

Identifying Key Determinants of Childhood Obesity: A Narrative Review of Machine Learning Studies

Madison N. LeCroy, PhD,¹ Ryung S. Kim, PhD,¹ June Stevens, PhD,^{2,3}
David B. Hanna, PhD,¹ and Carmen R. Isasi, MD, PhD¹

Abstract

Machine learning is a class of algorithms able to handle a large number of predictors with potentially nonlinear relationships. By applying machine learning to obesity, researchers can examine how risk factors across multiple settings (*e.g.*, school and home) interact to best predict childhood obesity risk. In this narrative review, we provide an overview of studies that have applied machine learning to predict childhood obesity using a combination of sociodemographic and behavioral risk factors. The objective is to summarize the key determinants of obesity identified in existing machine learning studies and highlight opportunities for future machine learning applications in the field. Of 15 peer-reviewed studies, approximately half examined early childhood (0–24 months of age) determinants. These studies identified child's weight history (*e.g.*, history of overweight/obesity or large increases in weight-related measures between birth and 24 months of age) and parental overweight/obesity (current or prior) as key risk factors, whereas the remaining studies indicated that social factors and physical inactivity were important in middle childhood and late childhood/adolescence. Across age groups, findings suggested that race/ethnic-specific models may be needed to accurately predict obesity from middle childhood onward. Future studies should consider using existing large data sets to take advantage of the benefits of machine learning and should collect a wider range of novel risk factors (*e.g.*, psychosocial and sociocultural determinants of health) to better predict childhood obesity. Ultimately, such research can aid in the development of effective obesity prevention interventions, particularly ones that address the disproportionate burden of obesity experienced by racial/ethnic minorities.

Keywords: childhood obesity; machine learning; minority health; social determinants of health

Introduction

Despite efforts to curtail the obesity epidemic, most behavioral interventions to prevent and treat childhood obesity have produced only small reductions in BMI.¹ Thus, there is a need to reassess the design and targets of childhood obesity prevention interventions, particularly for racial/ethnic minorities.² Obesity is the result of complex interactions between numerous biological, behavioral, sociocultural, and environmental factors.³ Machine learning is a relatively novel statistical approach in the field of epidemiology suitable for identifying predictors of obesity due to its ability to handle a large number of potentially collinear predictors.^{4,5} Defini-

tions of machine learning vary, but in this review, machine learning is defined as any statistical approach that identifies the most predictive model for a given outcome using methods that are capable of analyzing large data sets (*e.g.*, number of predictors > number of observations).^{6,7}

By applying machine learning to obesity epidemiology, researchers have the opportunity to examine a variety of predictors across multiple settings (*e.g.*, school and home) describe how these predictors interact to affect obesity risk, and potentially identify previously undiscovered determinants of obesity.⁴ The purpose of this narrative review is to summarize the findings of studies that have used machine learning to examine risk factors for childhood obesity across multiple domains (*i.e.*, sociodemographic and

¹Department of Epidemiology and Population Health, Albert Einstein College of Medicine, Bronx, NY, USA.

Departments of ²Nutrition and ³Epidemiology, Gillings School of Global Public Health, University of North Carolina at Chapel Hill, Chapel Hill, NC, USA.

behavioral risk factors). To be included in this review, studies must have (1) applied at least one machine learning method to predict obesity, (2) examined a combination of sociodemographic and behavioral risk factors, (3) predicted obesity in youth 0–18 years old, and (4) have been peer-reviewed, published, and available in PubMed or Google Scholar before August 2020. Reference lists of the selected articles were reviewed to identify additional relevant publications. Our goal is to promote the development of more effective childhood obesity prevention interventions by identifying key determinants of childhood obesity and to highlight opportunities for future machine learning applications. This review is organized according to the age at which risk factors were assessed, given that the determinants of childhood obesity vary across the life course.

Machine Learning Applications to Childhood Obesity Epidemiology

Although machine learning methods are capable of examining large data sets, the studies reviewed have largely consisted of small to moderately sized data sets and a limited set of variables conceived to be causal by experts (Supplementary Tables S1–S3). This is likely due to the limited size and scope of data sets available to the authors at the time as well as concerns over model overfitting and interpretability—concerns that characterize machine learning methods in general.⁸ For example, when a long list of variables is placed into a machine learning model, the model will identify the best predictors of obesity, irrespective of whether they could plausibly be related to or cause the outcome. If statistical safeguards such as cross-validation are not used, the resulting model may (1) not be generalizable (overfitting) or (2) not make sense for predicting the outcome (poor interpretability).

