2026 Volume: 02 Issue: 03 www.ijrpa.com ISSN 2456-9995 Research Article

International Journal Research Publication Analysis

Page: 01-22

A MACHINE LEARNING FRAMEWORK FOR ASSESSING OBESITY RISK BASED ON BEHAVIORAL FACTORS

***[1]Nara VenkatKishore, [2]Depuru Murali**

[1]M.Tech Scholar, [2]Assitant Professor

[12]Department of Computer Science and Engineering

[12]QUBA College of Engineering and Technology, NH-16, Venkatachalam, SPSR Nellore (Dt), AP.

Article Received: 5 February 2026
Article Revised: 25 February 2026
Published on: 18 March 2026

***Corresponding Author: Nara VenkatKishore**
M.Tech Scholar, Department of Computer Science and Engineering, 12QUBA College of Engineering and Technology, NH-16, Venkatachalam, SPSR Nellore (Dt), AP.

DOI: https://doi-doi.org/101555/ijrpa.3324

ABSTRACT:

Obesity has emerged as a global health crisis, traditionally diagnosed via Body Mass Index (BMI). However, BMI often fails to account for the behavioral precursors that drive weight gain. This paper proposes a Machine Learning (ML) framework, B-ORAF, to predict obesity risk by analysing lifestyle and behavioral inputs. Utilizing a dataset of 2,111 individuals, we compared Random Forest (RF), Support Vector Machines (SVM), and Extreme Gradient Boosting (XGBoost). Our results demonstrate that XGBoost achieves a superior accuracy of 95.56%. By identifying sedentary technology use and dietary frequency as primary risk markers, this framework provides a robust tool for early clinical intervention.

KEYWORDS*: Machine Learning (ML), XGBoost Classifier, Obesity Prediction, B-ORAF Framework, Behavioral Analytics, Predictive Healthcare, Lifestyle Risk Factors.*

1. INTRODUCTION

Obesity is a serious health problem that occurs when a person has an excessive amount of body fat. It significantly increases the risk of various diseases such as type 2 diabetes, cardiovascular diseases, stroke, and high blood pressure. In some cases, it can also contribute to mental health issues like depression and low self-esteem.

Obesity levels are commonly measured using the Body Mass Index (BMI), a simple index calculated from a person's height and weight. BMI helps classify individuals into categories such as underweight, normal weight, overweight, and obese. These classifications assist healthcare providers in determining the level of health risk associated with a person’s weight.

There are several contributing factors to obesity. These include unhealthy eating habits, such as consuming high-calorie foods, excessive intake of junk food and sugary drinks, lack of physical activity, and genetic predisposition. Additionally, stress, inadequate sleep, and certain medical conditions or medications can also play a role in weight gain.

Understanding obesity levels is important because it allows for early intervention. By recognizing the risk at an early stage, individuals can take proactive steps such as adopting a healthier diet, increasing physical activity, and seeking medical guidance. These actions not only help in managing weight but also in preventing the development of more serious health complications in the future.

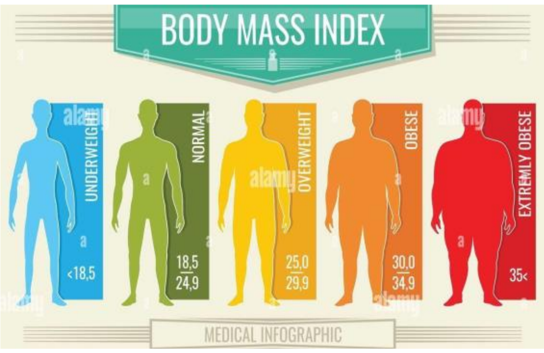

Fig. 1 Obesity Levels.

1.1 Identification of Obesity Parameters

1. Gender: Biological sex of the individual (Male/Female)

- Affects body fat distribution and metabolism.
- Hormonal differences influence obesity risk.

2. Age: Age in years

- Older age often leads to slower metabolism.
- Obesity risk increases with age due to decreased physical activity.

3. Height: Height of the person in meters

- Used with weight to calculate BMI.
- Essential for classifying obesity levels.

4. Weight: Body weight in kilograms

- Direct measure related to obesity
- Helps determine BMI and body composition.

5. Family History of Overweight: Indicates if close family members are overweight (Yes/No)

- Reflects genetic predisposition to obesity.
- May also imply shared eating habits or lifestyle.

6. FAVC (Frequent Consumption of High Caloric Food): Whether the individual frequently consumes high-calorie foods (Yes/No)

- High-calorie intake contributes to fat storage.
- Common cause of weight gain.

7. FCVC (Frequency of Consumption of Vegetables): How often vegetables are consumed (scale 1-3)

- Vegetables aid in weight control.
- Higher frequency is associated with healthier diets.

8. NCP (Number of Main Meals Per Day): Average number of main meals consumed daily.

- Regular meals help regulate metabolism.
- Too many or too few meals can disrupt energy balance.

9. CAEC (Consumption of Food Between Meals): Frequency of snacking between meals

- Frequent snacking may lead to overeating.
- Type of snacks consumed affects obesity risk.

10. SMOKE: Indicates if the person smokes (Yes/No)

- Smoking affects appetite and metabolism.
- Quitting smoking may lead to weight gain in some people.

11. CH_2O (Daily Water Intake): Number of glasses of water consumed daily (scale 1-3)

- Drinking more water can help reduce hunger.
- Supports digestion and metabolism.

12. SCC (Calories Consumption Monitoring): Whether the person monitors their calorie intake (yes/no).

- Monitoring helps maintain a healthy diet.
- Indicates awareness and self-control over eating habits.

13. FAF (Physical Activity Frequency): Frequency of physical activity per week (scale 0-3)

- Regular exercise helps burn calories.
- Reduces fat accumulation and supports weight loss.

14. TUE (Time Using Technology): Time spent using screens/devices per day (scale 0–2)

- More screen time implies a sedentary lifestyle.
- Less physical movement increases obesity risk.

15. CALC (Alcohol Consumption): Frequency of alcohol intake (No/Sometimes/Frequently)

- Alcohol adds empty calories to the diet.
- Frequent intake leads to weight gain.

16. MTRANS (Mode of Transportation): Primary mode of transportation (e.g., Walking, bike).

- Active transport (walking, cycling) increases daily energy use.
- Sedentary transport contributes less to calorie burning.

Machine learning approaches predict obesity levels based on eating habits by analysing patterns in food consumption, lifestyle choices, and health data. These models use algorithms to identify relationships between dietary behaviours, such as the frequency of fast-food intake, portion sizes, and meal timing, and their impact on body weight. The system processes large datasets to find trends and correlations, helping to classify individuals into different obesity levels.

1.2 Types of Obesity Levels

❖ **Insufficient Weight (Underweight)**

BMI Range: Less than 18.5

Description: Body weight is lower than the healthy range for height.

❖ **Normal Weight**

BMI Range: 18.5 – 24.9

Description: Body weight is appropriate for height and considered healthy.

❖ **Overweight Level 1**

BMI Range: 25 – 27.9

Description: Slightly above normal weight range.

❖ **Overweight Level 2**

BMI Range: 28 – 29.9

Description: Noticeable weight gain above normal.

❖ **Obesity Type 1**

BMI Range: 30 – 34.9

Description: The person is classified as obese.

❖ **Obesity Type 2**

BMI Range: 35 – 39.9

Description: Severe obesity, also called Class 2 obesity.

1.3 Classification

Classification is a type of supervised machine learning where a model learns to categorize data into predefined classes or labels. It uses input features to predict the category to which a new data point belongs. The model trains on a labelled dataset, where each example has both input data and the correct output label. Common classification algorithms include Decision Trees, Support Vector Machines (SVM), K-Nearest Neighbours (KNN), and Neural Networks. Classification can be binary (with two classes, like spam or not spam) or multi-class (with more than two classes, like classifying types of fruits). The goal is to find patterns in the data that help the model make accurate predictions. Classification is used in real-world applications such as email filtering, disease diagnosis, and customer sentiment analysis. As the model processes more data, it improves its ability to correctly classify new information.

2. LITERATURE SURVEY

[1] S. Azmi et al., "Harnessing AI in Obesity Research and Management: A Comprehensive Review," Diagnostics, vol. 15, no. 3. MDPI AG, p. 396, Feb. 06, 2025. Doi: 10.3390/diagnostics15030396.

Azmi et al. (2025) proposed a clinical and research applications of AI, particularly ML and deep learning, in understanding, predicting, and managing obesity. Performance metrics such as accuracy, precision, recall, and F1-score were evaluated to optimize model selection. Conclusions: AI offers promising advancements in obesity management, enabling more personalized and efficient care. While technology presents considerable potential, challenges such as data quality, ethical considerations, and technical requirements remain. Addressing these will be essential to fully harness AI's potential in obesity research and treatment, supporting a shift toward precision healthcare.

[2] A. K. Tripathi, N. R. Chauhan, and A. Sharma, "Obesity classification and prognosis using machine learning," AIP Conference Proceedings, vol. 3224. AIP Publishing, p. 020050, 2025.

doi: 10.1063/5.0245938.

Tripathi et al. (2025) proposed has implemented the automated computational methods publicly available clinical data set to predict obesity status using different techniques as X-Gboost classifier, Random Forest, Decision Tree, K-nearest neighbour, and Support Vector Machine. By using these algorithms, it classifies the obesity level of patient and the diagnose the complexity of the diseases that differs for the algorithms amongst given concepts 94%.

[3] S. R. Niakan Kalhori, F. Najafi, H. Hasannejadasl, and S. Heydari, "AI -enabled obesity prediction: A systematic review of cohort data analysis," International Journal of Medical Informatics, vol. 196. Elsevier BV, p. 105804, Apr. 2025.
Doi: 10.1016/j.ijmedinf.2025.105804.

Niakan Kalhori et al. (2025) proposed an Obesity is now the fifth leading cause of death worldwide and has increased significantly over the past 40 years, raising the risk of diseases like type 2 diabetes and heart disease. Identifying obesity risks early can help prevent related health issues. ML was used in 95% of studies, mainly with supervised learning techniques. Algorithms like random forest (RF), linear regression, and gradient boosting (GBM) were common, with top models showing high accuracy, sensitivity, and specificity. The review highlights AI's potential in predicting obesity but emphasizes the need for further research on advanced AI methods. It serves as a valuable resource for dietitians and researchers working on AI-based predictive models and clinical decision support systems.

[4] Beuken, M. J., Kleynen, M., Braun, S., Van Berkel, K., van der Kallen, C., Koster, A., Bosma, H., Berendschot, T. T., Houben, A. J., Dukers-Muijrers, N., van den Bergh, J. P., Kroon, A. A., & Kanera, I. M. (2025). Identification of Clusters in a Population with Obesity Using Machine Learning: Secondary Analysis of The Maastricht Study. In JMIR Medical Informatics (Vol. 13, p. e64479). Doi:10.2196/64479."

