

# Reliable prediction of childhood obesity using only routinely collected EHRs may be possible

Mehak Gupta<sup>a</sup>, Daniel Eckrich<sup>b</sup>, H. Timothy Bunnell<sup>b</sup>, Thao-Ly T. Phan<sup>b,1</sup>,  
Rahmatollah Beheshti<sup>c,1,\*</sup>

<sup>a</sup> Southern Methodist University, Dallas, TX, USA

<sup>b</sup> Nemours Children's Health, Wilmington, DE, USA

<sup>c</sup> University of Delaware, Newark, DE, USA

## ARTICLE INFO

### Keywords:

Childhood obesity  
Electronic health records  
Deep learning

## ABSTRACT

**Background:** Early identification of children at high risk of obesity can provide clinicians with the information needed to provide targeted lifestyle counseling to high-risk children at a critical time to change the disease course.

**Objectives:** This study aimed to develop predictive models of childhood obesity, applying advanced machine learning methods to a large unaugmented electronic health record (EHR) dataset. This work improves on other studies that have (i) relied on data not routinely available in EHRs (like prenatal data), (ii) focused on single-age predictions, or (iii) not been rigorously validated.

**Methods:** A customized sequential deep-learning model to predict the development of obesity was built, using EHR data from 36,191 diverse children aged 0–10 years. The model was evaluated using extensive discrimination, calibration, and utility analysis; and was validated temporally, geographically, and across various subgroups.

**Results:** Our results are mostly better or comparable to similar studies. Specifically, the model achieved an AUROC above 0.8 in all cases (with most cases around 0.9) for predicting obesity within the next 3 years for children 2–7 years of age. Validation results show the model's robustness and top predictors match important risk factors of obesity.

**Conclusions:** Our model can predict the risk of obesity for young children at multiple time points using only routinely collected EHR data, greatly facilitating its integration into clinical care. Our model can be used as an objective screening tool to provide clinicians with insights into a patient's risk for developing obesity so that early lifestyle counseling can be provided to prevent future obesity in young children.

## 1. Introduction

Childhood obesity is a major public health problem across the globe and in the US. Obesity affects almost 1 in 5 and about 14.7 million children and adolescents, for children and adolescents of 2–19 years of age [1]. Despite evidence of potentially modifiable risk factors that contribute to the development of obesity in children, children are often referred for obesity interventions after their obesity is well established and when intervention is less likely to be successful [2]. Prior studies show that less than 30 % of children with overweight or obesity are

identified by their provider at primary care visits and less than 10 % of children have a diagnoses code for overweight or obesity placed during visits [3–6]. While pediatric providers frequently use recommended CDC or WHO age and sex-specific BMI charts, they do not often recognize or address weight concerns until children cross overweight and obesity thresholds on these charts [7].

Primary barriers to effective diagnoses and management of obesity in pediatric systems are reported as lack of time, limited resources, and uncertainties about the level of risk or challenges in having conversations about risk with families due to differences in perception about a child's weight [6,8–10]. Tools that can reliably predict a patient's risk

\* Corresponding author.

E-mail addresses: [mehakg@smu.edu](mailto:mehakg@smu.edu) (M. Gupta), [daniel.eckrich@nemours.org](mailto:daniel.eckrich@nemours.org) (D. Eckrich), [tim.bunnell@nemours.org](mailto:tim.bunnell@nemours.org) (H.T. Bunnell), [thaoly.phan@nemours.org](mailto:thaoly.phan@nemours.org) (T.-L.T. Phan), [rbi@udel.edu](mailto:rbi@udel.edu) (R. Beheshti).

<sup>1</sup> Equal Supervision.

<https://doi.org/10.1016/j.obpill.2024.100128>

Received 26 May 2024; Received in revised form 21 August 2024; Accepted 22 August 2024

Available online 10 September 2024

2667-3681/© 2024 The Author(s). Published by Elsevier Inc. on behalf of Obesity Medicine Association. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

### Abbreviations

AUPRC	Area Under Precision-Recall (PR) Curve
AUROC	Area Under the Receiver Operating Characteristic
BMI	Body Mass Index
CDC	Centers for Disease Control and Prevention
EHR	Electronic Health Record
US	United States
WFL	Weight-for-length
WHO	World Health Organization

for obesity can help primary care providers identify patients at risk of developing obesity and have these conversations with families to prevent the development of obesity before it has been established [11–15]. Further, predictive models can also identify risk factors affecting a patients' rate of weight gain to help providers address certain risk factors as part of prevention efforts.

Clinical predictive models for obesity, designed using artificial intelligence and machine learning methods, are being considered to understand the contributing factors to the obesity epidemic and inform more effective interventions [16–24]. One major limitation of existing obesity prediction models is that they use features that, despite their importance, are generally not available in EHRs or are difficult and time-consuming to collect, such as genetic background, parental data, and children's lifestyle behaviors. This limits their utility in real-world settings where this type of data may not be readily available. Another limitation is that they mostly focus on predicting the risk of obesity at only one age point (e.g., at age five) and are not flexible in allowing obesity screening for patients at different ages. Additionally, large-scale and rigorously validated predictive models of obesity are rare [18,23].

In this work, we aim to demonstrate the possibility of achieving reliable estimates of the future risks of childhood obesity using commonly available (unaugmented) EHR elements across multiple age ranges. As a child's obesity is associated with obesity into adulthood [25], we focus our study on children up until the age of 10 years. This age range also includes a few years after the "adiposity rebound" patterns observed in children's weight [26,27]. We rigorously evaluate our model through an extensive series of temporal, geographic, and subgroup validations and explore the most important predictors of obesity in our model. The overarching aim of our study is to offer a general (practice-agnostic) predictive model that can be integrated within any healthcare system's EHR to reliably predict children's risk for obesity throughout early childhood to support clinical decision-making and obesity prevention at the point of care.

## 2. Materials and methods

### 2.1. Data source and study cohort selection

The EHR data used in this analysis was extracted from Nemours Children's Health; a large pediatric healthcare network in the United States (US) spanning the states of Delaware, Florida, Maryland, New Jersey, and Pennsylvania. The IRB panel at Nemours approved this research study. We used encounters between January 1, 2002, to December 31, 2019, obtained prior to the pandemic to avoid any short-term influence that the pandemic had on weight gain patterns [28]. A total of 68,029 children with 44,401,791 encounters were selected after excluding children with especially complex diseases that require multiple medications and hospitalizations (type 1 diabetes, cancer, sickle cell disease). Of them, 37,844 children were included having at least two routine infant checkups and at least one checkup between 2 to age 10 with recorded weight and length/height measurements. Of them, 1,653 children were excluded whose year of birth could not be verified,

leaving 36,191 children. These exclusions were necessary for the accuracy of our supervised machine learning models, which require enough data points in the observation and prediction windows to train the model. On analysis of the demographics of patients excluded, we found that they were not significantly different than the patients we included in our study. The final cohort was separately divided for temporal and geographic validation. More details about the Nemours EHR dataset are provided in Supplementary A (including the steps we took to extract our cohort of 36,191 patients for the model construction).

### 2.2. Feature selection and data preprocessing

We extracted clinically-relevant features to childhood obesity, by using a data-driven approach coupled with input from clinical experts [29]. Details about this process are presented in Supplementary B. The final EHR features consisted of 138 diagnoses (patient conditions and family-history conditions associated with obesity), 84 ATC3 (Anatomical Therapeutic Chemical Classification Level-3) medication groups, and 51 measurements (vitals and labs).

Following prior studies [15], the EHR features before age 2, were segmented into 5 windows. These 5 sliding windows correspond to 0–4 months, 4–8 months, 8–12 months, 12–18 months, and 18–24 months. All EHR features after age 2 were segmented into 1-year windows due to the lower frequency of medical visits (compared to before 2).

Weight and height data were converted to Weight-for-length (WFL %) which is used to classify nutritional and adiposity status in children below 2 years of age and BMI% for children above 2 years of age as defined by WHO and CDC, respectively. These values were considered Missing Completely at Random (MCAR) and imputed with carry-forward of the most recent value. WFL% and BMI% were categorized into underweight, normal, overweight, and obesity categories. We defined cutoffs for normal weight, overweight, and obesity per the CDC's standard thresholds of the 85th and 95th percentiles for overweight and obesity, respectively. Because WFL% and BMI% trajectories can help determine the risk of obesity [30–33], we also engineered a WFL% change feature that calculates the change in WFL% values between 5 windows defined above for model input. Table 1 shows the characteristics of the final cohort used.

The EHR data also included demographic information about sex (male or female), race (categorized as White, Black, Asian, and other), ethnicity (categorized as Hispanic or non-Hispanic), and payer (categorized as public and private insurance). We also included the Child Opportunity Index (COI) by geolocating the last address of patients before age 2 [34]. COI combines indicators of educational opportunity, health and environment, and socioeconomic opportunity for all US neighborhoods. COI ranges from 1 to 100, with higher numbers representing neighborhoods with more opportunities.

A binary representation (not MCAR) was used for all features, where the presence of a value for the variable was captured with a 1, and 0 otherwise. We used quintile binning for all numeric measurement features to divide each measurement into 5 categorical features. In total, we generated 506 binary features: 138 (diagnoses) + 84 (medications) + 51\*5 (measurements in 5 percentile bins) + 10 (demographic categories) + 4 (underweight, normal, overweight, and obesity) + 5 (WFL% changes in 5 percentile bins) + 10 (COIs in 10 percentile bins). A detailed process to select features is discussed in Supplementary B.

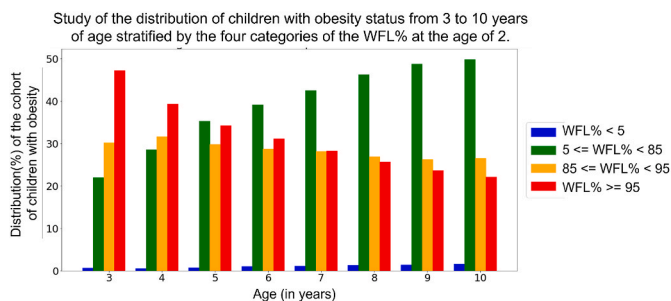
### 2.3. Descriptive analysis

To study the transition of obesity status, we analyzed the distribution of children with BMI%  $\geq 95$  at age 3 to 10 that transitioned from different WFL% status (WFL%  $< 5$ ,  $5 \leq$  WFL%  $< 85$ ,  $85 \leq$  WFL%  $< 95$ , and WFL%  $\geq 95$ ) at age 2. Fig. 1 shows that 22 % of children with obesity at age 3 had normal weight at infancy and this percentage increases with increasing age, where nearly half (49 %) of children with obesity at age 10 had normal weight at infancy.

**Table 1**

Statistics for the study cohort (full), as well as the cohorts used for temporal and geographic validation. Counts with the percentage prevalence in the respective cohort for last obesity status before age 2, sex, race, ethnicity, and payer are shown. In the bottom part, the mean (standard deviation) for the number of times that weight, height, diagnoses, medications, and lab measurements were recorded before and after age 2 are shown.

		Temporal Train (n = 26,786)	Temporal Test (n = 9405)	GeographicTrain (n = 32,848)	Geographic Test (n = 3343)	All (n = 36,191)
<b>Last obesity Status before age 2</b>						
Underweight (WFL < 5 %)	Count (%)	376 (1.4)	128 (1.36)	399 (1.2)	105 (3.14)	504 (1.39)
Normal (5 % ≤ WFL < 85 %)	Count (%)	13,781 (51.44)	5108 (54.31)	17,067 (51.9)	1822 (54.5)	18,889 (52.19)
Overweight (85 % ≤ WFL < 95 %)	Count (%)	5811 (21.69)	1979 (21.04)	7164 (21.8)	626 (18.72)	7790 (21.52)
Obesity (WFL ≥ 95 %)	Count (%)	6818 (25.45)	2190 (23.28)	8218 (25.01)	790 (23.62)	9008 (24.89)
<b>Sex:</b>						
Female	Count (%)	1,2643 (47.20)	4,355 (46.30)	15,549 (47.33)	1,449 (43.34)	16,998 (46.96)
Male	Count (%)	14,143 (52.79)	5,050 (53.69)	17,299 (52.66)	1,894 (56.65)	19,193 (53.03)
<b>Ethnicity:</b>						
Hispanic	Count (%)	3,475 (12.97)	1,323 (14.06)	4,277 (13.02)	521 (15.58)	4,798 (13.25)
Non-Hispanic	Count (%)	23,197 (86.60)	8,012 (85.18)	28,432 (86.55)	2,777 (83.06)	31,209 (86.23)
<b>Race:</b>						
Asian	Count (%)	472 (1.76)	191 (2.03)	610 (1.85)	53 (1.58)	663 (1.83)
Black	Count (%)	11,524 (43.02)	3,721 (39.56)	14,694 (44.73)	551 (16.48)	15,245 (42.12)
White	Count (%)	11,352 (42.38)	4,088 (43.46)	13,496 (41.08)	1,944 (58.15)	15,440 (42.66)
Other	Count (%)	3048 (11.37)	1,216 (12.92)	3,812 (11.60)	452 (13.52)	4,264 (11.78)
<b>Payer:</b>						
Private	Count (%)	11,114 (41.49)	3,850 (40.93)	13,483 (41.04)	1,481 (44.30)	14,964 (41.34)
Public	Count (%)	15,624 (58.32)	5,539 (58.89)	19,314 (58.79)	1,849 (55.30)	21,163 (58.47)
<hr/>						
Weight and height measurements before age 2	Mean (SD)	9.80 (2.30)	10.11 (2.24)	10.04 (2.20)	8.28 (2.53)	9.88 (2.29)
Weight and height measurements between age 2 and 10	Mean (SD)	22.10 (22.23)	17.07 (19.16)	21.15 (20.56)	17.86 (15.47)	20.83 (21.36)
Diagnoses available before age 2	Mean (SD)	2.95 (3.41)	3.32 (3.59)	2.92 (3.38)	4.46 (4.0)	3.03 (3.44)
Diagnoses available between age 2 and 10	Mean (SD)	4.82 (7.52)	3.87 (6.16)	4.49 (7.18)	5.81 (7.80)	4.58 (7.20)
Medications available before age 2	Mean (SD)	7.66 (5.40)	9.97 (7.87)	8.51 (6.25)	5.27 (5.25)	8.27 (6.23)
Medications available between age 2 and 10	Mean (SD)	11.33 (8.19)	10.75 (8.87)	11.45 (8.35)	7.99 (7.93)	11.18 (9.36)
Labs available before age 2	Mean (SD)	25.04 (29.54)	26.81 (29.54)	24.77 (28.67)	32.70 (35.66)	25.50 (29.47)
Labs available between age 2 and 10	Mean (SD)	60.00 (45.56)	43.95 (31.46)	54.69 (40.41)	66.99 (61.44)	55.83 (42.93)



**Fig. 1.** Study of the distribution of children with obesity (BMI% ≥ 95) at age 3 to 10. We analyze what percentage of children with obesity at ages 3 to 10 had WFL% < 5, 5 ≤ WFL% < 85, 85 ≤ WFL% < 95, and WFL% ≥ 95 at age 2. X-axis shows the distribution of the cohort with obesity at every age from 3 to age 10. Four categories of WFL% at age 2 are shown on the right.