Early Childhood

One of the earliest applications of machine learning to childhood obesity was in a 2004 prospective cohort study by Agras et al.⁹ The authors created a decision tree to predict childhood overweight/obesity at 9.5 years of age using data collected before 5 years of age in 150 predominately non-Hispanic white middle-class families living in the United States.^{9,10} In brief, decision trees are flowchart-like structures built through a series of if–then statements. The most important feature for predicting obesity is the first node of the tree.⁸ Different algorithms can be used to determine node sequence, but in this study, the method was not specified.⁹ The authors limited their examination to five predictors that were significant in a logistic regression model. Greatest risk for childhood overweight/obesity was seen when the parent had a BMI >27.5 kg/m², had low concern about child’s thinness, and had a child enrolled in the study with a highly emotional personality.

Another early application of machine learning by Toschke et al.¹¹ used classification and regression trees (CART)¹² to identify early childhood risk factors for

overweight and obesity, separately. CART is one type of algorithm used to create decision trees; notably, it allows for continuous and categorical features and missing data.¹³ This retrospective cohort study predicted obesity among 4289 5–6-year-olds in Germany. The highest prevalence of overweight corresponded to children with a combination of, in order of feature importance, high early weight gain, parents with normal weight, parental education <10 years, and birth weight ≥3800 g. Youth with high early weight gain and at least one parent with obesity had the second highest prevalence of overweight, irrespective of any other risk factors. Results for obesity were similar, with the primary difference being the identification of maternal smoking as an additional determinant.

Zhang et al.¹⁴ was the first to assess the utility of different machine learning techniques for predicting childhood obesity but only included sex, gestational age, and anthropometric measures as features due to their limited database. Dugan et al.¹⁵ expanded on this study using longitudinal data collected from a clinical decision support system to compare different machine learning techniques using 167 features, spanning child’s physical health, development, and sleep; mother’s anxiety and relationship; and financial stability during the first 24 months of life. In their sample of 7519 primarily racial minority, low-income US children, they found that an Iterative Dichotomiser 3 (ID3) decision tree¹⁶—a simple decision tree algorithm that requires complete categorical data¹³—had the highest accuracy of predicting obesity across the 2–10 year age span. A history of overweight before 24 months of age was the most important predictor. Children who were overweight before 24 months of age and had short stature before 6 months of age had a high prevalence of obesity. Among children who were not overweight before 24 months of age, the child’s race/ethnicity was an important determinant of obesity, with the highest prevalence of obesity being seen among individuals of Asian Pacific Islander descent, followed by individuals of Hispanic/Latino ethnicity. Individuals who were of Asian Pacific Islander descent and had a mother with evidence of depression had the highest prevalence of obesity.

Three additional studies used US-based cohorts. Kitasantas and Gaffney¹⁷ examined 6540 non-Hispanic white, non-Hispanic black, and Hispanic children enrolled in the Early Childhood Longitudinal Study-Birth Cohort. Of 12 features, 7 were associated with overweight/obesity at 4 years of age using CART: overweight/obesity at 24 months of age (most important), elevated pregravid BMI, low/middle socioeconomic status (SES), maternal Hispanic ethnicity, elevated birthweight, low duration of breastfeeding, and low parity. Children with overweight/obesity at 24 months had the highest prevalence of overweight/obesity, but among children without overweight/obesity at 24 months, high pregravid BMI and maternal Hispanic ethnicity predicted the highest risk.

Robson et al.¹⁸ examined 166 Latino mother–child dyads from pregnancy through 5 years of age in a prospective cohort study. A random forest classifier¹⁹ was used to

identify key determinants of childhood obesity. Random forests are an ensemble approach for decision trees, meaning they combine parameter estimates from multiple iterations of a decision tree to produce a more stable or efficient model than could be produced from a single iteration.⁸ From 10 prenatal/maternal and early postnatal features, 5 were identified as important. High pregravid BMI was the most important predictor, followed by high birth weight, large weight gain from birth to 6 months, nonexclusive breastfeeding 4–6 weeks of age, and low maternal age.