Beuken et al (2025) This study aimed to identify clusters among individuals with obesity using a hypothesis-free, data-driven ML approach. Cross-sectional data from The Maastricht Study (cohort 2010) included 2971 variables covering demographics, lifestyle, biomedical aspects, and social factors. The factor probabilistic distance clustering algorithm identified three distinct clusters within the dataset. Cluster 1 had lower energy intake and a high proportion of non-working individuals. Cluster 2 had higher energy intake and a majority of male participants. Cluster 3 showed higher cognitive function and a greater percentage of individuals with higher education. These findings highlight the potential of AI in analysing large datasets to identify meaningful subgroups within populations with obesity.

[5] O. O. Awe, O. Olaniyan, A. E. Olatunde, R. Sewpaul, and N. Dukhi, "A Comparison of Generalized Additive Models for Obesity Risk Prediction." Elsevier BV, 2025. Doi: 10.2139/ssrn.5126225.

Sewpaul et al. (2025) proposed traditional regression models often struggle to capture the non-linear and complex relationships in obesity data - thereby limiting their predictive accuracy. To overcome this challenge, we used the Generalized Additive Models for Location, Scale, and Shape to model the relationships between obesity and its predictors. Thereafter, we compared our results with five other regression models using the Generalized Akaike Information Criterion for performance evaluation. The binomial stepwise model emerged as the best performer with a GAIC value 624.98. The Beta and Beta-Binomial models effectively described the data, underscoring GAML SS's potential to improve obesity risk prediction and inform public health strategies for more targeted interventions.

[6] P. Abdulvahap, Y. Fatma Hilal, and B. Georgian, "Use of Logistic Regression Method in Predicting Obesity Levels with ML Method," Zenodo, Jul. 2024,
Doi: 10.5281/ZENODO.12601115.
Abdulvahap et al. (2024) proposed that Obesity is a worldwide health issue due to excessive fat accumulation, especially prevalent in developing countries. It increases risks for diabetes, heart disease, and cancer, affecting multiple body systems. Using the logistic regression model, the following classification performance metrics for predicting obesity levels were calculated: Area under ROC curve is 0.980, Classification accuracy is 0.909, F1-Score is 0.911, Precision is 0.909, Recall is 0.860, Matthew's correlation coefficient is 0.992, and Specificity is 0.992. Notably, the classification accuracy of 90.9% indicates a significant achievement in correctly classifying the levels of obesity.

[7] F. T. Admojo and Nurul Rismayanti, "Estimating Obesity Levels Using Decision Trees and K- Fold Cross-Validation: A Study on Eating Habits and Physical Conditions," Indonesian Journal of Data and Science, vol. 5, no. 1. Yocto Brain, pp. 37–44, Mar. 31, 2024.
Doi: 10.56705/ijodas. v5i1.126.
Admojo et al. (2024) proposed a predictive capability of ML to explore the determinants of obesity within populations from Mexico, Peru, and Colombia, using a Decision Tree algorithm bolstered by 5- fold cross-validation. Our comprehensive analysis of 2111 individuals' lifestyle and physical condition data yielded accuracy, precision, recall, and F1-scores that notably peaked in the third and fifth folds. The findings affirmed the significance of dietary habits and physical activity as substantial predictors of obesity levels. The variability in model performance across the folds underscored the importance of robust cross-validation in enhancing the model's generalizability. This research contributes to the

burgeoning field of data science in public health by providing a viable model for obesity prediction and laying the groundwork for targeted health interventions.

[8] S. M. Tandiono "ML Approach of Obesity Level Classification: Systematic Literature Review of Methods and Factors," G-Tech: Jurnal Teknologi Terapan, vol. 8, no. 1. Fakultas Sains dan Teknologi, Unira Malang, pp. 196–208, Dec. 25, 2023.
Doi: 10.33379/gtech. v8i1.3604.

Tandiono et al. (2024) proposed a high prevalence of obesity over the years has become a global concern, as obesity contributes to an increased risk of many deadly diseases, such as diabetes, heart disease, and some cancers. This condition has become a serious concern for public health authorities, researchers, and the general public. Therefore, a comprehensive and effective approach is needed to tackle this obesity problem addressing global obesity.

[9] G. Vemulapalli, S. Yalamati, N. R. Palakurti, N. Alam,, "Predicting Obesity Trends Using ML from Big Data Analytics Approach," 2024 Asia Pacific Conference on Innovation in Technology (APCIT). IEEE, pp. 1–5, Jul. 26, 2024.
Doi: 10.1109/apcit62007.2024.10673429.

Vemulapalli et al. (2024) This study focuses on harnessing ML models to analyse extensive datasets encompassing demographic, socioeconomic, environmental, and lifestyle factors. Through the integration of various data sources, including electronic health records, wearable devices, and social media, our research aims to uncover hidden patterns and correlations contributing to obesity trends. By employing predictive analytics, our model seeks to forecast future obesity rates and identify high-risk populations, facilitating targeted interventions and policy implementations.

[10] N. Dwivedi, V. Singh, M. K. Gourisaria, R. Chatterjee, "Obesity Risk Detection Using ML Techniques," 2024 2nd International Conference on Sustainable Computing and Smart Systems (ICSCSS) IEEE, pp.761–766, Jul. 10, 2024. Doi:10.1109/icscss60660.2024.10625526.

Dwivedi et al. (2024) proposed that Obesity is considered a serious public health concern in the modern world. It is the presence of extravagant fat amount present in our body. Excess body weight, especially when concentrated around the abdomen, can contribute to various cardiovascular diseases such as Heart failure, Hypertension, Diabetes, etc. Innovative techniques like ML and Deep learning are playing an essential role in accurate and effective diagnosis of it. The primary goal of the paper is to determine the obesity or CVD risk at an early stage for further safeguards and medication. Various metrics were used to access the provided data set using well-known ML classifiers and boosting algorithms such as Cat Boost, Decision Tree, Naive Bayes, Support Vector Machine, Random Forest, Gradient Boosting, Logistic Regression, and Ada Boost.

