfast frequency, MVPA, sleep duration, fruit intake, screen time, and vaping. It also demonstrates how machine learning models can complement traditional regression by highlighting key behavioral and dietary factors, despite their modest predictive performance. By targeting these behaviors, public health professionals can develop more effective interventions to reduce adolescent obesity and improve long-term health outcomes.

Supporting information

S1. 2021 YRBS Data Working V1 10.1.
(XLSX)

Acknowledgments

The authors sincerely appreciate all the students who participated in this study, especially the dedicated fieldworkers, for their invaluable contributions to data collection.

Author contributions

Conceptualization: Yuhai Peng, Songjian Du.

Data curation: Yuhai Peng.

Formal analysis: Yuhai Peng.

Methodology: Zehan Xu.

Project administration: Zehan Xu.

Resources: Zehan Xu, Tianyuan Hou.

Software: Zehan Xu, Tianyuan Hou.

Supervision: Tianyuan Hou, Jin Yan.

Validation: Songjian Du, Tianyuan Hou, Jin Yan.

Visualization: Songjian Du, Tianyuan Hou, Jin Yan.

Writing – original draft: Songjian Du, Jin Yan.

Writing – review & editing: Jin Yan.

References

1. WHO. Obesity and overweight. 2024. [cited 2024 11-1]. <https://www.who.int/news-room/fact-sheets/detail/obesity-and-overweight>
2. NCD Risk Factor Collaboration (NCD-RisC). Worldwide trends in body-mass index, underweight, overweight, and obesity from 1975 to 2016: a pooled analysis of 2416 population-based measurement studies in 128·9 million children, adolescents, and adults. *Lancet*. 2017;390(10113):2627–42. [https://doi.org/10.1016/S0140-6736\(17\)32129-3](https://doi.org/10.1016/S0140-6736(17)32129-3) PMID: 29029897
3. Powell LM, Han E, Chaloupka FJ. Economic contextual factors, food consumption, and obesity among U.S. adolescents. *J Nutr*. 2010;140(6):1175–80. <https://doi.org/10.3945/jn.109.111526> PMID: 20392882
4. Cynthia LO, Carroll M, Hannah GL, Fryar C, Kruszon-Moran D, Brian KK, et al. Trends in obesity prevalence among children and adolescents in the United States, 1988-1994 through 2013-2014. *JAMA*. 2016;315(21):2292–9.