Hammond et al.⁴ used data from 3449 children before 24 months of age to predict obesity at 5 years of age in a racially/ethnically diverse community. They merged longitudinal electronic health records with census data, resulting in 1509 features. Of the six machine learning techniques they examined, Least Absolute Shrinkage and Selection Operator (LASSO) regression²⁰ performed best in both males and females. LASSO regression is a type of linear regression that implements feature selection to improve model interpretability and reduce overfitting concerns.⁷ Although nonanthropometric measures were included as candidate features, weight- and height-related features (*e.g.*, average weight-for-length z-score, 13–16 months of age; BMI gain, 0–24 months of age) accounted for 28 of the 35 features selected for females and 122 of the 144 features selected for males. High postpartum weight was also an important predictor for females.

Finally, Lee et al.²¹ examined 1,001,775 children aged 24–80 months in South Korea using 33 features across four national databases including sociodemographics, child's diet, maternal behaviors/health, paternal health, child's weight history, and health claims that were merged for a retrospective cohort study. A J48 decision tree, an approach similar to CART¹³ and that is also referred to as a C4.5 decision tree,^{15,22} was used in this study. The authors identified 11 features that predicted obesity at 24–80 months, with pregravid obesity being the strongest predictor, followed by paternal obesity and whether the family received public assistance for health care (a proxy for low SES). The highest prevalence of obesity was seen when children had mothers with pregravid obesity, had fathers with obesity, were not part of a family receiving public assistance for health care, and had mothers with gestational hypertension.

Middle Childhood

A limited number of studies have applied machine learning to examine risk factors for obesity during middle childhood (8–12 years of age). One of these, a conference article by Abdullah et al.,²³ explored the best combination of feature selection, attribute evaluators, and classifiers to predict BMI category in 12-year-olds in Malaysia using cross-sectional data. Depending on the combination of classifiers and feature selection, 17–29 features were selected that spanned sociodemographics, family medical

history, school information, and individual-level child characteristics, including weight history, birth and physical activity, diet, and early childhood factors. A J48 decision tree with 29 features was the best predictor of obesity.

Van Hulst et al.⁵ examined 512 non-Hispanic white 8–10-year-olds in Canada from the Quebec Adipose and Lifestyle Investigation in Youth (QUALITY) study.²⁴ Using CART, they found that, of the 11 features examined, 6 determined obesity cross-sectionally at 8–10 years of age. The highest prevalence of obesity was seen when the child had ≥ 1 parent with obesity (most important predictor), followed by not meeting physical activity guidelines and having 2 parents with abdominal obesity, average to high neighborhood disadvantage, and zero neighborhood parks.

Ortega Hinojosa et al.²⁵ used a random forest classifier in cross-sectional data to predict obesity in ~800,000 ninth, seventh, and ninth graders in the United States according to 124 features representing individual, school, neighborhood, and school county risk factors assessed during the mandatory annual fitness test in California. The top three contributors to obesity across all grades included low academic performance index, a high percentage of individuals learning English as a second language, and young age. Additional predictors across all grades included having more crime that is violent, a high diversity index, and more teachers per student and being of Hispanic/ethnicity and male.

Finally, Gray et al.²⁶ examined 3847 9–10-year-olds from the Adolescent Brain Cognitive Development (ABCD) study in the United States. They used cross-sectional data including 43 features related to demographics, psychological health, lifestyle behaviors, and cognition to predict the percentage of the 95th BMI percentile (see Flegal et al. for a definition of “percentage of the 95th BMI percentile”²⁷) with ridge, LASSO, and elastic net regression. These three regression approaches all penalize overfitting by shrinking regression coefficients using varying regularization approaches.⁷ LASSO performed best, resulting in the selection of 25 features. The most important features were no stimulant medication use followed by being of Hispanic ethnicity, nonwhite race, male sex, and lower SES and having unmarried parents. The authors also examined sex- and race/ethnicity-specific models, finding similar predictors for males and females (nonwhite race, stimulant medication use, and Hispanic ethnicity were top five predictors in both). In exploratory race/ethnicity-specific models, models for Hispanic and non-Hispanic black participants performed significantly worse than those for non-Hispanic whites. For Hispanic youth, key predictors were no stimulant medication use; being male; and having high screen use time, unmarried parents, and low matrix reasoning (*i.e.*, nonverbal abstract problem solving and inductive reasoning). For non-Hispanic black youth, key predictors were no stimulant medication use and having high behavioral inhibition, sleep initiation problems, and low matrix reasoning.