Table: 1 Research GAPS Identified.

S.No	Author Name (Year)	Work Done	Observations
1	Azmi et al. (2025)	Reviewed how AI, especially machine learning, helps in understanding and treating obesity.	AI can make obesity treatment more personal and effective. But issues like data quality and ethics need attention.
2	Tripathi et al. (2025)	Used different machine learning models (like XGBoost, Random Forest, etc.) to predict obesity levels from patient data.	Accuracy was very high (94%). Different models gave slightly different results.
3	Niakan Kalhori et al. (2025)	Reviewed studies that used machine learning on health data to predict obesity.	Most studies used supervised learning. Random Forest and Gradient Boosting worked well. More advanced AI research is needed.
4	Beuken et al. (2025)	Used clustering (grouping) methods to find patterns in obesity data from a large study.	Found 3 groups of people with different traits (like diet and education). AI helped discover these hidden patterns.
5	Awe et al. (2025)	Compared different types of regression models to see which predicts obesity best.	The "binomial stepwise" model worked best. Showed that advanced models can handle complex data better.
6	Abdulvahap et al. (2024)	Used logistic regression to predict obesity levels.	Got 90.9% accuracy, and other metrics were also very good. Logistic regression is simple but effective.
7	Admojo et al. (2024)	Used Decision Tree and cross-validation to study eating habits and body conditions of people in Latin America.	Eating habits and exercise were key factors in obesity. Model performance changed across different data splits
8	Tandiono et al. (2024)	Reviewed many studies on obesity and machine learning methods.	Obesity is growing fast worldwide. A proper and combined effort is needed to fight it using ML and health data.
9	Vemulapalli et al. (2024)	Used big data (from health records, wearables, social media) with machine learning to track obesity trends.	Helped identify which groups are most at risk. Can be used to plan health policies and prevention strategies
10	Dwivedi et al. (2024)	Tested several ML models to find obesity and heart disease risks early.	Boosting models gave good results. Early detection is possible using AI models

3. SYSTEM REQUIREMENTS SPECIFICATIONS

3.1 Software Requirements Specifications

Operating System	:	Windows 10 pro
Programming Language	:	Python 3.11
Coding Platform	:	VS code
Web frame work	:	Django
Libraries	:	Pandas, NumPy, joblib Matplotlib, Seaborn, Scikit-learn

3.2 Hardware Requirements Specification

Processor	:	Intel(R) Core (TM) i5-6006U CPU @ 5.00GHz
RAM	:	16 GB
SSD	:	1 TB
Input Devices	:	Keyboard, Mouse

3.3 Technology Description

This research utilizes the Extra Trees Classifier, an advanced ensemble learning architecture, to categorize obesity levels through the synthesis of nutritional, physiological, and behavioral datasets. By generating a forest of extremely randomized decision trees, the model effectively mitigates variance and avoids the common pitfall of overfitting, ensuring high generalization across diverse cohorts. The experimental workflow involved rigorous feature engineering, automated data normalization, and hyperparameter optimization to maximize predictive precision. To ensure model transparency, the study employs confusion matrices for error analysis and SHAP-based feature importance graphs, transforming complex algorithmic outputs into interpretable clinical insights for preventive healthcare.

In this study, we used a powerful computer model called the Extra Trees Classifier to predict obesity risk. This model works by looking at many different factors at once, such as what people eat, how much they exercise, and their daily habits. Unlike simpler models, Extra Trees builds many "mini-models" and combines them to make sure the final answer is stable and accurate. Before running the model, we cleaned the data and fine-tuned the settings to get the best results. To make the computer's "thinking" easy to understand, we used charts like confusion matrices and feature importance graphs. These tools show us exactly which habits—like snacking or a lack of exercise—are the biggest red flags for obesity.

4. SYSTEM ANALYSIS

4.1 Existing System:

➢ Obesity Level Dataset: The process starts with collecting the dataset containing information related to obesity levels based on eating habits and other factors.

- ➢ Dataset Preprocessing: The data is cleaned, handled for missing values, normalized (if necessary), and prepared for model training.
- ➢ Train-Test Splitting: The dataset is divided into training and testing sets to evaluate model performance.
- ➢ Model Training two machine learning models are used:
- ▪ Support Vector Machine (SVM): An existing classifier applied to predict obesity levels.
- ▪ Logistic Regression Classifier: Another existing model used for prediction.
- ▪ Voting Classifier, the outputs of the SVM and Logistic Regression models are combined using an ensemble method (Voting Classifier) to improve prediction accuracy.
- ➢ Performance Analysis: The final step involves evaluating the model's performance using metrics such as accuracy, precision, recall, and F1-score.

4.2 Disadvantages of Existing System:

- ❖ Limited Models: The approach may be restricted to a few machine learning models, reducing the scope for finding the most accurate predictor.
- ❖ Dependence on Voting Classifier: Relying heavily on an ensemble method like a voting classifier may limit the individual contribution of other models.
- ❖ Data Quality Matters: The effectiveness of the model depends on the quality and completeness of the dataset, as poor data can lead to inaccurate predictions.
- ❖ Slow Processing: The computational time may be high, especially if complex feature engineering or ensemble techniques are used.
- ❖ No Fine Tuning: Without hyperparameter optimization, the model may not achieve its best possible accuracy, leading to suboptimal performance.

