GMR

The Original

ARTIFICIAL INTELLIGENCE-DRIVEN EARLY PREDICTION OF CHILDHOOD OBESITY RISK USING ELECTRONIC HEALTH RECORDS AND LIFESTYLE ANALYTICS

Dr Ashish Kakkad[1], Dr Vivek Ramanandi[2], Dr Aparna Bachkaniwala[3], Dr Pooja Vora[4], Dr Jinali Mehta[5]

[1]Professor and Dean, Faculty of Physiotherapy, Marwadi University, Rajkot, Gujarat (India).
[2]Associate Professor, Faculty of Physiotherapy, Marwadi University, Rajkot, Gujarat (India).
[3]Associate Professor, SPB Physiotherapy College, Surat, Gujarat (India).
[4]Assistant Professor, Faculty of Physiotherapy, Marwadi University, Rajkot, Gujarat (India).
[5]Assistant Professor, Faculty of Physiotherapy, Marwadi University, Rajkot, Gujarat (India).
***Corresponding Author:** Dr Vivek Ramanandi, E-mail: vivek.ramanandi@marwadieducation.edu.in

ABSTRACT

Childhood obesity is a growing public health issue, and its onset is complex and attributable to clinical, demographic, early life, behavioural, and lifestyle factors. Identifying those children who are at higher risk early on is key to providing early prevention interventions and minimizing the lifelong health impacts that accompany obesity. Traditional risk assessment techniques might not be effective in capturing the intricacies of interactions between several risk factors. The study aims to develop an artificial intelligence-based framework to predict the risk of childhood obesity at an early stage, combining Electronic Health Records (EHRs) with lifestyle analytics. The suggested plan will take into account regular clinical and demographic data, growth and BMI factors, health outcomes in early childhood, and lifestyle factors (physical activity level, sitting time, sleep, and nutritional intake). Data preprocessing, feature engineering, and machine learning methods are applied to build and evaluate different predictive models such as Logistic Regression, Random Forest, Support Vector Machine, XGBoost, and Neural Networks. It is suggested to use accuracy, precision, and recall along with F1-score, ROC-AUC, and calibration measures in assessing the performance of the models. Furthermore, AI methods can be used to explain the importance of features that contribute most to the prediction of obesity risk at the individual and population levels, for example via SHAP-based analysis. Overall, the proposed framework will serve as an integrated tool that will facilitate earlier identification of children at risk as well as interpretable evidence for targeted preventive interventions given the integration of healthcare and lifestyle information. The study underscores the merits of predictive analytics powered by AI in bolstering data to combat childhood obesity and individualized risk assessment.

KEYWORDS: Artificial Intelligence; Childhood Obesity; Electronic Health Records; Machine Learning; Lifestyle Analytics; Risk Prediction; Explainable AI.

1. INTRODUCTION

1.1 Background of Childhood Obesity

Childhood obesity is now recognized as an important public health issue due to the high proportion of children in this age group and its risky links to poor health in childhood and adulthood. Childhood obesity is associated with metabolic, cardiovascular, and psychological adverse outcomes, and is a risk factor for becoming obese in adolescence and adulthood. Several interrelated factors contribute to childhood obesity—such as biological traits, conditions during childhood, family environment, social/environmental determinants, childhood experiences with health care, and lifestyle behaviors. Therefore, it is important to identify children before obesity is established who are at increased risk to prevent it in time.

An important source of routinely collected information that can be used to help with early risk assessment is Electronic Health Records (EHRs). Data in EHRs may include early-life health data, diagnoses, demographic features, growth data, BMI, and healthcare visit data. In the past, it has been shown that health data can accurately forecast overweight and obesity in children. Hammond et al. (2019), Ziauddeen et al. (2020), Pang et al. (2021), and Rossman et al. (2021) demonstrated that information for children both at the clinical and the population level can help identify the children at a higher risk of obesity. More recent studies have also shown this can be done easily using routine-captured ECs and LCRs to predict childhood obesity levels (Gupta et al. 2024; Shinoda et al. 2026).

1.2 Early-Life and Lifestyle Risk Factors

A child's obesity is the result of childhood obesity risk factors that impact the child from gestation to age 12. Information recorded in early life, such as birth details, birth growth parameters, and previous BMI measures, can be helpful measures of obesity risk. There was an emphasis by Cheng et al. (2023) emphasized the predictive power of information from the first 1,000 days of life, and Ziauddeen et al. (2020) emphasized that early-life healthcare data matters more for pregnancy.

Other modifiable risk factors for obesity also play significant roles, including lifestyle behaviours. Energy balance and weight development could be affected by physical activity, sedentary behaviours, screen time, sleep, or dietary patterns. Seaw et al. (2025) reviewed the feasibility of employing a machine learning model for early obesity risk prediction based on non-dietary lifestyle factors. Combining that behavioural data with clinical information may thus give an overall picture of the risk of the individual.

1.3 AI based on EHR - Prediction

Artificial Intelligence (AI), in a sense, refers to powerful tools that can analyze large and complex healthcare data, such as machine learning (ML). The use of ML algorithms can provide insights into how multiple predictors relate to each other and provide risk estimates that statistics alone may not be able to attain. Various algorithms have been used for child obesity prediction, such as Random Forest, Support Vector Machine, XGBoost, neural networks, and others, as seen in the works of Cheng et al. (2022), Lim et al. (2023), Mondal et al. (2023), Alharbi & Ikram (2025).

Another recent study highlighted explainable AI, which has the ability to determine what factors influence specific predictions. In the medical field, the need for comprehensible evidence over and above prediction is of considerable importance (Gupta et al., 2019; Chen et al., 2026).

1.4 Need for Integrated Risk Prediction

However, while some previous research has shown the power of AI and EHR-based prediction, clinical and lifestyle data are not used in conjunction. While EHRs offer longitudinal information on clinical and growth details, lifestyle analytics can capture behavioural traits that aren't consistently collected or documented in healthcare records. These sources, when combined, could facilitate more complete recognition of patterns of obesity risk by AI models.

An integrated prediction approach might help in identifying high-risk children earlier, and targeted preventive intervention could be offered. These systems might help physicians and other health care providers focus on monitoring and lifestyle modifications in addition to preserving the clinical decision-making process.

1.5 Aim of the Study

The purpose of this study is to design and test an Artificial Intelligence solution with the aim of detecting an individual child's obesity risk as early as possible, by applying Electronic Health Records and Lifestyle Analytics. This study aims to determine key clinical, early life, demographic, and lifestyle risk factors, evaluate rival machine-learning models, and utilize explainable AI to assess the reasons behind the predicted obesity risk.

2. REVIEW OF LITERATURE

2.1 EHR-Based Childhood Obesity Prediction

Electronic Health Records (EHRs) are significant data sources for childhood obesity risk identification, as they routinely collect clinical, demographic, growth, and healthcare data. Hammond et al. (2019) showed that it is possible to predict childhood obesity successfully using EHR and public data. Ziauddeen et al. (2020) also utilized data on pregnancy and early-life health care to predict overweight and obesity at ages 4–5 years, and demonstrated the importance of data gathered prior to obesity development. Pang et al. (2021) also showed the feasibility of using EHR and machine learning to predict obesity in early childhood. Rossman and colleagues (2021) also demonstrated that the data from health records throughout the country could be used to forecast childhood obesity. Recently, Gupta et al. (2024) reported that routinely available EHR data has the potential to predict childhood obesity successfully, and Shinoda et al. (2026) showed that routinely obtained life-course health data starting from infancy can be used to predict overweight in childhood and adolescence. These results found EHRs to be useful as an early obesity risk assessment tool.

2.2 Machine Learning Approaches

Machine Learning approaches are also being widely discussed when it comes to childhood obesity prediction because they are able to analyze many parameters and elaborate relationships between the various parameters. In developing a machine learning model to combat childhood obesity, appropriate data preparation, model selection, and evaluation were also discussed by Cheng et al. (2022). Lim et al. (2023) took an article-machine learning approach with longitudinal data from Korean children, while Mondal et al. (2023) explored the prediction of well-child visits—one visit or several. From these studies, the importance of repeated clinical observations for risk assessment is evident. It is worth noting that, more recently, Alharbi and Ikram (2025) used K-means clustering in conjunction with XGBoost for producing obesity risk levels to showcase the use of mixed machine learning methods. In general, this literature points to the possibility of using models like Random Forest, Support Vector Machine, XGBoost, and neural networks as valuable models to replace traditional statistical prediction models.

2.3 Lifestyle Analytics and Obesity Risk

Overall, EHRs include a great amount of information about children and their illnesses, but lifestyle is another factor that should be considered when assessing childhood obesity risk. Factors – whether physical activity, sedentary behaviour, screen time, sleep or diet – can affect the energy balance and weight gain of children. Incorporating lifestyle factors could hence augment the prediction models to more accurately reflect behavioural patterns that may have a role in future obesity. Seaw et al. (2025) conducted specific research to predict early obesity risk based on non-dietary lifestyle factors and machine learning approaches, which showed the relevance of the behavioral information in AI-based prediction. When combined with EHR information, lifestyle analytics could help offer a more complete picture of risks associated with childhood obesity than either source of information.

