Journal of Science Engineering Technology and Management Science
Volume 02, Issue 04, April 2025 *ISSN: 3049-0952*
www.jsetms.com *DOI:10.63590/jsetms.2025.v02.i04.57-63*

PREDICTING OBESITY TRENDS: THE ROLE OF DIETARY HABITS THROUGH MACHINE LEARNING MODELS

Sk. Asiff, V. Bharathi, D. Ramesh, B. Poojitha
Department of Computer Science and Engineering (AI&ML), Geethanjali Institute of Science and Technology, Nellore, Andhra Pradesh, India.

To Cite this Article
Sk. Asiff, V. Bharathi, D. Ramesh, B. Poojitha,“ Predicting Obesity Trends: The Role Of Dietary Habits Through Machine Learning Models”, Journal of Science Engineering Technology and Management Science, Vol. 02, Issue 04, April 2025,pp: 57-63, DOI: http://doi.org/10.63590/jsetms.2025.v02.i04.pp57-63
Submitted: 09-03-2025 *Accepted: 18-04-2025* *Published: 26-04-2025*

ABSTRACT
Obesity is a growing global health concern, influenced by a complex interplay of genetic, environmental, and lifestyle factors. Among these, dietary habits are a key contributor, making it essential to understand how eating behaviors impact obesity levels. This research investigates the use of machine learning (ML) techniques to predict obesity based on individuals' dietary patterns and demographic characteristics. A comprehensive dataset containing information on age, gender, eating frequency, food type, physical activity, and obesity levels is utilized. Several machine learning algorithms, including Decision Trees, Support Vector Machines (SVM), Random Forests are implemented to develop predictive models. Feature selection techniques are applied to identify the most influential variables affecting obesity. Model performance is evaluated using standard classification metrics such as accuracy, precision, recall, and F1-score. The results demonstrate that ML algorithms can effectively predict obesity levels, with some models outperforming others in terms of accuracy and generalization. Notably, feature importance analysis reveals specific dietary behaviors and food groups—such as high consumption of fast food, sugary beverages, and low physical activity levels—as strong predictors of obesity. This research underscores the potential of machine learning in public health by offering data-driven insights into the dietary factors contributing to obesity. The findings can inform the development of personalized nutrition recommendations, targeted health interventions, and evidence-based public health policies. Ultimately, this work contributes to the advancement of predictive healthcare analytics and supports efforts to combat the global obesity epidemic through informed decision-making and lifestyle modifications.

Keywords: Obesity, Machine Learning, Dietary habits, Extra Trees Classifier, Health.

1.INTRODUCTION

Obesity has become one of the most urgent global health issues of the 21st century, with its prevalence rising at an alarming rate across both developed and developing nations. Defined as an abnormal or excessive accumulation of body fat that presents a risk to health, obesity is more than a cosmetic concern—it is a serious medical condition that significantly increases the likelihood of life-threatening diseases such as type 2 diabetes, cardiovascular disorders, hypertension, sleep apnea, certain cancers, and musculoskeletal problems. According to the World Health Organization (WHO), worldwide obesity has nearly tripled since 1975, and it continues to affect millions of adults and children globally.

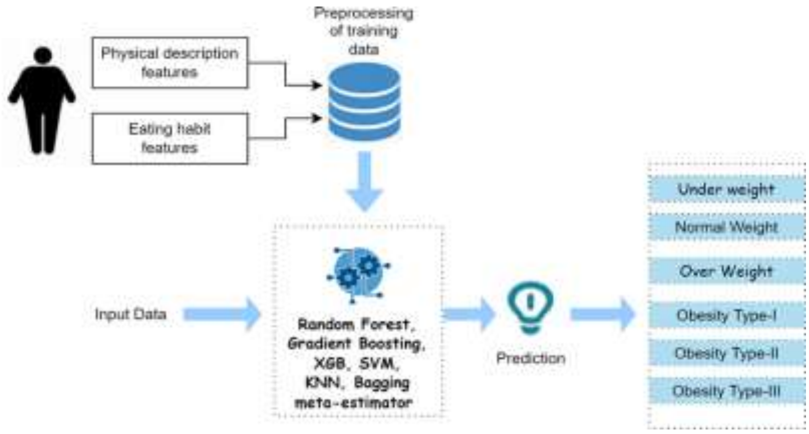

Fig. 1: Obesity prediction ML model workflow.