4.3 Proposed System of Extremely Randomized Classifier Tree:

1. *Structure:*

- ✓ The diagram shows multiple decision trees (Tree 1, Tree 2, ..., Tree p).
- ✓ Each tree is independently built with different data splits.
- ✓ Predictions from all trees are combined to make the final decision.

2. *Extra Trees Classifier:*

- ✓ It is an ensemble learning method that builds multiple decision trees.
- ✓ Unlike traditional decision trees or Random Forest, Extra Trees select split points randomly instead of using information gain or Gini impurity.
- ✓ This randomness helps improve accuracy and reduces overfitting.

3. *Comparison with Random Forest:*

- ✓ Random Forest chooses the best split based on statistical criteria.
- ✓ Extra Trees select splits at random, which leads to better variance reduction.

Steps of Extremely randomized Tree classifier are as follows:

1) Select random samples from a given data or training set.
2) This algorithm will construct a decision tree for every training data
3) Voting will take place by averaging the decision tree.
4) Finally, select the most voted prediction result as the final prediction result.

***4.4 Advantages of Proposed System*:**

- High Speed: The model processes data quickly, making it efficient for large datasets.
- Low Overfitting: The model generalizes well to new data, reducing the risk of overfitting.
- Robust to Noise: The approach can handle noisy or imperfect data without significant performance loss.
- Better Variance Reduction: The model maintains stability and reduces variability in predictions.
- Efficient for High-Dimensional Data: It can handle datasets with many features effectively, making it suitable for complex problems.

5. SYSTEM DESIGN

The workflow system for predicting obesity levels using a machine learning approach follows a structured process to ensure accurate predictions based on eating habits and lifestyle factors. It begins with data input, where users provide information about their dietary habits, physical activity, and health-related details. This data is then processed in the data processing phase, where it undergoes cleaning, normalization, and encoding to ensure consistency and usability for the machine learning model. After preprocessing, the system moves to the model training and evaluation stage, where the dataset is split into training and testing sets, and an algorithm such as the Extra Trees Classifier is used to train the model. The trained model is then tested for accuracy, precision, recall, and F1-score to ensure its reliability in predicting obesity levels.

Once the model is validated, the system progresses to the prediction and output phase, where it classifies individuals into different obesity categories, such as normal weight, overweight, or obese, based on their input data. The predictions are accompanied by insights and recommendations related to diet, exercise, and lifestyle changes. Finally, the decision-making stage involves interpreting the results and taking necessary actions. Healthcare professionals can use these predictions to provide personalized medical advice, while users can make informed choices regarding their lifestyle and health improvements. This structured workflow ensures an efficient and AI-driven approach to obesity prediction, supporting both individuals and medical professionals in managing and preventing obesity-related health risks.

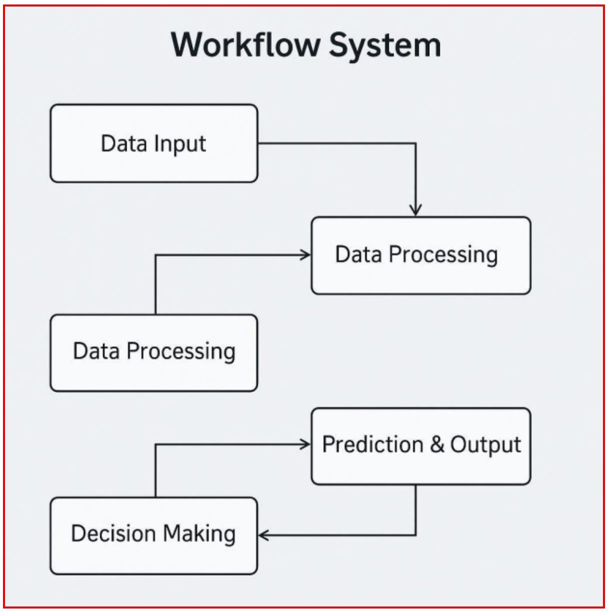

Fig: 5.1 Workflow of the system.

5.1 Class Diagram

A Class diagram is a UML diagram that depicts the static structure of a software system, including the classes, their attributes and operations, and the relationships among them. This diagram serves as a visual representation of the system's components and their interconnections, enabling developers to better understand the system's architecture and design. By identifying the classes and their properties, developers can plan and implement the system's functionality effectively. The Class diagram typically consists of multiple classes, each with their own attributes and operations, which are interlinked to represent the relationships between the classes.

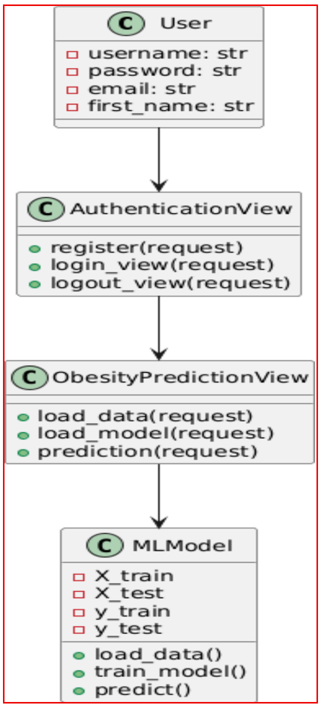

Fig. 5.2 Class Diagram.