**2.4. Deep learning model training**

We adopted our encoder-decoder<sup>2</sup> deep neural network model and the training procedure presented in detail in prior work [24,35]. The encoder part consists of long short-term memory (LSTM)<sup>3</sup> cells and the decoder consists of a feed-forward network with two fully-connected layers.

To train the model, we first trained the encoder network on the input EHR data from 0 to 2 years. EHR data for every year after age 2 until the age that the prediction was performed (i.e., the end of the observation

<sup>2</sup> An encoder-decoder refers to a design, where an initial network (encoder) receives input and maps (compresses) that to a lower dimension representation. Then a second network (decoder) learns from the encoder output to generate output.

<sup>3</sup> LSTMs are a type of (deep) neural network used for processing sequential data types.

window) were then combined (concatenated) with the  $n$ -dimensional vector representation derived from the encoder for 0–2 years of EHR data. The number of years of data after age 2 depends on the length of the patient’s recorded medical history. Using this design, all the prior data for each patient was used for model training to predict the risk of obesity for the next 3 years. A detailed description of the architecture design and parameter settings is provided in Supplementary C.

### 3. Experiments and results

#### 3.1. Setup

We extensively evaluated our predictive model by studying its discrimination power, calibration, and robustness. We validated our model temporally and geographically and studied the model’s performance across subpopulations (which can also capture the model’s fairness across these groups), using separate test datasets. When using the entire cohort, the data was split with an 80:20 train and test regime, with 5 % of the training data as a validation set to fix the best model. Model performance was reported exclusively on the test dataset. The confidence intervals (CI) are calculated using 100 bootstrapped replicates.

**Baseline comparison** – Similar to prior work [13,32], we consider only using the last WFL% (below 2 years) and BMI% (above 2 years) available in the observation window as a baseline. This scenario mimics what is generally used in clinical practice for screening children (i.e., viewing weight, height, and BMI data at each time point on growth curves) [36–39].

**Discrimination power** – We report prediction performance using Area Under Receiver Operating Characteristic (AUROC), Area Under Precision-Recall Curve (AUPRC), sensitivity, and specificity. We provide sensitivity and specificity at different binary classification thresholds, demonstrating the trade-off between false negatives (patients who develop obesity but are not predicted to develop obesity by the model) and false positives (patients who do not develop obesity but are predicted to develop obesity by the model). Specifically, we fixed either sensitivity or specificity at 90 % and 95 % and measured the value of the other metric.

**Utility and calibration** – We perform decision curve analysis [40] to demonstrate the tradeoffs (costs and benefits) of using the prediction model to analyze the clinical utility at various thresholds. Related to the above approach, we provide calibration metrics to help quantify how well the predicted probabilities of an outcome match the true probabilities observed in the data [41,42].

**Temporal validation** – To study our model’s robustness across time shifts, we divided our cohort according to the date of birth of the children. The data for 26,786 children who were born between January 1, 2002, and December 31, 2009, were included as a training set; and the data for 9405 children who were born between January 1, 2010, and December 31, 2015, were included in the test set. We report AUROC for the temporal validation in Table 2.

**Geographic validation** – We additionally validated our model across two different geographic regions in the US. We used 32,848 children seen in Delaware Valley sites, located in the northeastern US, as a training set and 3343 children seen in Florida, located in the southeastern US, as a test set. We report AUROC for the geographic validation in Table 2.

**Robustness across subpopulations** – Robustness across subpopulations (group fairness) is evaluated by comparing model AUROCs in the test dataset across five groups determined by the: last WFL% category before age 2 (underweight, normal weight, overweight, obesity), race (Black, White, Asian, Other), ethnicity (Hispanic, Non-Hispanic), sex (female, male), and payer (private, public).

**Net-benefit Analysis** – Net-benefit analysis allows the analysis of the clinical utility of adopting a clinical prediction model at different probability thresholds. Here, a threshold such as 10 % means that for every 10 children evaluated for risk of obesity by the predictive model,

the health service is willing to risk the cost of mistakenly prescribing 9 children [32]. The health service provider can now use the net benefit analysis to decide if the net benefit of utilizing our predictive model is more than other treatment strategies such as intervention for all (treatment for all) and intervention for none (no treatment for all) or baseline strategy (only looking at WFL and BMI growth curves).

#### 3.2. Evaluation results

We trained our deep learning model using 36,191 eligible children to predict the risk of obesity using 6 different lengths of observation windows from age 0–2 to 0–7. The observed AUROC, sensitivity, specificity, and net benefit for all prediction ages are presented in Table 2. Focusing on a popular setting that is heavily studied in the literature [13,17,32], we specifically present discrimination results for 0–2 years observation window and obesity prediction at age 5 in Fig. 2. Notably, our model outperforms the presented baseline based on a child’s last WFL% with an AUROC and AUPRC of 0.81 and 0.61 compared with 0.71 and 0.56, respectively. Our model also dominates over other strategies (baseline, intervention for all, and intervention for none as explained in Section 3.1) across various net benefit threshold probabilities, with significant margins above the 15 % threshold probability regime.

Fig. 3 compares the AUROC of our model on different subpopulations of children for prediction at age 5, across the 5 groups mentioned in Section 3.1. AUROCs among each group show minimal deviation of 0.04, 0.04, 0.01, 0.001, and 0.008, respectively, showing the robustness of our model across each group.

#### 3.3. Analyzing model predictors

We investigated the risk factors identified by the model by analyzing which predictors most attribute to the model’s prediction. We used the attention scores [43,44] obtained from the LSTM cells, as a way to determine which features are given more attention (importance) by the model to predict the output (more details are provided in Supplementary D.1).

We present the mean of the importance scores of all features inside each of the 7 input feature categories (diagnoses, family-history diagnoses, medications, measurements, demographics, last obesity status before age 2, and WFL% changes before age 2 for the entire cohort) in Fig. 4A. Clinical diagnoses and WFL% changes were the top 2 important predictor groups. We also present the ranking for the top 20 predictors based on the mean importance scores for the predictors in the entire cohort in Fig. 4B. Previous weight percentile measurements and weight gain patterns were among the top 5 features; in addition to weight-related diagnoses and diagnoses of elevated blood pressure, gastroesophageal reflux disease, developmental delay, hypothyroidism, and asthma. Among the family history diagnoses, a family history of hypertension, cardiac disorder, and depression was important. Our engineered feature (WFL% change before age 2) was also among the top 10 predictors.

We evaluated partial dependency plots (PDP<sup>4</sup>) for the WFL% change before age 2, and child opportunity index, as a function of the predicted risk of (the probability of developing) obesity at age 5 for the test dataset. Specifically, Fig. 4C shows that an increase in the WFL% before age 2 increases the risk of obesity at age 5. Fig. 4D shows that the risk of obesity decreases with increasing COI score.

