iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 13 **Issue:** III **Month of publication:** March 2025

DOI: https://doi.org/10.22214/ijraset.2025.67506

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 13 Issue III Mar 2025- Available at www.ijraset.com

Early Prediction of Obesity Levels Using Machine Learning

Dr. Satya Prasad Ravi[1], Yasaswini Sri Gonnabhaktula[2], Dharani Pakalapati[3], Ravi Kumar Sunkesula[4], Likhith Vital Naragani[5], Sanjay Nutalapati[6]
[1]*Professor,* [2, 3, 4, 5, 6] *Student, Department of Computer Science and Engineering, Dhanekula Institute of Engineering and Technology, Ganguru, AP, India*

Abstract: ***Obesity is a growing global health issue, closely linked to various chronic diseases. Accurate estimation of obesity levels is essential for early intervention and effective management. This project focuses on using machine learning techniques to estimate obesity levels based on diverse health and lifestyle factors, such as age, gender, physical activity, and dietary habits. The model predicts an individual's obesity level. The dataset used for this research contains various attributes related to health, nutrition, and physical activity. To prepare the data, preprocessing techniques, including label encoding for categorical features, are applied. We implement the XGBoost algorithm to predict obesity levels. Our findings demonstrate that the XGBoost algorithm achieves greater accuracy, making it the most suitable for estimating obesity levels. Model performance is evaluated using accuracy, precision, recall, and F1-score metrics. This project highlights the importance of integrating machine learning in public health initiatives, aiming to address obesity more effectively and reduce its associated health risks.***
Keywords: Obesity, Lifestyle factors, Machine Learning, XGBoost algorithm, Health management.

I. INTRODUCTION

Obesity is a chronic disease characterized by excess accumulation of body fat leading to serious health consequences, given that its major providers of displacement include heart disease, diabetes, and even some forms of cancer. It is mainly caused by an imbalance between caloric intake and energy expenditure, affected by factors such as diet, exercise, and genetic composition. According to statistics, about 16 percent of the adult population (aged 18 years and above) globally was classified as obese in 2022. Worldwide, obesity has more than doubled since 1990, making it one of the most serious public health problems of the 21st century. There is a need for identification of obesity at an early stage, which allows for timely action in preventing such severe health complications. We are using multiple lifestyle-related variables along with hereditary variables in this project to predict obesity levels of individuals using machine learning. Traditional approaches, like BMI calculations, give a rough estimation of the extent of obesity without saying a word about other determinants such as dietary habits, levels of physical activity, water intake, and genetic history. Hence, machine learning-based predictive models gain importance in the analysis of the complex patterns seen in the health data to give better obesity predictions. Early Prediction of Obesity Levels using Machine Learning is a system developed to provide a solution to this condition. This approach capitalizes on advanced techniques drawn from data to analyze multifactorial lifestyle-hereditary correlations for an accurate prediction of an individual's obesity levels. This project employs a dataset comprised of 2111 records from Kaggle, containing 17 crucial features like age, gender, dietary habits, and levels of physical activities. For the classification of obesity levels, we use the XGBoost classifying model, one of the highly efficient and powerful machine-learning algorithms for structured data and high accuracy. The predictions show personalized preventive measures from the system; this project is aimed at empowering individuals with the knowledge of actionable insights to take informed choices regarding health-related decisions and to minimize health risks arising from obesity.

II. LITERATURE REVIEW

In [1], factors related to health and lifestyle were examined as potential predictors of obesity through machine-learning techniques. It was concluded that selection of features improves accuracy and efficiency. In [2], the authors emphasize the benefits deep-learning methods have to offer in the classification of obesity over traditional methods, specifically casting emphasis on neural networks for extracting complex patterns from health-related datasets.

The role of nutrition in obesity has been well reviewed in [3], which analyzed the relationship between dietary habits and obesity prevalence. This study gives the possible associations of excessive calorie consumption and low vegetable intake with weight gain. Another view on dietary-related factors is presented in [4], where ultra-processed foods are examined for their effects on metabolic health and obesity, warranting better dietary guidelines for preventing obesity-related ailments.