The factors contributing to obesity are diverse and complex, encompassing a wide range of biological, behavioral, environmental, social, and cultural elements. While genetics and hormonal imbalances may predispose individuals to obesity, the most influential and modifiable factors are dietary habits, physical inactivity, sedentary lifestyles, and poor nutrition choices. The global shift toward energy-dense, high-fat, and high-sugar foods—combined with reduced physical activity due to increasingly automated and screen-oriented lives—has created an environment conducive to rapid weight gain and chronic obesity.

The growing concern over obesity has prompted researchers, clinicians, and policymakers to seek more effective strategies for prevention, early detection, and management. Traditional approaches to studying obesity, such as clinical trials and population-based surveys, though valuable, often fall short in capturing the complex interactions among multiple lifestyle and health variables. In this context, the emergence of data science and machine learning has opened new avenues for analyzing health data more efficiently and accurately. These technologies offer the ability to detect hidden patterns, predict health outcomes, and personalize interventions based on individual behaviors and risk profiles. By integrating demographic, behavioral, and nutritional data, modern computational models can aid in understanding how specific factors influence obesity. This can support healthcare providers in identifying at-risk individuals earlier and tailoring prevention strategies accordingly. Furthermore, such predictive tools hold promise in the fields of public health, clinical decision-making, and personalized wellness, helping to combat the obesity epidemic through informed, data-driven action.

2.LITERATURE SURVEY

S. Maria [1] proposed that approximately about two billion peoples are affected by obesity that has drawn significant attention on social media. As the sedentary lifestyle which includes consumption of junk foods, no physical activities,spending more on screen,etc are one of the causes of obesity.Obesity generally refers to that a person's body possessing an excessive amount of fat.There is a huge increase in obesity cases which resulting cardiac problems,stroke,insomnia, breathing problems,etc.Type-2 diabetes has been detected in the patients suffering from obesity recently. The studies showing that there are lot of young individuals and children's who has been suffering from overweight and obesity issues in Bangladesh. Here, a strategy for predicting the risk of obesity is proposed that makes use of various machine learning methods. The dataset Obesity and Lifestyle taken from Kaggle site which is collection of different data based on the eating habits and physical conditions,such as height, weight,calorie intake,physical activities are just a few of the 17 different categories in the dataset that reflect the elements that cause obesity. Several machine learning methods include Gradient Boosting Classifier, Adaptive Boosting (ADA boosting), K-nearest Neighbor (K-NN), Support Vector Machine (SVM), Random Forest, and Decision Tree.

T. Cui [2] proposed in recent decades, there has been increasing concern about obesity in adolescents and adults. Obesity can cause many physical health problems and affect people's quality of life. So people are starting to look at the factors that lead to obesity and predict the emergence of obesity. This research presents an estimation of obesity levels based on eating habits, physical condition, and other factors, using a dataset found on UCI Machine Learning Repository. This dataset contains 17 attributes and 2111 records. The labels of this dataset are classified as Insufficient Weight, Normal Weight, Overweight Level I, Overweight Level II, Obesity Type I, Obesity Type II and Obesity Type III. In this research, three major methods are chosen for prediction: Decision Trees, Logistic Regression, and K Nearest Neighbor. Finally, the result obtained by Decision Trees has the best accuracy.

N. P. Sable [3] encountered More than 2.1 billion people worldwide are shuddering from overweightness or obesity, which represents approximately 30% of the world's population. Obesity is a serious global health problem. By 2030, 41% of people will likely be overweight or obese, if the current trend continues. People who show indications of weight increase or obesity run the danger of contracting life-threatening conditions including type 2 diabetes, respiratory issues, heart disease, and stroke. Some intervention strategies, like regular exercise and a balanced diet, might be essential to preserving a healthy lifestyle. Thus, it is crucial to identify obesity as soon as feasible. We have collected data from sources like schools and colleges within our organization to create our dataset. A vast range of ages is considered and the BMI value is examined in order to determine the level of obesity. The dataset of people with normal BMI and those at risk has an inherent imbalance. The outcomes are collected and showcased via a website which also includes various preventive measures and calculators. The outcomes are promising, and clock an accuracy of about 90%

