Proceedings of CONF-CDS 2026 Symposium: Data-Centric AI Security: Securing Models, Learning Agents, and Autonomous Systems
DOI: 10.54254/2755-2721/2026.CH33263

Balancing Interpretability and Predictive Power: A Comparative Study of Machine Learning Models for Obesity Risk Prediction

Siyu Zhu

Faculty of Arts & Science, University of Toronto, Toronto, Canada
shzsy0903@126.com

Abstract. Obesity has become a major global health issue associated with severe chronic diseases such as diabetes and cardiovascular disease, making early and accurate risk prediction crucial for public health interventions. With the rapid growth of health data, complex machine learning models are increasingly used to predict obesity risk. However, many high-accuracy models lack the interpretability required for clinical trust. This paper explores obesity risk prediction and systematically examines the trade-off between predictive accuracy and model interpretability. Therefore, the study compares a traditional parametric model, logistic regression, with a non-linear ensemble method, random forest, through a comprehensive obesity dataset featuring demographic, dietary, and physical activity variables. The paper finds that while the random forest model achieves a superior F1-score, performing better in balancing precision and recall, by capturing complex feature interactions, logistic regression provides necessary interpretability by clearly quantifying specific risk factors. Therefore, the study concludes that both predictive performance and model transparency must be simultaneously prioritized in health data analysis to develop diagnostic tools that are both highly accurate and practically actionable for medical professionals.

Keywords: Obesity Prediction, Machine Learning, Model Interpretability, Logistic Regression, Random Forest

1. Introduction

Obesity has become a critical global public health crisis, making it essential to have early risk prediction systems based on demographic and lifestyle data. In recent years, researchers have increasingly applied machine learning to clinical diagnostics to uncover hidden patterns in health records [1-3]. For instance, studies have demonstrated the exceptional accuracy of complex ensemble methods in identifying disease patterns [4]. However, a fundamental challenge in medical machine learning is the trade-off between predictive power and model transparency. Recent research highlights that highly accurate models frequently lack the interpretability required for safe, ethical, and practical clinical decision-making [5,6].

© 2026 The Authors. This is an open access article distributed under the terms of the Creative Commons Attribution License 4.0 (https://creativecommons.org/licenses/by/4.0/).

DOI: 10.54254/2755-2721/2026.CH33263

This paper explores the trade-off between model interpretability and predictive accuracy by conducting a comparative analysis of logistic regression and random forest on an obesity risk dataset. Specifically, this study evaluates how a linear, parametric model provides clear and quantifiable insights into individual lifestyle risk factors through odds ratios, while a non-linear, ensemble method uses complex feature interactions to maximize overall classification performance.

The findings of this research contribute to the field of healthcare data analytics by providing actionable insights into balancing algorithmic precision with clinical transparency. By clearly defining the operational advantages of each approach, this paper offers valuable guidance for medical practitioners in selecting appropriate machine learning tools, helping develop diagnostic systems that are both highly accurate and practically interpretable.

2. Literature review

With the development from basic statistical assessments to advanced machine learning frameworks, a significant transformation has taken place in predicting obesity risk. Early research primarily relied on Body Mass Index (BMI) and linear models, which often failed to capture the multifaceted nature of obesity driven by genetics, environment, and lifestyle. Recent studies demonstrate a pivotal shift toward high-dimensional data analysis, utilizing ensemble methods like XGBoost, LightGBM, and Random Forests [3,7]. These models have achieved remarkable predictive accuracies in classifying obesity levels by processing complex interactions between dietary habits and physical activity.

However, the latest progress in the field highlights a critical challenge. As models become more complex, their decision-making processes become increasingly opaque, limiting their utility in clinical settings where transparency is important. Consequently, the most recent frontier in health data science is the integration of Explainable AI techniques, such as SHapley Additive exPlanations (SHAP) and LIME, to decode model outputs [7-9]. Current trends also emphasize personalized medicine, where predictive tools are not only diagnostic but also designed to provide actionable, individualized intervention strategies.

Despite these advancements, a gap remains in balancing the computational efficiency of non-linear algorithms with the direct, interpretable risk quantification provided by traditional parametric models. While recent literature showcases the raw power of ensemble learning, there is an urgent need for comparative research that evaluates these algorithms against established statistical benchmarks within a clinical context. This study contributes to this evolving field by bridging the gap between high-performance predictive analytics and the interpretability required for public health trust.

3. Methodology and empirical analysis

3.1. Data description and preprocessing

This study utilizes comprehensive health data sourced from the National Health and Nutrition Examination Survey (NHANES) 2021–2023 cycle [10]. The dataset integrates multiple dimensions of patient information, specifically designed to capture the complex demographic, dietary, and physical activity variables associated with obesity.

To systematically evaluate obesity risk, the independent variables were categorized into three core domains, namely demographic and socioeconomic factors, lifestyle and behavioral metrics, and dietary and physiological indicators. Demographic and socioeconomic factors include age, gender, educational attainment, and the family monthly poverty level index. Lifestyle and behavioral metrics

DOI: 10.54254/2755-2721/2026.CH33263

capture daily sedentary time, nightly sleep duration, lifetime smoking status, and annual alcohol consumption. Dietary and physiological indicators cover total caloric intake on the first day of examination and resting pulse rate, which acts as an important proxy for cardiovascular fitness and autonomic nervous system regulation.

Data preprocessing was implemented to ensure the integrity of the machine learning models. NHANES survey data frequently contain special numerical codes indicating that a respondent "refused to answer" or "did not know." These invalid responses were systematically recoded as missing values (NA). Following this, a complete-case analysis approach was applied to remove instances with missing data across the selected features. To fulfill the mathematical prerequisites of the algorithms, all categorical variables, such as gender and smoking status, were transformed into dummy variables through one-hot encoding.

After applying these data cleaning procedures, the final analytic sample consists of 2621 individuals (1172 males and 1449 females). The primary target variable is a binary indicator for obesity risk, derived from BMI. Individuals with a BMI of 30 or higher were classified as "Yes" (N = 1073), while those below 30 were classified as "No" (N = 1548). This well-balanced distribution provides an optimal foundation for training robust machine learning classifiers.

3.2. Model constructions

This study employed two distinct machine learning algorithms to systematically evaluate the trade-off between predictive power and interpretability. Before model training, the dataset was randomly partitioned into a 70% training set to construct the models and a 30% testing set to evaluate their performance on unseen data.

As a traditional parametric model, logistic regression was constructed to provide a baseline for predictive performance while prioritizing high clinical transparency. Logistic regression predicts the probability that a patient is at risk of obesity, modeling the log-odds as a linear combination of the independent variables: $\ln\left(\frac{p}{1-p}\right) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_k X_k$, where p is the probability of obesity, X_k is the specific demographic or lifestyle features, and β_k is the estimated coefficient. To ensure the reliability of the estimated coefficients, the model's linear assumptions were analyzed, and prior diagnostic checks were conducted to prevent severe multicollinearity among the demographic and lifestyle predictors. The model parameters were optimized using maximum likelihood estimation. A critical step in this construction was addressing the model's sensitivity to multicollinearity. Before modeling, variance inflation factors were monitored to ensure that closely related socioeconomic predictors, such as educational attainment and the family poverty index, did not exhibit severe collinearity that could distort the estimated coefficients and violate the model's fundamental linear assumptions.

To capture complex, non-linear relationships that parametric models intrinsically miss, a random forest classifier was implemented. Random forest is an ensemble learning method that trains a multitude of decision trees. A primary advantage of this non-parametric method is its ability to natively handle mixed data types without requiring strict linear assumptions or feature scaling. Each decision tree in the forest splits the data by evaluating features to minimize impurity. In this study, the Gini impurity (G) was utilized as the splitting criterion: $G = 1 - \sum_{i=1}^{C} p_i^2$, where C is the number of classes, which is binary in this case, and p_i is the probability of an item being classified into a particular class. To reduce the risk of overfitting associated with complex decision trees, the random forest model was constructed using the bootstrap aggregating technique. An ensemble of 500 independent decision trees was generated, with each tree trained on a random bootstrap sample

DOI: 10.54254/2755-2721/2026.CH33263

of the dataset. The final classification for obesity risk was determined by majority voting across all 500 trees, ensuring robust generalization to unseen data. Furthermore, out-of-bag error estimation was used during training to monitor the ensemble's performance and prevent overfitting, ensuring that the hyperparameter settings yielded a robust, generalizable classifier.

4. Model comparison and results

4.1. Performance evaluation

The predictive capabilities of logistic regression and random forest were evaluated on a 30% test set by comparing their accuracy, AUC, and F1-score. Table 1 illustrates that both models demonstrated remarkably similar baseline accuracy (59.75% for logistic regression and 59.49% for random forest) and overall discriminatory power (AUC: 0.613 for logistic regression and 0.603 for random forest). However, evaluating clinical prediction models solely on accuracy can be misleading, especially when the cost of false negatives is substantial, failing to identify a patient at risk of obesity. A significant divergence emerged in the F1-score. The non-parametric random forest model achieved an F1-score of 0.388, substantially outperforming the parametric logistic regression, which only achieves 0.298. This substantial improvement in the F1-score indicates that by relaxing rigid linear assumptions, the random forest model was significantly more effective at identifying true positive obesity risks without causing a disproportionate increase in false alarms. This makes the ensemble model structurally superior for initial patient screening.

Table 1. Predictive performance of machine learning models

Model	Accuracy	AUC	F1-score
Logistic Regression	0.597	0.613	0.298
Random Forest	0.595	0.603	0.388

4.2. The interpretability trade-off and clinical implications

While the random forest model shows superior ability in capturing complex feature interactions, evidenced by its higher F1-score, this predictive power inherently sacrifices clinical transparency. Random forest provides a global ranking of feature importance, but it cannot quantify the exact directional impact of a single variable.

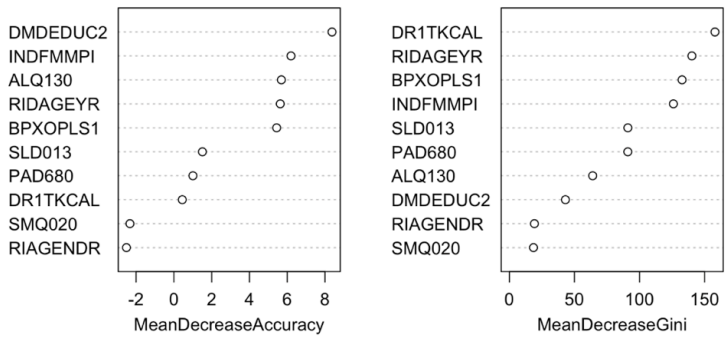


Figure 1. Random forest feature importance

DOI: 10.54254/2755-2721/2026.CH33263

Analysis of the random forest's mean decrease Gini metric (Figure 1) highlights total caloric intake (DR1TKCAL), resting pulse rate (BPXOPLS1), and age (RIDAGEYR) as the top physiological predictors of obesity. Notably, when evaluated by mean decrease accuracy, socioeconomic factors such as educational attainment (DMDEDUC2) and the family poverty income index (INDFMMPI) emerged as dominant features. This empirically validates the sociological understanding that obesity is not merely a metabolic failure, but a complex outcome heavily influenced by socioeconomic disparities and environmental constraints.

Conversely, the logistic regression model, despite a lower F1-score, offers complete algorithmic transparency. It yields actionable odds ratios that allow healthcare providers to interpret exactly how an increase in a specific factor multiplies a patient's baseline obesity risk. For example, logistic regression coefficients can directly inform a patient that reducing daily sedentary time by a specific number of hours corresponds to an exact percentage decrease in their obesity risk.

Ultimately, the choice between these models represents a fundamental trade-off. Random forests offer robust screening capabilities by maximizing diagnostic recall through complex non-linear pattern recognition. In contrast, logistic regression provides the interpretability and transparency necessary for physicians to design targeted, evidence-based behavioral interventions.

5. Conclusion

This study systematically compared logistic regression and random forest using a comprehensive lifestyle and dietary dataset to predict obesity risk. The findings clearly demonstrate the classic machine learning trade-off between predictive power and model transparency. Random forest achieved superior predictive balance, successfully capturing complex, non-linear feature interactions to deliver a substantially higher F1-score. On the other hand, logistic regression provided essential interpretability, clearly quantifying specific behavioral risk factors through straightforward odds ratios. This research highlights the critical need to balance algorithmic precision with clinical interpretability.