Apart from diet, it has been established that lifestyle and behavior are important determinants of obesity. Studies conducted indicate that a sedentary lifestyle combined with excessive screen time and less physical exercise triggers the obesity epidemic in many people, as discussed in [5], where the correlation between modern lifestyle choices, physical inactivity, and deteriorating metabolic health is examined.

The work analyzed in [6] has gone a step further and considered public health strategies for obesity prevention, with the conclusion that early intervention, through awareness programs and lifestyle changes, would help arrest obesity rates. From a predictive analytics perspective, [7] describe contrasting machine learning models that classify obesity according to demographic and behavioral parameters. It discusses feature engineering techniques that improve model interpretability and accuracy.

The study in [8] extends this by integrating environmental variables such as urbanization and mode of transportation into the models predicting obesity. In contrast, [9] pertains to health development, integration approaches, and operationalization of AI in decision-making to improve patient outcomes through personalized recommendation systems. Lastly, regarding [10], ethical problems with obesity data analysis are discussed and highlight bias reduction in the predictive models to ensure fairness and accuracy.

III. PROPOSED SYSTEM

The Early Prediction of Obesity Levels Using Machine Learning system aims to classify people into different levels of obesity according to certain parameters such as lifestyle, diet, and family history. This is followed by a structured machine learning workflow comprising data preprocessing, training, evaluation, and prediction for the best possible results, as well as effective preventive recommendations.

This data set retrieved from Kaggle carries 2111 records comprising of 17 major features which include age, gender, diet, physical activity, water intake, and family history of being overweight. From this, preprocessing technique such as filling missing values, encoding categorical variables, and normalizing numerical features will be involved to ensure that the model receives quality data.

When the preprocessing is complete, it will split into training and testing sets, and the model will then be trained using the XGBoost classifier, a very powerful machine-learning algorithm that proves efficient for structured data. Performance metrics like accuracy, precision, recall, and F1-score will help evaluate the model's performance and ensure reliability.

For prediction, users will interact with the system via an interface designed specifically for entry of personal and lifestyle information; then, the input will be run against the trained model to provide an obesity level prediction on preset patterns learned. After classifying the obesity level, the system offers personalized preventive measures associated with the obesity threshold, such as dietary modification, increased physical activity, and increased water intake, helping users make informed health decisions. In effect, with the help of machine learning, the system detects obesity at an early stage and encourages proactive healthcare in life, therefore minimizing the risk of various obesity-associated diseases such as diabetes, cardiovascular, or hypertension.