Singh, B [4] observed individuals developing signs of weight gain or obesity are also at a risk of developing serious illnesses such as type 2 diabetes, respiratory problems, heart disease and stroke. Some intervention measures such as physical activity and healthy eating can be a fundamental component to maintain a healthy lifestyle. Therefore, it is absolutely essential to detect childhood obesity as early as possible. This paper utilises the vast amount of data available via UK's millennium cohort study in order to construct a machine learning driven model to predict young people at the risk of becoming overweight or obese. The childhood BMI values from the ages 3, 5, 7 and 11 are used to predict adolescents of age 14 at the risk of becoming overweight or obese. There is an inherent imbalance in the dataset of individuals with normal BMI and the ones at risk. The results obtained are encouraging and a prediction accuracy of over 90% for the target class has been achieved. Various issues relating to data preprocessing and prediction accuracy are addressed and discussed.

Cheng [5]. used 11 classification algorithms (logistic regression, radial basis function (RBF), naïve Bayes, classification via regression (CVR), local k-nearest neighbors (k-NN), a decision table, random subspace, random tree, a multi-objective evolutionary fuzzy classifier, and a multilayer perceptron) to predict obesity in adults and achieved a highest overall accuracy of 70% with a random subspace algorithm [5]

Cervantes. [6]. developed decision tree (DT), k-means, and support vector machine (SVM)-based data mining techniques to identify obesity levels among young adults between 18 and 25 years of age so that interventions could be undertaken to maintain a healthier lifestyle in the future [6]. Gupta [7]. developed a deep learning model (long short-term memory (LSTM)), which predicted obesity between 3 and 20 years of age with 80% accuracy using unaugmented electronic health record (EHR) data from 1 to 3 years prior [7]. Marcos-Pasero [8] used random forest (RF) and gradient boosting to predict the BMI from 190 multidomain variables (data collected from 221 children aged 6 to 9 years) and determined the relative importance of the predictors [8]

Zare [9]. used kindergarten-level BMI information, demographic, socioeconomic information such as family income, poverty level, race, ethnic compositing, housing, parent education, and family structure to predict obesity at the fourth grade and achieved an accuracy of about 87% by using logistic regression and an artificial neural network. Zare [9] used kindergarten-level BMI information, demographic, socioeconomic information such as family income, poverty level, race, ethnic compositing, housing, parent education, and family structure to predict obesity at the fourth grade and achieved an accuracy of about 87% by using logistic regression and an artificial neural network. Fu,et.al [10]. developed an ML-based framework to predict childhood obesity by using health examination, lifestyle and dietary habits, and anthropometric measurement-related data [10]. Pang [11]. proposed ML models to predict childhood obesity from EHR data [11].

3.PROPOSED METHODOLOGY

The system architecture is designed as a structured pipeline encompassing data collection, preprocessing, model training, and evaluation as shown in Fig.2. It integrates various machine learning algorithms to analyze dietary and demographic data for predicting obesity levels.

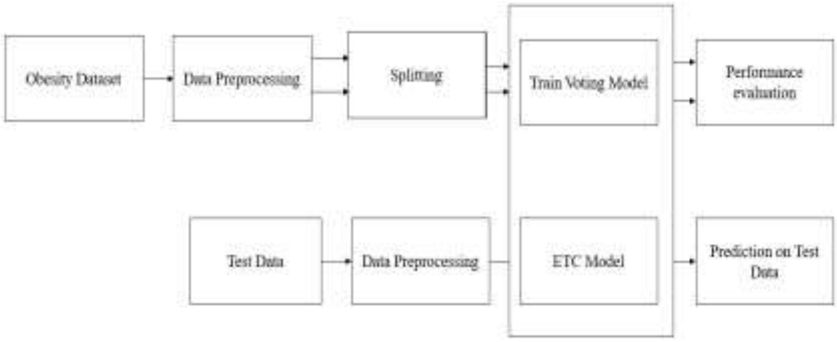

Fig. 2: Block diagram of proposed system architecture of prediction of obesity levels.

3.1 Extremely Randomized Trees

Extremely Randomized Trees, also known as Extra Trees, construct multiple trees like RF algorithms during training time over the entire dataset. During training, the ET will construct trees over every observation in the dataset but with different subsets of features as shown in Fig.3.