## 4. Discussion

In this study, we developed and extensively validated a model that can predict obesity reliably in early childhood using a large EHR dataset

<sup>4</sup> Demonstrates the marginal effect of the different values of the feature of interest on the predicted outcome.

**Table 2**

Predictive performance across six different observation windows. Results are shown for prediction for the next 1, 2, and 3 years. The fixed values for sensitivity and specificity are highlighted. The net benefits are shown at three thresholds of 20, 40, and 60%.

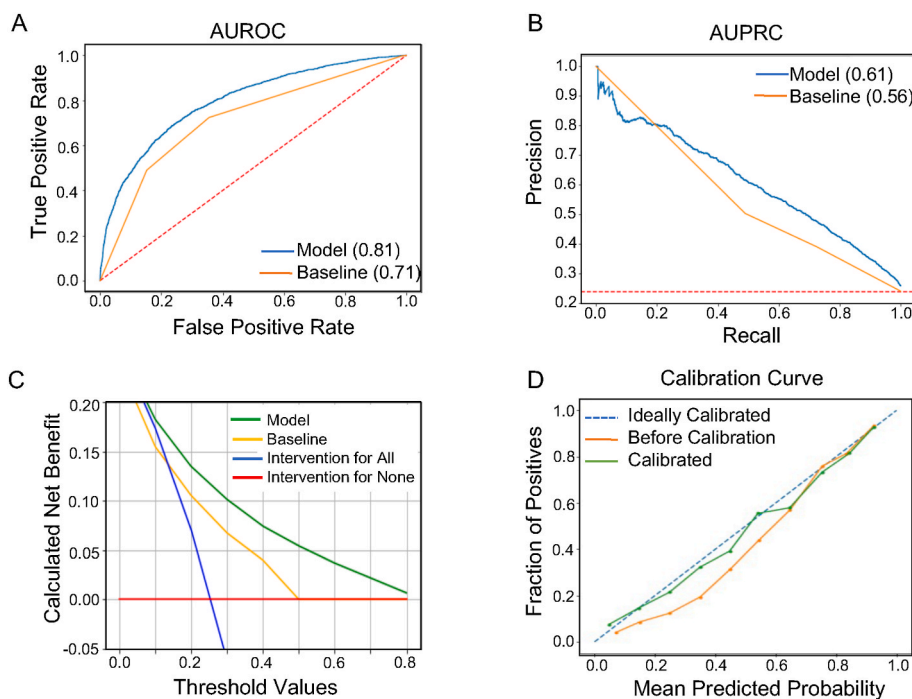
Input age range	Prediction Age		
	3 (n=7 224)	4 (n=8 875)	5 (n=8 982)
0-2			
AUROC(95%CI)	0.87(0.85-0.88)	0.84(0.81-0.86)	0.81(0.79-0.83)
Temporal validation	0.85(0.84-0.86)	0.83(0.82-0.84)	0.81(0.80-0.82)
Geographic validation	0.82(0.80-0.83)	0.80(0.79-0.82)	0.77(0.76-0.79)
Baseline	0.77(0.75-0.78)	0.75(0.73-0.77)	0.71(0.68-0.73)
Sensitivity/Specificity	0.90 /0.56	0.90 /0.52	0.90 /0.43
	0.95 /0.40	0.95 /0.37	0.95 /0.30
	0.62/ 0.90	0.57/ 0.90	0.48/ 0.90
	0.47/ 0.95	0.42/ 0.95	0.37/ 0.95
Net Benefit(20/40/60)	0.14/0.09/0.05	0.12/0.07/0.04	0.12/0.08/0.04
0-3			
AUROC(95%CI)	0.90(0.87-0.92)	0.88(0.86-0.90)	0.84(0.82-0.86)
Temporal validation	0.89(0.87-0.91)	0.86(0.84-0.88)	0.83(0.81-0.85)
Geographic validation	0.85(0.84-0.87)	0.83(0.81-0.84)	0.79(0.77-0.80)
Baseline	0.82(0.81-0.84)	0.78(0.76-0.80)	0.74(0.72-0.75)
Sensitivity/Specificity	0.90 /0.70	0.90 /0.60	0.90 /0.55
	0.95 /0.55	0.95 /0.46	0.95 /0.40
	0.68/ 0.90	0.62/ 0.90	0.54/ 0.90
	0.54/ 0.95	0.46/ 0.95	0.40/ 0.95
Net Benefit(20/40/60)	0.17/0.12/0.08	0.14/0.09/0.05	0.12/0.07/0.03
0-4			
AUROC(95%CI)	0.91(0.89-0.92)	0.88(0.86-0.90)	0.86(0.84-0.87)
Temporal validation	0.90(0.88-0.92)	0.86(0.85-0.88)	0.84(0.82-0.86)
Geographic validation	0.87(0.85-0.89)	0.84(0.82-0.85)	0.82(0.80-0.84)
Baseline	0.84(0.82-0.83)	0.82(0.81-0.83)	0.78(0.75-0.80)
Sensitivity/Specificity	0.90 /0.71	0.90 /0.59	0.90 /0.53
	0.95 /0.58	0.95 /0.43	0.95 /0.38
	0.71/ 0.90	0.62/ 0.90	0.57/ 0.90
	0.58/ 0.95	0.51/ 0.95	0.45/ 0.95
Net Benefit(20/40/60)	0.16/0.11/0.08	0.12/0.09/0.05	0.12/0.08/0.05

0-5	6 (n=8 974)	7 (n=5 256)	8 (n=2 300)
AUROC(95%CI)	0.92(0.90-0.93)	0.90(0.88-0.92)	0.87(0.85-0.88)
Temporal validation	0.91(0.89-0.92)	0.88(0.87-0.89)	0.85(0.84-0.86)
Geographic validation	0.89(0.87-0.91)	0.85(0.83-0.86)	0.83(0.82-0.84)
Baseline	0.83(0.82-0.84)	0.80(0.79-0.81)	0.78(0.75-0.80)
Sensitivity/Specificity	0.90 /0.71	0.90 /0.69	0.90 /0.54
	0.95 /0.57	0.95 /0.53	0.95 /0.39
	0.75/ 0.90	0.72/ 0.90	0.64/ 0.90
	0.64/ 0.95	0.59/ 0.95	0.52/ 0.95
Net Benefit(20/40/60)	0.17/0.12/0.08	0.15/0.11/0.08	0.16/0.12/0.07
0-6	7 (n=5 272)	8 (n=2 314)	9 (n=1 249)
AUROC(95%CI)	0.92(0.90-0.95)	0.90(0.88-0.91)	0.88(0.86-0.90)
Temporal validation	0.92(0.91-0.94)	0.89(0.87-0.90)	0.86(0.85-0.88)
Geographic validation	0.90(0.88-0.91)	0.87(0.85-0.89)	0.84(0.83-0.86)
Baseline	0.84(0.82-0.85)	0.82(0.81-0.83)	0.79(0.78-0.80)
Sensitivity/Specificity	0.90 /0.74	0.90 /0.61	0.90 /0.54
	0.95 /0.60	0.95 /0.46	0.95 /0.34
	0.79/ 0.90	0.70/ 0.90	0.56/ 0.90
	0.67/ 0.95	0.59/ 0.95	0.45/ 0.95
Net Benefit(20/40/60)	0.16/0.11/0.08	0.17/0.11/0.08	0.16/0.09/0.07
0-7	8 (n=2 316)	9 (n=1 150)	10 (n=539)
AUROC(95%CI)	0.93(0.92-0.94)	0.91(0.88-0.92)	0.87(0.85-0.89)
Temporal validation	0.92(0.90-0.93)	0.90(0.88-0.91)	0.85(0.82-0.87)
Geographic validation	0.90(0.89-0.92)	0.88(0.86-0.90)	0.84(0.81-0.85)
Baseline	0.84(0.82-0.85)	0.82(0.81-0.83)	0.79(0.78-0.80)
Sensitivity/Specificity	0.90 /0.70	0.90 /0.63	0.90 /0.56
	0.95 /0.54	0.95 /0.47	0.95 /0.40
	0.75/ 0.90	0.67/ 0.90	0.59/ 0.90
	0.65/ 0.95	0.55/ 0.95	0.46/ 0.95
Net Benefit(20/40/60)	0.16/0.11/0.09	0.16/0.09/0.09	0.17/0.09/0.08

