

A Clinically Interpretable Machine Learning Framework for Knowledge Discovery and Early Outcome Prediction in Pediatric Obesity Intervention

Lichuan Liu ^{1,*}, Beth Moxley ² and Nicole Klinkhamer ³

¹ Department of Electrical Engineering, Northern Illinois University, DeKalb, IL 60115, USA

² School of Nursing, Northern Illinois University, DeKalb, IL 60115, USA

³ Pivotal Health Advisors, Geneva, IL 60134, USA

* Correspondence: liu@niu.edu

How To Cite: Liu, L.; Moxley, B.; Klinkhamer, N. A Clinically Interpretable Machine Learning Framework for Knowledge Discovery and Early Outcome Prediction in Pediatric Obesity Intervention. *Journal of Artificial Intelligence for Automation* **2026**, *1*(2), 14. <https://doi.org/10.53941/jaia.2026.100014>

Received: 10 July 2026

Revised: 8 September 2026

Accepted: 20 September 2026

Published: 30 September 2026

Abstract: Pediatric obesity remains a major public health challenge, yet substantial variability in treatment response limits the effectiveness of standardized intervention programs. Early assessment of short-term weight-change trajectories may provide useful information during an intervention. This study presents a clinically interpretable machine learning framework for predicting short-term pediatric obesity intervention outcomes using routinely collected demographic, anthropometric, body-composition, and early-response measurements from the ProActive Kids program. A systematic workflow including data understanding, clinical data preprocessing, feature engineering, and predictive modeling is developed. Three complementary machine learning paradigms, Logistic Regression, Random Forest, and Extreme Gradient Boosting (XGBoost), are comparatively evaluated to investigate predictive performance and feature importance. Baseline prediction models demonstrate limited predictive performance using only pre-intervention measurements, whereas incorporation of Week-3 early-response information substantially improves prediction across all three classifiers. Week-3 percentage weight change emerges as the dominant predictor of the Week-8 weight-reduction outcome. Waist-to-Hip Ratio, Body Fat Percentage, Body Mass Index, and Fat-Free Mass are also identified as influential baseline variables, although their individual importance should be interpreted cautiously because of correlations among anthropometric and body-composition measures. These findings demonstrate that machine learning can support short-term outcome prediction while identifying potentially clinically relevant patterns in routinely collected pediatric obesity data. The results highlight the predictive importance of early intervention response and may inform future studies of pediatric obesity intervention outcomes.

Keywords: pediatric obesity; machine learning; early intervention; body composition; predictive modeling; personalized healthcare

1. Introduction

Pediatric obesity has become one of the most significant public health challenges worldwide and is associated with increased risks of cardiovascular disease, type 2 diabetes, metabolic disorders, and numerous long-term health complications [1,2]. Although multidisciplinary lifestyle intervention programs have demonstrated encouraging outcomes, treatment responses vary considerably among individual children. Some participants achieve substantial improvements in body weight and body composition, whereas others exhibit only limited response despite receiving similar interventions [3,4]. Understanding why children respond differently remains an important clinical question,

as early identification of likely participants with and without weight decrease could facilitate more effective treatment planning, improve allocation of healthcare resources, and ultimately support more personalized pediatric obesity management [5].

Current clinical evaluation of pediatric obesity interventions primarily focuses on assessing overall program effectiveness using conventional statistical analyses. These approaches have provided valuable evidence regarding intervention efficacy and the influence of individual clinical variables [6]. However, pediatric obesity is a multifactorial condition influenced by complex interactions among demographic characteristics, anthropometric measurements, body composition, and physiological response to treatment. Conventional statistical approaches generally evaluate variables individually or test predefined hypotheses, making it difficult to fully characterize the complex relationships embedded within routinely collected clinical data. Consequently, clinically meaningful patterns associated with the Week-8 weight-reduction outcome may remain undiscovered.

Recent advances in machine learning provide an opportunity to address this challenge. Unlike traditional statistical methods that primarily quantify associations between individual variables and clinical outcomes, machine learning can analyze multiple interacting clinical measurements and identify complex predictive relationships directly from the data [7–11]. Beyond prediction, machine learning provides a powerful analytical framework for discovering latent patterns that may not be readily apparent by using conventional statistical approaches [12]. Such knowledge has the potential to improve our understanding of treatment response while providing clinically interpretable evidence to support individualized intervention strategies.

The objective of this study is therefore not only to develop predictive models for short-term pediatric obesity intervention outcomes, but also to investigate whether machine learning can reveal clinically meaningful relationships embedded within routinely collected clinical measurements. Three complementary machine learning algorithms, representing linear, ensemble bagging, and gradient boosting learning paradigms, are investigated. Rather than relying on a single prediction model, the proposed framework compares multiple fundamentally different learning mechanisms to determine whether they consistently identify similar clinical predictors of weight-decrease. Such cross-model agreement would increase confidence that the discovered relationships reflect meaningful characteristics of the underlying clinical data rather than artifacts of a particular machine learning algorithm.

The proposed framework is evaluated by using participant data collected from the ProActive Kids (PAK) program, a multidisciplinary lifestyle intervention designed to improve health outcomes among children and adolescents with overweight or obesity [5]. The dataset contains routinely collected demographic, anthropometric, body composition, and longitudinal weight measurements acquired throughout the intervention program. Because these measurements are commonly obtained during routine clinical practice, the resulting machine learning framework is designed to utilize readily available clinical information without requiring specialized laboratory testing or advanced imaging. Unlike many publicly available benchmark datasets, the PAK dataset contains longitudinal clinical measurements collected throughout an active pediatric obesity intervention program, enabling investigation of both baseline physiological characteristics and early treatment response. Unlike previous studies that primarily evaluated the overall effectiveness of the intervention program, the present study investigates why individual children respond differently by combining longitudinal clinical measurements with complementary machine learning paradigms for knowledge discovery and early outcome prediction.

An overview of the proposed machine learning framework is illustrated in Figure 1. The workflow consists of data understanding, clinical data preprocessing, feature engineering, predictive model development, performance evaluation, and cross-model feature importance analysis. The resulting framework provides both quantitative prediction of intervention outcome and clinically interpretable insights into the physiological characteristics associated with pediatric obesity intervention. Beyond improving prediction accuracy, the proposed framework illustrates how machine learning can transform routinely collected clinical data into actionable knowledge that supports clinical decision making and future intelligent healthcare systems [13].

The primary contributions of this work can be summarized as follows:

A clinically interpretable machine learning framework is developed that integrates systematic data understanding, clinical data preprocessing, and feature engineering using routinely collected demographic, anthropometric, body composition, and early-response measurements. The framework is designed to utilize information readily available in routine clinical practice without requiring specialized laboratory testing or advanced imaging.

Three complementary machine learning paradigms representing fundamentally different learning mechanisms, namely Logistic Regression, Random Forest, and Extreme Gradient Boosting, are comparatively investigated [10,14–16]. Rather than relying on a single prediction algorithm, the proposed framework examines whether different learning paradigms consistently discover similar relationships between clinical measurements and intervention outcomes.