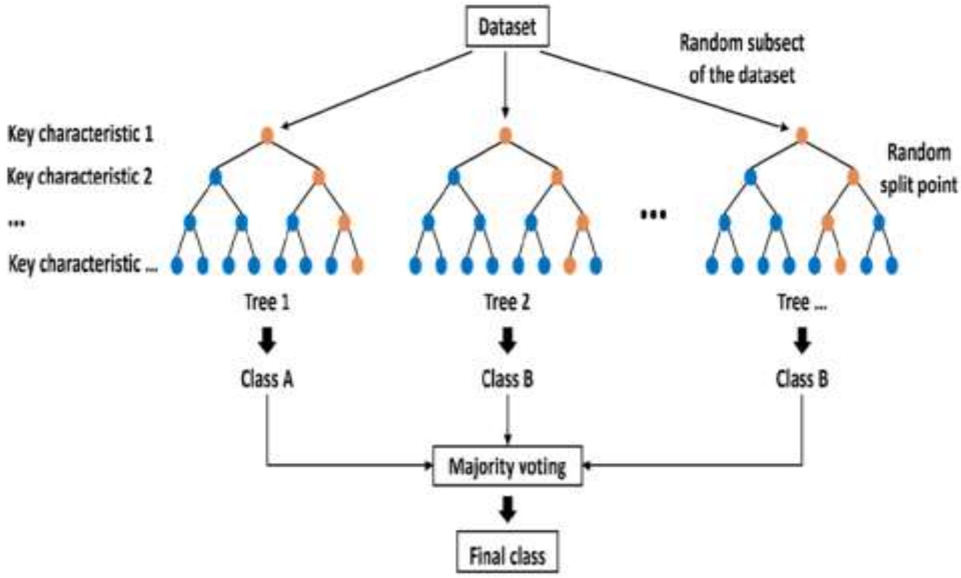

Fig. 3: Diagram of Extremely Randomized Tree Classifier.

It is important to note that although bootstrapping is not implemented in ET's original structure, we can add it in some implementations. Furthermore, when constructing each decision tree, the ET algorithm splits nodes randomly. The main advantage of Extra Trees is the reduction in bias. This is in terms of sampling from the entire dataset during the construction of the trees. Different subsets of the data may introduce different biases in the results obtained, hence Extra Trees prevents this by sampling the entire dataset.

Another advantage of Extra Trees is that they reduce variance. This is a result of the randomized splitting of nodes within the decision trees, hence the algorithm is not heavily influenced by certain features or patterns in the dataset.

4.RESULTS AND DISCUSSION

The research provides a comprehensive overview of the entire machine learning pipeline, including data loading, preprocessing, model training, evaluation, and predictions. It uses various classifiers and ensemble methods to predict obesity levels based on eating habits. The results are presented through accuracy scores, classification reports, and confusion matrices.