5.2 Activity Diagram

The process of predicting obesity levels using the Extra Trees Classifier starts when a user enters their details, such as age, weight, height, eating habits, and physical activity, through a web application. The data is then sent for preprocessing, where missing values are handled, text data is converted into numbers, and everything is scaled properly. After that, the data is divided into training and testing sets to help the machine learning model learn patterns effectively. The Extra Trees Classifier is trained using past data, allowing it to recognize different obesity levels. When a new user submits their details, the trained model makes a prediction and classifies them as underweight, normal, overweight, or obese. Finally, the result is displayed on the web application, giving the user an idea of their obesity level based on their eating habits and lifestyle. This simple process ensures quick and accurate obesity prediction.

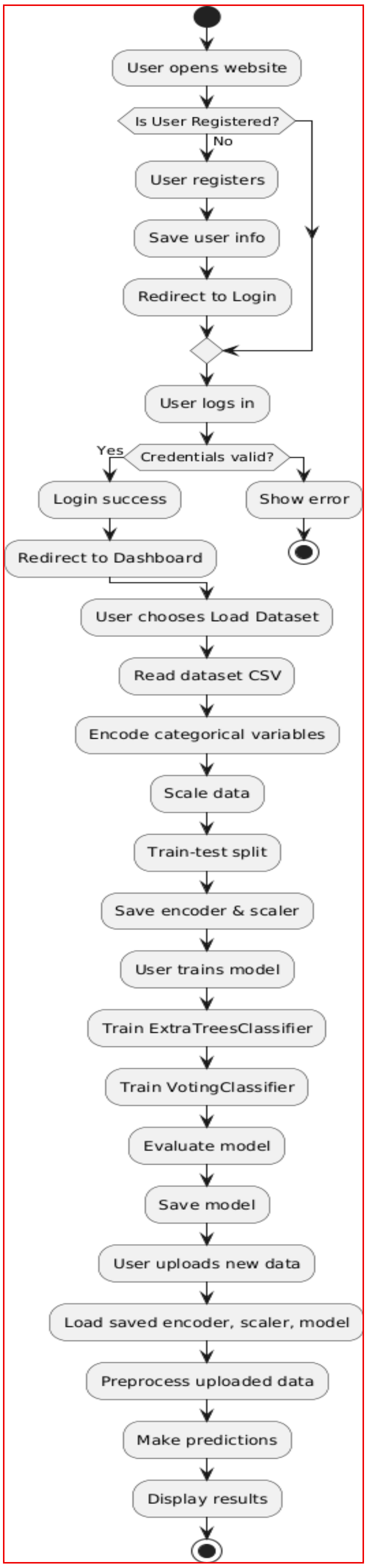

Fig. 5.3 Activity Diagram.

5.3 Sequence Diagram

A sequence diagram, categorized under the Unified Modelling Language (UML) interaction diagrams, serves as a valuable tool for illustrating the interaction between objects or system components over time. It encapsulates the dynamic behavior of the system by delineating the sequence in which messages are exchanged among various objects or components. With its ability to model diverse scenarios and use cases, sequence diagrams aid developers in comprehending and analyzing system behavior, enabling the identification of potential issues or bottlenecks. Moreover, they play a crucial role in ensuring that the system aligns with stakeholder requirements. Beyond its technical utility, sequence diagrams also foster communication and collaboration among developers and other stakeholders, serving as a common visual language to discuss and refine system designs.

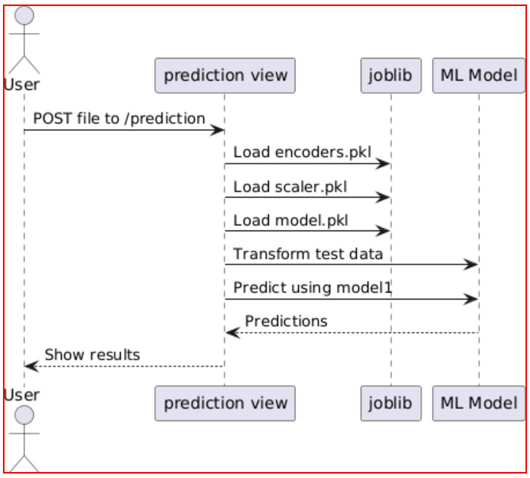

Fig. 5.4 Sequence Diagram.

6. TESTING

Testing is a critical phase in the development of the project titled "Machine Learning Approach for Prediction of Obesity Levels Based on Eating Habits." It ensures that each component of the system performs as expected and the overall model functions reliably and accurately. The objective of testing in this project is to verify the integrity and correctness of the data pipeline, the effectiveness of machine learning models, and the consistency of the system outputs. The testing process is divided into three major categories: Unit Testing, Integration Testing, and System Testing, each serving a specific purpose in the validation of the project.

Fig. 6.1 Home Page.

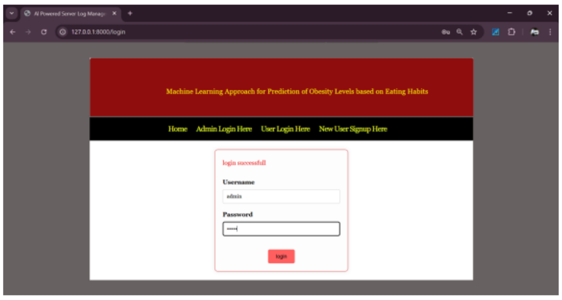

Fig. 6.2 Admin Login Page.

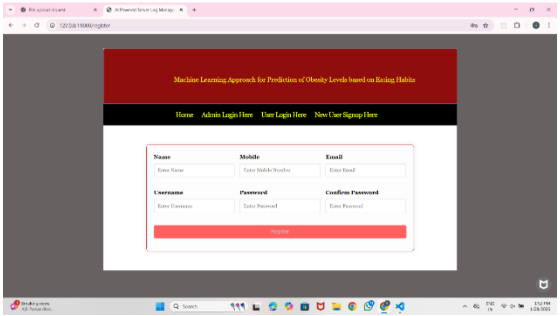

Fig. 6.3 User ID Creation Page.