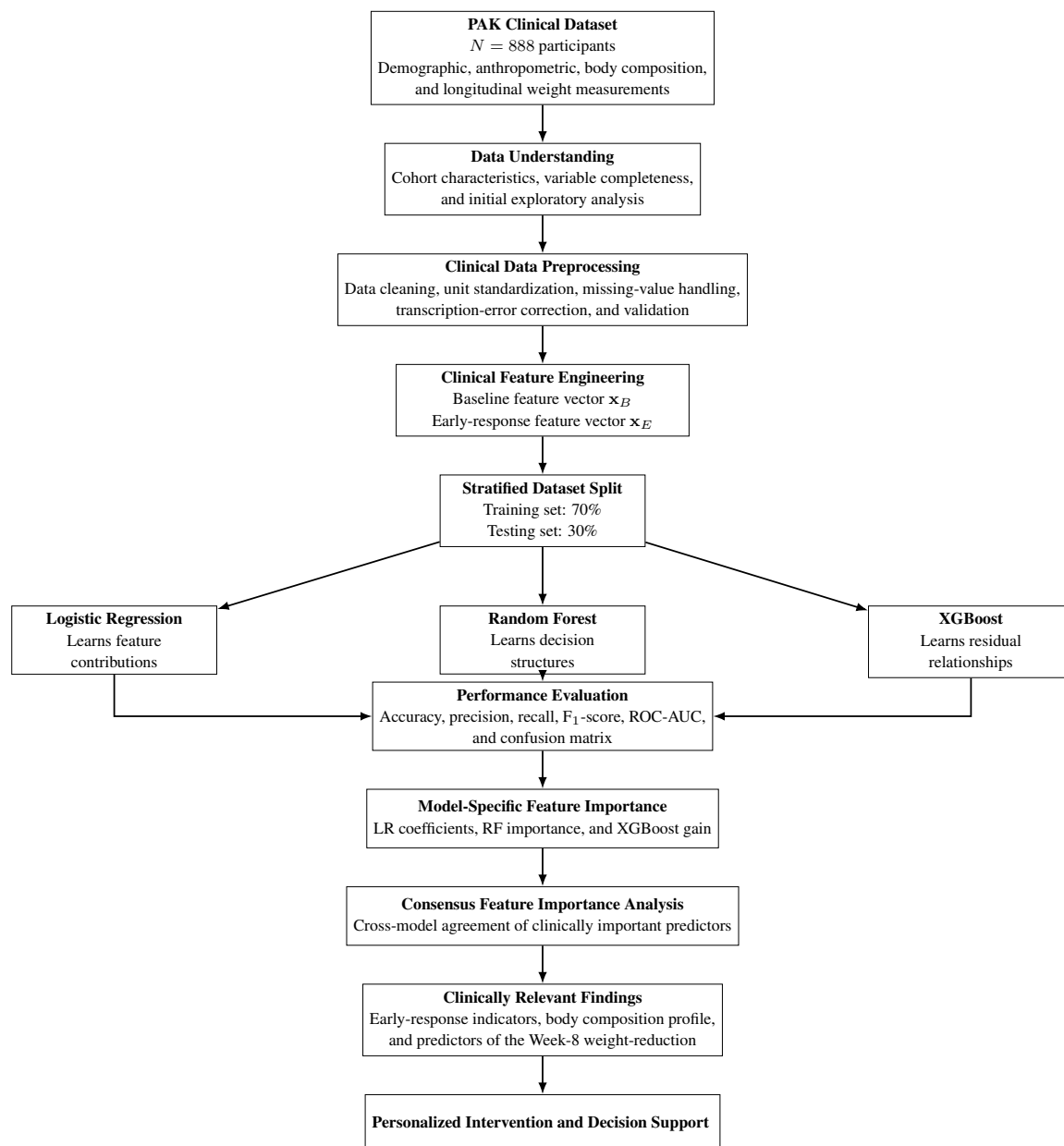


Figure 1. Overview of the proposed clinically interpretable machine learning framework for knowledge discovery and early prediction of pediatric obesity intervention outcomes. The framework integrates data understanding, clinical preprocessing, feature engineering, complementary learning paradigms, model evaluation, consensus feature importance analysis, and Clinically Relevant Findings to support future personalized intervention planning.

Third, comparison across multiple learning paradigms identifies a set of variables with relatively consistent predictive importances associated with the Week-8 weight-reduction outcome. The results demonstrate that machine learning serves not only as an accurate prediction tool but also as a knowledge discovery framework capable of revealing potentially clinically relevant information embedded within routinely collected longitudinal clinical data. These findings provide new insights into the roles of body composition profile and early treatment response in supporting future personalized pediatric obesity interventions.

2. Materials and Methods

This study employed a data-driven machine learning framework to investigate whether routinely collected clinical measurements and early intervention responses can predict short-term pediatric obesity intervention outcomes. The overall methodology consisted of six sequential stages: (1) study cohort and dataset description; (2) data understanding; (3) data preprocessing; (4) clinical feature engineering; (5) machine learning model development; and (6) model evaluation. An overview of the proposed framework is illustrated in Figure 1.

2.1. Study Cohort and Dataset

The data analyzed in this study were obtained from the ProActive Kids (PAK) Foundation, a non-profit pediatric obesity intervention program established in 2008. The program aimed to improve body composition, physical fitness, nutritional awareness, and overall health through a structured multidisciplinary intervention designed for overweight and obese children and adolescents [5].

The ProActive Kids (PAK) program primarily serves children and adolescents with overweight or obesity. Participants enrolled in the program were required to have a body mass index (BMI) at or above the 85th percentile for their age and sex. Prior to enrollment, anthropometric measurements were verified by healthcare providers to ensure consistency with program eligibility requirements.

The intervention consisted of an intensive eight-week program integrating supervised physical activity, nutritional education, lifestyle modification, and family participation. Participants attended structured exercise and educational sessions throughout the week, while parents or primary caregivers actively participated in family-based educational activities designed to reinforce healthy behavioral changes outside the program [1,5].

The original database contains demographic information, anthropometric measurements, body composition variables, attendance records, and longitudinal weight measurements collected throughout the intervention period [17]. These routinely collected clinical variables served as the foundation for subsequent data preprocessing, feature engineering, and predictive modeling.

The PAK dataset was collected from 2010 to 2017 as part of an eight-week pediatric obesity intervention program. Program inclusion criteria were: (1) BMI at or above the 85th percentile for age and sex; (2) age 5–17 years; and (3) residence in Cook, DuPage, Will, or Kane County, Illinois. Age and weight eligibility were validated by a healthcare provider. No additional exclusion criteria were specified for participation in the PAK program. The analytical dataset included participants ranging from 5 to 19 years of age, with a small number of observations above the program's intended age range. Height, weight, body fat percentage, and fat-free mass were collected by PAK-trained personnel using standardized measurement procedures. Body composition, including body fat percentage, was assessed using bioelectrical impedance analysis (BIA). Waist circumference was measured horizontally at the midpoint between the lower edge of the last palpable rib and the iliac crest, and hip circumference was measured horizontally around the widest portion of the buttocks using a flexible, non-stretchable measuring tape. Right Fit Sport Wellness trained PAK personnel in the standardized body-composition measurement procedures. Measurements were collected at Weeks 1 and 8, with selected measurements also collected at Weeks 3 and 5.