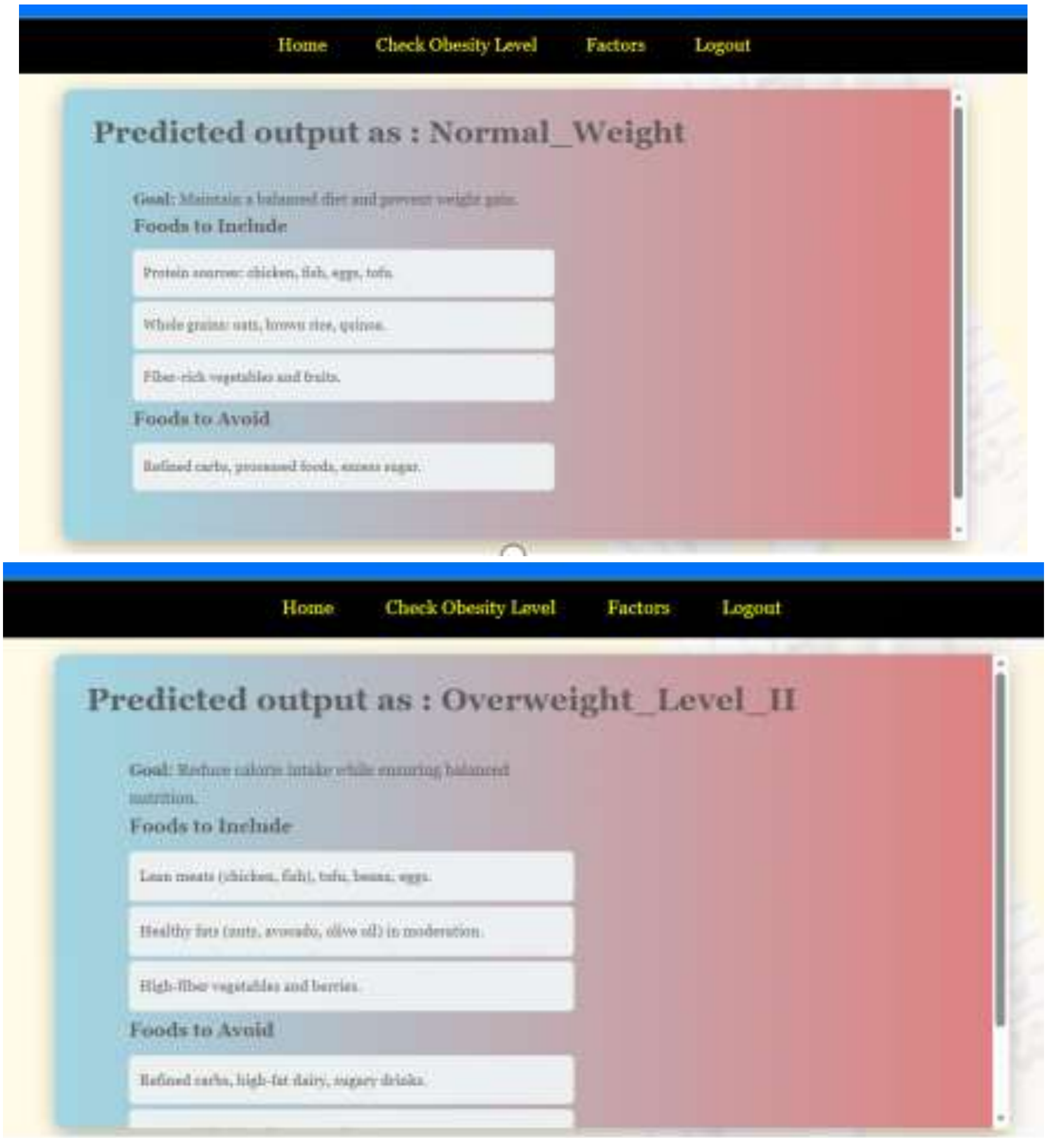

Fig. 4: Prediction on test dataset.

Fig. 4 show the output of an obesity prediction system that provides personalized dietary recommendations based on the predicted obesity level. In the first case, the system predicts Normal Weight and advises maintaining a balanced diet with protein sources, whole grains, and fiber-rich foods while avoiding refined carbs and processed items. In the second case, the prediction is Overweight Level II, where the system suggests reducing calorie intake by including lean meats, healthy fats, and high-fiber vegetables, and avoiding high-fat dairy, sugary drinks, and refined carbs.

Table. 1: Performance comparison of existing and proposed ML models.

Model	Accuracy (%)	Precision	Recall	F1-score
Voting Classifier	0.93	0.93	0.93	0.93
ETC model	0.95	0.95	0.95	0.95

In Table. 1, The performance evaluation of the models shows that the Voting Classifier achieved an accuracy of 0.93, indicating a strong match between actual and predicted values. Its precision, recall, and F1-score are also 0.93, reflecting a balanced and consistent performance across all evaluation metrics. In comparison, the Extra Tree Classifier performed slightly better, with an accuracy, precision, recall, and F1-score all at 0.95, demonstrating a higher overall predictive capability and reliability in classifying obesity levels based on the dataset.

5.CONCLUSION

In conclusion, this research successfully explored the connection between eating habits and obesity through the application of machine learning techniques, offering a data-driven approach to predicting obesity levels. By leveraging a comprehensive dataset containing a wide range of demographic, behavioral, and dietary variables, multiple predictive models—such as Decision Trees, Support Vector Machines (SVM), and Random Forests—were developed and evaluated. The models demonstrated strong performance in classifying individuals according to obesity levels, with several algorithms achieving high accuracy, precision, recall, and F1-scores. These results indicate the capability of machine learning not only to make reliable predictions but also to generalize effectively across diverse data samples.