5. Zheng W, Shen H, Belhaidas MB, Zhao Y, Wang L, Yan J. The Relationship between Physical Fitness and Perceived Well-Being, Motivation, and Enjoyment in Chinese Adolescents during Physical Education: A Preliminary Cross-Sectional Study. *Children (Basel)*. 2023;10(1):111. <https://doi.org/10.3390/children10010111> PMID: [36670661](#)
6. Shi J, Gao M, Xu X, Zhang X, Yan J. Associations of muscle-strengthening exercise with overweight, obesity, and depressive symptoms in adolescents: Findings from 2019 Youth Risk Behavior Surveillance system. *Front Psychol*. 2022;13:980076. <https://doi.org/10.3389/fpsyg.2022.980076> PMID: [36160591](#)
7. Twig G, Yaniv G, Levine H, Leiba A, Goldberger N, Derazne E, et al. Body-Mass Index in 2.3 Million Adolescents and Cardiovascular Death in Adulthood. *New England Journal of Medicine*. 2016;374(25):2430–40.
8. Gordon-Larsen P, The NS, Adair LS. Longitudinal trends in obesity in the United States from adolescence to the third decade of life. *Obesity (Silver Spring)*. 2010;18(9):1801–4. <https://doi.org/10.1038/oby.2009.451> PMID: [20035278](#)
9. Li H, Zhang W, Yan J. Physical activity and sedentary behavior among school-going adolescents in low- and middle-income countries: insights from the global school-based health survey. *PeerJ*. 2024;12:e17097. <https://doi.org/10.7717/peerj.17097> PMID: [38680891](#)
10. Chooi YC, Ding C, Magkos F. The epidemiology of obesity. *Metabolism*. 2019;92:6–10. <https://doi.org/10.1016/j.metabol.2018.09.005> PMID: [30253139](#)
11. Cynthia LO, Carroll M. Prevalence of obesity among children and adolescents: United States, trends 1963–1965 through 2007–2008. 2010.
12. Lee EY, Yoon K-H. Epidemic obesity in children and adolescents: risk factors and prevention. *Front Med*. 2018;12(6):658–66. <https://doi.org/10.1007/s11684-018-0640-1> PMID: [30280308](#)
13. Al-Khudairy L, Loveman E, Colquitt JL, Mead E, Johnson RE, Fraser H, et al. Diet, physical activity and behavioural interventions for the treatment of overweight or obese adolescents aged 12 to 17 years. *Cochrane Database Syst Rev*. 2017;6(6):CD012691. <https://doi.org/10.1002/14651858.CD012691> PMID: [28639320](#)
14. Goran MI, Treuth MS. Energy expenditure, physical activity, and obesity in children. *Pediatric Clinics of North America*. 2001;48(4):931–53.
15. Kansra AR, Lakkunarajah S, Jay MS. Childhood and Adolescent Obesity: A Review. *Frontiers in Pediatrics*. 2021;8:581461.
16. Khan LK, Sobush K, Keener D, Goodman K, Lowry A, Kakietek J, et al. Recommended community strategies and measurements to prevent obesity in the United States. *MMWR Recomm Rep*. 2009;58(RR-7):1–26. PMID: [19629029](#)
17. Swinburn B. Obesity prevention in children and adolescents. *Child and Adolescent Psychiatric Clinics of North America*. 2009;18(1):209–23.
18. Nohara Y, Inoguchi T, Nojiri C, Nakashima N. Explanation of Machine Learning Models of Colon Cancer Using SHAP Considering Interaction Effects. *ArXiv*. 2022. <https://doi.org/abs/2208.03112>