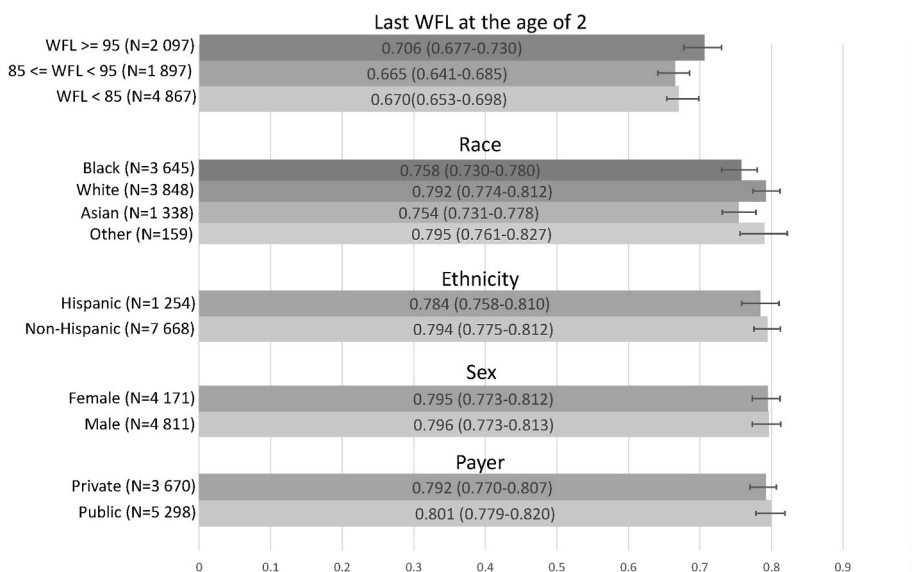
with over 36,000 children using deep learning methods. Our descriptive analysis in Section 2.3 underscores the importance of having a tool that includes other elements beyond a static WFL% or BMI% to identify children at high-risk for developing obesity. Our model used around 500 clinically relevant features from the EHR and demonstrated strong performance across multiple age ranges, chronological time periods, geographies, and demographic subgroups. Because our model only leverages unaugmented EHR data collected as part of routine clinical care, which is what is practically and widely available to clinicians without the burden of collecting additional information, it can be integrated into

common EHRs. Our model also provides flexibility in the age at which the model is applied to support the implementation of preventive measures before a child develops obesity between 3 to age 10.

As an example of how the information from this model might be used clinically, we provide the following case scenario (Fig. 5). A pediatric primary care provider is seeing a 4-year-old girl with no significant past medical history for her well-child visit. The girl’s BMI percentile has increased from the 74th to 77th percentile from 2 to 3 years and from the 77th to the 84th percentile from 3 to 4 years. The primary care provider is not sure whether to bring up healthy lifestyle recommendations to the



**Fig. 2.** Evaluation of obesity prediction model to predict obesity at age 5 using 0–2 years of data: **A.** AUROC curve of the model (blue) and a baseline model based on the last available WFL% or BMI% measurement in the observation window (orange line), **B.** AUPRC curve of the model (blue) and a baseline model based on the last available WFL% measurement before age 2 (orange line), **C.** Decision curve analysis for different strategies of treatment, **D.** Calibration curve. The dotted line represents ideal calibration and the orange line for before calibration and the green line after calibration. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