Late Childhood/Adolescence

Two of the four articles that applied machine learning to studying obesity in adolescence were conference articles, both of which used cross-sectional data from the National Youth Risk Behavior Survey (YRBS) in the United States. Pochini et al.²⁸ examined nine risk factors related to sleep, diet, physical activity, tobacco use, and screen time in 15,425 high school students. Using an unspecified decision tree, they found that obesity was best predicted by not being active during all previous 7 days, followed by tobacco use in the past 30 days and not consuming breakfast during all previous 7 days. Using similar lifestyle variables, Zheng and Ruggiero²⁹ compared the performance of different machine learning approaches in a subset of the YRBS ($n=5127$). All models indicated that frequent physical activity and breakfast intake were inversely associated with obesity, whereas frequent consumption of sugar-sweetened beverages and excessive computer use were directly associated with risk for obesity.

Nau et al.³⁰ examined 10–18-year-olds ($n=22,497$) living in the United States to predict average community-level BMI z-score. Of the 44 cross-sectional community-level variables entered into their conditional random forests model,³¹ 13 features consistently contributed to prediction of obesity across 50 different runs, 6 of which were social factors, including the top 3 determinants: unemployment, greater population density, and social disorganization. Removal of the social features, leaving food and physical activity features only, resulted in similar model accuracy.

Finally, Kim et al.³² used cross-sectional data from adolescents in South Korea to examine the performance of multiple machine learning methods for determining associations with BMI category. They used data from 11,206 individuals in the Korean Youth Behavior Survey and examined 10 risk factors for obesity. Risk factors included traditional SES measures (parental education and wealth), novel SES measures (pocket money, smartphone service, and academic performance), a novel measure of stress, and novel measures of sedentary behavior (time spent studying while seated, time on smartphone, and sleep quality). Using Bayesian Network³³ with Markov Blanket feature selection—a machine learning approach where class membership is predicted by probabilities and features are assumed to be dependent¹⁴—they found that a moderate amount of pocket money, followed by limited time sleeping, moderate time studying while seated, modest academic performance, and middle family economic level were the most important factors.

Discussion

Childhood Obesity Determinants by Age Group

The studies included in this review identified combinations of features that are most important for identifying childhood obesity, as well as their relative importance for predicting obesity. Approximately half of these studies examined early childhood determinants of obesity during

“the first 1000 days” (conception through 24 months of age). This is a critical period for obesity prevention, particularly with respect to diet as infants transition from formula/breastfeeding to solid foods and begin to develop dietary preferences and habits that may affect lifelong obesity risk.^{34,35} The majority of these early childhood studies predicted obesity between 4 and 7 years of age, a period referred to as the “adiposity rebound”³⁶ that is associated with risk of persistent overweight/obesity.³⁷ Across these early studies, elevated weight or weight gain (as assessed by varying weight-related measures),^{4,11,15,17,18} parent overweight/obesity,^{9,11,21} and maternal weight history (*i.e.*, high pregravid BMI^{17,18,21} and high postpartum weight⁴) were generally identified as important determinants of childhood obesity. These findings are consistent with previous research.^{38–42}

Fewer studies examined obesity determinants during middle childhood, an important period for obesity intervention,³ as children spend as much time in school as they do at home during the weekdays.⁴³ One of the four studies in this review focused on the school environment, finding that social environment features, such as learning English as a second language in school and more teachers per student, were particularly important for predicting obesity risk.²⁵ The other studies focused on a combination of individual, family, and neighborhood characteristics, and found that parental obesity^{5,23} and child’s history of obesity²³ were again key determinants of childhood obesity risk, along with physical inactivity.⁵ This is consistent with existing literature,^{44,45} including the importance of physical activity in childhood obesity prevention trials.³

Adolescence is a critical period for obesity prevention,⁴⁶ but studies among adolescents were scarce. One of the four studies examined novel predictors of obesity specific to this age range and reported that having some financial independence (*i.e.*, adolescents’ pocket money) was an important determinant of obesity, particularly for individuals from low-income families.³² The remaining three studies highlighted the importance of physical activity for obesity prevention,^{28–30} as would be expected given existing literature.³ Interestingly, findings from Nau et al.³⁰ indicated that, although social factors were the most important determinants of childhood obesity risk, they appeared to operate through physical activity and diet factors.