19. Ahmed S, Kaiser MS, Hossain MS, Andersson K. A comparative analysis of LIME and SHAP interpreters with explainable ML-based diabetes predictions. *IEEE Access*. 2024:1–1.
20. Zhang L, Wang Y, Niu M, Wang C, Wang Z. Machine learning for characterizing risk of type 2 diabetes mellitus in a rural Chinese population: the Henan Rural Cohort Study. *Sci Rep*. 2020;10(1):4406. <https://doi.org/10.1038/s41598-020-61123-x> PMID: [32157171](#)
21. Ab NL, Syahid A. Machine Learning Modelling for Imbalanced Dataset: Case Study of Adolescent Obesity in Malaysia. *Journal of Advanced Research in Applied Sciences and Engineering Technology*. 2023;36(1):189–202.
22. LeCroy MN, Kim RS, Stevens J, Hanna DB, Isasi CR. Identifying Key Determinants of Childhood Obesity: A Narrative Review of Machine Learning Studies. *Child Obes*. 2021;17(3):153–9. <https://doi.org/10.1089/chi.2020.0324> PMID: [33661719](#)
23. Jeon J, Lee S, Oh C. Age-specific risk factors for the prediction of obesity using a machine learning approach. *Front Public Health*. 2023;10:998782. <https://doi.org/10.3389/fpubh.2022.998782> PMID: [36733276](#)
24. Colmenarejo G. Machine Learning Models to Predict Childhood and Adolescent Obesity: A Review. *Nutrients*. 2020;12(8):2466. <https://doi.org/10.3390/nu12082466> PMID: [32824342](#)
25. DeGregory KW, Kuiper P, DeSilvio T, Pleuss JD, Miller R, Roginski JW, et al. A review of machine learning in obesity. *Obes Rev*. 2018;19(5):668–85. <https://doi.org/10.1111/obr.12667> PMID: [29426065](#)
26. Pang X, Forrest CB, Lê-Scherban F, Masino AJ. Prediction of early childhood obesity with machine learning and electronic health record data. *International Journal of Medical Informatics*. 2021;150:104454.
27. Safaei M, Sundararajan EA, Driss M, Boulila W, Shapi'i A. A systematic literature review on obesity: Understanding the causes & consequences of obesity and reviewing various machine learning approaches used to predict obesity. *Computers in Biology and Medicine*. 2021;136:104754.
28. Underwood JM, Brener N, Thornton J, Harris WA, Bryan LN, Shanklin SL, et al. Overview and Methods for the Youth Risk Behavior Surveillance System - United States, 2019. *MMWR Supplements*. 2020;69(1):1–10.
29. Brener ND, Mpfu JJ, Krause KH, Everett Jones S, Thornton JE, Myles Z, et al. Overview and Methods for the Youth Risk Behavior Surveillance System - United States, 2023. *MMWR Suppl*. 2024;73(4):1–12. <https://doi.org/10.15585/mmwr.su7304a1> PMID: [39378301](#)
30. Burns RD. Energy balance-related factors associating with adolescent weight loss intent: evidence from the 2017 National Youth Risk Behavior Survey. *BMC Public Health*. 2019;19(1):1206. <https://doi.org/10.1186/s12889-019-7565-8> PMID: [31477084](#)
31. Brener ND, Kann L, Kinchen SA, Grunbaum JA, Whalen L, Eaton D, et al. Methodology of the youth risk behavior surveillance system. *MMWR Recomm Rep*. 2004;53(RR-12):1–13. PMID: [15385915](#)