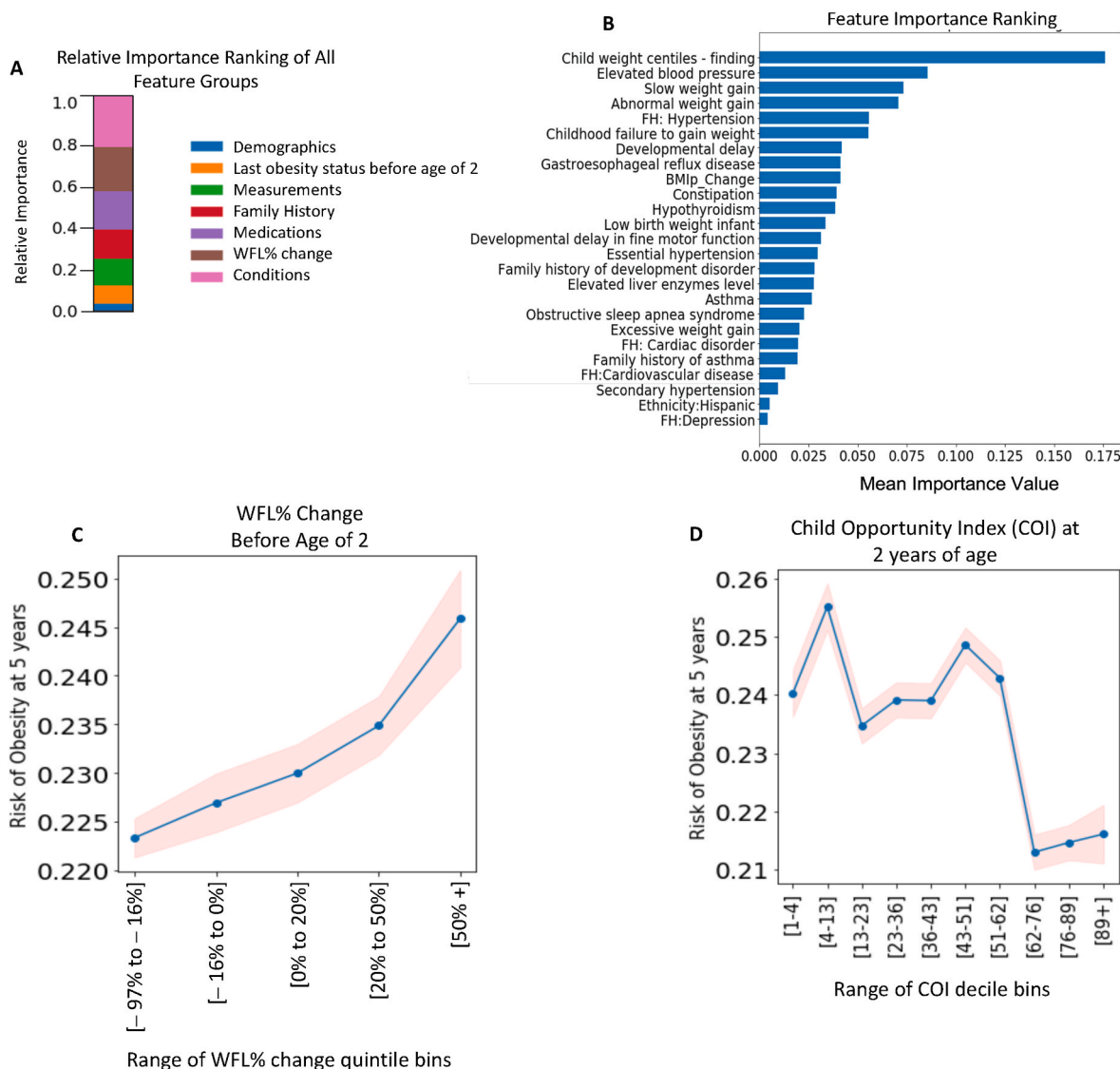


**Fig. 3.** Evaluating the predictive model's robustness by comparing AUROCs (in the test dataset) across five groups (13 subgroups): last WFL before age 2 (3 categories), race (Asian, Black, White, Other), ethnicity (Hispanic, Non-Hispanic), sex (female, male), and payer (private, public).

family. Using our model, the provider learns that the girl's risk for developing obesity in the next three years is 85% and that her main risk factors for developing obesity are a rapid increase in BMI percentile, family history of diabetes, and living in a disadvantaged neighborhood. This gives the primary care provider the confidence he needs to proceed with talking to the family about healthy lifestyle behaviors, framed in the context of health and diabetes risk and addressing social risk factors like food insecurity.

To date, there has been limited research on using deep learning methods with large longitudinal EHR data to predict childhood obesity

[18,24,45]. There is a large body of studies that use logistic regression to predict the development of obesity at a certain age, most commonly 5 years. Three of the most comparable ones are presented by Liebert et al. [46], Robson et al. [47], and Hammond et al. [13] with AUROC of 0.67, 0.78, and 0.81, respectively. While these studies used a smaller set of features, they also used features like parental weight, parental smoking habits, prenatal history, and infant dietary habits which are not available in many EHRs. Several advanced ML-based models were compared by Dugan et al. [14] and Pang et al. [15] using 167 features (anthropometric measures and questionnaire data), and 102 features (maternal



**Fig. 4.** Feature Importance Analysis for all samples to predict obesity at age 5 using 0–2 years of data. **(A)** Feature importance ranking of 7 feature groups, **(B) Top 20 feature importance ranking**, and **(C)** Change in WFL% before age 2 over 5 percentile categories (Dividing the numeric data (WFL%) into 5 bins (aka quintiles) such that there are an equal number of observations in each bin. This would produce a categorical object indicating quantile membership for each data point. The x-axis shows the range of each bin.), **(D)** child opportunity index (COI) at age 2 over COI-decile (Dividing the numeric data into 10 bins (aka deciles) such that there are an equal number of observations in each bin. This would produce a categorical object indicating decile membership for each data point. The x-axis shows the range of each bin.) categories.

and EHR data), respectively for children before age 2. They reported an accuracy of 0.85 using a decision-tree method at age 3 and an AUROC of 0.81 b using XGBoost for ages 2–7.

Our study improves on prior studies in several ways. Our predictive model achieves an AUROC of 0.85 (0.84–0.86), 0.83 (0.82–0.84), and 0.81 (0.80–0.82) for predicting the risk of obesity at 3, 4, and age 5, respectively, using 0–2 years of data. Our predictive model also demonstrates high accuracy for estimating the risk of obesity for the next 3 years across a wide age range from 2 to 7 years of age (Table 2). Indeed, our model provides the flexibility of learning from as much data as available for patients before the time of screening, which is a method referred to as a “flexible window design” that our team developed in a prior paper [35]. Because of this flexibility, it provides a tool to screen children at different ages and with enough of a time window before the development of obesity for preventive interventions to be effective [48, 49].

Beyond demonstrating the accuracy and flexibility of our model, we demonstrated the validity of our model across temporal, geographic

shifts, multiple demographic strata and the last WFL% before age 2 and demonstrated our model’s sensitivity in predicting obesity. As a screening tool, having a highly sensitive model is preferred to ensure that young children at risk for developing obesity are identified and preventive measures like parenting interventions, lifestyle behavior counseling, and weight checks can be implemented on time. Notably, the sensitivity of the model does mean that there is a possibility of false positives; however, because obesity prevention measures are generally low-cost and beneficial for all children no matter their weight status, we believe this is an acceptable trade-off as long as discussion about a child’s risk for developing obesity is done in a family-centered, thoughtful, and positive way and prevent any weight-related stigma that could otherwise occur.

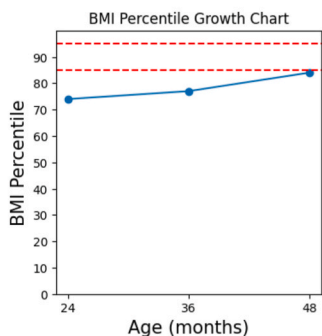
Another important strength of our model is that it utilizes only EHR data that is collected as part of routine clinical care. While the data used is already available in the EHR to providers, our model uniquely examines the complex and longitudinal relationships between this data to predict an individual child’s risk of obesity. Our feature selection



Below we have a patient data who is being screened at Age 4 years

```

Patient ID: 836178954
Demographics of the Patient:
Race: Some Other Race
Ethnicity: Hispanic
Gender: Female
COI: [23-36]
Insurance: Medicaid/sCHIP
    
```



Risk(%) of developing obesity (BMI<sub>p</sub> ≥ 95)

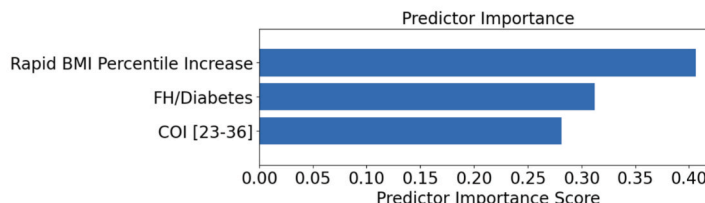
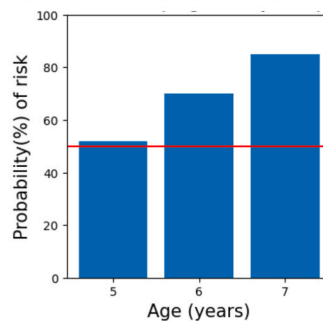


Fig. 5. The example output for a patient who is being screened at the age of 4 years. The left side shows the patient demographic information and patient’s BMI Percentile growth chart. The right side shows the output produced by the model.

process can be applied to any EHR dataset that contains common data elements of basic demographics, diagnosis codes, medications, and measurements. This is advantageous compared to prior models which may require extensive data collection or data linkages that are not feasible outside of the research setting. Only using available EHR data increases the feasibility of deploying these predictive models in the clinical setting and increases the generalizability of our model to other clinical settings that collect similar EHR data. For example, our team is developing prototype tools applying this model to EHR data to inform clinical decision-making. Such a tool could be very valuable in helping primary care providers decide who to initiate conversations about healthy lifestyles with given the time constraints of primary care visits. It can also help justify to payers the need for early healthy lifestyle interventions to prevent future obesity as recommended by the AAP Clinical Practice Guidelines.