Raw weight and BMI measurements were retained as predictive variables to characterize within-participant changes during the eight-week intervention. For clinical assessment and program eligibility, BMI was interpreted using age- and sex-specific pediatric growth charts.

Participant data were de-identified by PAK before being provided to the researchers. All participants or their parents/legal guardians provided informed consent prior to participation. The de-identified dataset was therefore considered exempt from IRB review.

2.2. Data Understanding

Before predictive modeling, an exploratory analysis was conducted to understand the characteristics, completeness, and overall quality of the dataset. Variables were grouped into four major categories:

- Demographic variables (age and sex);
- Baseline anthropometric measurements (height, weight, BMI, waist circumference, hip circumference, body fat percentage, fat mass, fat-free mass, and waist-to-hip ratio);
- Longitudinal intervention measurements (weekly body weight);
- Clinical outcome variables derived from the intervention.

Following preliminary quality assessment, a total of 888 participants were retained for analysis. The cohort exhibited an almost balanced sex distribution, consisting of 446 males and 440 females, with only two records containing invalid sex entries that were subsequently corrected during preprocessing.

Overall, the dataset demonstrated high completeness for baseline anthropometric measurements, with only a limited number of missing values observed across individual variables. Early exploratory analysis also revealed several data quality issues, including inconsistent measurement units, formatting inconsistencies, typographical errors, and incomplete records. The full dataset contained 888 participants. Sex information was available for 886 participants (446 male and 440 female) and missing for two participants. Missing values were present in several predictor variables, particularly age, early-response measurements, and anthropometric/body-composition measurements, as shown in Table 1. Complete-case analysis was applied separately to the variables required for model development. For the early-response models, 600

participants had complete data across the 12 predictors and outcome and were included in the analysis. The analytical cohort was divided using a stratified 70/30 split, resulting in 420 training and 180 testing participants. These observations motivated the systematic preprocessing procedures described in the following subsection.

Table 1. Missing data among variables used in the early-response model.

Variable	Missing, n	Missing, %
Age	189	21.3
Early Response Percentage	105	11.8
Waist-to-Hip Ratio	95	10.7
Hip Circumference	77	8.7
Waist Circumference	69	7.8
Fat Mass	53	6.0
Fat-Free Mass	53	6.0
Body Fat Percentage	52	5.9
Starting Height	7	0.8
Starting Weight	7	0.8
Starting BMI	4	0.5
Sex	2	0.2

Percentages are calculated relative to the full cohort of 888 participants. Complete-case analysis across all 12 predictors and the Week-8 outcome resulted in 600 participants available for the early-response model.

2.3. Data Preprocessing

High-quality data are essential for developing reliable machine learning models. Therefore, a comprehensive preprocessing procedure was performed before feature engineering and model development.

Table 2 summarizes the overall characteristics of the study cohort and the major categories of variables available for machine learning analysis. In addition to demographic information, the dataset includes baseline anthropometric measurements, body composition variables, longitudinal weight measurements collected throughout the intervention, and the final intervention outcome. Together, these variables provide the foundation for subsequent data preprocessing, clinical feature engineering, and predictive model development. The preprocessing process consisted of four major stages. First, demographic and anthropometric variables were examined individually to identify missing values, inconsistent formatting, impossible measurements, duplicated records, and transcription errors. Second, measurement units were standardized across the entire dataset. Height values recorded using mixed feet-inch and inch formats were converted into a consistent unit (inches), while weight measurements were converted into numerical values suitable for computation.

Table 2. Summary of the study cohort following initial data screening.

Characteristic	Value
Participants included	888
Male	446
Female	440
Age range	5–19 years
Intervention duration	8 weeks
Demographic variables	Age, Sex
Baseline anthropometric variables	Height, Weight, BMI
Body composition variables	Body Fat, Fat Mass, Fat-Free Mass
Body shape variables	Waist, Hip, Waist-to-Hip Ratio
Longitudinal measurements	Weekly body weight
Outcome variable	Week-8 weight-decrease

Third, obvious data entry errors were manually reviewed using longitudinal consistency across repeated measurements whenever possible. Invalid categorical entries, such as incorrectly coded sex values, were corrected or removed after verification. Finally, derived variables including BMI were independently recalculated from cleaned height and weight measurements to validate data consistency.

Rather than relying solely on automated cleaning procedures, manual inspection was incorporated throughout the preprocessing process to preserve clinically meaningful observations while minimizing the influence of transcription and formatting errors. The major preprocessing procedures are summarized in Table 3.

Table 3. Summary of data preprocessing procedures performed before feature engineering and machine learning model development.

Data Category	Quality Issue	Preprocessing Procedure	Purpose
Demographic variables	Invalid categorical entries and missing information	Standardized categorical encoding and verified inconsistent records	Improve data consistency
Height measurements	Mixed recording formats (feet/inches and inches)	Converted all measurements to inches and corrected obvious transcription errors	Standardize measurement units
Weight measurements	Formatting inconsistencies and transcription errors	Converted to numerical values and verified using longitudinal measurements	Improve reliability
Body Mass Index (BMI)	Potential calculation inconsistencies	Independently recalculated from cleaned height and weight measurements	Data validation
Body composition variables	Incomplete observations	Complete-case analysis for model development	Maintain data quality
Derived outcome variables	Not directly available	Calculated from cleaned longitudinal weight measurements	Prediction target generation
Feature encoding	Categorical variables	Numerical encoding for machine learning models	Model compatibility

2.4. Clinical Feature Engineering

Feature engineering plays a critical role in transforming routinely collected clinical measurements into informative predictors for machine learning analysis. Rather than utilizing every available variable directly, clinically meaningful features were selected and engineered to characterize participants' baseline body composition profiles and their early physiological response to the intervention [17]. The feature engineering process was guided by clinical relevance, data completeness, and the objective of developing prediction models that remain interpretable for healthcare providers.

Two complementary prediction scenarios were investigated in this study. The first focused exclusively on baseline measurements collected before the intervention to determine whether treatment outcome could be predicted during the initial clinical assessment. The second incorporated early intervention response observed during the first three weeks of the program to evaluate the additional predictive value of longitudinal information.

2.4.1. Baseline Prediction Features

The baseline prediction model utilized routinely collected demographic information and anthropometric measurements obtained prior to intervention. These variables represent information immediately available to clinicians during the initial evaluation and therefore provide the earliest opportunity for predicting treatment outcome.