32. Foti K, Balaji A, Shanklin S. Uses of Youth Risk Behavior Survey and School Health Profiles data: applications for improving adolescent and school health. *J Sch Health*. 2011;81(6):345–54. <https://doi.org/10.1111/j.1746-1561.2011.00601.x> PMID: 21592130
33. Bell BA, Onwuegbuzie AJ, Ferron JM, Jiao QG, Hibbard ST, Kromrey JD. Use of design effects and sample weights in complex health survey data: a review of published articles using data from 3 commonly used adolescent health surveys. *Am J Public Health*. 2012;102(7):1399–405. <https://doi.org/10.2105/AJPH.2011.300398> PMID: 22676502
34. von Elm E, Altman DG, Egger M, Pocock SJ, Gøtzsche PC, Vandenbroucke JP, et al. The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: guidelines for reporting observational studies. *Lancet*. 2007;370(9596):1453–7. [https://doi.org/10.1016/S0140-6736\(07\)61602-X](https://doi.org/10.1016/S0140-6736(07)61602-X) PMID: 18064739
35. Prevention, C.f.D.C.a. Child and Teen BMI Categories. 2024. <https://www.cdc.gov/bmi/child-teen-calculator/bmi-categories.html>
36. Guanlan S. Comparison of prediction of obesity status based on different machine learning approaches with different factor quantities. in *Proc. SPIE*. 2022.
37. Khushi Joshi E al. Comparison of Different Machine Learning and Self-Learning Methods for Predicting Obesity on Generalized and Gender-Segregated Data. *IJRITCC*. 2023;11(10):464–71. <https://doi.org/10.17762/ijritcc.v11i10.8510>
38. Dharmawan H, Sartono B, Kurnia A, Hadi AF, Ramadhani E. A study of machine learning algorithms to measure the feature importance in class-imbalance data of food insecurity cases in Indonesia. *Communications in Mathematical Biology and Neuroscience*. 2022:101.
39. Huang T, Le D, Yuan L, Xu S, Peng X. Machine learning for prediction of in-hospital mortality in lung cancer patients admitted to intensive care unit. *PLoS One*. 2023;18(1):e0280606. <https://doi.org/10.1371/journal.pone.0280606> PMID: 36701342
40. Sewpaul R, Awe OO, Dogbey DM, Sekgala MD, Dukhi N. Classification of Obesity among South African Female Adolescents: Comparative Analysis of Logistic Regression and Random Forest Algorithms. *Int J Environ Res Public Health*. 2023;21(1):2. <https://doi.org/10.3390/ijerph21010002> PMID: 38276791
41. Ren G, Xie Z, Wang Y, Liu L, Wang P, Zhang W, et al. Machine learning with interpretability predict surgical site infection after posterior cervical surgery. 2021.
42. Chae S-M, Kim MJ, Park CG, Yeo J-Y, Hwang J-H, Kwon I, et al. Association of Weight Control Behaviors with Body Mass Index in Korean Adolescents: A Quantile Regression Approach. *J Pediatr Nurs*. 2018;40:e18–25. <https://doi.org/10.1016/j.pedn.2018.01.021> PMID: 29398318
43. Negrea MO, Negrea GO, Săndulescu G, Neamtu B, Solomon A, Popa ML, et al. Assessing Lifestyle Patterns and Their Influence on Weight Status in Students from a High School in Sibiu, Romania: An Adaptation of ISCOLE Questionnaires and the Child Feeding Questionnaire. *Nutrients*. 2024;16(10):1532. <https://doi.org/10.3390/nu16101532> PMID: 38794770