Childhood Obesity Determinants among Racial/Ethnic Minorities

Six of the studies reviewed here examined whether race/ethnicity predicted childhood obesity risk in a US population, and four found that Hispanic/Latino ethnicity of the child^{15,25,26} or mother¹⁷ was a top predictor of obesity risk. This echoes national statistics in the United States that indicate a disproportionate burden of obesity among racial/ethnic minorities.² Existing literature has shown that psychosocial and sociocultural determinants of health, such as weight perception and acculturative stress, may play a greater role in determining obesity risk in racial/ethnic

minority groups compared with non-Hispanic whites.^{47–49} The lack of inclusion of sociocultural variables in Gray et al.²⁶ may explain why the obesity prediction models for non-Hispanic black individuals and Hispanic/Latino individuals were less accurate than those for matched non-Hispanic white individuals.

Social determinants of health were not measured in the three studies that examined predominately racial/ethnic minorities in the United States.^{4,15,18} However, in contrast to the race/ethnic-specific models of Gray et al.,²⁶ their models performed well in terms of overall accuracy. One potential explanation for the discrepancy is that the aforementioned three studies focused on early childhood determinants (to predict obesity at 5 years of age^{4,18} or 2–10 years of age¹⁵), whereas Gray et al.²⁶ examined middle childhood determinants (to predict obesity at 9–10 years of age). Traditional early childhood risk factors may accurately predict childhood obesity at ~5 years of age across race/ethnic groups, but more novel sociocultural measures may be needed to accurately predict obesity during middle childhood and onward in racial/ethnic minorities. It is also possible that the discrepancies in accuracy were due to differences in study design (cross-sectional vs. longitudinal).

Irrespective of the race/ethnicity examined, studies included in this review primarily confirmed existing knowledge on the sociodemographic, behavioral, and environmental risk factors that predict obesity rather than identifying novel determinants of obesity, perhaps due to the fact that most studies examined a limited range of risk factors (≤ 12).^{5,9,11,17,18,28,29,32} An examination of a spectrum of sociocultural and psychosocial risk factors (*e.g.*, social support, stress, and acculturation) was notably lacking. Future studies should aim to incorporate these measures as well as more biological factors, such as the microbiome, genetics, or metabolomics, which have only been examined in machine learning studies that do not include features from other domains.^{50–52}

Evaluation of Existing Machine Learning Algorithms

Only 4 of the 15 studies reported sensitivity or specificity,^{15,17,18,29} with findings indicating good predictive validity (sensitivity: 62.2%–89.0%, specificity: 76.7%–99.5%). Some studies reported overall accuracy or area under the curve (53.7%–93%) in addition to or in place of sensitivity and specificity,^{15,17,23,29,30,32} but these measures conflate sensitivity and specificity and can thus lead to misleading conclusions about the accuracy of obesity prediction.⁵³ Risk factors identified from a model with a high sensitivity, rather than high specificity, will be most relevant for designing future obesity interventions as higher sensitivity indicates that youth with obesity were accurately identified as such.⁵⁴

The ideal data set for future machine learning applications would examine a wide range of features longitudinally, but researchers should also be mindful that including hundreds of variables may lead to overfitting or issues of interpretability.⁸ By only including variables that could plausibly be inter-

vened on or that could reasonably be associated with obesity, researchers could avoid this issue while also identify novel determinants important for effective childhood obesity interventions. Furthermore, using longitudinal instead of cross-sectional data in future research will improve interpretability of the machine learning models by establishing clear temporality between the risk factors and outcomes. Such studies should also consider using ensemble approaches that compare multiple machine learning and classical approaches in one algorithm. For example, SuperLearner in R statistical software^{55,56} compares 12 approaches and selects prediction function with the best cross-validated mean squared error.⁵⁷ Using such an approach helps researchers avoid committing to a poor model for their data *a priori* and produces the best prediction model for their data.⁵⁷

Conclusions

During early to middle childhood, child's weight history (*e.g.*, history of overweight/obesity or large increases in weight-related measures)^{4,11,15,17,18,23} and parental overweight/obesity (current or prior)^{4,5,9,11,17,18,21,23} are key determinants of childhood obesity risk, whereas social factors^{25,26,30} and physical inactivity^{5,28–30} appear to be important risk factors for obesity during middle childhood to adolescence. However, given the relative paucity of research in middle childhood and late childhood/adolescence, we caution against drawing firm conclusions on the most important determinants of obesity during these periods.