The baseline prediction model utilized eleven routinely collected clinical variables. Let

$$\begin{aligned}
 x_1 &: \text{Sex,} \\
 x_2 &: \text{Age,} \\
 x_3 &: \text{Height,} \\
 x_4 &: \text{Weight,} \\
 x_5 &: \text{Body Mass Index (BMI),} \\
 x_6 &: \text{Body Fat Percentage,} \\
 x_7 &: \text{Fat Mass,} \\
 x_8 &: \text{Fat-Free Mass,} \\
 x_9 &: \text{Waist Circumference,} \\
 x_{10} &: \text{Hip Circumference,} \\
 x_{11} &: \text{Waist-to-Hip Ratio (WHR).}
 \end{aligned} \tag{1}$$

The baseline feature vector is therefore expressed as

$$\mathbf{x}_B = [x_1, x_2, \dots, x_{11}]^T. \quad (2)$$

These variables collectively describe each participant's baseline body composition profile from multiple perspectives. Demographic variables characterize patient background, anthropometric measurements describe overall body size, while body composition and fat distribution variables provide additional information beyond BMI alone. Such complementary measurements allow machine learning models to evaluate interactions among multiple clinical characteristics that may contribute to the Week-8 weight-reduction outcome.

2.4.2. Early-Response Prediction Features

To investigate the additional predictive value of early intervention progress, an engineered variable was constructed from body weight measurements collected during the first three weeks of the program.

Let x_{12} be the Early Response Percentage,

$$x_{12} = \frac{W_3 - W_0}{W_3} \times 100\%. \quad (3)$$

The early-response prediction model therefore extended the baseline feature vector as

$$\mathbf{x}_E = [\mathbf{x}_B^T, x_{12}]^T. \quad (4)$$

These engineered variables capture early physiological response to the intervention and enable evaluation of whether short-term progress substantially improves prediction of the final intervention outcome.

2.4.3. Outcome Definition

The primary objective of this study was to predict short-term weight-reduction outcome at the completion of the eight-week pediatric obesity program.

Accordingly, the prediction task was formulated as a supervised binary classification problem. Participants meeting the predefined clinical criterion for the Week-8 weight-reduction outcome were assigned a class label of one, while all remaining participants were assigned a class label of zero. The resulting binary outcome variable served as the prediction target for both the baseline and early-response machine learning models.

2.5. Machine Learning Framework

The objective of this study was to develop machine learning models capable of predicting short-term pediatric obesity intervention outcomes using routinely collected clinical measurements. Two prediction scenarios were investigated. The first utilized only baseline clinical measurements collected prior to intervention, while the second incorporated early treatment response observed during the first three weeks of the program. Both prediction tasks were formulated as supervised binary classification problems [18].

Let

$$\mathbf{x} = [x_1, x_2, \dots, x_p]^T, \quad (5)$$

where x_j denotes the j^{th} clinical feature and p represents the number of input variables. The corresponding intervention outcome is represented as

$$y_i \in \{0, 1\}, \quad (6)$$

where $y_i = 1$ indicates the Week-8 weight-reduction outcome and $y_i = 0$ otherwise.

The complete dataset is therefore expressed as

$$\mathcal{D} = \{(\mathbf{x}_i, y_i)\}_{i=1}^N, \quad (7)$$

where $N = 888$ participants were included in the present study. The notation used in the machine learning formulation is summarized in Table 4.

Three complementary machine learning algorithms were investigated: Logistic Regression, Random Forest, and Extreme Gradient Boosting (XGBoost). These classifiers represent linear statistical learning, bagging-based ensemble learning, and boosting-based ensemble learning, respectively. Comparing these fundamentally different learning paradigms provides insight into both predictive performance and the consistency of clinically important predictors identified by different modeling approaches.

Table 4. Notation used in the machine learning formulation.

Symbol	Description
N	Number of participants
p	Number of input features
$\mathbf{x}_i$	Feature vector of participant i
x_{ij}	j^{th} feature of participant i
y_i	Binary intervention outcome
$\hat{y}_i$	Predicted outcome
$P(y = 1 \mathbf{x})$	Predicted probability of the Week-8 weight-reduction

2.5.1. Logistic Regression

Logistic Regression (LR) was selected as the baseline classification model because of its simplicity, computational efficiency, and high interpretability [19,20]. The underlying assumption of Logistic Regression is that the intervention outcome can be described by a weighted combination of multiple clinical measurements. Specifically, the relationship between the engineered clinical features and the intervention outcome is first modeled as

$$z = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_p x_p, \quad (8)$$

where x_i denotes the i^{th} clinical feature and β_i represents its corresponding coefficient learned from the training data. Each coefficient quantifies the relative contribution of an individual clinical measurement to the prediction of weight-decrease and no-weight-decrease. Positive coefficients indicate that increasing the corresponding feature increases the likelihood of the Week-8 weight-reduction, whereas negative coefficients indicate the opposite.

Because the intervention outcome is binary (the Week-8 weight-reduction or the Week-8 weight-reduction), the weighted linear combination cannot be directly interpreted as a probability. Therefore, the linear output is transformed using the logistic (sigmoid) function,

$$P(y = 1|\mathbf{x}) = \frac{1}{1 + \exp(-z)}, \quad (9)$$

which maps the continuous linear response into a probability between 0 and 1. During model training, the regression coefficients are automatically estimated by minimizing the prediction error over the training dataset. Consequently, Logistic Regression provides not only outcome prediction but also an interpretable quantitative representation of the relationships between routinely collected clinical measurements and weight-decrease.

2.5.2. Random Forest

Unlike Logistic Regression, which models the intervention outcome as a weighted combination of all clinical features, Random Forest (RF) learns a collection of hierarchical decision rules directly from the data. Each decision tree recursively partitions participants into progressively more homogeneous groups by selecting the clinical feature and decision threshold that provide the greatest reduction in uncertainty [21].

The quality of each candidate split is evaluated using the information gain,

$$IG(D, x_j) = H(D) - \sum_{k=1}^m \frac{|D_k|}{|D|} H(D_k), \quad (10)$$

where D denotes the set of training samples at the current node, D_k represents the resulting subsets after splitting on feature x_j , and $H(\cdot)$ is the entropy defined as

$$H(D) = - \sum_{c=1}^C p_c \log_2(p_c), \quad (11)$$

where p_c denotes the proportion of samples belonging to class c .

Entropy measures the uncertainty of the intervention outcome within a group of participants. During model training, every candidate feature and splitting threshold are evaluated, and the split that maximizes the information gain is selected. Consequently, each decision node identifies the clinical measurement that most effectively separates participants with different intervention outcomes, thereby progressively revealing the underlying decision structure embedded in the clinical data.

A single decision tree may overfit the training data because of sampling variability. Therefore, Random Forest constructs multiple decision trees using bootstrap sampling of participants together with randomly selected subsets of predictor variables. Final classification is determined by majority voting across all decision trees, resulting in a more robust and stable prediction model.

From a clinical perspective, Random Forest discovers combinations of clinical characteristics that repeatedly distinguish weight-decrease and no-weight-decrease outcomes. Clinical variables that consistently produce informative decision rules across many decision trees receive higher feature importance scores. Therefore, Random Forest feature importance reflects the ability of each clinical measurement to partition participants into groups with different probabilities of weight-decrease and no-weight-decrease intervention outcomes throughout the ensemble.

2.5.3. Extreme Gradient Boosting

Extreme Gradient Boosting (XGBoost) extends the concept of decision tree learning by constructing decision trees sequentially rather than independently [22]. Unlike Random Forest, where each tree is developed independently, XGBoost continuously improves the prediction model by allowing each newly generated tree to correct the prediction errors made by the existing ensemble. Consequently, the model progressively discovers additional relationships between the clinical features and intervention outcome that were not captured by the previous trees.