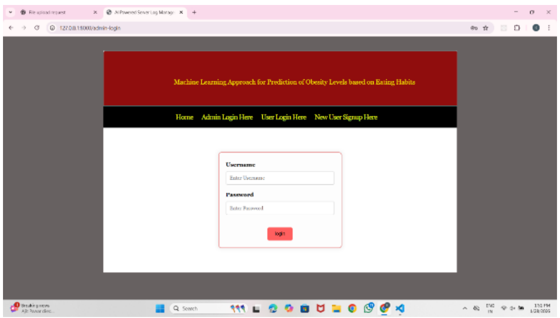

Fig. 6.4 User Login Page.

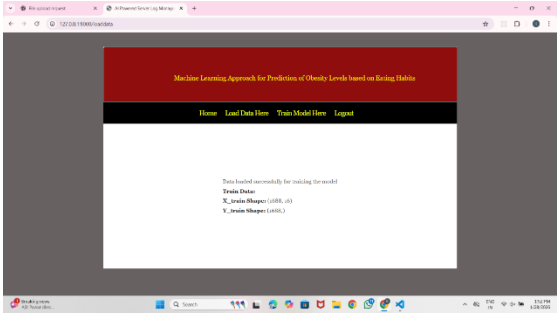

Fig. 6.5 Data Preprocessing.

Fig. 6.6 Performance Comparison.

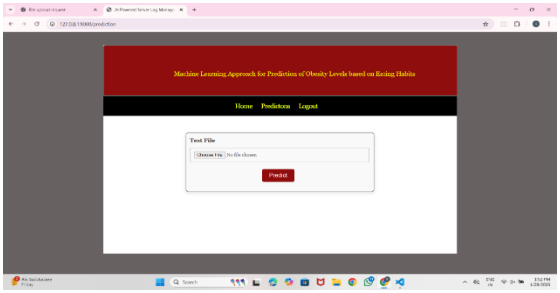

Fig. 6.7 Prediction Page.

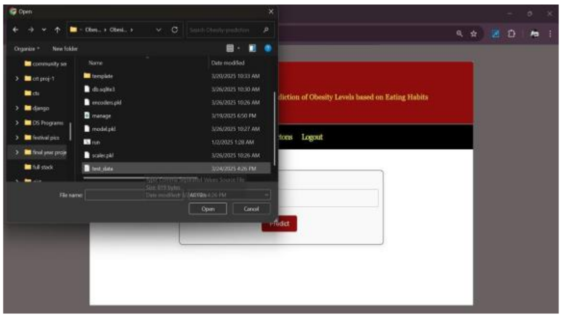

Fig. 6.8 Upload Dataset Page.

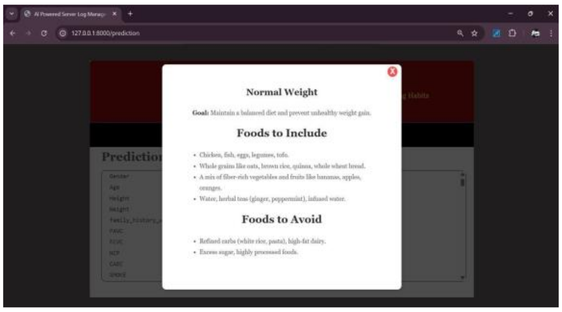

Fig. 6.9 Precautions Page.

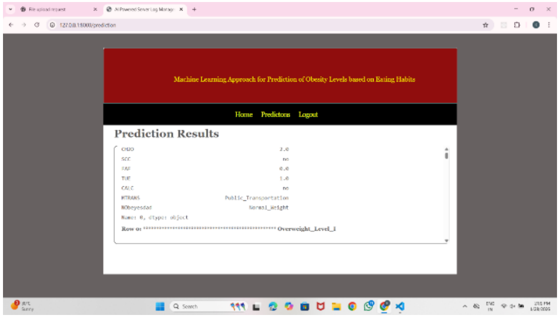

Fig. 6.10 Result 1.

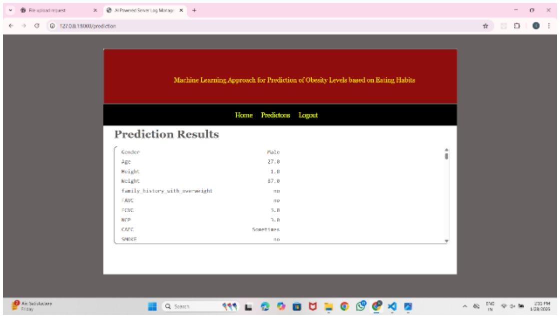

Fig. 6.11 Result 2.

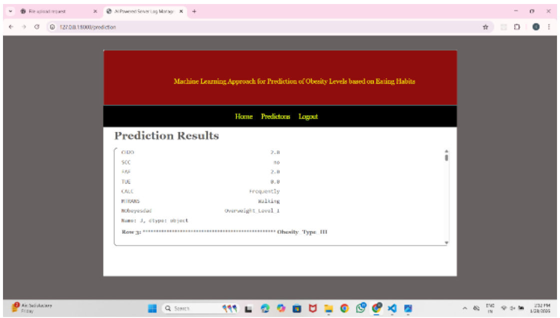

Fig. 6.12 Result 3.

7. RESULTS AND DISCUSSION

7.1 Performance Evaluation Metrics:

The efficacy of the proposed machine learning architectures was evaluated using a multi-metric approach, including Accuracy, Precision, Recall, and the F1-score. These metrics were selected to provide a granular assessment of the models' ability to categorize obesity levels accurately while minimizing Type I and Type II errors.