Future research should focus on using machine learning to (1) examine determinants of childhood obesity during middle childhood and adolescence, (2) identify how traditional risk factors for childhood obesity interact with novel sociocultural and psychosocial risk factors, and (3) identify the most pertinent determinants of obesity for racial/ethnic minorities. Such research should consider using existing large data sets to examine a wide range of features longitudinally and thus take full advantage of the benefits of machine learning approaches. Ultimately, this future research will aid in the development of effective obesity prevention interventions that address the disproportionate burden of obesity experienced by racial/ethnic minorities.

Disclaimer

The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Heart, Lung, and Blood Institute or the National Institutes of Health.

Funding Information

The research described was supported by the National Institutes of Health/National Center for Advancing Translational Science (NCATS) Einstein-Montefiore Clinical and Translational Science Award (UL1 TR002556). Additional support was provided by the Life Course Methodology Core (LCMC) at Albert Einstein College of

Medicine and the New York Regional Center for Diabetes Translation Research (P30 DK111022-8786 and P30 DK111022) through funds from the National Institute of Diabetes and Digestive and Kidney Diseases. Support for M.N.L. was provided by a National Heart, Lung, and Blood Institute training grant (T32HL144456). Support for D.B.H. was also provided by the National Heart, Lung, and Blood Institute (K01HL137557).

Author Disclosure Statement

No competing financial interests exist.

Supplementary Material

Supplementary Table S1
Supplementary Table S2
Supplementary Table S3

References

1. Ells LJ, Rees K, Brown T, et al. Interventions for treating children and adolescents with overweight and obesity: An overview of Cochrane reviews. *Int J Obes (Lond)* 2018;42:1823–1833.
2. Hales CM, Carroll MD, Fryar CD, Ogden CL. Prevalence of obesity among adults and youth: United States, 2015-2016. *NCHS Data Brief* 2017;288:1–8.
3. Wang Y, Cai L, Wu Y, et al. What childhood obesity prevention programmes work? A systematic review and meta-analysis. *Obes Rev* 2015;16:547–565.
4. Hammond R, Athanasiadou R, Curado S, et al. Predicting childhood obesity using electronic health records and publicly available data. *PLoS One* 2019;14:e0215571.
5. Van Hulst A, Roy-Gagnon M-H, Gauvin L, et al. Identifying risk profiles for childhood obesity using recursive partitioning based on individual, familial, and neighborhood environment factors. *Int J Behav Nutr Phys Act* 2015;12:17.
6. Bi Q, Goodman KE, Kaminsky J, Lessler J. What is machine learning: A primer for the epidemiologist. *Am J Epidemiol* 2019; 188:2222–2239.
7. Hastie T, Tibshirani R, Friedman J. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*, 2nd ed. Springer Series in Statistics. Stanford, Springer: California, 2009.
8. Yoo I, Alafaireet P, Marinov M, et al. Data mining in healthcare and biomedicine: A survey of the literature. *J Med Syst* 2012;36: 2431–2448.
9. Agras WS, Hammer LD, McNicholas F, Kraemer HC. Risk factors for childhood overweight: A prospective study from birth to 9.5 years. *J Pediatr* 2004;145:20–25.
10. Hammer LD, Bryson S, Agras WS. Development of feeding practices during the first 5 years of life. *Arch Pediatr Adolesc Med* 1999;153:189–194.
11. Toschke AM, Beyerlein A, von Kries R. Children at high risk for overweight: A classification and regression trees analysis approach. *Obes Res* 2005;13:1270–1274.
12. Breiman L, Friedman JH, Olshen RA, Stone CJ. *Classification and Regression Trees*. First Edition. Chapman and Hall/CRC: Boca Raton, FL, 1984.
13. Singh S, Gupta P. Comparative study ID3, CART and C4.5 decision tree algorithm: A survey. *Int J Adv Inf Sci Technol* 2014; 3:47–52.
14. Zhang S, Tjortjis C, Zeng X, et al. Comparing data mining methods with logistic regression in childhood obesity prediction. *Inf Syst Front* 2009;11:449–460.
15. Dugan TM, Mukhopadhyay S, Carroll A, Downs S. Machine learning techniques for prediction of early childhood obesity. *Appl Clin Inform* 2015;6:506–520.
16. Quinlan JR. Induction of decision trees. *Mach Learn* 1986;1: 81–106.
17. Kitsantas P, Gaffney KF. Risk profiles for overweight/obesity among preschoolers. *Early Hum Dev* 2010;86:563–568.
18. Robson JO, Verstraete SG, Shiboski S, et al. A risk score for childhood obesity in an Urban Latino Cohort. *J Pediatr* 2016;172: 29–34.e1.
19. Breiman L. Random forests. *Mach Learn* 2001;45:5–32.
20. Tibshirani R. Regression shrinkage and selection via the Lasso. *J R Stat Soc Series B Stat Methodol* 1996;58:267–288.
21. Lee I, Bang K-S, Moon H, Kim J. Risk factors for obesity among children aged 24 to 80 months in Korea: A decision tree analysis. *J Pediatr Nurs* 2019;46:e15–e23.
22. Quinlan JR. *C4.5: Programs for Machine Learning*. Morgan Kaufman Publishers, Inc.: San Francisco, CA, 1993.
23. Abdullah FS, Manan NSA, Ahmad A, et al. Data mining techniques for classification of childhood obesity among year 6 school children. In: Herawan T, Ghazali R, Nawi NM, Deris MM (eds), *Recent Advances on Soft Computing and Data Mining*. Advances in Intelligent Systems and Computing. Springer International Publishing: Bandung, Indonesia, 2017, pp. 465–474.
24. Lambert M, Van Hulst A, O’Loughlin J, et al. Cohort profile: The Quebec Adipose and Lifestyle Investigation in Youth cohort. *Int J Epidemiol* 2012;41:1533–1544.
25. Ortega Hinojosa AM, MacLeod KE, Balmes J, Jerrett M. Influence of school environments on childhood obesity in California. *Environ Res* 2018;166:100–107.
26. Gray JC, Schvey NA, Tanofsky-Kraff M. Demographic, psychological, behavioral, and cognitive correlates of BMI in youth: Findings from the Adolescent Brain Cognitive Development (ABCD) study. *Psychol Med* 2020;50:1539–1547.
27. Flegal KM, Wei R, Ogden CL, et al. Characterizing extreme values of body mass index-for-age by using the 2000 Centers for Disease Control and Prevention growth charts. *Am J Clin Nutr* 2009;90: 1314–1320.
28. Pochini A, Wu Y, Hu G. Data mining for lifestyle risk factors associated with overweight and obesity among adolescents. In: *2014 IIAI 3rd International Conference on Advanced Applied Informatics*. IEEE: Kita-Kyushu, Japan, 2014, pp. 883–888.
29. Zheng Z, Ruggiero K. Using machine learning to predict obesity in high school students. In: *2017 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*. IEEE: Kansas City, MO, 2017, pp. 2132–2138.
30. Nau C, Ellis H, Huang H, et al. Exploring the forest instead of the trees: An innovative method for defining obesogenic and obesoprotective environments. *Health Place* 2015;35:136–146.
31. Strobl C, Boulesteix A-L, Kneib T, et al. Conditional variable importance for random forests. *BMC Bioinformatics* 2008;9: 307.
32. Kim C, Costello FJ, Lee KC, et al. Predicting factors affecting adolescent obesity using general Bayesian Network and what-if analysis. *Int J Environ Res Public Health* 2019;16:4684.