44. Helgadóttir B, Baurén H, Kjellenberg K, Ekblom Ö, Nyberg G. Breakfast Habits and Associations with Fruit and Vegetable Intake, Physical Activity, Sedentary Time, and Screen Time among Swedish 13-14-Year-Old Girls and Boys. *Nutrients*. 2021;13(12):4467. <https://doi.org/10.3390/nu13124467> PMID: 34960017
45. Janssen X, Basterfield L, Parkinson KN, Pearce MS, Reilly JK, Adamson AJ, et al. Non-linear longitudinal associations between moderate-to-vigorous physical activity and adiposity across the adiposity distribution during childhood and adolescence: Gateshead Millennium Study. *Int J Obes (Lond)*. 2019;43(4):744–50. <https://doi.org/10.1038/s41366-018-0188-9> PMID: 30108270
46. Elmesmari R, Martin A, Reilly JJ, Paton JY. Comparison of accelerometer measured levels of physical activity and sedentary time between obese and non-obese children and adolescents: a systematic review. *BMC Pediatr*. 2018;18(1):106. <https://doi.org/10.1186/s12887-018-1031-0> PMID: 29523101
47. Wang Z, Xu F, Ye Q, Tse LA, Xue H, Tan Z, et al. Childhood obesity prevention through a community-based cluster randomized controlled physical activity intervention among schools in china: the health legacy project of the 2nd world summer youth olympic Games (YOG-Obesity study). *Int J Obes (Lond)*. 2018;42(4):625–33. <https://doi.org/10.1038/s41366-017-243> PMID: 28978975
48. Lister NB, Baur LA, Felix JF, Hill AJ, Marcus C, Reinehr T, et al. Child and adolescent obesity. *Nat Rev Dis Primers*. 2023;9(1):24. <https://doi.org/10.1038/s41572-023-00435-4> PMID: 37202378
49. Klein DH, Mohamoud I, Olanisa OO, Parab P, Chaudhary P, Mukhtar S. Impact of school-based interventions on pediatric obesity: A systematic review. *Cureus*. 2023;15(8):e43153.
50. Christofaro DGD, De Andrade SM, Mesas AE, Fernandes RA, Farias Júnior JC. Higher screen time is associated with overweight, poor dietary habits and physical inactivity in Brazilian adolescents, mainly among girls. *Eur J Sport Sci*. 2016;16(4):498–506. <https://doi.org/10.1080/17461391.2015.1068868> PMID: 26239965
51. Al-Hazzaa HM. Lifestyle behaviors and obesity: brief observations from the Arab teens lifestyle study (ATLS) findings. *Obesity: Open Access*. 2018;4(1).
52. Ramsey Buchanan L, Rooks-Peck CR, Finnie RKC, Wethington HR, Jacob V, Fulton JE, et al. Reducing Recreational Sedentary Screen Time: A Community Guide Systematic Review. *Am J Prev Med*. 2016;50(3):402–15. <https://doi.org/10.1016/j.amepre.2015.09.030> PMID: 26897342
53. Jonsson KR, Bailey CK, Corell M, Löfstedt P, Adjei NK. Associations between dietary behaviours and the mental and physical well-being of Swedish adolescents. *Child Adolesc Psychiatry Ment Health*. 2024;18(1):43. <https://doi.org/10.1186/s13034-024-00733-z> PMID: 38555430
54. Fiore G, Scapaticci S, Neri CR, Azaryah H, Escudero-Marín M, Pascuzzi MC, et al. Chrononutrition and metabolic health in children and adolescents: a systematic review and meta-analysis. *Nutr Rev*. 2024;82(10):1309–54. <https://doi.org/10.1093/nutrit/nuad122> PMID: 37944081