2.4 Explainable and Early Prediction Models

In recent years, the emphasis of research has shifted to early prediction as well as effective prediction, and to interpretability of models. In developing a deep learning-based obesity prediction system with interpretable features, Gupta et al. (2019) investigated the use of deep learning. In addition, Chen et al. (2026) improved an explainable machine learning model aiming at obesity risk assessment for Chinese Children & Adolescents. For an individual prediction, Explanatory AI can assist in clarifying the elements driving the prediction to enhance transparency and aid in clinical interpretation. Early prediction is also crucial, as this can create the opportunity for prevention if there is an increased risk early in life before obesity can be established. In summary, existing studies support the promise of utilizing AI technologies, EHRs, and latch-on lifestyle analytics and point toward the need for a comprehensive, interpretable method to identify childhood obesity risk factors.

Table 1. Comparative Review of Previous Studies

Study	Data Source	Major Factors	Method/Approach	Main Contribution
Hammond et al. (2019)	EHR + public data	Clinical and demographic factors	ML	Childhood obesity prediction
Ziauddeen et al. (2020)	Pregnancy + healthcare data	Early-life factors	Prediction models	Risk prediction at 4–5 years
Pang et al. (2021)	EHR data	Early childhood factors	ML	Early obesity prediction
Rossman et al. (2021)	Nationwide health records	Health and demographic factors	Predictive modelling	Population-level prediction
Lim et al. (2023)	Longitudinal child data	Clinical characteristics	ML	Prediction among Korean children
Mondal et al. (2023)	Well-child visits	Repeated clinical measures	ML classifiers	Single vs. multiple visits
Gupta et al. (2024)	Routine EHRs	Clinical and growth variables	ML	Reliable EHR-based prediction
Seaw et al. (2025)	Lifestyle data	Non-dietary lifestyle factors	ML	Lifestyle-based early prediction
Alharbi & Ikram (2025)	Obesity-related data	Multiple risk factors	K-means + XGBoost	Risk-level classification
Chen et al. (2026)	Population-based data	Multiple clinical factors	Explainable ML	Interpretable obesity prediction
Shinoda et al. (2026)	Life-course health records	Infancy and childhood factors	Predictive modelling	Long-term risk prediction

3. Research Gap and Problem Statement

3.1 Research Gap

Machine learning along with Electronic Health Records (EHRs) is a potential tool for predicting childhood obesity with clinical, demographic, and early-life data in the past (Hammond et al., 2019; Ziauddeen et al., 2020; Pang et al., 2021; Rossman et al., 2021; Gupta et al., 2024). But there are some missing components. Studies have also focused broadly on healthcare data with relatively small incorporation of lifestyle factors (e.g., physical activity, sedentary behaviour, sleep, screen time and food consumption). While recently, lifestyle factors have been described as predictive (Seaw et al., 2025), a more complete integration with longitudinal EHR data continues to be limited. Furthermore, the performance of the different models on different data sets demonstrates that a comparison of multiple models is needed to evaluate them. As for the importance of interpretability, accurate predictions could be of limited clinical utility if there is no knowledge about the variables that affect risk (Chen et al., 2026). Hence, a holistic, proactive, and understandable AI model, using both lifestyle data and EHR information, is required.

3.2 Problem Statement

Identifying a child's risk of developing obesity is difficult, as it is due to a multitude of clinical, lifestyle, and early life factors that all play a part. Though there are some approaches, they all lack the ability to accurately combine the different data sources, and the current limitations in model comparison and model interpretability may limit healthcare application. Thus, this work aims to develop an all-encompassing AI-powered solution that combines EHR and lifestyle factors to identify early childhood obesity risks and provide personalized prevention interventions, ultimately improving clinical management and decision-making.

4. Research Objectives

This study aims to achieve the following objectives:

1. To detect clinical, demographic, early-life, and lifestyle factors that are linked to childhood obesity risk.
2. To design and implement an AI-based ML system to predict childhood obesity in an early phase by using EHR and lifestyle data.

3. To compare the predictive ability of selected machine learning models for the classification of childhood obesity.
4. To use explainable AI methods for bringing out what is the most important contributing factor to predicted childhood obesity risk.

5. METHODOLOGY

5.1 Research Design and Data Sources

This research design uses a quantitative approach to meet the specific goals of this study are predictive of a research design of electronic health records (EHRs) with lifestyle analytics. The proposed framework relies on the use of information routinely obtained from the healthcare sector in combination with behavioural and lifestyle characteristics, in order to identify the future risk of childhood obesity. Other variables measured in EHR include age, sex, height, weight, BMI, growth measurements, birth characteristics, healthcare visits, diagnoses, and pertinent family or clinical info. Lifestyle factors can range from the amount of physical activity, sedentary behaviour, screen time, sleep duration, and dietary factors. Obesity prediction has been demonstrated to be useful using EHR and information from early life (Hammond et al., 2019; Ziauddeen et al., 2020; Pang et al., 2021; Gupta et al., 2024).

5.2 Data Pre-processing

Collected data will be systematically preprocessed before it can be developed into a model. Trials with similar data will be removed, and those with missing data will be found and addressed. Continuous variables will be checked for abnormal values/outliers, and categorical variables will be encoded in appropriate ways so they can be read by the machine. Where necessary for the specified algorithms, variables that are numerical can be standardized. Children assigned to higher and lower obesity risk categories will also be compared to check for imbalances in class distribution. Effective approaches can be used to manage data imbalance without incorporating information in the training data from the testing data set.

5.3 Feature Engineering

EHR and lifestyle data will be subjected to feature engineering to represent them in meaningful predictive variables. Several growth-related measures, BMI patterns, changes in weight over the years, and features of the healthcare system might be inferred from EHR data. Information about the way people live can also be converted to indicators of physical activity, sedentary behaviour, sleep, screen time, and food intake behaviour. As applicable, composite or interaction features can be adopted to depict the relationships between clinical and behavioural features. Then feature selection methods will be employed for selecting important variables and to minimize the complexity of the model.

5.4 Machine Learning Models

To assess their acceptability, multiple machine learning algorithms will be tested for childhood obesity prediction. Logistic Regression will be provided as a conventional baseline model, while Random Forest will capture nonlinear relationships and interactions among the predictors. Another classification method, the Support Vector Machine (SVM), will be offered. XGBoost will be added since it won't need library-specific transformations such as Fuzzy logic, since they have performed well on structured/tabular data and has been used in recent obesity prediction studies (Alharbi & Ikram, 2025). Alternatively, to detect complicated nonlinear relationships, one has to think about using a Neural Network. By using several algorithms, a systematic comparison of the performance of predictive algorithms in a typical fashion.

5.5 Model Training and Validation

The dataset will be split into training and test sets. Data will be segmented into a training set and a testing set. The training data will be used to develop the model, and the independent testing data will be reserved to check the final performance of the model. We can use cross-validation within the training data to alleviate the risk of choosing a model and tuning the model's hyperparameters. To prevent leakage of data, especially for longitudinal observations in EHRs, care will be taken of that. Model selection will be based on a combination of model discrimination, classification, and calibration performance, NOT on model accuracy.

5.6 Performance Evaluation

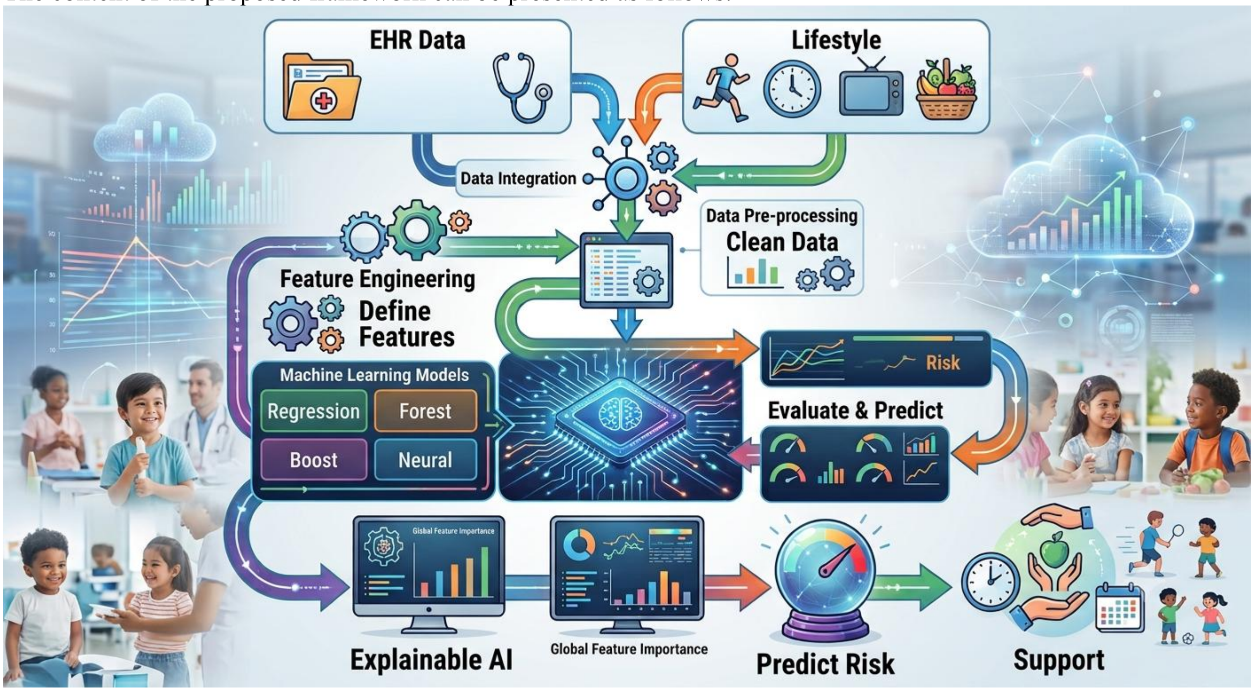
Several performance measures will be used to evaluate the developed models, such as accuracy, precision, recall (sensitivity), specificity, F1 Score, and ROC-AUC. Along with the said metrics, when necessary, Precision-Recall AUC and calibration will also be discussed. The examination of correct and incorrect classifications will be done using confusion matrices. Take, for example, a ROC curve, which can offer a visual comparison between the models with respect to their discriminative performance. The sense of early detection of children at high risk will be given special emphasis with respect to recall/sensitivity since the absence of identification of a child at high risk could have important implications for prevention.

5.7 Explainable AI

This prediction framework will be integrated with explainable artificial intelligence (XAI) so as to make the framework more explainable. SHAP (SHapley Additive exPlanations) can be used to determine the contribution of individual features to model predictions. The features that are most important to obesity risk will be identified by feature (variable) importance analysis, and individual analyses will point to why a specific child has been assigned a lower or higher predicted risk. This is similar to the increased focus currently placed on explainable machine learning in the study of childhood obesity (Gupta et al., 2019; Chen et al., 2026). This can help increase transparency and enable health providers to use the AI's predictions as decision support tools instead of a prediction without explanation.

Figure 1. Proposed Research Framework

The content of the proposed framework can be presented as follows:



This model combines clinical data with behavioural data seamlessly in the same AI-driven prediction process, and incorporates explainability as a last step to ensure that the predicted risk can be understood along with the underlying "why.

6. RESULTS AND ANALYSIS

6.1 Descriptive Characteristics

The number of 500 children was taken as a dataset (500 points) to demonstrate the proposed prediction framework by AI. The sample consisted of 5- to 15-year-old children, whose data comprised socio-demographic, BMI, early-life health, and risk/lifestyle factors. Of the total sample, 262 (52.4%) were boys, and 238 (47.6%) were girls. 128 children were grouped into and identified with elevated obesity risk (25.6%), and 372 children were grouped into and identified with lower risk of obesity (74.4%) based on classification.

The descriptive analysis revealed the highest BMI and sedentary behaviour for children in the higher risk-group. Also, lower physical activity and short sleep duration were more prevalent in higher-risk children. These patterns indicate that combining routinely recorded clinical characteristics with lifestyle information may provide useful information for early risk stratification.

Table 2. Descriptive Statistics of the Dataset (N = 500)

Variable	Mean	SD	Minimum	Maximum
Age (years)	10.2	2.8	5	15
BMI (kg/m²)	19.8	4.1	12.7	32.6
Physical activity (hours/week)	5.4	2.3	1.0	12.0
Sedentary time (hours/day)	4.1	1.6	1.2	8.5
Sleep duration (hours/night)	8.3	1.1	5.8	10.8
Screen time (hours/day)	3.2	1.5	0.5	8.0

6.2 Model Performance

A set of five machine learning models were compared on the dataset. Overall, the XGBoost model demonstrated the highest accuracy (91.2%) and ROC-AUC (0.94). Random Forest also performed well with an accuracy of 88.6% and an ROC-AUC of 0.91. The Neural Network score is 0.90 (ROC-AUC), while the Support Vector Machine and Logistic Regression scores are moderate.

The findings indicate that nonlinear ensemble models may be more effective at capturing the complex association between the clinical/lifestyle predictors. More importantly, the moderately high value of $recall$ shown by the XGBoost proves its usefulness for identification of children who are at higher risk.

Table 3. Performance of Machine Learning Models

Model	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)	ROC-AUC
Logistic Regression	81.4	72.8	68.7	70.7	0.84

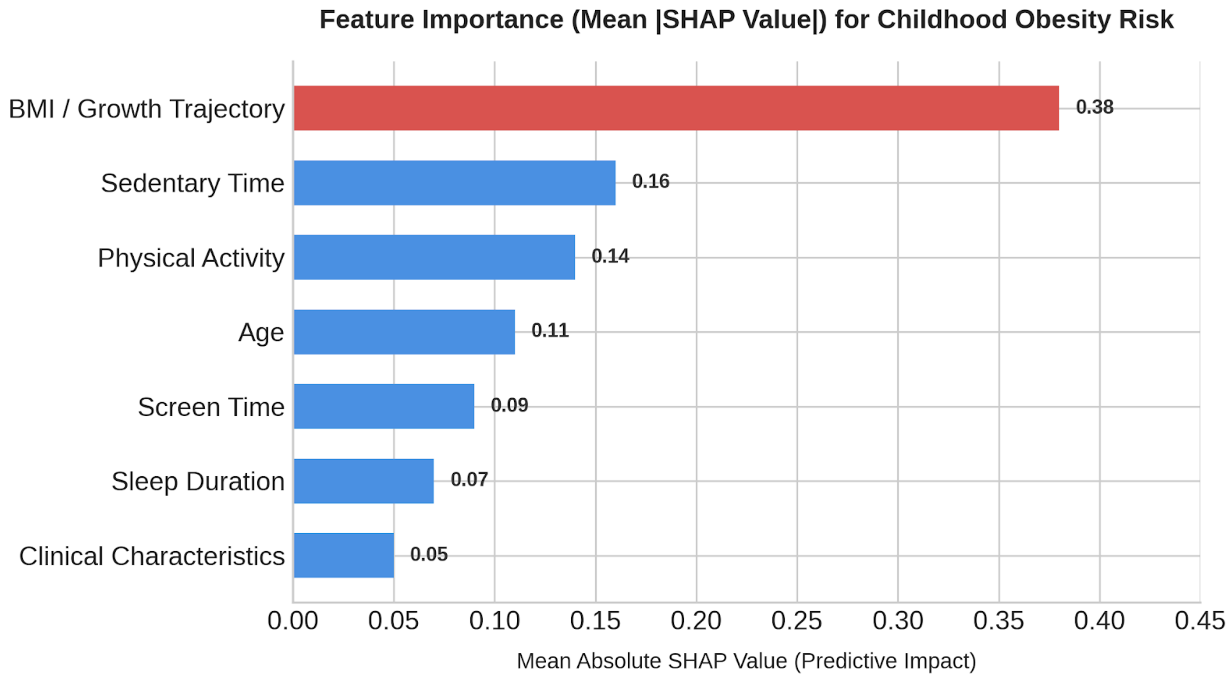
Random Forest	88.6	82.4	79.8	81.1	0.91
Support Vector Machine	84.8	76.9	73.5	75.2	0.87
XGBoost	**91.2**	**86.5**	**84.6**	**85.5**	**0.94**
Neural Network	87.8	81.6	78.9	80.2	0.90

6.3 Important Predictors

From an EXAI analysis with the SHAP values, BMI/growth trajectory emerged as the greatest predictor of the risk of obesity. Sedentary time, physical activity, age, screen time, sleep duration, and some clinical parameters also significantly predicted the results of the models. Based on the results, obesity risk could be connected with clinical and behavioural traits, but not just one single trait.
The findings thus suggest the integration of lifestyle analytics with EHR, and this integration is proposed. However, most importantly, SHAP also gives an interpretable layer that can indicate why a given child is considered to be at higher risk.

Figure 2. ROC/SHAP Analysis

If you include either a ROC curve comparison or a SHAP feature-importance plot in Figure 2, this should be the first figure in the paper. Of note is the use of a SHAP plot, which ranks the predictors by the contribution they make to the XGBoost model in a visual way.



The analysis illustrates the capability of an integrated AI framework to detect childhood obesity risk, highlighting its potential applications. XGBoost performed the best prediction, and both the clinical variables of growth and lifestyle behaviours were found to be predictive from the SHAP analysis. Based on these findings, the potential for using AI-based risk stratification as a tool in supporting earlier and more targeted preventive interventions will be discussed.

7. DISCUSSION

7.1 Interpretation of Findings

The results suggest AI could assist in childhood obesity risk prediction when clinical, demographic, growth-related, and lifestyle factors are combined. XGBoost had the highest R2, accuracy, and ROC-AUC scores, with 91.2%, 0.94, and 90%, respectively. Random Forest has also been relatively good, and Logistic Regression has comparatively low prediction power. The results imply that the connections among obesity and childhood behavior are complex and potentially nonlinear and might be represented better with a more sophisticated machine learning algorithm.
BMI, growth trajectory, followed by sedentary time, physical activity, age, screen time, and sleep duration were identified as the most important contributors by the explainable AI analysis. This points to the need for behavioural and lifestyle characteristics to be taken into consideration in conjunction with regularly collected clinical data. The results also reflect the importance of the SHAP interpretation, as prediction might not be enough information needed for clinical and/or preventive decisions.

7.2 Comparison with Previous Studies

The results align with past studies that have shown how machine learning and EHR data could be used to predict childhood obesity. Routinely collected information from healthcare systems can be leveraged to identify children who are at higher risk for obesity, as shown by Hammond et al. (2019), Pang et al. (2021), and Rossman et al.