A key strength of this study lies in its use of feature selection methods to identify the most significant dietary and lifestyle factors influencing obesity. This not only improved the efficiency and interpretability of the models but also provided deeper insights into the underlying behaviors associated with excessive weight gain. Patterns such as meal frequency, snack consumption, physical activity, and water intake emerged as critical predictors, suggesting areas where targeted interventions and behavior modification strategies could be most effective.

The findings of this work have practical implications for both personalized health management and broader public health policies. By identifying high-risk individuals early and providing insights into their contributing lifestyle factors, such systems could inform tailored nutritional guidance and preventive programs. Ultimately, this study highlights the transformative potential of machine learning in the field of healthcare analytics, paving the way for more intelligent, proactive, and personalized solutions to address the growing obesity epidemic.

REFERENCES

[1]. S. Maria, R. Sunder and R. S. Kumar, "Obesity Risk Prediction Using Machine Learning Approach," 2023 International Conference on Networking and Communications (ICNWC), Chennai, India, 2023, pp. 1-7, doi: 10.1109/ICNWC57852.2023.10127434.

[2]. T. Cui, Y. Chen, J. Wang, H. Deng and Y. Huang, "Estimation of Obesity Levels Based on Decision Trees," 2021 International Symposium on Artificial Intelligence and its Application on Media (ISAIAM), Xi'an, China, 2021, pp. 160-165, doi: 10.1109/ISAIAM53259.2021.00041.

[3]. N. P. Sable, R. Bhimanpallewar, R. Mehta, S. Shaikh, A. Indani and S. Jadhav, "A Machine Learning approach for Early Detection and Prevention of Obesity and Overweight," 2023 IEEE 8th International Conference for Convergence in Technology (I2CT), Lonavla, India, 2023, pp. 1-5, doi: 10.1109/I2CT57861.2023.10126346.

[4]. Singh, B., Tawfik, H. (2020). Machine Learning Approach for the Early Prediction of the Risk of Overweight and Obesity in Young People. In: Krzhizhanovskaya, V.V., et al. Computational

Science – ICCS 2020. ICCS 2020. Lecture Notes in Computer Science(), vol 12140. Springer, Cham.

[5]. Cheng, X.; Lin, S.-y.; Liu, J.; Liu, S.; Zhang, J.; Nie, P.; Fuemmeler, B.F.; Wang, Y.; Xue, H. Does physical activity predict obesity—A machine learning and statistical method-based analysis. Int. J. Environ. Res. Public Health 2021, 18, 3966

[6]. Cervantes, R.C.; Palacio, U.M. Estimation of obesity levels based on computational intelligence. Inform. Med. Unlocked 2020, 21, 100472.

[7]. Gupta, M.; Phan, T.-L.T.; Bunnell, H.T.; Beheshti, R. Obesity Prediction with EHR Data: A deep learning approach with interpretable elements. ACM Trans. Comput. Healthc. (HEALTH) 2022, 3, 1–19.

[8]. Marcos-Pasero, H.; Colmenarejo, G.; Aguilar-Aguilar, E.; Ramírez de Molina, A.; Reglero, G.; Loria-Kohen, V. Ranking of a wide multidomain set of predictor variables of children obesity by machine learning variable importance techniques. Sci. Rep. 2021, 11, 1910

[9]. Zare, S.; Thomsen, M.R.; Nayga Jr, R.M.; Goudie, A. Use of machine learning to determine the information value of a BMI screening program. Am. J. Prev. Med. 2021, 60, 425–433

[10]. Fu, Y.; Gou, W.; Hu, W.; Mao, Y.; Tian, Y.; Liang, X.; Guan, Y.; Huang, T.; Li, K.; Guo, X. Integration of an interpretable machine learning algorithm to identify early life risk factors of childhood obesity among preterm infants: A prospective birth cohort. BMC Med. 2020, 18, 184

[11]. Pang, X.; Forrest, C.B.; Lê-Scherban, F.; Masino, A.J. Prediction of early childhood obesity with machine learning and electronic health record data. Int. J. Med. Inform. 2021, 150, 104454.