The prediction at the t^{th} boosting iteration can be expressed as

$$\hat{y}^{(t)} = \hat{y}^{(t-1)} + f_t(\mathbf{x}), \quad (12)$$

where $\hat{y}^{(t-1)}$ denotes the prediction produced by the existing ensemble, $f_t(\mathbf{x})$ represents the newly constructed decision tree, and $\mathbf{x}$ is the input features. Rather than rebuilding the entire model, each additional tree contributes only the information required to reduce the remaining prediction errors.

The objective function optimized during training is

$$\mathcal{L} = \sum_{i=1}^N l(y_i, \hat{y}_i) + \sum_k \Omega(f_k), \quad (13)$$

where $l(\cdot)$ measures the discrepancy between the predicted and observed intervention outcomes, and $\Omega(\cdot)$ is a regularization term that controls model complexity and reduces overfitting. During each boosting iteration, the algorithm searches for the new decision tree that minimizes the objective function, thereby progressively reducing the prediction error while maintaining good model generalization.

From a learning perspective, XGBoost continually discovers residual relationships that remain unexplained by the existing ensemble. As successive trees are added, increasingly complex nonlinear interactions among the clinical measurements can be identified. Consequently, feature importance in XGBoost reflects the extent to which individual clinical variables repeatedly contribute to reducing the remaining prediction errors throughout the boosting process. Clinical features with consistently high importance therefore contain complementary predictive information beyond that explained by other variables in the dataset [23].

2.5.4. Model Training

The dataset was randomly divided into training (70%) and testing (30%) subsets while preserving the class distribution using stratified sampling. Model development was performed exclusively using the training dataset, whereas the testing dataset was reserved for independent performance evaluation.

Default hyperparameters implemented in the Scikit-learn and XGBoost libraries were initially adopted to establish baseline performance [14]. For the XGBoost classifier, the principal hyperparameters were selected as

$$N_{\text{tree}} = 300, \quad d_{\text{max}} = 3, \quad \eta = 0.05,$$

where N_{tree} denotes the number of boosting trees, d_{max} is the maximum tree depth, and η is the learning rate. To improve model generalization and reduce overfitting, stochastic sampling was employed with

$$r_{\text{sub}} = 0.8, \quad r_{\text{col}} = 0.8,$$

where r_{sub} represents the fraction of training samples randomly selected for constructing each tree (subsample), and r_{col} denotes the fraction of predictor variables randomly selected for tree construction (colsample_bytree). These hyperparameters were selected to provide a balance between prediction accuracy, model complexity, and computational efficiency without extensive hyperparameter optimization.

The primary objective of this study was not hyperparameter optimization, but rather to investigate whether clinically meaningful predictors could be consistently identified across different machine learning paradigms.

2.6. Performance Evaluation

Model performance was evaluated using five commonly adopted classification metrics: accuracy, precision, recall, F1-score, and the area under the receiver operating characteristic curve (ROC-AUC). These complementary metrics collectively characterize overall classification performance while accounting for potential class imbalance.

Beyond predictive accuracy, feature importance analysis was performed to investigate the clinical variables contributing most strongly to intervention outcome prediction. Logistic Regression coefficients were used to quantify the direction and magnitude of linear associations between predictors and weight-decrease. Random Forest feature importance was calculated from the average impurity reduction across all decision trees, whereas XGBoost feature importance was obtained from the cumulative gain contributed by each predictor during gradient boosting.

To assess model stability and reduce dependence on a single train-test split, stratified five-fold cross-validation was additionally performed. Accuracy, precision, recall, F1-score, and ROC-AUC were summarized as the mean $\pm$ standard deviation across the five folds. For ROC-AUC, 95% confidence intervals for the mean cross-validation performance were calculated using the five-fold-specific ROC-AUC values, as shown in Table 5.

Table 5. Five-fold cross-validation performance of baseline and early-response models. Values are mean $\pm$ standard deviation across five folds. LR: Logistic Regression; RF: Random Forest.

Features	Model	Acc.	Prec.	Recall	F1	AUC	95% CI
Baseline	LR	0.596 $\pm$ 0.021	0.612 $\pm$ 0.012	0.903 $\pm$ 0.027	0.729 $\pm$ 0.014	0.577 $\pm$ 0.033	0.536–0.619
Baseline	RF	0.546 $\pm$ 0.042	0.603 $\pm$ 0.028	0.721 $\pm$ 0.043	0.657 $\pm$ 0.032	0.518 $\pm$ 0.044	0.463–0.572
Baseline	XGBoost	0.535 $\pm$ 0.022	0.597 $\pm$ 0.017	0.698 $\pm$ 0.037	0.644 $\pm$ 0.021	0.528 $\pm$ 0.024	0.498–0.558
Early	LR	0.720 $\pm$ 0.027	0.742 $\pm$ 0.018	0.827 $\pm$ 0.049	0.781 $\pm$ 0.025	0.780 $\pm$ 0.031	0.741–0.819
Early	RF	0.713 $\pm$ 0.016	0.739 $\pm$ 0.018	0.816 $\pm$ 0.025	0.775 $\pm$ 0.011	0.734 $\pm$ 0.036	0.690–0.779
Early	XGBoost	0.712 $\pm$ 0.043	0.748 $\pm$ 0.035	0.794 $\pm$ 0.049	0.769 $\pm$ 0.034	0.740 $\pm$ 0.046	0.682–0.797

Finally, feature rankings obtained from the three machine learning algorithms were compared to identify predictors that consistently demonstrated high importance across fundamentally different learning paradigms. Such cross-model agreement increases confidence that the identified variables represent meaningful clinical characteristics rather than model-specific artifacts.

3. Results

This section evaluates the proposed machine learning framework from two complementary perspectives. First, the predictive performance of the baseline and early-response models is compared using three representative machine learning algorithms. Second, feature importance analysis is performed to investigate the clinical variables most strongly associated with week-8 weight-decrease and no-weight-decrease outcomes. Rather than focusing solely on prediction accuracy, the objective is to determine whether different machine learning paradigms consistently identify clinically meaningful predictors embedded within routinely collected pediatric obesity data.

3.1. Prediction Performance

Three supervised machine learning classifiers, including Logistic Regression (LR), Random Forest (RF), and Extreme Gradient Boosting (XGBoost), were evaluated under two prediction scenarios. The baseline model utilized only pre-intervention clinical measurements, whereas the early-response model additionally incorporated engineered features describing weight change during the first three weeks of the intervention.

Five-fold cross-validation demonstrated stronger predictive performance for models incorporating early-response information than for models using baseline features alone. Among the baseline models, logistic regression achieved the highest ROC-AUC of 0.577 $\pm$ 0.033 (95% CI: 0.536–0.619). With the inclusion of Week-3 early-response information, logistic regression achieved a ROC-AUC of 0.780 $\pm$ 0.031 (95% CI: 0.741–0.819), while Random Forest and XGBoost achieved ROC-AUC values of 0.734 $\pm$ 0.036 and 0.740 $\pm$ 0.046, respectively. These cross-validation results indicate that Week-3 early-response information provides substantial predictive information for the Week-8 weight-reduction outcome.

