Journal of Engineering Excellence
ISSN: 3108-2793 | Volume 2 Issue 1 January-March 2026
www.jeejournal.com JEE-09

TRANSPARENT AND ACCURATE OBESITY RISK CLASSIFICATION USING ENSEMBLE LEARNING AND LIME EXPLANATIONS

[*1]**RAHIMAN KHAJA HUSSAIN,** ***M.Tech Student,***
[*2]**K CHANDRA PRASAD,** ***Assistant Professor,***
Department of Computer Science & Engineering,
Srinivasa Institute of Technology &Science(Autonomous), Kadapa, AP.

ABSTRACT: The purpose of this method is to make sure that forecasts are clear and easy to use for planning healthcare, while also accurately assessing the risk of obesity. It makes use of the advantages of several classifiers, such as Random Forest, Gradient Boosting, and Logistic Regression, by employing a hybrid ensemble learning methodology. In order to address the complex nature of obesity, the methodology takes into account a wide range of clinical and lifestyle factors, such as age, BMI, eating habits, physical activity, and hereditary susceptibility. The implementation of thorough preprocessing, feature improvement, and imbalance management sustains the learning process and reduces biased results. The goal of LIME is to provide instance-level explanations that highlight the characteristics that have the biggest impact on each person's risk score. This clarifies predictions. In terms of accuracy, precision, recall, F1-score, and confusion matrices, models generally outperform individual models. Strong generalization across a range of behavioral and demographic characteristics is shown by the ensemble. It ensures that doctors can rely on the results by striking a balance between forecasting accuracy and understandability. The method combines transparency and dependability to allow for efficient, real-time obesity risk assessment. To sum up, it is a reliable instrument for making decisions in medical environments that prioritize prevention.

Keywords: *Obesity Risk Prediction, Ensemble Learning, Interpretable Machine Learning, LIME Explanations, Transparent AI, Healthcare Analytics, Explainable Artificial Intelligence (XAI)*

I. INTRODUCTION

The epidemic of obesity is quickly becoming one of the most pressing public health issues of the modern era, affecting people all over the world. It has a clear correlation to a host of long-term health problems, such as heart disease, high blood pressure, diabetes type 2, and many cancers. These problems reduce living standards and necessitate additional funding to fix. Therefore, it is critical to quickly detect individuals who are at risk of becoming obese so that they can provide individualized treatment plans and appropriate preventive services. There is a growing amount of publicly available data on health and lifestyle choices. To evaluate the risk of obesity and help doctors make informed decisions, data-driven methods are useful.

The potential for machine learning algorithms to discover intricate, nonlinear correlations between demographic, behavioural, and physiological data holds great promise for the definition of obesity risk. Traditional statistical methods are

This is an open access article under the creative commons license https://creativecommons.org/licenses/by-nc-nd/4.0/

DOI: https://doi.org/10.5281/zenodo.19233064

inadequate for unraveling the intricate webs of relationships that exist among obesity risk factors such as age, gender, food habits, exercise levels, genetic predisposition, and metabolic markers. Although more complex models, such as those that use ensemble learning, tend to make quite accurate predictions, their decision-making processes are usually not made clear. Because of this, it is more difficult to build trust and accept responsibility in therapeutic settings.

Ensemble learning methods like XGBoost, Gradient Boosting, and Random Forests use many base learners to improve the generalizability, robustness, and accuracy of predictions. Due to their ability to handle different characteristics, missing data, and nonlinear interrelationships between factors, these models excel at assessing the risk of obesity. In contrast to individual models, ensemble approaches reduce bias and variation by combining the unique decision-making styles of multiple learners. This improves the accuracy and reliability of predictions over a wide range of demographics and healthcare contexts.

In the healthcare industry, where comprehension is paramount, the opacity of ensemble models is a topic that is sometimes raised in criticism. But making predictions is what they excel at. Understanding the factors that identify an individual as high-risk is vital for clinicians and patients alike to maintain confidence, make educated decisions, and promote effective therapies. The regulatory and ethical frameworks in the healthcare sector highlight the importance of explainable artificial intelligence (XAI) when algorithmic judgments impact diagnosis, treatment planning, or preventive initiatives.

One way to make predictions that people can comprehend is by using Local Interpretable Model-Agnostic Explanations (LIME). When approximating the complicated model locally, LIME makes use of an interpretable surrogate model. This helps to highlight the most important aspects that influence a specific forecast. In clinical contexts, LIME can improve the accuracy and understandability of obesity risk classification models by showing how variables like food, lifestyle, and physiological markers affect a person's predicted risk.

Combining ensemble learning with LIME-based explanations provides a solid basis for developing accurate and explicit algorithms for obesity risk categorization. Clinicians can make better use of model outputs thanks to this paradigm, which allows for precise forecast creation while maintaining individual interpretability. By providing explanations that are relevant to clinical reasons, this combination also makes tailored risk assessment easier. This encourages systems that predict the likelihood of obesity, allow for early intervention, and foster trust to wisely apply machine learning algorithms.

II. LITERATURE SURVEY

Cervantes & Palacio (2020) The prevalence of obesity is evaluated in this research by combining AI with data on lifestyle and demographics. To decipher the intricate web of relationships between food, exercise, and body composition, we examine a number of machine learning classifiers. The results are not as precise as

This is an open access article under the creative commons license https://creativecommons.org/licenses/by-nc-nd/4.0/
DOI: https://doi.org/10.5281/zenodo.19233064

they could be using traditional statistical approaches. Behavioral components are found to be crucial when looking at feature significance. The research highlights the significance of data in identifying early indicators of obesity risk.

Singh & Tawfik (2020) Machine learning techniques are employed by the researchers to evaluate the probability that young individuals will develop obesity or overweight. Sociodemographic, lifestyle, and health-related factors are evaluated using supervised learning algorithms. Machine learning techniques outperform baseline methods in terms of predicting accuracy, according to the research. In public health, the framework encourages the use of early intervention tactics. The importance of analytics aimed at preventing obesity in adolescents is highlighted.

Lim, Xue& Wang (2020) Obesity trends and the various causes impacting different populations around the world are explored in this research. It compiles epidemiological data on lifestyle, urbanization, and dietary trends. Cost inequalities are brought to light by age and geographical disparities. Worldwide, the incidence of diseases associated with obesity is on the rise, according to the data. Findings highlight the need for predictive and preventative analytics to fully comprehend obesity, according to the research.

Thamrin et al. (2021) Machine learning methodologies are implemented in this investigation to forecast adult obesity using data from the national health survey. Considerations of demographic and behavioral characteristics are used to evaluate different classifiers. The main risks, such as unhealthy eating and lack of exercise, are accurately predicted by the model. The model verifies the procedures for screening populations. This research proves that public health datasets are good candidates for machine learning.

Ferdowsy et al. (2021) The authors use machine learning approaches to predict the likelihood of obesity by looking at behavioral and physiological characteristics. Machine learning demonstrates superior performance as compared to conventional approaches. It is widely acknowledged that calorie consumption and exercise intensity are crucial factors. Those at heightened risk can be identified more easily using this strategy. The research's main focus is on analytics that use machine learning for preventative healthcare.

Peng et al. (2021) SHAP and machine learning are combined in this work to identify modifiable predictors of adolescent obesity. Feature attributions are provided alongside each model prediction. Obesity categorization influences food and behavior, according to the findings. Clarity boosts confidence among doctors and makes treatment more feasible. In targeted therapy, the research shows that XAI has a function.

Jeon, Lee & Oh (2022) Theresearch uses machine learning approaches to predict obesity by looking at age-associated risk factors. We look at health, lifestyle, and demographic factors in people of all ages. The results demonstrate that predictors exhibit substantial variability related to age, which supports tailored risk modeling. Analyzing the value of features improves comprehension. The method advocates for personalized preventive treatments that

This is an open access article under the creative commons license https://creativecommons.org/licenses/by-nc-nd/4.0/
DOI: https://doi.org/10.5281/zenodo.19233064

meet the unique requirements of people at different points in their lives.

Lin et al. (2022) Predicting obesity using machine learning by analyzing interactions throughout the genome and epigenome is the goal of this work. Accuracy is improved by simulating complex interactions between genes and food. The research emphasizes how environmental and biological factors interact to increase the likelihood of obesity. The results show that AI can improve precision medicine. Merging genetic and lifestyle data models is the end game.

Kaur, Kumar & Gupta (2022) The authors support the use of artificial intelligence (AI) to develop customized nutrition plans and determine a person's likelihood of developing obesity. Machine learning algorithms categorize people according to their health-related characteristics and lifestyle choices. In order to lessen the risk, the system incorporates dietary suggestions. The results of the experiments indicate that the accuracy of the predictions is favorable. The use of AI to encourage behavioral changes is the primary focus of the research.

Lin et al. (2023) Machine learning algorithms that can be understood by humans have been developed to assess the likelihood of obesity in those who suffer from overweight problems. In order to evaluate various algorithms, explainability techniques are employed. According to the findings, the features work as intended and offer something new. Predictive models are made more useful in clinical situations by this strategy. Individualized monitoring and preventative actions are consistent with the outcomes.

Fernandes et al. (2023) An AI method for predicting the effectiveness of weight loss is presented in the research. Results from machine learning models that are easy to understand can help people make better health care decisions. Methods for attributing features make it easier to understand forecasts. Individualized treatment plans can be more easily created with the help of this tool. Within the context of patient-centered care, the research shows that XAI has benefits.

Choudhuri (2023) A hybrid machine learning methodology is suggested for evaluating obesity levels in this research. The use of many classifiers enhances the accuracy of predictions. Performance evaluation has shown that using different models is advantageous. The research shows that ensemble methods work. The methodology provides solid groundwork for a rigorous categorization of obesity levels.

Khater&Elhajj (2024) The authors employ explainable AI to investigate the impact of lifestyle factors on the prediction of obesity. Model-independent explanations highlight the significance of factors like physical exercise and dietary quality. Predictions made by machine learning are easier to understand. The results provide useful information for both patients and doctors. The research finds that risk modeling is more accurate when it is easy to understand.

Wang (2024) Obesity risk assessment using data on lifestyle behaviors is the focus of this research, which evaluates machine learning methods. We test the robustness and efficiency of several classifiers. Important behavioral traits that impact risk are recognized. The results

This is an open access article under the creative commons license https://creativecommons.org/licenses/by-nc-nd/4.0/

DOI: https://doi.org/10.5281/zenodo.19233064

show that screening approaches based on machine learning can work. The research lends credence to preventative healthcare strategies that are data-driven.

Özkurt (2024) Theresearch assesses obesity risk indicators using AI methods that can be understood and explained. Analyzing machine learning models reveals the leading predictions. Predictions made by computers are more reliable and easier to understand. The results provide practical advice on how to change one's way of living. Explainability is crucial in health analytics, according to this research.

Azad, Khan & El-Ghany (2025) An improved stacking-based ensemble model is suggested to include LIME explanations in this work. Regardless of the dataset, the ensemble consistently produces better predictions. LIME provides accurate predictions based on relevant, locally relevant data. There is an improvement in clinical trust and responsibility. The research proves without a reasonable doubt that obesity risk forecasts are accurate.

Görmez et al. (2025) An interpretable machine learning framework for anticipating trends in obesity is established by the authors using SHAP and LIME. The physiological and activity characteristics of the models are used to evaluate them. The most important forecasts for each person are identified using explainability techniques. A lot less work goes into understanding and keeping tabs on the clinic's operations. The results support the idea of tailored risk messaging.

Ganie, Reddy &Rege (2025) Combining ensemble learning algorithms with living data from this research's participants could help predict obesity in a wide range of groups. The ensemble enhances the robustness and accuracy of every categorization. Important aspects of lifestyle and behavior are highlighted in the explainability analysis. This approach allows for quick and accurate risk categorization. Interpretable ensemble models are shown to be useful in this research.

III. METHODOLOGY

Accurate and transparent obesity risk classification based on ensemble learning and LIME-based explainable AI has five main components: data gathering, preparation, training, understanding, and testing.

DATASET COLLECTION

Datasets including demographics, lifestyle, and physical characteristics are the backbone of the obesity research. This dataset is used to create the predictive model.

Pre-processing Stage

The raw data is suitably prepared for the model's training process using this approach.

- **Data Analysis:**The primary goal of the initial evaluation is to comprehend the data distribution and identify outliers, noise, and missing values.
- **Data Mapping:**Gender, eating habits, and levels of physical activity are examples of category qualities that can be numerically represented using encoding techniques.
- **Data Normalization:**The numerical features are normalized to a common range to reduce the influence of out-of-range values on the model's learning process.
- **Feature Selection:**Reduce dimensionality and improve prediction

This is an open access article under the creative commons license https://creativecommons.org/licenses/by-nc-nd/4.0/
DOI: https://doi.org/10.5281/zenodo.19233064

accuracy by selecting critical obesity risk factors using statistical and model-based methodologies.

Data Splitting and Balancing

After preprocessing:

- **Data Splitting:**The dataset is separated into subsets for testing and training in order to provide an objective assessment of the model's performance.
- **Data Balancing:**Training will adequately represent minority obesity classes if data balancing methods like SMOTE are used to correct class imbalances in obesity categories.

TRAINING PROCESS

This method yields a very effective model for prediction:

- **Model Selection:**Machine learning models like Support Vector Machines, Gradient Boosting, and Random Forest are used by novice learners.
- **Model Ensemble and Cross-Validation:**The goal of using ensemble methods, such as stacking and voting, is to enhance generalization by combining predictions from several models. Afterwards, k-fold cross-validation is used to assess the results.
- **Hyperparameter Tuning:**The best settings for each model to improve performance can be found via grid search or Bayesian optimization.
- **Final Training:**The final predictive model is developed by training the ensemble model with optimal outcomes on the whole training set.

EXPLAINABLE AI USING LIME

In order to guarantee the accuracy and precision of forecasts:

- **Explainable AI Framework (LIME):**The trained ensemble model uses LIME to provide contextualized explanations for each prediction.
- **Features Influencing the Decision:**Patients and doctors can better understand the model thanks to LIME, which highlights the most important factors (such physical activity, family history of obesity, and body mass index) that correspond to a certain risk category for obesity.

MODEL EVALUATION

Traditional measures of model performance are used to assess the finished product:

- **Accuracy:**The number of times a model correctly recognized all predictions is a measure of its accuracy. Although it could be misleading if the classes aren't well-balanced, it does provide a fundamental understanding of how the concept works.
- **Precision**: The accuracy of an obesity classification prediction is defined as the proportion of real cases that fit that prediction. A lower rate of false positive predictions is shown by increased precision in that category.
- **Recall:**The recall rate is a measure of the model's ability to correctly identify each actual instance of a specific kind of obesity. Reduced occurrence of forgotten cases (false negatives) is indicative of a high recall rate, which aids in the detection of health problems.
- **F1-Score:**A balanced ratio of false positives to false negatives is represented by the F1-score, which is the harmonic mean of recall and

This is an open access article under the creative commons license https://creativecommons.org/licenses/by-nc-nd/4.0/

DOI: https://doi.org/10.5281/zenodo.19233064

Journal of Engineering Excellence
ISSN: 3108-2793 | Volume 2 Issue 1 January-March 2026
www.jeejournal.com JEE-09

accuracy. Even when classes aren't distributed evenly, it provides a more reliable measure of success.

- **Confusion Matrix**: Accurate counts of positive and negative results, as well as false positives and negatives, can be seen in a confusion matrix. Using this metric, we can see how often the model misclassifies a given category and how well it does overall.

Fig 1: Flowchart illustrating the entire workflow, including data preprocessing, model training, evaluation, and explainability with LIME.

IV. RESULTS

Fig 2: Login Page

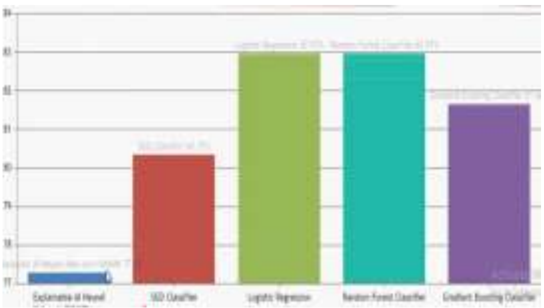

Fig 3: Model Accuracy Comparison page

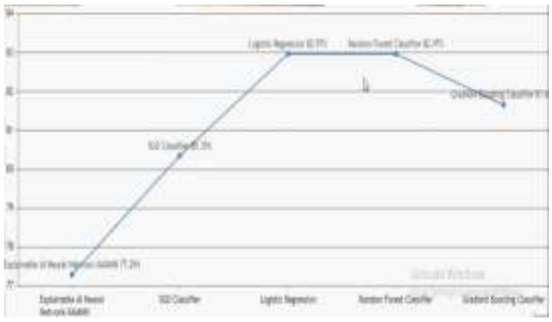

Fig 4: Classifier Accuracy Trend page

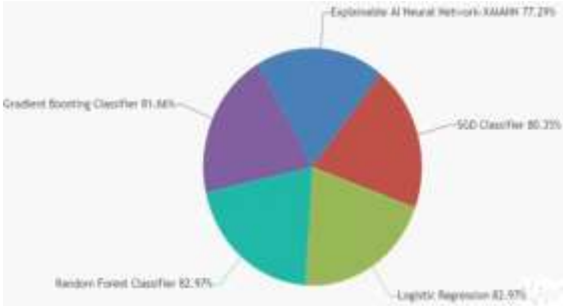

Fig 5: Classifier Accuracy Distribution.png

Fig 6: User Registration Page

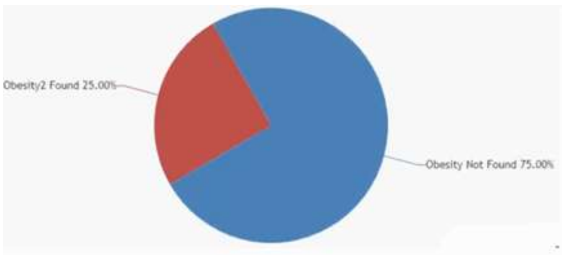

Fig 7: Obesity Prediction Pie Chart

V. CONCLUSION

An accurate evaluation of obesity risk is essential for facilitating timely, customized, and dependable healthcare interventions in the escalating worldwide obesity epidemic. By incorporating

This is an open access article under the creative commons license https://creativecommons.org/licenses/by-nc-nd/4.0/
DOI: https://doi.org/10.5281/zenodo.19233064

complicated, nonlinear interactions across demographic, physiological, and lifestyle factors, ensemble learning methods considerably enhance predictive performance. To rely on and trust a model in a therapeutic setting, people need to know how it makes decisions, not just that it's really accurate. This limitation is addressed by LIME explanations, which provide understandable, instance-specific insights into the reasoning behind a prediction.

This transparency expedites the examination of the model's results by physicians and aids individuals in understanding their personal risk factors. Improved forecasts facilitate the implementation of targeted behavioral and lifestyle modifications and enhance the efficiency of risk assessment systems. Transparent models also aid in the resolution of ethical and legal concerns regarding justice and accountability in AI-powered healthcare. The precision of machine learning risk predictions for obesity is enhanced by the combination of LIME and ensemble learning. These technologies can be advantageous to platforms and initiatives that prioritize digital health. Additionally, they facilitate the evaluation and enhancement of models with interpretable input over time.

REFERENCES

1. Cervantes, R. C., & Palacio, A. L. H. (2020). Estimation of obesity levels based on computational intelligence. Information Medicine Unlocked, 21, 100472.
2. Singh, B., & Tawfik, H. (2020). Machine learning approach for the early prediction of the risk of overweight and obesity in young people. Computational Science – ICCS 2020, Springer.
3. Lim, H. J., Xue, H., & Wang, Y. (2020). Global trends in obesity. In Handbook of Eating and Drinking: Interdisciplinary Perspectives.
4. Thamrin, S. A., Arsyad, D. S., Kuswanto, H., Lawi, A., & Nasir, S. (2021). Predicting obesity in adults using machine learning techniques: an analysis of Indonesian basic health research 2018. Frontiers in Nutrition, 8, 669155.
5. Ferdowsy, F., Rahi, K. S. A., Jabiullah, M. I., & Habib, M. T. (2021). A machine learning approach for obesity risk prediction. Current Research in Behavioral Sciences, 2, 100053.
6. Peng, Y., Xu, Z., Du, S., Hou, T., & Yan, J. (2021). SHAP-enhanced machine learning identifies modifiable obesity predictors across adolescent weight groups.
7. Jeon, J., Lee, S., & Oh, C. (2022). Age-specific risk factors for the prediction of obesity using a machine learning approach. Frontiers in Public Health, 10:998782.
8. Lin, Y-C., Christensen, J. J., Parnell, L. D., Smith, C. E., Shao, J., McKeown, N. M., et al. (2022). Using machine learning to predict obesity based on genome-wide and epigenome-wide gene–gene and gene–diet interactions. Frontiers in Genetics, 12:783845.
9. Kaur, R., Kumar, R., & Gupta, M. (2022). Predicting risk of obesity and meal planning to reduce obesity in adulthood using artificial intelligence. Endocrine, 78(3), 458–469.

This is an open access article under the creative commons license https://creativecommons.org/licenses/by-nc-nd/4.0/
DOI: https://doi.org/10.5281/zenodo.19233064

10. Lin, W., Shi, S., Huang, H., Wen, J., & Chen, G. (2023). Predicting risk of obesity in overweight adults using interpretable machine learning algorithms. Frontiers in Endocrinology, 14, 1292167.
11. Fernandes, G. J., et al. (2023). An explainable artificial intelligence software tool for early prediction of weight loss success using ML models. Journal of Medical Internet Research, 25, e42047.
12. Choudhuri, A. (2023). A Hybrid Machine Learning Model for Estimation of Obesity Levels. International Conference on Data Management, Analytics and Innovation, Springer.
13. Khater, T., &Elhajj, I. (2024). Explainable artificial intelligence for investigating the effect of lifestyle factors on obesity. Intelligent Systems with Applications, 23, 200427.
14. Wang, X. (2024). Predicting obesity risk through lifestyle habits: comparative analysis of machine learning models. E3S Web of Conferences, 385:05037.

This is an open access article under the creative commons license https://creativecommons.org/licenses/by-nc-nd/4.0/

DOI: https://doi.org/10.5281/zenodo.19233064