Similar to other studies [43,50], our evaluation of the importance of the EHR features to our model also allows us to better understand what risk factors are important to a child’s risk of obesity, which can facilitate the provision of more personalized interventions. For example, stressing healthy lifestyles for infants whose family has a strong cardiovascular history could be an effective intervention based on our model. In addition to certain diagnoses and family history, our results demonstrated that change in WFL% in infancy is an important feature to calculate from EHR measurements to determine a child’s risk of obesity. Being able to learn from these longitudinal engineered and other EHR features is an advantage of our deep learning model.

We also found that the risk of obesity decreases with the increasing child opportunity index (COI), corroborating the importance of social determinants of health on childhood obesity [51,52]. While COI may not be part of routine EHR data, it is easily calculated with information that is in the EHR (through ZIP codes) and our study demonstrates the importance of accounting for the influence of neighborhood environments on child health outcomes.

Our study does have some limitations. First, our study is retrospective and may be affected by changes in healthcare delivery over time. However, we were able to temporally validate the model to account for time shifts between children born between 2002-2009 and 2010–2015. Second, the dataset included information from a single healthcare system. Despite this, the dataset did include children from five different

states in different geographic regions and we were able to validate our model between these geographic regions. Third, a large number of children were excluded due to inadequate or missing height and weight measurements before age 2 and this may have introduced sample bias. Finally, our dataset did not contain information on lifestyle behaviors, prenatal variables, or other sociodemographic variables, which we know are important to a child’s risk for the development of obesity. However, our focus was to use an EHR dataset that included only data collected during routine clinical care, to increase the potential that the model could be integrated into existing EHR systems to provide clinical decision support for providers in identifying children at risk for obesity without the need to collect additional data to facilitate early prevention efforts across many pediatric healthcare settings. Our team is actively working on partnering with other healthcare institutions to validate the model across multiple healthcare systems and creating a CDS (clinical decision support) tool on the SMART on FHIR (Fast Healthcare Interoperability Resources) platform [53], allowing its wider adoption and integration in clinical practice and increasing its interoperability. An interactive demo of our model is available at [fhir-obesity.com](http://fhir-obesity.com).

### 5. Conclusion

This study presents an extensively validated predictive model for identifying obesity risk in early childhood using a large electronic health record (EHR) dataset and customized deep learning techniques. Compared to existing research, our study demonstrates advancements in predictive accuracy, robustness, and flexibility, particularly through the use of a “flexible window design.” Integrating our model with pediatric workflows can meaningfully inform prevention intervention of childhood obesity.

### Author contributions

MG conceived and carried out experiments. TP and RB conceived the experiments and helped in the study design. MG, DE, and HB worked on data extraction, cleaning, and linkage. All authors were involved in writing the paper and had final approval of the submitted and published versions.

## Funding

Our study was supported by NIH awards P20GM103446 and U54-GM104941.

## Data availability

Our code, containing the model with parameter (weight) values, is publicly available on GitHub at <https://github.com/healthylaife/ObesityPrediction>. Interested scholars can access the data by contacting Nemours Biomedical Research Informatics Center and signing a data use agreement.

## Declaration of competing interest

The authors declare no competing financial interests.

## Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.obpill.2024.100128>.