Table 6 summarizes the prediction performance on the independent testing dataset.

Table 6. Performance comparison of baseline and early-response prediction models.

Prediction Model	Classifier	Accuracy	Precision	Recall	F ₁	ROC-AUC
Baseline	Logistic Regression	0.590	0.612	0.863	0.716	0.572
	Random Forest	0.559	0.618	0.692	0.653	0.559
	XGBoost	0.564	0.619	0.709	0.661	0.554
Early Response	Logistic Regression	0.744	0.774	0.817	0.795	0.795
	Random Forest	0.728	0.754	0.817	0.784	0.772
	XGBoost	0.733	0.761	0.817	0.788	0.753

The Logistic Regression early-response model correctly identified 89 participants who experienced a Week-8 weight decrease and 45 participants who did not experience a Week-8 weight decrease, while producing 26 false positives and 20 false negatives. The confusion matrix for the Logistic Regression early-response model is presented in Table 7.

Table 7. Confusion matrix for the Logistic Regression early-response model.

	Predicted No	Predicted Yes
Actual No	45	26
Actual Yes	20	89

3.2. Value of Early-Response Information

To evaluate the additional predictive value of early treatment response, the prediction performance of the baseline and early-response models was compared across the three machine learning algorithms.

Incorporating the engineered early-response features substantially improved prediction performance for all classifiers. Logistic Regression exhibited the largest improvement in overall accuracy, increasing from 58.97% using baseline measurements alone to 74.4% after incorporating early-response information. Similar improvements were observed for Random Forest and XGBoost.

These results indicate that body weight changes observed during the first three weeks of intervention contain clinically meaningful information regarding subsequent weight-decrease. More importantly, the consistent performance improvement across three fundamentally different machine learning paradigms suggests that the Week-3 percentage weight-change measure captured most of the predictive information in the early-response model, while the addition of baseline variables provided limited incremental predictive benefit.

3.3. Consensus Feature Importance Analysis

Logistic Regression, Random Forest, and XGBoost employ different learning mechanisms and measures of feature importance. Nevertheless, all three models identified Early Response (%) as the most influential predictor. Several anthropometric and body-composition variables also appeared among the more influential baseline predictors, although their exact rankings varied across models.

Table 8 summarizes the consensus ranking based on the mean rank of each feature across the three classifiers. Lower mean ranks indicate greater overall importance.

Early Response (%) was ranked first by all three machine learning models, indicating that the magnitude of the initial response to the intervention contained the strongest predictive information for the week-8 outcome. This agreement is particularly notable because the three models quantify feature importance differently.

Among the baseline clinical measurements, Waist-to-Hip Ratio achieved the highest consensus ranking. BMI, Age, Body Fat Percentage, Waist Circumference, and Fat-Free Mass also appeared among the more influential baseline predictors. However, their model-specific rankings varied, indicating that conclusions about the relative importance of individual baseline variables should be interpreted cautiously. Sex had the lowest consensus ranking, suggesting limited additional predictive contribution within the present dataset.

Recent machine-learning studies of pediatric weight status have similarly emphasized the importance of evaluating the stability of identified predictors across feature-selection approaches and models [24]. In the present study, correlations among anthropometric and body-composition variables may distribute predictive importance across related features and reduce the stability of their individual rankings. Accordingly, the consensus rankings should be regarded as exploratory indicators of predictive relevance rather than evidence of independent clinical effects.

Table 8. Consensus ranking of feature importance across Logistic Regression (LR), Random Forest (RF), and XGBoost (XGB). The mean rank was calculated from the model-specific rankings after removal of the binary early-response variable. Lower mean rank indicates greater overall importance.

Clinical Feature	LR	RF	XGB	Mean Rank	Consensus Rank	Clinical Interpretation
Early Response (%)	1	1	1	1.00	1	Early treatment response
Waist-to-Hip Ratio	2	4	7	4.33	2	Body fat distribution
BMI	5	8	4	5.67	3	Overall obesity severity
Age	12	2	3	5.67	3	Growth and developmental stage
Body Fat (percentage)	3	5	11	6.33	5	Body composition
Waist Circumference	8	9	2	6.33	5	Central adiposity
Fat-Free Mass	11	3	6	6.67	7	Lean body mass
Hip Circumference	4	10	8	7.33	8	Body shape
Height	7	11	5	7.67	9	Physical growth
Fat Mass	9	6	10	8.33	10	Absolute fat burden
Weight	10	7	9	8.67	11	Overall body size
Sex	6	12	12	10.00	12	Biological sex

Logistic Regression represents predictor contributions through regression coefficients, whereas Random Forest and XGBoost quantify the contributions of variables to tree-based prediction. Despite these methodological differences, the models consistently identified Early Response (%) as the dominant predictor and collectively highlighted body size, body composition, and fat-distribution measures as potentially informative baseline characteristics.

To complement the consensus ranking, Table 9 reports the raw and normalized feature-importance values obtained from each model. Because the importance measures have different meanings and scales across the three algorithms, normalized values facilitate within-model comparisons but should not be interpreted as directly comparable effect sizes across models.

Table 9. Model-specific feature importance values for predicting the week-8 outcome. Raw importance values are reported together with normalized magnitudes, for which the highest importance within each model was scaled to 100. For Logistic Regression, normalized importance was based on the absolute magnitude of the standardized coefficient.

Feature	LR Raw	LR Norm.	RF Raw	RF Norm.	XGB Raw	XGB Norm.
Early Response (%)	0.927	100.0	0.210	100.0	0.122	100.0
Waist-to-Hip Ratio	0.836	90.2	0.073	34.9	0.051	41.8
Body Fat (percentage)	0.198	21.4	0.072	34.1	0.046	37.7
Fat-Free Mass	0.008	0.9	0.077	36.6	0.051	41.8
BMI	-0.123	13.3	0.069	32.9	0.052	42.6
Waist Circumference	-0.029	3.1	0.063	30.0	0.054	44.3
Hip Circumference	0.136	14.7	0.062	29.7	0.050	41.0
Age	0.005	0.5	0.085	40.4	0.052	42.6
Fat Mass	0.019	2.0	0.071	33.9	0.048	39.3
Weight	0.010	1.1	0.071	33.7	0.050	41.0
Height	-0.060	6.5	0.055	26.1	0.051	41.8
Sex	0.060	6.5	0.016	7.5	0.045	36.9

3.4. Clinical Characteristics Associated with Week-8 Weight Reduction

Beyond their predictive performance, the machine learning models identified several clinically relevant characteristics associated with the week-8 outcome.

Early Response (%) was the dominant predictor in all three models, suggesting that the magnitude of change observed early in the intervention may provide useful information about subsequent weight reduction. Because this variable is measured after the intervention begins, however, it should be interpreted as an early monitoring indicator rather than a baseline predictor.

Among the baseline variables, Waist-to-Hip Ratio achieved the highest consensus ranking, suggesting that body fat distribution may provide useful predictive information beyond body weight alone. BMI, Waist Circumference, Body Fat Percentage, and Fat-Free Mass also contributed to prediction, although their relative importance differed

across models. Collectively, these findings suggest that body size, fat distribution, and body composition may each provide relevant information for characterizing intervention outcomes.

