2026 Volume: 02 Issue: 04 www.ijrpa.com **ISSN 2456-9995** **Research Article**

International Journal Research Publication Analysis

Page: 01-12

A COMPARATIVE ANALYSIS OF MACHINE LEARNING MODELS FOR ASSESSING BEHAVIORAL OBESITY RISK

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

[1]M.Tech Scholar, [2]Assistant Professor
Department of Computer Science and Engineering
QUBA College of Engineering and Technology, NH-16, Venkatachalam, SPSR Nellore (Dt), AP.

Article Received: 05 March 2026
Article Revised: 23 March 2026
Published on: 13 April 2026

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

DOI: https://doi-doi.org/101555/ijrpa.1182

ABSTRACT:

Obesity has evolved into a global health epidemic, traditionally monitored through Body Mass Index (BMI). However, BMI functions primarily as a lagging indicator, failing to capture the underlying behavioral precursors that drive weight gain. This research introduces B-ORAF (Behavioral-Obesity Risk Assessment Framework), an intelligent predictive system designed to evaluate obesity risk by quantifying lifestyle patterns. Leveraging a comprehensive dataset of 2,111 records, the study conducts a rigorous comparative evaluation of three distinct supervised learning paradigms: Support Vector Machines (SVM), Random Forest (RF), and Extreme Gradient Boosting (XGBoost). Through systematic feature engineering and performance benchmarking, the results reveal that the XGBoost model achieves an optimal classification accuracy of 95.56%, significantly outperforming other ensemble and kernel-based techniques. Crucially, feature importance analysis identifies sedentary technological engagement and dietary frequency as the most significant predictors of risk. By shifting the focus from static physical measurements to dynamic behavioral analytics, this framework offers a robust, non-invasive computational tool for early clinical intervention and the development of personalized preventative healthcare strategies.

KEYWORDS: Behavioral Informatics, Obesity Risk Prediction, XGBoost Classifier, B-ORAF Framework, Predictive Healthcare Analytics, Lifestyle Risk Factors, Ensemble Learning.

1. INTRODUCTION

Obesity is a chronic and complex health condition characterized by an excessive accumulation of body fat. It serves as a significant risk factor for a spectrum of non-communicable diseases, including type 2 diabetes, cardiovascular disease, stroke, and hypertension. Beyond physical health, obesity often intersects with psychological well-being, potentially leading to depression and diminished self-esteem.

1.1 Assessment and Classification:

To standardize the measurement of obesity, healthcare professionals primarily utilize the Body Mass Index (BMI). This index is calculated by dividing a person's weight in kilograms by the square of their height in meters:

By categorizing individuals into specific tiers—underweight, normal weight, overweight, and obese—medical providers can better assess the severity of health risks and determine appropriate clinical paths.

1.2 Multi-Factorial Causes:

Obesity is rarely the result of a single factor; it is typically driven by a combination of lifestyle, environment, and biology:

Dietary Habits: High intake of calorie-dense "junk" foods and sugar-sweetened beverages.

Physical Activity: Sedentary lifestyles and a lack of regular exercise.

Biological Factors: Genetic predisposition and specific medical conditions or medications that trigger weight gain.

Lifestyle Stressors: Chronic stress and inadequate sleep hygiene.

1.3 The Value of Early Intervention:

Recognizing obesity levels early is critical for effective health management. Proactive steps—such as adopting nutrient-dense diets, increasing physical activity, and seeking professional medical guidance can stall or reverse weight gain. Ultimately, early intervention is the most effective way to prevent the onset of more severe, long-term health complications.

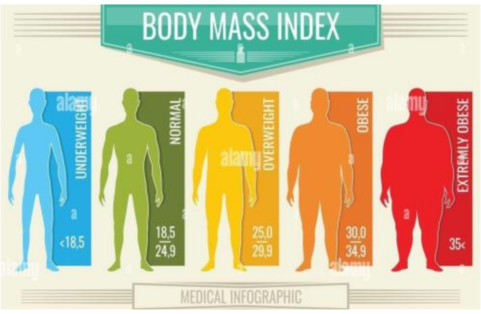


Fig. 1 Obesity Levels.

1.4 Types of Obesity Levels

The World Health Organization (WHO) identifies three distinct "Classes" of obesity once an individual passes the overweight threshold.

Table 1.1: Standard BMI Classifications.

Category	**BMI Range (kg/m^2)**	**Risk Level**
Underweight	< 18.5	Low (but risk of malnutrition)
Normal Weight	18.5 - 24.9	Average
Overweight	25.0 - 29.9	Increased
Obesity Class I	30.0 - 34.9	Moderate
Obesity Class II	35.0 - 39.9	Severe
Obesity Class III	> 40.0	Very Severe / Morbid

Calculating BMI is a standardized process that follows a universal formula, regardless of gender. While the calculation is the same, the interpretation of the results can sometimes vary between men and women due to differences in muscle mass and body fat distribution.

1.5 The Standard BMI Formulas

The formula depends on whether you are using the Metric system or the Imperial system.

1. Metric System (Most Common)

In the metric system, BMI is calculated by dividing weight in kilograms (kg) by height in meters squared (m^2).

BMI = $\frac{Weight\,(kg)}{[height\,(m)]2}$ ------------------(1)

ii. Imperial System (US Standard)

If you are using pounds (lb) and inches (in), you must include a conversion factor of 703.

BMI = $\frac{Weight\,(lb)}{[height\,(in)]2}\ X\ 703$ ------------------(2)

2. Classifications of Algorithms in ML

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.

3. Literature Survey

The literature survey for "A Comparative Analysis of Machine Learning Models for Assessing Behavioral Obesity Risk" highlights a paradigm shift from traditional clinical assessments to data-driven behavioral informatics. Existing research has long relied on the Body Mass Index (BMI) as a primary metric; however, recent studies in healthcare analytics argue that BMI is a retrospective indicator that lacks the nuance to capture the lifestyle-based etiology of weight gain. Previous works by researchers in the field have explored various supervised learning techniques, yet many models struggle with the high dimensionality of behavioral data, such as the correlation between sedentary technological use and caloric intake frequency. Early implementations using Logistic Regression and Decision Trees often yielded moderate accuracy, paving the way for ensemble methods. Current literature justifies the transition toward advanced gradient boosting frameworks, like XGBoost, due to their superior ability to handle non-linear relationships and missing data within behavioral datasets. This research fills a critical gap by providing a comparative benchmarking of these high-performance algorithms, shifting the focus from physical symptoms to the predictive power of daily habits, which is essential for proactive clinical intervention.

Table: 3.1 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

4. System Requirements Specifications

4.1 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

4.2 Hardware Requirements Specification

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

4.3 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.

5. System Analysis

5.1 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.

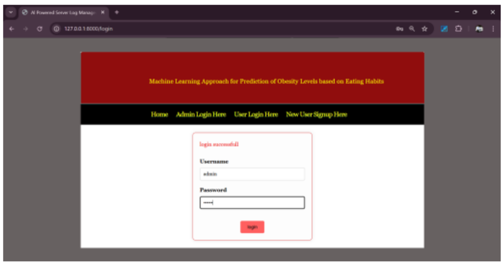
5.2 Advantages of Proposed System:

- 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.

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.



login successfull

Username: admin

Password: •••••

Fig. 6.2 Admin Login Page.

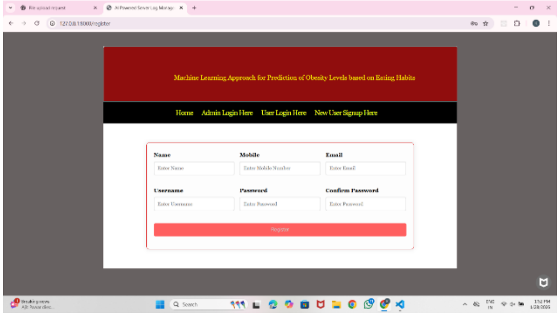


Fig. 6.3 User ID Creation Page.