## References

- [1] CDC. Centers for disease control and prevention. childhood obesity facts 2023. June 2023.
- [2] Loth Katie A, Lebow Jocelyn, Uy Marc James Abrigo, et al. First, do No harm: understanding primary care providers' perception of risks associated with discussing weight with pediatric patients. *Glob Pediatr Health* 2021;8: 2333794X2111040979.
- [3] Klein Jonathan D, Sesselberg Tracy S, Johnson Mark S, et al. Adoption of body mass index guidelines for screening and counseling in pediatric practice. *Pediatrics* 2010;125(2):265–72.
- [4] Kuhle Stefan, Kirk Sara FL, Ohinmaa Arto, et al. Comparison of ICD code-based diagnosis of obesity with measured obesity in children and the implications for health care cost estimates. *BMC Med Res Methodol* 2011;11(1):1–5.
- [5] Liang Lan, Meyerhoefer Chad, Wang Justin. Obesity counseling by pediatric health professionals: an assessment using nationally representative data. *Pediatrics* 2012; 130(1):67–77.
- [6] Rhee Kyung E, Kessl Stephanie, Lindback Sarah, et al. Provider views on childhood obesity management in primary care settings: a mixed methods analysis. *BMC Health Serv Res* 2018;18(1):55.
- [7] Eliana M Perrin, Finkle Joanne P, Benjamin John T. Obesity prevention and the primary care pediatrician's office. *Curr Opin Pediatr* 2007;19(3):354–61.
- [8] Conrad Rausch John, Rothbaum Perito Emily, Hametz Patricia. Obesity prevention, screening, and treatment: practices of pediatric providers since the 2007 expert committee recommendations. *Clin Pediatr* 2011;50(5):434–41.
- [9] Halford Jason CG, Bereket Abdullah, Bin-Abbas Bassam, et al. Misalignment among adolescents living with obesity, caregivers, and healthcare professionals: action teens global survey study. *Pediatric obesity* 2022;17(11):e12957.
- [10] Hill Samareh G, Thao-Ly T Phan, Datto George A, et al. Integrating childhood obesity resources into the patient-centered medical home: provider perspectives in the United States. *J Child Health Care* 2019;23(1):63–78.
- [11] Scheinker David, Valencia Areli, Rodriguez Fatima. Identification of factors associated with variation in us county-level obesity prevalence rates using epidemiologic vs machine learning models. *JAMA Netw Open* 2019;2(4):e192884.
- [12] Ferdowsy Faria, Rahi Kazi Samsul Alam, Jabiullah Md Ismail, et al. A machine learning approach for obesity risk prediction. *Current Research in Behavioral Sciences* 2021;2:100053.
- [13] Hammond Robert, Athanasiadou Rodoniki, Curado Silvia, et al. Predicting childhood obesity using electronic health records and publicly available data. *PLoS One* 2019;14(4):e0215571.
- [14] Dugan Tamara M, Mukhopadhyay S, Carroll Aaron, et al. Machine learning techniques for prediction of early childhood obesity. *Appl Clin Inf* 2015;6(3): 506–20.
- [15] Pang Xueqin, Forrest Christopher B, Lê-Scherban Félice, et al. Prediction of early childhood obesity with machine learning and electronic health record data. *Int J Med Inf* 2021;150:104454.
- [16] Monteiro POA, Victora CG. Rapid growth in infancy and childhood and obesity in later life—a systematic review. *Obes Rev* 2005;6(2):143–54.
- [17] Ziauddeen Nida, Roderick Paul J, Macklon Nicholas S, et al. Predicting childhood overweight and obesity using maternal and early life risk factors: a systematic review. *Obes Rev* 2018;19(3):302–12.
- [18] Colmenarejo Gonzalo. Machine learning models to predict childhood and adolescent obesity: a review. *Nutrients* 2020;12(8):2466.
- [19] Campbell Elizabeth A, Qian Ting, Miller Jeffrey M, et al. Identification of temporal condition patterns associated with pediatric obesity incidence using sequence mining and big data. *Int J Obes Aug* 2020;44(8):1753–65.
- [20] Siddiqui Hera, Rattani Ajita, Woods Nikki K, et al. A survey on machine and deep learning models for childhood and adolescent obesity. *IEEE Access* 2021;9: 157337–60.
- [21] Nirmala Devi K, Krishnamoorthy N, Jayanthi P, et al. Machine learning based adult obesity prediction. In: 2022 international conference on computer communication and Informatics (ICCCI). IEEE; 2022. p. 1–5.
- [22] Zhou Xiaobei, Chen Lei, Liu Hui-Xin. Applications of machine learning models to predict and prevent obesity: a mini-review. *Front Nutr* 2022;9.
- [23] Ferreras Antonio, Sumalla-Cano Sandra, Martínez-Licort Rosmeri, et al. Systematic review of machine learning applied to the prediction of obesity and overweight. *J Med Syst* 2023;47(1):1–11.
- [24] Gupta Mehak, Thao-Ly T Phan, Bunnell H Timothy, et al. Obesity Prediction with EHR Data: a deep learning approach with interpretable elements. *ACM Trans Comput Healthc* 2022;3(3):1–19.
- [25] Horesh Adi, Tsur Avishai M, Bardugo Aya, et al. Adolescent and childhood obesity and excess morbidity and mortality in young adulthood—a systematic review. *Curr Obes Rep* 2021;10(3):301–10.
- [26] Zhou Jixing, Zhang Fu, Qin Xiaoyun, et al. Age at adiposity rebound and the relevance for obesity: a systematic review and meta-analysis. *Int J Obes* 2022;46 (8):1413–24.
- [27] Freedman David S, Goodwin-Davies Amy J, Kompaniyets Lyudmyla, et al. Interrelationships among age at adiposity rebound, bmi during childhood, and bmi after age 14 years in an electronic health record database. *Obesity* 2022;30(1): 201–8.
- [28] Anderson Laura N, Yoshida-Montezuma Yulika, Dewart Nora, et al. Obesity and weight change during the covid-19 pandemic in children and adults: a systematic review and meta-analysis. *Obes Rev* 2023:e13550.
- [29] Rhodes Erinn T, Thao-Ly T Phan, Earley Elizabeth R, et al. Patient-reported outcomes to describe global health and family relationships in pediatric weight management. *Child. Obes.* 2023.
- [30] Aris Izzuddin M, Rifas-Shiman Sheryl L, Li Ling-Jun, et al. Association of weight for length vs body mass index during the first 2 years of life with cardiometabolic risk in early adolescence. *JAMA Netw Open* 2018;1(5):e182460.
- [31] Taveras Elsie M, Rifas-Shiman Sheryl L, Belfort Mandy B, et al. Weight status in the first 6 months of life and obesity at 3 years of age. *Pediatrics* 2009;123(4):1177–83.
- [32] Rossman Hagai, Shilo Smadar, Barbash-Hazan Shiri, et al. Prediction of childhood obesity from nationwide health records. *J Pediatr* 2021;233:132–40.
- [33] Grossman David C, Bibbins-Domingo Kirsten, Curry Susan J, et al. Screening for obesity in children and adolescents: us preventive services task force recommendation statement. *JAMA* 2017;317(23):2417–26.
- [34] Institute for Child, Youth, and Family Policy. Child opportunity index, 2022. Heller school for social policy and management. Brandeis University; Nov 2022.
- [35] Gupta Mehak, Poulain Raphael, Phan Thao-Ly T, et al. Flexible-window predictions on electronic health records. *Proc AAAI Conf Artif Intell Jun.* 2022;36(11): 12510–6.
- [36] Grummer-Strawn Laurence M, Reinold Christopher M, Funnemark Krebs Nancy, et al. Use of world health organization and cdc growth charts for children aged 0–59 months in the United States. Department of Health and Human Services, Centers for Disease Control and Prevention; 2010.
- [37] De Onis Mercedes, Garza Cutberto, Onyango Adelheid W, et al. Comparison of the who child growth standards and the cdc 2000 growth charts. *J Nutr* 2007;137(1): 144–8.
- [38] Kuczumski Robert J. CDC growth charts: United States. US Department of Health and Human Services, Centers for Disease Control and Prevention; 2000.
- [39] Mei Zuguo, Ogdan Cynthia L, Flegal Katherine M, et al. Comparison of the prevalence of shortness, underweight, and overweight among us children aged 0 to 59 months by using the cdc 2000 and the who 2006 growth charts. *J Pediatr* 2008; 153(5):622–8.
- [40] Vickers Andrew J, Elkin Elena B. Decision curve analysis: a novel method for evaluating prediction models. *Med Decis Making* 2006;26(6):565–74.
- [41] Matheny Michael, Thadaneys Israni S, Ahmed Mahnoor, et al. Artificial intelligence in health care: the hope, the hype, the promise, the peril. Washington, DC: National Academy of Medicine; 2019.
- [42] Van Calster Ben, McLernon David J, Van Smeden Maarten, et al. Calibration: the achilles heel of predictive analytics. *BMC Med* 2019;17(1):1–7.
- [43] Choi Edward, Taha Bahadori Mohammad, Sun Jimeng, et al. Retain: an interpretable predictive model for healthcare using reverse time attention mechanism. *Adv Neural Inf Process Syst* 2016;29.
- [44] Du Juan, Zeng Dajian, Li Zhao, et al. An interpretable outcome prediction model based on electronic health records and hierarchical attention. *Int J Intell Syst* 2022; 37(6):3460–79.
- [45] Triantafyllidis Andreas, Polychronidou Eleftheria, Alexiadis Anastasios, et al. Computerized decision support and machine learning applications for the prevention and treatment of childhood obesity: a systematic review of the literature. *Artif Intell Med* 2020;104:101844.
- [46] Redsell Sarah A, Weng Stephen, Swift Judy A, et al. Validation, optimal threshold determination, and clinical utility of the infant risk of overweight checklist for early prevention of child overweight. *Child Obes* 2016;12(3):202–9.
- [47] Jacob O Robson, Verstraete Sofia G, Shiboski Stephen, et al. A risk score for childhood obesity in an urban latino cohort. *J Pediatr* 2016;172:29–34.
- [48] Hampl Sarah E, Hassink Sandra G, Skinner Ashley C, et al. Clinical practice guideline for the evaluation and treatment of children and adolescents with obesity. *Pediatrics* 2023;151(2):e2022060640.
- [49] Baidal Jennifer A Woo, Taveras Elsie M. Childhood obesity: shifting the focus to early prevention. *Arch Pediatr Adolesc Med* 2012;166(12):1179–81.

- [50] Ma Fenglong, Chitta Radha, Zhou Jing, et al. Dipole: diagnosis prediction in healthcare via attention-based bidirectional recurrent neural networks. In: Proceedings of the 23rd ACM SIGKDD international conference on knowledge discovery and data mining; 2017. p. 1903–11.
- [51] Aris Izzuddin M, Perng Wei, Dabelea Dana, et al. Associations of neighborhood opportunity and social vulnerability with trajectories of childhood body mass index and obesity among us children. *JAMA Netw Open* 2022;5(12):e2247957.
- [52] Gupta Mehak, Thao-Ly T Phan, Lê-Scherban Félice, et al. Associations of longitudinal BMI percentile classification patterns in early childhood with neighborhood-level social determinants of health. medRxiv 2023. 2023–06.
- [53] Mandel Joshua C, Kreda David A, Mandl Kenneth D, et al. SMART on FHIR: a standards-based, interoperable apps platform for electronic health records. *J Am Med Inf Assoc* 2016;23(5):899–908.