Sex had the lowest consensus ranking, indicating limited additional predictive value within the present dataset after the other variables were included. This result should not be interpreted as evidence that biological sex has no clinical relationship with treatment response, because feature importance is model- and sample-dependent and may also be affected by correlations among predictors.

Overall, routinely collected anthropometric and body-composition measurements may provide clinically meaningful information beyond conventional indicators considered individually. These exploratory findings motivate further investigation of their potential role in monitoring and individualizing pediatric obesity interventions.

4. Discussion

The primary objective of this study was not only to develop accurate machine learning models for predicting pediatric obesity intervention outcomes, but also to investigate whether machine learning can reveal clinically meaningful relationships embedded within routinely collected clinical measurements. By comparing three fundamentally different machine learning paradigms, Logistic Regression, Random Forest, and XGBoost, this study demonstrated that clinically interpretable predictors can be consistently identified despite substantial differences in the underlying learning mechanisms. Rather than viewing machine learning solely as a prediction tool, the proposed framework provides a data-driven approach for discovering clinical characteristics associated with weight-decrease.

Multicollinearity is an important consideration when interpreting machine-learning predictors derived from related anthropometric and body-composition measures. Recent explainable-AI research in childhood and adolescent obesity has explicitly addressed correlated predictors when identifying determinants of obesity [25].

In our data, high VIF values among weight, BMI, fat mass, fat-free mass, waist, and hip measures indicate substantial redundancy among baseline anthropometric variables. Accordingly, individual predictor rankings should not be interpreted as independent effects. The partial agreement across models suggests that several variables contribute predictive information; however, individual feature rankings should be interpreted cautiously because of substantial correlations among baseline anthropometric and body-composition measures.

4.1. Machine Learning Reveals Clinically Meaningful Relationships

This study finds the agreement among three fundamentally different machine learning algorithms regarding the variables most strongly associated with the Week-8 weight-reduction outcome. Logistic Regression estimates the contribution of each predictor through regression coefficients, Random Forest identifies informative decision rules by maximizing information gain, and XGBoost iteratively discovers residual relationships by minimizing prediction error. Despite these distinct learning strategies, all three models consistently identified a similar group of relevant predictors.

This agreement potentially increases confidence that the identified variables represent meaningful clinical characteristics rather than artifacts of a particular machine learning algorithm. In particular, the engineered early-response variable was consistently ranked as the most influential predictor, demonstrating that physiological changes observed during the first three weeks of treatment contain valuable information regarding the subsequent Week-8 weight-reduction outcome. Among baseline measurements, Waist-to-Hip Ratio consistently emerged as one of the strongest predictors across all three learning paradigms, followed by Body Fat Percentage, Body Mass Index, Fat-Free Mass, and Waist Circumference.

4.2. Body Composition Profile and Intervention

One potentially clinically relevant finding of this study is that body composition variables contributed predictive information for the Week-8 weight-reduction outcome. Although Body Mass Index has long served as a primary indicator for obesity assessment, the present results suggest that additional measurements describing body fat distribution and body composition may provide complementary information for short-term outcome prediction.

In particular, Waist-to-Hip Ratio was consistently identified as an important baseline predictor across the three machine learning algorithms. Body Fat Percentage and Fat-Free Mass also ranked among the important baseline variables. However, because substantial multicollinearity was observed among the anthropometric and body-composition measures, the importance of these individual variables should be interpreted cautiously and should not be considered evidence of independent clinical effects. Rather, these findings suggest that body-composition profiles may warrant further investigation in relation to pediatric obesity intervention outcomes.

These findings provide preliminary evidence that routinely collected body-composition measurements may contribute information beyond conventional measures such as Body Mass Index (BMI) when examining short-term

intervention outcomes. Future studies should investigate whether these measurements provide reproducible incremental predictive value and whether distinct body-composition profiles are associated with differences in intervention response.

4.3. Clinical Value of Early Treatment Response

Early treatment response has long been recognized as an important indicator of subsequent the Week-8 weight-reduction outcome. The present study quantitatively confirms this observation using three independent machine learning paradigms. Early Response Percentage consistently ranked as the most influential predictor across all classifiers, indicating that even modest physiological changes during the first three weeks contain substantial predictive information.

From a clinical perspective, these findings suggest that early intervention response may provide a practical indicator for monitoring short-term treatment response. Children demonstrating limited early response could potentially benefit from timely modification of intervention strategies rather than waiting until completion of the entire treatment program. Consequently, integrating machine learning with routinely collected longitudinal measurements may provide healthcare professionals with objective decision-support tools for potential intervention planning.

5. Conclusions

This study presented a clinically interpretable machine learning framework for predicting short-term weight-change outcomes in a pediatric obesity intervention program using routinely collected anthropometric and body-composition measurements. Across Logistic Regression, Random Forest, and XGBoost models, baseline measurements alone provided limited predictive performance, whereas incorporating Week-3 early-response information substantially improved prediction of the Week-8 weight-reduction outcome. In particular, Week-3 percentage weight change emerged as the dominant predictor and captured much of the predictive information available at this early stage of the intervention.

Several baseline anthropometric and body-composition variables also contributed predictive information across the models. However, substantial multicollinearity was observed among these measures, and their individual feature rankings should therefore be interpreted cautiously rather than as independent clinical effects. Overall, the findings suggest that early changes observed during an intervention may provide useful information for assessing subsequent short-term weight-change trajectories, while the incremental contribution and clinical interpretation of individual baseline characteristics require further investigation.

The present analysis focused primarily on routinely collected anthropometric and body-composition measures. PAK also collected parent-reported information regarding nutrition, physical activity, family behaviors, and psychosocial outcomes, which may provide additional information for understanding variation in intervention outcomes. Future studies may examine these behavioral and family-context variables, additional longitudinal measurements, and model-interpretation approaches. External validation in independent pediatric obesity cohorts will also be important for evaluating model generalizability and the potential clinical utility of these predictive approaches.

Author Contributions

L.L.: methodology, software, formal analysis, visualization, writing—original draft preparation, and writing—review and editing; B.M.: conceptualization, validation, and writing—review and editing; N.K.: conceptualization, investigation, data curation, resources, project administration, supervision, and writing—review and editing. All authors have read and agreed to the published version of the manuscript.

Funding

This study did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors. All phases of the research were funded by the ProActive Kids Foundation, Wheaton, IL, USA.

Institutional Review Board Statement

Participant data were de-identified upon entry into the ProActive Kids program; therefore, the study was exempt from review by an Institutional Review Board.

Informed Consent Statement

Informed consent was obtained from the participants or their parents/legal guardians prior to participation.

Data Availability Statement

The data are not publicly available because they contain sensitive information concerning pediatric participants. For further information regarding data access, please contact the ProActive Kids Foundation.

Conflicts of Interest

The authors declare no conflicts of interest and have no financial disclosures.