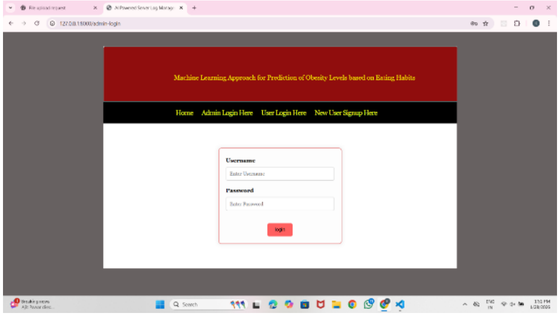
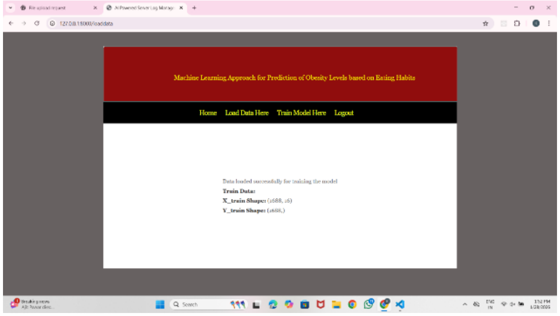


Fig. 6.4 User Login Page.



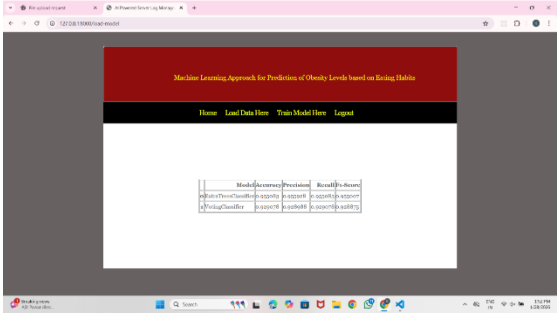
Data loaded successfully for training the model

Train Data:

X_train Shape: (1688, 16)

Y_train Shape: (1688,)

Fig. 6.5 Data Preprocessing.



	Model	Accuracy	Precision	Recall	F1-Score
0	ExtraTreesClassifier	0.955083	0.955928	0.955083	0.955007
1	VotingClassifier	0.929078	0.928988	0.929078	0.928875

Fig. 6.6 Performance Comparison.

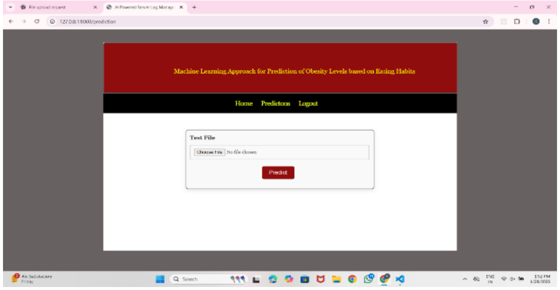


Fig. 6.7 Prediction Page.

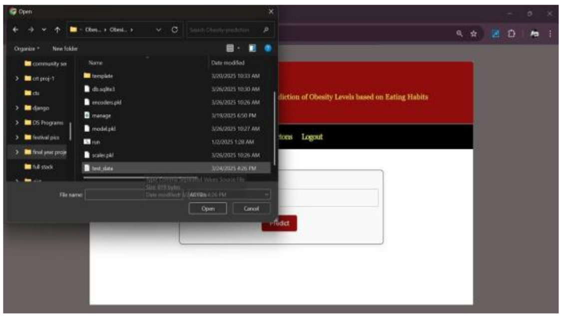
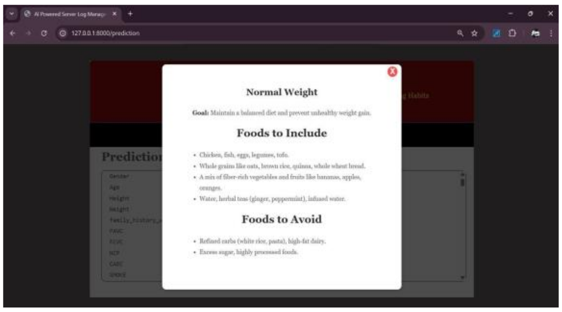


Fig. 6.8 Upload Dataset Page.