33. Pearl J. *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*, 1st ed. Morgan Kaufmann Publishers, Inc.: San Francisco, CA, 1988.
34. Blake-Lamb TL, Locks LM, Perkins ME, et al. Interventions for childhood obesity in the first 1,000 days: A systematic review. *Am J Prev Med* 2016;50:780–789.
35. Woo Baidal JA, Locks LM, Cheng ER, et al. Risk factors for childhood obesity in the first 1,000 days: A systematic review. *Am J Prev Med* 2016;50:761–779.
36. Rolland-Cachera MF, Deheeger M, Bellisle F, et al. Adiposity rebound in children: A simple indicator for predicting obesity. *Am J Clin Nutr* 1984;39:129–135.
37. Dietz WH. Critical periods in childhood for the development of obesity. *Am J Clin Nutr* 1994;59:955–959.
38. Wrotniak BH, Epstein LH, Paluch RA, Roemmich JN. Parent weight change as a predictor of child weight change in family-based behavioral obesity treatment. *Arch Pediatr Adolesc Med* 2004;158:342–347.
39. Andriani H, Liao C-Y, Kuo H-W. Parental weight changes as key predictors of child weight changes. *BMC Public Health* 2015;15:645.
40. Robinson CA, Cohen AK, Rehkopf DH, et al. Pregnancy and post-delivery maternal weight changes and overweight in preschool children. *Prev Med* 2014;60:77–82.
41. van Rossem L, Wijga AH, Gehring U, et al. Maternal gestational and postdelivery weight gain and child weight. *Pediatrics* 2015;136:e1294–1301.
42. Heslehurst N, Vieira R, Akhter Z, et al. The association between maternal body mass index and child obesity: A systematic review and meta-analysis. *PLoS Med* 2019;16:e1002817.
43. Epps EG, Smith SF. Chapter 7: School and Children: The Middle Childhood Years. In: Collins WA. *Development During Middle Childhood: The Years From Six to Twelve*. National Academies Press: Washington, DC, 1984, pp. 283–334.
44. Mamun AA, Lawlor DA, O’Callaghan MJ, et al. Family and early life factors associated with changes in overweight status between ages 5 and 14 years: Findings from the Mater University Study of Pregnancy and its outcomes. *Int J Obes (Lond)* 2005;29:475–482.
45. Francis LA, Ventura AK, Marini M, Birch LL. Parent overweight predicts daughters’ increase in BMI and disinhibited overeating from 5 to 13 years. *Obesity (Silver Spring)* 2007;15:1544–1553.
46. Taylor SA, Borzutzky C, Jasik CB, et al. Preventing and treating adolescent obesity: A position paper of the society for adolescent health and medicine. *J Adolesc Health* 2016;59:602–606.
47. Isasi CR, Rastogi D, Molina K. Health issues in Hispanic/Latino Youth. *J Lat Psychol* 2016;4:67–82.
48. D’Alonzo KT, Johnson S, Fanfan D. A biobehavioral approach to understanding obesity and the development of obesogenic illnesses among Latino immigrants in the United States. *Biol Res Nurs* 2012;14:364–374.
49. Hendley Y, Zhao L, Coverson DL, et al. Differences in weight perception among Blacks and Whites. *J Womens Health (Larchmt)* 2011;20:1805–1811.
50. Korpela K, Renko M, Vänni P, et al. Microbiome of the first stool and overweight at age 3 years: A prospective cohort study. *Pediatr Obes* 2020;15:e12680.
51. Gerl MJ, Klose C, Surma MA, et al. Machine learning of human plasma lipidomes for obesity estimation in a large population cohort. *PLoS Biol* 2019;17:e3000443.
52. Wang H-Y, Chang S-C, Lin W-Y, et al. Machine learning-based method for obesity risk evaluation using single-nucleotide polymorphisms derived from next-generation sequencing. *J Comput Biol* 2018;25:1347–1360.
53. Zou KH, O’Malley AJ, Mauri L. Receiver-operating characteristic analysis for evaluating diagnostic tests and predictive models. *Circulation* 2007;115:654–657.
54. Rautiainen I, Äyrämö S. Predicting overweight and obesity in later life from childhood data: A review of predictive modeling approaches. *arXiv:1911.08361*. Available at <https://arxiv.org/abs/1911.08361> Last accessed July 23, 2020.
55. Laan MJ van der, Polley EC, Hubbard AE. Super learner. *Stat Appl Genet Mol Biol* 2007;6:25.
56. Polley E, LeDell E, Kennedy C, Van der Laan M. SuperLearner: Super Learner Prediction, Package Version 2.0-26. 2019. Available at <https://cran.r-project.org/web/packages/SuperLearner/> Last accessed October 19, 2020.
57. Rose S. Mortality risk score prediction in an elderly population using machine learning. *Am J Epidemiol* 2013;177:443–452.

Address correspondence to:

Madison N. LeCroy, PhD

Department of Epidemiology and Population Health

Albert Einstein College of Medicine

1300 Morris Park Avenue

Bronx, NY 10461

USA

E-mail: madison.lecroy@einsteinmed.org