55. Szajewska H, Ruszczynski M. Systematic review demonstrating that breakfast consumption influences body weight outcomes in children and adolescents in Europe. *Crit Rev Food Sci Nutr*. 2010;50(2):113–9. <https://doi.org/10.1080/10408390903467514> PMID: [20112153](#)
56. Gordon-Larsen P, Nelson MC, Page P, Popkin BM. Inequality in the built environment underlies key health disparities in physical activity and obesity. *Pediatrics*. 2006;117(2):417–24. <https://doi.org/10.1542/peds.2005-0058> PMID: [16452361](#)
57. Wang K, Niu Y, Lu Z, Duo B, Effah CY, Guan L. The effect of breakfast on childhood obesity: a systematic review and meta-analysis. *Front Nutr*. 2023;10:1222536. <https://doi.org/10.3389/fnut.2023.1222536> PMID: [37736138](#)
58. Al-Hazzaa HM, Abahussain NA, Al-Sobayel HI, Qahwaji DM, Musaiger AO. Lifestyle factors associated with overweight and obesity among Saudi adolescents. *BMC Public Health*. 2012;12(1):354.
59. Givens DI. MILK Symposium review: The importance of milk and dairy foods in the diets of infants, adolescents, pregnant women, adults, and the elderly. *J Dairy Sci*. 2020;103(11):9681–99. <https://doi.org/10.3168/jds.2020-18296> PMID: [33076181](#)
60. Huh SY, Rifas-Shiman SL, Rich-Edwards JW, Taveras EM, Gillman MW. Prospective association between milk intake and adiposity in preschool-aged children. *J Am Diet Assoc*. 2010;110(4):563–70. <https://doi.org/10.1016/j.jada.2009.12.025> PMID: [20338282](#)
61. Barrea L, Di Somma C, Macchia PE, Falco A, Savanelli MC, Orio F, et al. Influence of nutrition on somatotrophic axis: Milk consumption in adult individuals with moderate-severe obesity. *Clinical Nutrition*. 2017;36(1):293–301.
62. Abreu S, Santos R, Moreira C, Santos PC, Vale S, Soares-Miranda L, et al. Milk intake is inversely related to body mass index and body fat in girls. *Eur J Pediatr*. 2012;171(10):1467–74. <https://doi.org/10.1007/s00431-012-1742-4> PMID: [22547119](#)
63. Malik VS, Pan A, Willett WC, Hu FB. Sugar-sweetened beverages and weight gain in children and adults: a systematic review and meta-analysis. *Am J Clin Nutr*. 2013;98(4):1084–102. <https://doi.org/10.3945/ajcn.113.058362> PMID: [23966427](#)
64. Rolls BJ. What is the role of portion control in weight management?. *Int J Obes (Lond)*. 2014;38(Suppl 1):S1–8. <https://doi.org/10.1038/ijo.2014.82> PMID: [25033958](#)
65. Wall CR, Stewart AW, Hancox RJ, Murphy R, Braithwaite I, Beasley R, et al. Association between Frequency of Consumption of Fruit, Vegetables, Nuts and Pulses and BMI: Analyses of the International Study of Asthma and Allergies in Childhood (ISAAC). *Nutrients*. 2018;10(3):316. <https://doi.org/10.3390/nu10030316> PMID: [29518923](#)
66. Demissie Z, Everett Jones S, Clayton HB, King BA. Adolescent risk behaviors and use of electronic vapor products and cigarettes. *Pediatrics*. 2017;139(2).
67. Cho B-Y, Seo D-C, Lin H-C, Lohrmann DK, Chomistek AK, Hendricks PS, et al. Adolescent Weight and Electronic Vapor Product Use: Comparing BMI-Based With Perceived Weight Status. *Am J Prev Med*. 2018;55(4):541–50. <https://doi.org/10.1016/j.amepre.2018.05.010> PMID: [30126669](#)
68. Jacobs M. Adolescent smoking: The relationship between cigarette consumption and BMI. *Addict Behav Rep*. 2018;9:100153. <https://doi.org/10.1016/j.abrep.2018.100153> PMID: [31193813](#)
69. Jakob J, Schwerdtel F, Sidney S, Rodondi N, Pletcher MJ, Reis JP, et al. Associations of cannabis use and body mass index-The Coronary Artery Risk Development in Young Adults (CARDIA) study. *Eur J Intern Med*. 2024;129:41–7. <https://doi.org/10.1016/j.ejim.2024.07.007> PMID: [38987097](#)
70. Davies-Owen J, Christiansen P, Roberts CA. Associations Between Motivations for Cannabis Use and “the Munchies”: Construct Validity of the Cannabinoid Eating Experience Questionnaire. *Subst Use Misuse*. 2025;60(1):20–7. <https://doi.org/10.1080/10826084.2024.2403121> PMID: [39279236](#)
71. Crane NA, Langenecker SA, Mermelstein RJ. Risk factors for alcohol, marijuana, and cigarette polysubstance use during adolescence and young adulthood: A 7-year longitudinal study of youth at high risk for smoking escalation. *Addict Behav*. 2021;119:106944. <https://doi.org/10.1016/j.add-beh.2021.106944> PMID: [33872847](#)
72. Smith JJ, Morgan PJ, Plotnikoff RC, Dally KA, Salmon J, Okely AD, et al. Smart-phone obesity prevention trial for adolescent boys in low-income communities: the ATLAS RCT. *Pediatrics*. 2014;134(3):e723–31. <https://doi.org/10.1542/peds.2014-1012> PMID: [25157000](#)
73. Bagherniya M, Mostafavi Darani F, Sharma M, Maracy MR, Allipour Birgani R, Ranjbar G, et al. Assessment of the Efficacy of Physical Activity Level and Lifestyle Behavior Interventions Applying Social Cognitive Theory for Overweight and Obese Girl Adolescents. *J Res Health Sci*. 2018;18(2):e00409. PMID: [29784890](#)
74. Cohen L, Manion L, Morrison K. Surveys, longitudinal, cross-sectional and trend studies. 2017. p. 334–60.
75. Gopinath B, Hardy LL, Baur LA, Burlutsky G, Mitchell P. Physical activity and sedentary behaviors and health-related quality of life in adolescents. *Pediatrics*. 2012;130(1):e167–74. <https://doi.org/10.1542/peds.2011-3637> PMID: [22689863](#)
76. Dong X, Ding L, Zhang R, Ding M, Wang B, Yi X. Physical activity, screen-based sedentary behavior and physical fitness in Chinese adolescents: A cross-sectional study. 2021;9.