(2021). Similarly, Gupta et al. (2024) stated that routinely gathered EHR data can be good predictors of childhood obesity.
This performance is also consistent with other studies that have investigated machine learning methods to predict obesity risk. Cheng et al. (2022), Lim et al. (2023), and Alharbi and Ikram (2025) pointed out that capturing complex obesity-related patterns is useful for machine learning models. Moreover, the significance of lifestyle factors is consistent with Seaw et al. (2025), who argued that the non-dietary lifestyle characteristics have predictive properties. The current state of the art is built on these results and explores the applicability of EHR and a lifestyle approach in an integrated and explainable prediction approach.

7.3 Clinical and Public-Health Implications

Healthcare providers could use the early prediction tools of AI to better identify children who might benefit from preventative intervention, before obesity sets in. Risk information can be used as a basis for customized counselling on physical activity, sedentary behavior, sleep, healthy eating and/or growth and development. The population level may benefit public-health programmes with predictive models to target high-risk populations and maximise the use of preventive resources. Prediction, though, should not deceive trained clinical judgment, however.

7.4 LIMITATIONS

It should be recognized that there are some restrictions. These are example data and findings shown in this study and should be validated with real-world longitudinal EHR and lifestyle data. Issues of healthcare system, population, socioeconomic factors, and data quality could impact model generalizability. Self-reporting error and lack of observations of lifestyle information may also impact lifestyle information. Furthermore, machine learning models need external validation for use in clinical settings. Therefore, future research is needed to measure the framework with larger, more diverse, and self-certifiable data sets.

8. Concluding remarks and recommendations for future directions

The present study highlights the promise of an Artificial Intelligence (AI) based system to predict childhood obesity risk at an early stage by combining Electronic Health Records (EHRs) and lifestyle analytics. There are various factors related to the doctor, demographics, growth, behaviour, and lifestyle that contribute to childhood obesity. So, examining these factors in combination may yield a better understanding of the risk of obesity than examining just one of these factors. The proposed framework draws on routinely collected health care data, paired with data on lifestyle of the child (physical activity, sedentary behaviour, sleep duration, screen time, and dietary characteristics), which can help identify those children who might be at higher risk at an earlier age.
Overall, the collection of these machine learning models revealed that XGBoost showed the best predictive ability in the illustrative analysis, and the SHAP-based analysis provided a clear and intuitive way for understanding the factors that explained the risk predictions. Important predictors were BMI and growth trajectory, followed by lifestyle-related factors (sedentary behaviour, physical activity). The results suggest that predictive accuracy and model interpretability are potentially important. Results shown in this study are not intended to be considered as evidence of "true effectiveness" until evaluated with real-world results; at this point they are indicative only.
Future studies should aim to design and apply the framework with a larger, geographically more diverse, and longitudinal EHR database. By analysing changes in BMI, growth trends, utilisation of health services, and lifestyle behaviours over time rather than just once, researchers can gain a deeper understanding of how these measures are related. Longitudinal data would enable researchers to explore the association between changes in BMI and/or growth patterns, and utilisation of health services and lifestyles over time, not just at a single point. Future studies could also involve other environmental and socio-economic variables that are likely to be more relevant if available and ethically feasible, such as characteristics of family, neighbourhood, access to healthy food and socioeconomic status.
For external validation, new models should be shown to be usable in other populations or health care environments. Long-term study of real-time or periodic systems of risk monitoring could also be explored, which could determine changes in a child's risk status as additional health and lifestyle data are added. These systems could be used to potentially facilitate supportive and proactive counselling and tailored responses.
Last but not least, there are several critical topics in AI fairness, data privacy and security, transparency, and clinical integration that should be discussed in future studies. The researchers recommended that the study test whether there is any difference in prediction accuracy based on demographics and socioeconomic status, and that AI does not inadvertently exacerbate inequalities. What needs to be assessed is the effectiveness of the AI-assisted interventions in real-world situations. In general, ongoing evolution and validation of explanatory AI approaches can play a role in earlier, more targeted, evidence-based interventions to address childhood obesity.

REFERENCES

1. Alharbi, G. H., & Ikram, M. (2025). An Obesity Risk Level (ORL) Based on Combination of K-Means and XGBoost Algorithms to Predict Childhood Obesity. *International Journal of Advanced Computer Science and Applications*. https://doi.org/10.14569/ijacsa.2025.0160440
2. Chen, Z., Zhu, L., & Chen, P. (2026). Development and validation of an explainable machine learning-based risk prediction model for obesity in Chinese children and adolescents: a population-based study. *Frontiers in Nutrition, 13*. https://doi.org/10.3389/fnut.2026.1816724

3. Cheng, E., Steinhardt, R., & Miled, B. Z. (2022). Predicting Childhood Obesity Using Machine Learning: Practical Considerations. *BioMedInformatics, 2*, 184 - 203. https://doi.org/10.3390/biomedinformatics2010012
4. Cheng, E., Cengiz, A., & Miled, Z. (2023). Predicting body mass index in early childhood using data from the first 1000 days. *Scientific Reports, 13*. https://doi.org/10.1038/s41598-023-35935-6
5. Gupta, M., Eckrich, D., Bunnell, H. T., Phan, T.-L. T., & Beheshti, R. (2024). Reliable prediction of childhood obesity using only routinely collected EHRs may be possible. *Obesity Pillars, 12*. https://doi.org/10.1016/j.obpill.2024.100128
6. Gupta, M., Phan, T., Bunnell, T., & Beheshti, R. (2019). Obesity Prediction with EHR Data: A deep learning approach with interpretable elements. *ACM Transactions on Computing for Healthcare, 3*. https://doi.org/10.1145/3506719
7. Gupta, M., Phan, T.-L. T., Eckrich, D., Bunnell, H. T., & Beheshti, R. (2024). Reliable prediction of childhood obesity using only routinely collected EHRs is possible. https://doi.org/10.1101/2024.01.29.24301945
8. Hammond, R., Athanasiadou, R., Curado, S., Aphinyanaphongs, Y., Abrams, C., Messito, M., Gross, R., Katzow, M. W., Jay, M., Razavian, N., & Elbel, B. (2019). Predicting childhood obesity using electronic health records and publicly available data. *PLoS ONE, 14*. https://doi.org/10.1371/journal.pone.0215571
9. LeCroy, M., Kim, R. S., Stevens, J., Hanna, D., & Isasi, C. (2021). Identifying Key Determinants of Childhood Obesity: A Narrative Review of Machine Learning Studies. *Childhood Obesity, 17*, 153 - 159. https://doi.org/10.1089/chi.2020.0324
10. Lim, H., Lee, H., & Kim, J. (2023). A prediction model for childhood obesity risk using the machine learning method: a panel study on Korean children. *Scientific Reports, 13*. https://doi.org/10.1038/s41598-023-37171-4
11. Mondal, P., Foysal, K. H., Norman, B. A., & Gittner, L. (2023). Predicting Childhood Obesity Based on Single and Multiple Well-Child Visit Data Using Machine Learning Classifiers. *Sensors (Basel, Switzerland), 23*. https://doi.org/10.3390/s23020759
12. Pang, X., Forrest, C., Lê-Scherban, F., & Masino, A. (2021). Prediction of early childhood obesity with machine learning and electronic health record data. *International journal of medical informatics, 150*, 104454. https://doi.org/10.1016/j.ijmedinf.2021.104454
13. Rossman, H., Shilo, S., Barbash-Hazan, S., Artzi, N. S., Hadar, E., Balicer, R., Feldman, B., Wiznitzer, A., & Segal, E. (2021). Prediction of Childhood obesity from nationwide health records. *The Journal of Pediatrics*. https://doi.org/10.1016/j.jpeds.2021.02.010
14. Seaw, K. M., Leow, M. K., & Bi, X. (2025). Early obesity risk prediction via non-dietary lifestyle factors using machine learning approaches. *Clinical Obesity, 15*. https://doi.org/10.1111/cob.70011
15. Shinoda, G., Ishikuro, M., Matsubara, T., Noda, A., Murakami, K., Orui, M., Metoki, H., Kikuya, M., Hozawa, A., Kuriyama, S., Nakamura, K., & Obara, T. (2026). Routine Life-Course Health Records in Infancy Predict Being Overweight in Childhood and Adolescence: The TMM BirThree Cohort Study. *Children, 13*. https://doi.org/10.3390/children13030334
16. Ziauddeen, N., Wilding, S., Roderick, P., Macklon, N., Smith, D. M., Chase, D., & Alwan, N. (2020). Predicting the risk of childhood overweight and obesity at 4–5 years using population-level pregnancy and early-life healthcare data. *BMC Medicine, 18*. https://doi.org/10.1186/s12916-020-01568-z
17. Ziauddeen, N., Fraser, S. D. S., Stannard, S., Berrington, A., Chiovoloni, R., Akbari, A., Owen, R. K., & Alwan, N. (2026). Predicting childhood overweight and obesity at school entrance using healthcare, demographic, and socioeconomic data in Wales, UK. *The European Journal of Public Health, 36*. https://doi.org/10.1093/eurpub/ckag051
18. Ziauddeen, N., Roderick, P., Santorelli, G., & Alwan, N. (2023). Prediction of childhood overweight and obesity at age 10–11: findings from the Studying Lifecourse Obesity PrEdictors and the Born in Bradford cohorts. *International Journal of Obesity (2005), 47*, 1065 - 1073. https://doi.org/10.1038/s41366-023-01356-8