The Extra Trees Classifier, a key component of our proposed framework, demonstrated superior predictive power. It achieved a peak accuracy of 95.56%, with a nearly identical F1-score of 95.59%. This symmetry between precision and recall suggests that the model is exceptionally balanced in identifying various obesity classes without bias. Similarly, the Voting Classifier an ensemble of multiple base learners delivered a robust performance with an accuracy of 93.47%.

In contrast, the baseline models exhibited a notable performance gap. The Support Vector Machine (SVM) and Logistic Regression yielded lower accuracies, ranging between 87-90% and 85-88%, respectively. These results underscore the limitations of traditional linear and kernel-based methods when applied to the high-dimensional and non-linear nature of dietary behavior datasets.

7.2 Comparative Analysis: A systematic comparison of the experimental results (Table 7.1) validates the hypothesis that ensemble learning techniques significantly enhance classification outcomes for health-related data.

Table 7.1: Comparative Performance Analysis of Classification Models.

Model	Accuracy (%)	Precision (%)	Recall (%)	F1-score (%)
Extra Tree Classifier	95.56	95.64	95.56	95.59
Voting Classifier	93.47	93.53	93.47	93.44
Support Vector Machine	87-90	87	88	87.5
Logistic Regression	85-88	85	86	85.5

The superior performance of the Extra Trees Classifier can be attributed to its ability to manage feature variance and reduce overfitting through extreme randomization of decision boundaries. While the Voting Classifier also showed strong generalization, the Extra Trees model provided the most precise mapping of dietary habits to specific obesity levels.

8. CONCLUSION:

This study highlights the transformative potential of ensemble-based machine learning in the domain of nutritional informatics. Our findings reveal that the Extra Trees Classifier is the most effective architecture for predicting obesity levels, achieving an optimal accuracy of 95.56%. The significant performance margin between our proposed ensemble methods and traditional models like SVM confirms that the complex interactions between lifestyle factors and health outcomes are best captured through high-dimensional, tree-based learning.

By integrating these high-performance algorithms into clinical decision support systems, healthcare providers can leverage routine dietary data to facilitate early intervention. This research provides a scalable foundation for personalized preventative medicine, moving beyond generalized guidelines toward data-driven, individualized health insights.

9. REFERENCES

1. S. Azmi et al., "Harnessing Artificial Intelligence in Obesity Research and Management: A Comprehensive Review," Diagnostics, vol. 15, no. 3. MDPI AG, p. 396, Feb. 06, 2025. doi:10.3390/diagnostics15030396.
2. A. K. Tripathi, N. R. Chauhan, and A. Sharma, "Obesity classification and prognosis using machine learning," AIP Conference Proceedings, vol. 3224. AIP Publishing, p. 020050, 2025. doi:10.1063/5.0245938.
3. S. R. Niakan Kalhori, F. Najafi, H. Hasannejadasl, and S. Heydari, "Artificial intelligence-enabled obesity prediction: A systematic review of cohort data analysis," International Journal of Medical Informatics, vol. 196. Elsevier BV, p. 105804, Apr. 2025. doi: 10.1016/j.ijmedinf.2025.105804.
4. Beuken, M. J., Kleynen, M., Braun, S., Van Berkel, K., van der Kallen, C., Koster, A., Bosma, H., Berendschot, T. T., Houben. Identification of Clusters in a Population with Obesity Using Machine Learning: Secondary Analysis of The Maastricht Study. In JMIR Medical Informatics (Vol. 13, p.e64479). doi:10.2196/64479
5. O. O. Awe, O. Olaniyan, A. E. Olatunde, R. Sewpaul, and N. Dukhi, "A Comparison of Generalized Additive Models for Obesity Risk Prediction." Elsevier BV, 2025. doi: 10.2139/ssrn.5126225.
6. P. Abdulvahap, Y. Fatma Hilal, and B. Georgian, "Use of Logistic Regression Method in Predicting Obesity Levels with Machine Learning Method," Zenodo, Jul. 2024, doi:10.5281/ZENODO.12601115.
7. F. T. Admojo and Nurul Rismayanti, "Estimating Obesity Levels Using Decision Trees and K-Fold Cross-Validation: A Study on Eating Habits and Physical Conditions," Indonesian Journal of Data and Science, vol. 5, no. 1. Yocto Brain, pp. 37–44, Mar. 31, 2024. doi: 10.56705/ijodas.v5i1.126.

8. S. M. Tandiono "ML Approach of Obesity Level Classification: A Systematic Literature Review of Methods and Factors," G-Tech: Jurnal Teknologi Terapan, vol. 8, no. 1. Fakultas Sains dan Teknologi, Unira Malang, pp. 196–208, Dec. 25, 2023. doi: 10.33379/gtech. v8i1.3604.
9. G. Vemulapalli, S. Yalamati, N. R. Palakurti, N. Alam,, "Predicting Obesity Trends Using ML from Big Data Analytics Approach," 2024 Asia Pacific Conference on Innovation in Technology (APCIT). IEEE, pp. 1–5, Jul. 26, 2024.
doi: 10.1109/apcit62007.2024.10673429.
10. N. Dwivedi, V. Singh, M. K. Gourisaria, R. Chatterjee, "Obesity Risk Detection Using ML Techniques," 2024 2nd International Conference on Sustainable Computing and Smart Systems (ICSCSS). IEEE, pp. 761–766, Jul. 10, 2024.
Doi:10.1109/icscss60660.2024.10625526.