A. Architecture

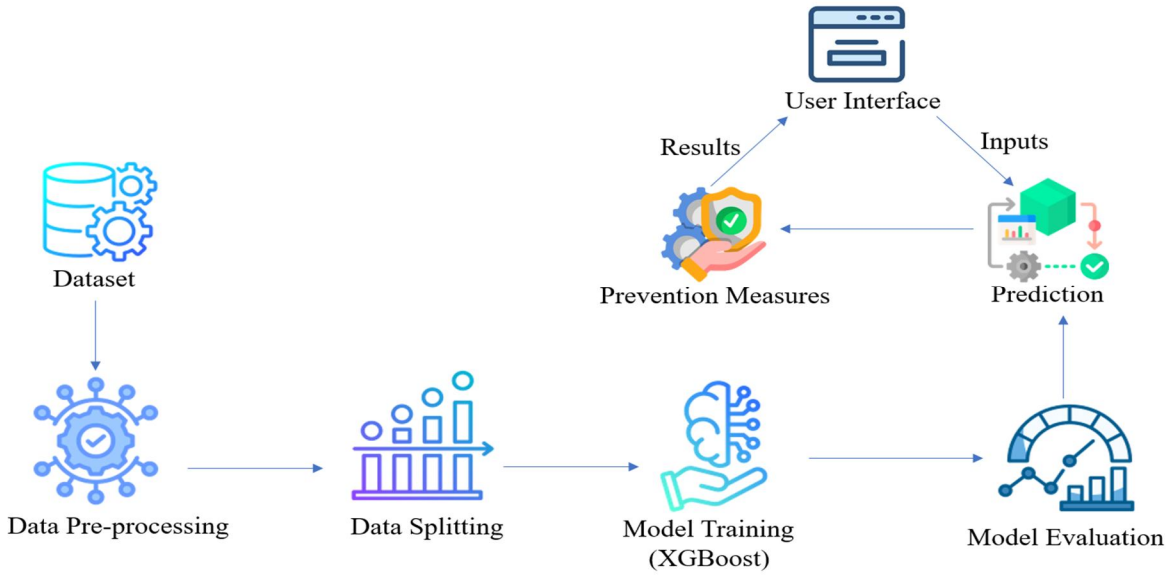


Fig. 1 Obesity Prediction System Architecture

B. Working of the System

The entire system is based on a well-structured workflow that involves data preprocessing, followed by data splitting, model training with XGBoost, performance evaluation, and finally prediction through a user interface. The end output includes obesity level classification along with a few prevention measures based on the user input.

1) Dataset Collection

The proposed system predicts obesity levels through machine learning based on a mixture of lifestyle, dietary, and hereditary factors. The source of the dataset used in the project is Kaggle, which consists of 2111 records with 17 essential features affecting the obesity levels. These features include age, gender, height, weight, dietary habits (caloric intake, vegetable consumption, meal frequency), lifestyle factors (physical activity, smoking, alcohol consumption, water intake, screen time, transportation mode), and hereditary factors (family history of overweight).

2) Data Preprocessing

Before training the model, the data set is put through a pre-processing phase where data is treated in terms of cleaning and preparation for machine learning. For any missing values, a proper solution is sought-whether assigning appropriate values via imputation or an overall delete of the missing value. Categorical variables such as gender and transportation mode are encoded into numerical values so that these values can be processed by the model.

3) Splitting the Dataset

After preprocessing, the dataset is divided into training and test sets, usually in a ratio of 80:20. The training set (80%) is utilized to train the model patterns and relationships between features, and the test set (20%) is utilized to test how well the model generalizes to new data.

4) Model Training Using XGBoost Classifier

The system uses the XGBoost Classifier, which is a very popular and highly efficient machine learning algorithm used for the classification of structured data. The functioning of XGBoost involves making many decision trees in sequence, each new one correcting the mistakes made by the previous ones. The iterative process of boosting minimizes loss and improves the accuracy of prediction substantially.

5) Model Evaluation

The model is evaluated using various performance metrics such as accuracy, precision, recall and F1-score to make sure that the forecasts thus rendered are accurate and powerful enough.

a) Accuracy (96.92%): Measures the proportion of correctly predicted instances out of total instances.
b) Precision (96.93%): Assesses the correctness of predicted obesity levels by calculating the ratio of true positives to total predicted positives.
c) Recall (97.00%): It examines the ability of the model in correctly identifying all actual cases of obesity.
d) F1-Score (96.95%): Provides a balance between precision and recall for overall performance.

The supported confusion matrix visually represents correct and incorrect classifications and validates the reliability of the model.