Use of AI and AI-Assisted Technologies

During the preparation of this work, the authors used ChatGPT (OpenAI) to assist with language editing and improve the clarity and presentation of the manuscript. All scientific analyses, interpretations, and conclusions were performed and verified by the authors. After using this tool, the authors reviewed and edited the content as needed and take full responsibility for the content of the published article.

References

1. Noiman, A.; Saif, N.; Afful, J. Prevalence of Overweight, Obesity, and Severe Obesity among Children and Adolescents Aged 2–19 Years: United States, 1963–1965 through August 2021–August 2023. Available online: <https://doi.org/10.15620/cdc/174645>, (accessed on 5 July 2026).
2. Hampl, S.E.; Hassink, S.G.; Skinner, A.C.; et al. Clinical Practice Guideline for the Evaluation and Treatment of Children and Adolescents with Obesity. *Pediatrics* **2023**, *151*, e2022060640. <https://doi.org/10.1542/peds.2022-060640>.
3. US Preventive Services Task Force. Interventions for High Body Mass Index in Children and Adolescents: US Preventive Services Task Force Recommendation Statement. *JAMA* **2024**, *332*, 226–232. <https://doi.org/10.1001/jama.2024.11146>.
4. Barlow, S.E. Expert Committee Recommendations regarding the Prevention, Assessment, and Treatment of Child and Adolescent Overweight and Obesity: Summary Report. *Pediatrics* **2007**, *120*, S164–S192. <https://doi.org/10.1542/peds.2007-2329c>.
5. Moxley, E.; Habtzghi, D.; Klinkhamer, N.; et al. Prevention and Treatment of Pediatric Obesity: A Strategy Involving Children, Adolescents and the Family for Improved Body Composition. *J. Pediatr. Nurs.* **2019**, *45*, 13–19. <https://doi.org/10.1016/j.pedn.2018.12.010>.
6. O'Connor, E.A.; Evans, C.V.; Henninger, M.; et al. Interventions for Weight Management in Children and Adolescents: Updated Evidence Report and Systematic Review for the US Preventive Services Task Force. *JAMA* **2024**, *332*, 233–248. <https://doi.org/10.1001/jama.2024.6739>.
7. Fayyaz, H.; Gupta, M.; Perez Ramirez, A.; et al. An Interoperable Machine Learning Pipeline for Pediatric Obesity Risk Estimation. In Proceedings of the 4th Machine Learning for Health Symposium, Vancouver, BC, Canada, 15–16 December 2024; pp. 308–324.
8. Chun, D.; Rhie, Y.J.; Sawyer, J.; et al. Machine Learning Prediction of Obesity Development in Children with Overweight Using Longitudinal Body Composition Data. *Pediatr. Obes.* **2026**, *21*, e70105. <https://doi.org/10.1111/ijpo.70105>.
9. Zhu, H.; Zhou, M.; Liu, G.; et al. NUS: Noisy-Sample-Removed Undersampling Scheme for Imbalanced Classification and Application to Credit Card Fraud Detection. *IEEE Trans. Comput. Soc. Syst.* **2024**, *11*, 1793–1804. <https://doi.org/10.1109/tcss.2023.3243925>.
10. Li, H.; Hu, G.; Li, J.; et al. Intelligent Fault Diagnosis for Large-Scale Rotating Machines Using Binarized Deep Neural Networks and Random Forests. *IEEE Trans. Autom. Sci. Eng.* **2022**, *19*, 1109–1119. <https://doi.org/10.1109/tase.2020.3048056>.
11. Lim, H.; Lee, H.; Kim, J. A Prediction Model for Childhood Obesity Risk Using the Machine Learning Method: A Panel Study on Korean Children. *Sci. Rep.* **2023**, *13*, 10122. <https://doi.org/10.1038/s41598-023-37171-4>.
12. Gupta, M.; Eckrich, D.; Bunnell, H.T.; et al. Reliable Prediction of Childhood Obesity Using Only Routinely Collected EHRs May Be Possible. *Obes. Pillars* **2024**, *12*, 100128. <https://doi.org/10.1016/j.obpill.2024.100128>.
13. Zhou, M.C. Artificial Intelligence and Automation for Better Life of Human Beings. *J. Artif. Intell. Autom.* **2026**, *1*, 7. <https://doi.org/10.53941/jaia.2026.100007>.
14. Pedregosa, F.; Varoquaux, G.; Gramfort, A.; et al. Scikit-Learn: Machine Learning in Python. *J. Mach. Learn. Res.* **2011**, *12*, 2825–2830. <https://doi.org/10.5555/1953048.2078195>.
15. Gaucherot, A.; Beraud, D.; Lonjou, P.; et al. Identification of Weight Loss Predictors Using Machine Learning Approaches in Adolescents with Obesity. *Pediatr. Res.* **2026**. <https://doi.org/10.1038/s41390-026-05359-9>.
16. Gou, H.; Song, H.; Tian, Z.; et al. Prediction Models for Children/Adolescents with Obesity/Overweight: A Systematic Review and Meta-Analysis. *Prev. Med.* **2024**, *179*, 107823. <https://doi.org/10.1016/j.ypmed.2023.107823>.
17. Wells, J.C.K. The Evolution of Human Body Composition. *Proc. Nutr. Soc.* **2012**, *71*, 593–602.
18. Bishop, C.M. *Pattern Recognition and Machine Learning*; Springer: New York, NY, USA, 2006. <https://doi.org/10.1007/978-0-387-45528-0>.
19. Hosmer, D.W.; Lemeshow, S.; Sturdivant, R.X. *Applied Logistic Regression*, 3rd ed.; John Wiley & Sons, Inc.: Hoboken, NJ, USA, 2013. <https://doi.org/10.1002/9781118548387>.

20. Bhupathiraju, S.N.; Hu, F.B. Epidemiology of Obesity and Diabetes and Their Cardiovascular Complications. *Circ. Res.* **2016**, *118*, 1723–1735. <https://doi.org/10.1161/circresaha.115.306825>.
21. Breiman, L. Random Forests. *Mach. Learn.* **2001**, *45*, 5–32. <https://doi.org/10.1023/a:1010933404324>.
22. Chen, T.; Guestrin, C. XGBoost: A Scalable Tree Boosting System. In Proceedings of the KDD '16: The 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, San Francisco, CA, USA, 13–17 August 2016; pp. 785–794. <https://doi.org/10.1145/2939672.2939785>.
23. Tjoa, E.; Guan, C. A Survey on Explainable Artificial Intelligence (XAI): Toward Medical XAI. *IEEE Trans. Neural Netw. Learn. Syst.* **2021**, *32*, 4793–4813. <https://doi.org/10.1109/tnnls.2020.3027314>.
24. Liu, H.; Leng, Y.; Wu, Y.C.; et al. Robust Identification Key Predictors of Short- and Long-Term Weight Status in Children and Adolescents by Machine Learning. *Front. Public Health* **2024**, *12*, 1414046. <https://doi.org/10.3389/fpubh.2024.1414046>.
25. Gan, Q.; Han, J.; Vitiello, D. Determinants of Childhood and Adolescent Obesity: An Explainable AI Approach Using the ICAD Database. *Public Health* **2026**, *257*, 106366. <https://doi.org/10.1016/j.puhe.2026.106366>.