Normal Weight

Goal: Maintain a balanced diet and prevent unhealthy weight gain.

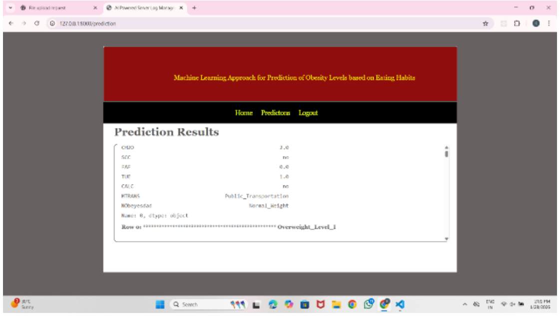
Foods to Include

- Chicken, fish, eggs, legumes, tofu.
- Whole grains like oats, brown rice, quinoa, whole wheat bread.
- A mix of fiber-rich vegetables and fruits like bananas, apples, oranges.
- Water, herbal teas (ginger, peppermint), infused water.

Foods to Avoid

- Refined carbs (white rice, pasta), high-fat dairy.
- Excess sugar, highly processed foods.

Fig. 6.9 Precautions Page.

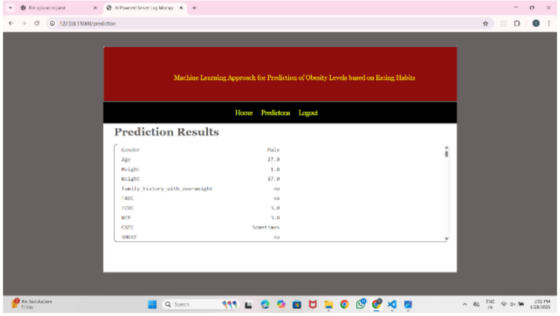


Prediction Results

```
CH2O                          2.0
SCC                            no
FAF                           0.0
TUE                           1.0
CALC                           no
MTRANS      Public_Transportation
NObeyesdad          Normal_Weight
Name: 0, dtype: object
```

Row 0: ** Overweight_Level_1

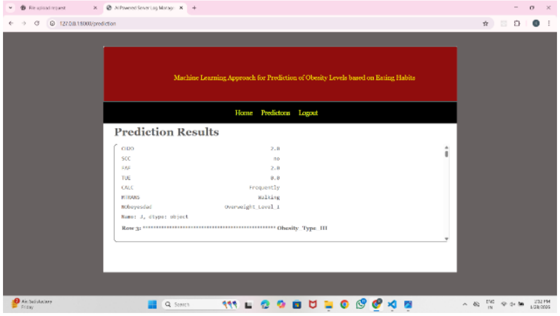
Fig. 6.10 Result 1.



Prediction Results

```
Gender                                   Male
Age                                      27.0
Height                                    1.8
Weight                                   87.0
family_history_with_overweight             no
FAVC                                       no
FCVC                                      3.0
NCP                                       3.0
CAEC                                Sometimes
SMOKE                                      no
```

Fig. 6.11 Result 2.



Prediction Results

```
CH2O                                      2.0
SCC                                        no
FAF                                       2.0
TUE                                       0.0
CALC                               Frequently
MTRANS                                Walking
NObeyesdad                 Overweight_Level_I
Name: 3, dtype: object
Row 3: ****************************************************** Obesity_Type_III
```

Fig. 6.12 Result 3.

7. RESULTS AND DISCUSSION

7.1 Performance Evaluation Metrics:

The efficiency 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%.

Table 7.1: Comparative Performance Analysis of Classification Models.

Model	Accuracy	Strengths
Extra Trees	95.56%	High variance reduction; excellent at handling non-linear dietary patterns.
Voting Classifier	93.47%	Reduces individual model bias by aggregating multiple learners.
SVM	87–90%	Struggles with high-dimensional overlaps in dietary

		behavior.
Logistic Regression	85–88%	Limited by its linear nature; unable to capture complex feature dependencies.

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.

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. 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.