Classification Report:

	precision	recall	f1-score	support
0	0.95	1.00	0.97	56
1	0.95	0.92	0.93	62
2	0.97	0.96	0.97	78
3	0.97	0.98	0.97	58
4	1.00	1.00	1.00	63
5	0.95	0.95	0.95	56
6	1.00	0.98	0.99	50
accuracy			0.97	423
macro avg	0.97	0.97	0.97	423
weighted avg	0.97	0.97	0.97	423

Fig. 2 Classification Report

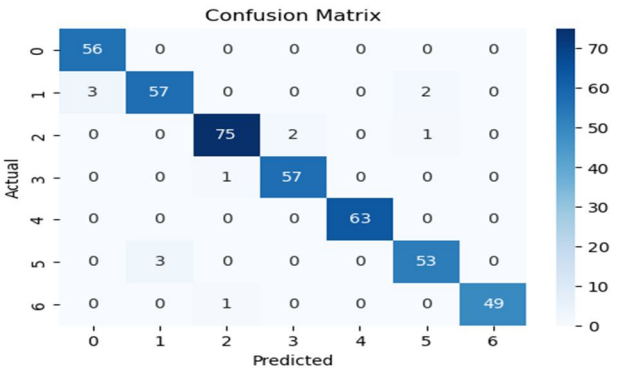


Fig. 2 Confusion Matrix

6) Prediction and Classification

A user-friendly interface is provided for users to insert the input for prediction. That is a model trained with XGBoost to predict the obesity level of the user. On the prediction basis, the system gives prevention measures as encouragement for taking active steps into a healthy lifestyle.

IV. ADVANTAGES OF THE PROPOSED SYSTEM

1) Early Detection of Obesity: The system predicts obesity levels at an early stage, allowing individuals to take preventive measures before severe health issues arise.
2) High Accuracy with XGBoost: Much effectiveness and efficiency in obesity level classification using XGBoost classifier than the other traditional approaches.
3) Data-Driven Decision Making: Apart from using BMI alone, the system includes different lifestyle factors, which make the prediction comprehensive and reliable.
4) User-Friendly: The system simplifies obesity classification, making it accessible to users with minimal technical knowledge.
5) Scalability and Flexibility: The model can be expanded to include additional health parameters and can be adapted for different population groups.

V. FUTURE SCOPE

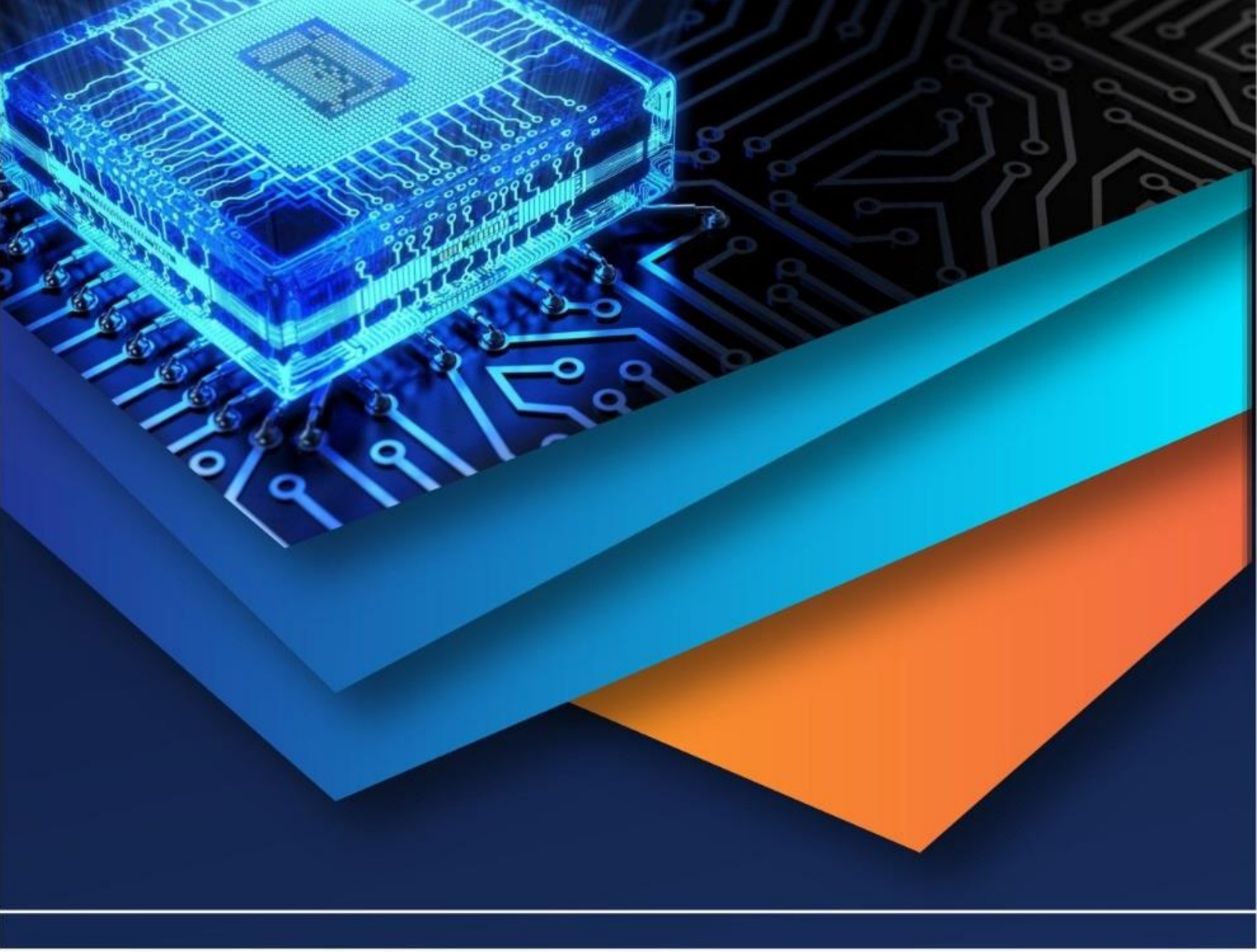
For future enhancements, the system can involve real-time monitoring of health and merging of intelligent devices with wearables and AI-driven personalized recommendations. The system can be extended to larger datasets, multilingual support, and mobile applications, thus enhancing reach. Its incorporation with telemedicine platforms will improve proactive preventive healthcare and increase the precision of obesity prediction and accessibility.