A primary limitation of this study is its reliance on a static, secondary dataset. Since the data represent a single cross-sectional snapshot, it lacks longitudinal tracking of patients' changing habits over time. Therefore, the models cannot capture the dynamic trajectory of weight gain, which would require primary survey data. Future research should address these gaps by incorporating real-time, longitudinal clinical data. Additionally, to improve the transparency of highly accurate models, subsequent studies should explore Explainable AI techniques to interpret the decision-making processes explicitly.

In conclusion, the future of healthcare data science lies not only in maximizing algorithmic accuracy but also in developing hybrid predictive systems that foster clinical credibility and trust. Given the widespread integration of AI into public health practice, predictive models are required to offer high precision while providing transparent, explainable reasoning to guide personalized treatment plans. With both predictive power and interpretability, medical practitioners can solve the obesity problem more effectively through these advanced analytics.

References

[1] Mamillapalli, E. K., & Sharma, N. (2025). A Micro-Macro Machine Learning Framework for Predicting Childhood Obesity Risk Using NHANES and Environmental Determinants (arXiv: 2512.22758). arXiv. https: //doi.org/10.48550/arXiv.2512.22758

[2] Riveros Perez, E., & Avella-Molano, B. (2025). Learning from the machine: Is diabetes in adults predicted by lifestyle variables? A retrospective predictive modelling study of NHANES 2007–2018. BMJ Open, 15(3),

DOI: 10.54254/2755-2721/2026.CH33263

e096595.
[3] Syahidah, H., Irsandi, N., Ajizah, A. N., & Amelia, A. (2025). Obesity prediction using machine learning algorithms. IJATIS: Indonesian Journal of Applied Technology and Innovation Science, 2(1), 53–62.
[4] DeepaLakshmi, S. (2025). Leveraging ensemble methods for robust multiclass classification in obesity prediction. IAPQR Transactions, 158.
[5] Atrey, P., Brundage, M. P., Wu, M., & Dutta, S. (2025). Demystifying the accuracy-interpretability trade-off: A case study of inferring ratings from reviews. arXiv Preprint arXiv: 2503.07914. https: //arxiv.org/abs/2503.07914
[6] Majumdar, P. (2025). The accuracy–interpretability dilemma: A strategic framework for navigating the trade-off in modern machine learning. American Journal of Information Science and Technology, 9(3), 211–224.
[7] Görmez, Y., Yagin, F. H., Yagin, B., Aygun, Y., Boke, H., Badicu, G., De Sousa Fernandes, M. S., Alkhateeb, A., Al-Rawi, M. B. A., & Aghaei, M. (2025). Prediction of obesity levels based on physical activity and eating habits with a machine learning model integrated with explainable artificial intelligence. Frontiers in Physiology, 16, 1549306.
[8] Nandan, M., Banerjee, J. S., Chakraborty, A., & Sarigiannidis, P. (2026). XAI4Obesity: Explainable AI for Obesity Risk Prediction. In S. Bhattacharyya, J. S. Banerjee, D. De, & M. Mahmud (Eds.), Intelligent Human Centered Computing (Vol. 1691, pp. 188–199). Springer Nature Singapore. https: //doi.org/10.1007/978-981-95-3671-9_17
[9] Nguemdjom, D. K. T., Mbayandjambe, A. M., Nkwimi, G. B., Oshasha, F., Muluba, C., Mbengandji, H. I., & Bazie, I. G. (2025). Explainable AI (XAI) for Obesity Prediction: An Optimized MLP Approach with SHAP Interpretability on Lifestyle and Behavioral Data. International Journal of Innovative Science and Research Technology, 3192–3200. https: //doi.org/10.38124/ijisrt/25apr1962
[10] Centers for Disease Control and Prevention (CDC), National Center for Health Statistics (NCHS). (2024). National Health and Nutrition Examination Survey Data, August 2021-August 2023. Retrieved from https: //wwwn.cdc.gov/nchs/nhanes/