VI. CONCLUSION

The Early Prediction of Obesity Levels Using Machine Learning system provides an efficient and reliable way to identify the obesity levels of an individual by classifying multiple lifestyle and health factors. We used the XGBoost algorithm to train the model for prediction, which achieved an accuracy of 96.92% that outperforms the traditional BMI-based assessments. The user-friendly interface makes it easy for individuals to check their obesity risk and take early preventive action. This system can go further by expanding into wearable devices and telemedicine, which would improve proactive healthcare and contribute immensely to preventing obesity.

REFERENCES

[1] IEEE Xplore, “Comparison Of Different Machine Learning Methods Applied To Obesity Classification”, [Cross Ref]
[2] ACM Digital Library, “Application of Machine Learning Techniques to Predict Teenage Obesity Using Earlier Childhood Measurements from Millennium Cohort Study”, [Cross Ref]
[3] Frontiers, “Predicting Obesity in Adults Using Machine Learning Techniques: An Analysis of Indonesian Basic Health Research 2018”, [Cross Ref]
[4] ELSEVIER, “A machine learning approach for obesity risk prediction”, [Cross Ref]
[5] IEEE Xplore, “Understanding Early Childhood Obesity via Interpretation of Machine Learning Model Predictions”, [Cross Ref]
[6] Frontiers, “Age-specific risk factors for the prediction of obesity using a machine learning approach”, [Cross Ref]
[7] IJAAI, “Obesity prediction using machine learning techniques”, [Cross Ref]
[8] MDPI, “Does Physical Activity Predict Obesity—A Machine Learning and Statistical Method-Based Analysis”, [Cross Ref]
[9] ELSEVIER, “Prediction of early childhood obesity with machine learning and electronic health record data”, [Cross Ref]
[10] MDPI, “Predicting Childhood Obesity Using Machine Learning: Practical Considerations”, [Cross Ref]



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY