International Journal of
DATA SCIENCE AND IOT MANAGEMENT SYSTEM
Peer Reviewed, Referred & Indexed Journal
ISSN: 3068-272X www.ijdim.com Original Research Paper

A Multi-Model Learning Approach for Early Identification of Obesity Risk from Lifestyle Factors

E Tharun[1], K Bhaskar[2]
[1] P.G Scholar, Department of MCA, Sri Venkatesa Perumal College of Engineering & Technology, Puttur,
E-mail: tharune143@gmail.com, ORC-ID: https://orcid.org/0009-0003-2841-7131
[2] Assistant Professor, Department of CSE(AI & ML),Sri Venkatesa Perumal College of Engineering &Technology,
Puttur, E-mail: bhaskark.mca@gmail.com , ORC-ID: https://orcid.org/0009-0000-3309-4240

Abstract: Obesity is still a major public health problem around the world, and correct risk classification is needed to help people get help early and avoid long-term problems. The study uses the UCI Machine Learning Repository dataset, which has information about eating habits, physical exercise, and anthropometric measurements. It gives a wide range of examples of how people live their lives. The suggested method uses StandardScaler to normalize the data, RFECV with Logistic Regression to choose the features, and SMOTENC to resample classes that aren't balanced. Stratified 5-Fold Cross-Validation is used to measure how well the method works. Adaboost, Perceptron, GaussianNB, SGD, SVM, KNN, MLP, Decision Tree, ExtraTrees, BaggingClassifier, RandomForest, GradientBoosting, LogisticRegressionCV, XGBoost, and LightGBM are some of the algorithms that are used. There is also a proposed Stacking model that uses ensemble learning and LIME to give local, model-agnostic explanations of individual predictions. In addition, to further enhance robustness and interpretability, an Extended Voting ensemble combining Gradient Boosting Classifier, XGB Classifier, LGBM Classifier and CatBoost Classifier is reserved as an extension with SHAP-based global explanations and deployment via a Flask framework. With 99.3% accuracy, a perfect ROC-AUC, and relatively high precision, recall, and F1-score, the combined pipeline does a great job of classifying data.

*“**Index Terms:** Obesity, machine learning, stacking, explainable AI”.*

1. INTRODUCTION

Obesity, which is described as having too much body fat, has become a major global health problem that affects millions of people and causes a lot of illness and death [1, 2]. The condition is often connected to a lack of physical activity, bad eating habits, and external factors. Together, these things cause more and more people to be overweight or obese around the world [3, 4]. Because obesity raises the chance of heart disease, diabetes, stroke, and some cancers, this epidemic puts a huge strain on healthcare systems.

Body mass index (BMI) and other anthropometric measurements are often used to find people who are overweight or obese, but they don't always show how complicated the situation is [5, 6]. BMI doesn't tell the difference between lean mass and body fat, and it doesn't take into account lifestyle or social factors that affect the risk of obesity. Using these old-fashioned metrics alone could lead to wrong classification and poor risk assessment, showing the need for broader review methods.

The goal of this study is to create a system for predicting obesity risk that can be understood by combining different types of lifestyle, dietary, and physiological data [7, 8]. The model can find patterns and risk factors that are often missed by traditional assessment methods because it uses large datasets. The framework also stresses openness, making sure that predictions can be understood and used by policymakers and healthcare experts.

The importance of this study lies in its ability to help shape public health plans and individualized care [9, 10]. Accurate and easy-to-understand predictions of obesity risk can help with early intervention, lower long-term complications, and lower the costs of healthcare linked to obesity. Furthermore, by finding the main factors that lead to obesity, this study improves our knowledge of how environment and lifestyle can affect people, which helps people and groups make decisions based on facts.

2. LITERATURE REVIEW

Because obesity is so common, a lot of work has gone into making predictions and using data to find groups of people who are at risk and plan specific treatments. Using the Indonesian Basic Health Research 2018 dataset, Thamrin et al. [11] looked into how machine learning methods might be able to identify adult obesity. Their method was to test how well different classification models could find adults who were overweight. This showed how useful data-driven methods can be in public health situations. The study stressed how important it is to combine anthropometric, demographic, and lifestyle factors.

It also showed that machine learning models can better understand complicated patterns of obesity than older methods like body mass index-based classification. In the same way, Zheng and Ruggiero [12] used machine learning to look into how to predict obesity in high school kids. Their work showed that predictive analytics can be used with teen groups, taking into account things like diet, exercise, and social and environmental factors. The study showed that using models with a lot of different behavioral and physiological inputs could make it a lot more accurate at figuring out who is at risk for obesity. This shows how important it is to have a lot of different datasets for predictive modeling.

Machorro-Cano et al. [13] created PISIoT, a new smart health tool that uses machine learning and the Internet of Things (IoT) to track and manage overweight and obese people. Wearable IoT devices were used in their system to collect real-time health data. These were then combined with predictive models to make personalized suggestions. This study showed that it is possible to combine real-time data collection with predictive analytics. This would make weight control interventions more timely and useful. Kibria et al. [14] suggested using a soft voting classifier along with explainable AI methods to make a group prediction for diabetes mellitus. While they were focused on diabetes, their method shows how ensemble and interpretable models can be used in a wider range of health risk prediction situations. It also gives us information that can be directly applied to frameworks for predicting obesity. The study showed that explainable AI methods can make model thinking clearer, which helps practitioners understand what factors affect predicted outcomes. This builds trust and makes the methods more useful in real life.

Raihan et al. [15] looked into how to find people with chronic kidney disease using an XGBoost classifier and SHAP explanations to see how different characteristics affect model predictions. Their main focus was on kidney disease, but their method can be used in obesity study too because it shows how to combine high-performance ensemble models with explainability frameworks to make sense of complicated health data. In the same way, Jahan et al. [16] used multimodal data to show how explainable AI can be used to predict Alzheimer's. Their method used a mix of different datasets to improve prediction accuracy while still giving results that were easy to understand. This study shows that using AI that can be explained is becoming more common in healthcare, and it stresses how important clear models are for encouraging clinical acceptance.

Dua and Graff [17], who managed the UCI Machine Learning Repository, which is a useful tool in health analytics, also say that large datasets are important for building models. The repository has organized datasets with demographic, behavioral, and physiological information that make it easier to test predictive models thoroughly. The work of Hastie, Tibshirani, and Friedman [18] laid the groundwork for statistical learning and predictive modeling. They gave both theory and practical advice on how to make strong classifiers, choose the right features, and test models. Their work is the basis for many modern studies that try to predict obesity. It shows how important strict statistical methods are to make sure that models are reliable and can be used in other situations.

Ke et al. [19] created LightGBM, a gradient boosting decision tree structure that makes the best use of training time and memory. LightGBM has become a standard for very fast prediction tasks, such as healthcare-related classification problems like figuring out if someone is overweight. The framework uses histogram-based and leaf-wise tree growth methods to make predictions more accurate and save time on computation. This is especially helpful for big datasets. In addition, Data Headhunters [20] did a comparison of gradient boosting methods, showing how XGBoost and LightGBM differ in terms of speed, scalability, and ability to be used with real-world datasets. Their talk brought up the trade-offs and things to think about when choosing the right boosting algorithms for health analytics. This showed how important algorithmic choice is for getting the best results from predictions.

3. MATERIALS AND METHODS

The suggested method uses the UCI Machine Learning Repository obesity dataset, which has information about people's behaviors, lifestyles, and body measurements. StandardScaler is used to normalize the data, RFECV with Logistic Regression is used to choose the best features, and SMOTENC is used to deal with category imbalance. Stratified 5-Fold Cross-Validation makes sure that the judgment is fair. Adaboost, Perceptron, GaussianNB, SGD, SVM, KNN, MLP, Decision Tree, ExtraTrees, Bagging, Random Forest, GradientBoosting, Logistic RegressionCV, XGBoost, and Light GBM are all part of the model suite. They are put together using Stacking and an Advanced Extended Voting ensemble that combines Gradient Boosting, XGB, LGBM, and CatBoost classifiers. To get explainability, LIME is used for local analysis that doesn't depend on the model and SHAP is used for global feature attribution. The system is set up using the Flask API, which allows for scalable updates, real-time inference, and interactive panels for explainability.

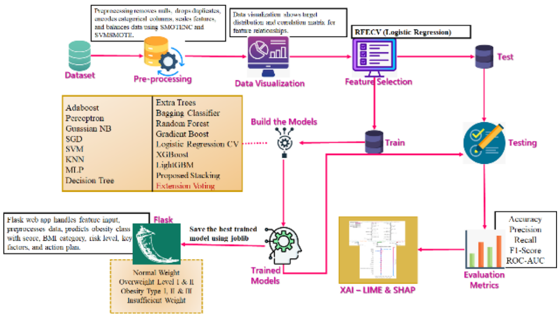
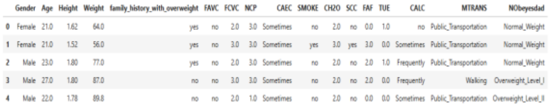


Fig.1 Proposed Architecture

As shown in Fig. 1, the system design starts with getting the dataset. Next comes preprocessing, which gets rid of duplicates and missing values, encodes categorical variables, and uses oversampling to fix class imbalance. Then, data visualization is used to look at how features are distributed and how they relate to each other. Before the data is split into training and testing sets, the input space is narrowed down using RFECV feature selection. Standard performance measures are used to build and test multiple machine learning models. The best learned model is saved until it is time to use it. Explainable AI techniques, such as LIME and SHAP, make things easier to understand, and a Flask-based web app makes it easier to predict obesity risk in real time and connect with users.

a) Dataset Collection:

The UCI Machine Learning Repository has a dataset on obesity with 2,111 entries and 17 characteristics that cover physical, behavioral, and lifestyle factors. It has 8 numerical variables, like age, height, weight, and eating habits, and 9 categorical variables, like gender, family background, smoking, and mode of transportation. The amount of obesity is shown by the target variable, NObeyesdad. This large set of data allows for a thorough study and classification of obesity risk using physiological, demographic, and lifestyle-related markers.



	Gender	Age	Height	Weight	family_history_with_overweight	FAVC	FCVC	NCP	CAEC	SMOKE	CH2O	SCC	FAF	TUE	CALC	MTRANS	NObeyesdad
0	Female	21.0	1.62	64.0	yes	no	2.0	3.0	Sometimes	no	2.0	no	0.0	1.0	no	Public_Transportation	Normal_Weight
1	Female	21.0	1.52	56.0	yes	no	3.0	3.0	Sometimes	yes	3.0	yes	3.0	0.0	Sometimes	Public_Transportation	Normal_Weight
2	Male	23.0	1.80	77.0	yes	no	2.0	3.0	Sometimes	no	2.0	no	2.0	1.0	Frequently	Public_Transportation	Normal_Weight
3	Male	27.0	1.80	87.0	no	no	3.0	3.0	Sometimes	no	2.0	no	2.0	0.0	Frequently	Walking	Overweight_Level_I
4	Male	22.0	1.78	89.8	no	no	2.0	1.0	Sometimes	no	2.0	no	0.0	0.0	Sometimes	Public_Transportation	Overweight_Level_II

Fig.2 UCI Machine Learning Repository

b) Pre-Processing:

The pre-processing phase ensures data quality and readiness for modeling by cleaning, transforming, and analyzing the dataset through structured steps, enabling accurate, balanced, and efficient obesity level prediction.

Exploratory Data Analysis (EDA): With exploratory data analysis, you can find patterns, trends, and links between the factors in the obesity dataset. It has statistical summaries, graphs like count plots, and a study of how the different types of obesity are distributed. EDA finds class imbalances, feature relationships, and possible outliers, which helps us learn more about the psychological and physiological factors that affect obesity. This step helps you make smart choices about feature engineering, model selection, and speed optimization in figuring out the risk of obesity.

Label Encoding: Label encoding turns categorical traits like gender, eating habits, and mode of transportation into numbers so that machine learning can work with them. An integer is given to each unique category so that algorithms can effectively handle categorical data. This step keeps the links between the data while making model input easier. By consistently encoding features, it makes sure that classifiers can correctly understand and analyze patterns based on lifestyle, which leads to more accurate predictions of obesity risk.

Feature Selection: Feature selection finds the most important characteristics that play a big role in classifying fat. It gets rid of factors that aren't needed or don't add much to the model, making it more efficient and easier to understand. Some techniques, such as association analysis, feature importance scores, and recursive selection, help keep the features that have the most predictive power. This optimization makes the model work better, stops it from fitting too well, and makes sure that only important behavioral, dietary, and anthropometric factors are used to accurately predict obesity levels.

Handling Class Imbalance: When different groups of obese people have different sample numbers, it is important to make sure that model learning is fair.

To make the minority and majority groups equal, methods like oversampling, undersampling, or class weighting are used. This stops bias toward the most common amounts of obesity and makes predictions more accurate in all groups. Taking care of imbalance the right way makes sure that the model finds small differences between types of fat, which makes the results of classification more reliable and general for a wide range of people.

c) Training and Testing:

During training and testing, the processed information is split into smaller groups to check how well the model works. During the training phase, algorithms learn how to find patterns and connections in features. The testing phase checks how well they can predict new data. As a result of this step, the model will be more general and won't fit too well. The best classifiers are chosen based on performance measures such as accuracy, precision, recall, and F1-score so that lifestyle and physiological factors can be used to reliably predict obesity levels.

d) Algorithms:

Adaboost: Combines weak learners and focuses on samples that were wrongly classified, which improves the ability to spot minor patterns of obesity risk. Its adaptive weighting takes into account complicated lifestyle factors, which makes it more reliable and helps tell the difference between obesity groups that are very similar.

$$H(x) = sign\left(\sum_{t=1}^{T} \propto_t h_t(x)\right) \quad (1)$$

Perceptron: Helps with baseline comparison by acting as a quick linear model that finds basic trends that can be separated in numeric and encoded features. It makes decisions better by updating weights over and over again and gives basic information for combining with more advanced ensemble models.

GaussianNB: Uses probabilistic modeling for continuous traits like age, height, and weight, which makes it easy to figure out the chances of each class. Its speed and ease of use make it possible to quickly guess the level of obesity while working with more complicated models in integrated ensemble systems.

SGD: Uses gradient-based optimization to learn scalable linear classifiers with mini-batches, which makes it easy to handle large amounts of health data. It quickly adjusts to new inputs, which helps build a strong model and makes predictions more accurate by updating parameters over and over again.

SVM: Uses optimal hyperplanes and kernel functions to maximize separation between obesity groups, making sure that boundaries are accurately detected. It finds both straight-line and curved connections between lifestyle factors, which makes it more accurate at classifying obesity levels that overlap or are closely linked.

$$minimize\ \frac{1}{2}\ ||W||^2 + C\sum_{i=1}^{n} \xi_i \quad (2)$$

KNN: Figures out the level of fat by looking at the closest neighbors and comparing their lifestyles and body measurements. Its non-parametric structure successfully captures local patterns, allowing intuitive, proximity-based predictions for different types of obesity.

$$distance(x, X_i) = \sqrt{\sum_{j=1}^{d} \left(x_j - X_{i_j}\right)^2} \quad (3)$$

MLP: Learns how dietary, behavioral, and physiological factors combine in complex, non-linear ways by using many hidden layers. It works well across different levels of obesity, finding complex feature relationships that make classification more accurate and flexible.

$$\hat{y} = f\left(W^L f\left(W^{L-1} \cdots f(W^1 X + b^1) + b^{(L-1)}\right) + b^L\right) \quad (4)$$

Decision Tree: Uses lifestyle and physical measures to divide data into hierarchical levels that can be used to make rules for categorizing obesity that are easy to understand. It points out important decision points and makes it clear which features have the most impact on classification results.

$$I\ (i) = 1 - \sum_{i=1}^{k} p_i^2 \quad (5)$$

ExtraTrees: Creates several randomized trees to record different ways that features interact while limiting overfitting. Through ensemble average of randomized decision boundaries, it improves the stability of predictions and finds important factors related to obesity.

BaggingClassifier: combines the results of several learners who were trained on bootstrapped groups to lower the range of results and improve stability. It makes balanced predictions about fat, which means that the results can be trusted even when the behavioral and physiological data are different.

Random Forest: Combines several Decision Trees to improve precision and stop overfitting. It handles complicated feature sets well, making them easy to understand by giving each feature a score that shows how important it is in determining obesity.

$$Gini = 1 - \sum_{i=1}^{C} (Pi)^2 \quad (6)$$

Gradient Boosting: builds models in a way that fixes mistakes made in the past while taking into account minor lifestyle and physiological connections. It makes predictions more accurate for hard-to-classify types of obesity and models complicated, non-linear feature interactions well.

$$\hat{y} = f\left(W^L f\left(W^{L-1} \cdots f(W^1 X + b^1) + b^{(L-1)}\right) + b^L\right) \quad (7)$$

LogisticRegressionCV: It does regularized logistic regression and automatically tunes the hyperparameters using cross-validation. It gives us understandable coefficients that connect living factors to the likelihood of being overweight, which makes sure that multi-class classification is reliable and general.

XGBoost: Uses regularization, parallel processing, and efficient gradient boosting to find complex lifestyle relationships. It has good generalization, fast computation, and feature significance that can be understood for finding factors that increase the chance of obesity.

$$\hat{y}_i = \sigma\left(\sum_{k=1}^{K} f_k(x_i)\right), f_k \in F \quad (8)$$

LightGBM: Improves gradient boosting by using leaf-wise growth and histogram-based learning to make classification quick and light on memory. It correctly models the factors that lead to obesity and can handle large amounts of different, complex health data.

Stacking: combines predictions from several base models using a meta-classifier, taking advantage of how their skills complement each other. It raises the accuracy generally, lowers the variation, and gives a full classification for all types of obesity.

$$\hat{y} = g(Y_{base}) = g\left(f_1(x), f_2(x), \dots, f_m(x)\right) \quad (9)$$

Extended Voting: Uses weighted consensus to combine results from the best boosting classifiers. It makes the system more stable, reduces bias, and makes sure that steady, high-accuracy multi-class obesity identification across a wide range of behavioral and physiological datasets.

e) Integration of XAI and Flask Framework:

The Explainable Artificial Intelligence (XAI) and Flask framework work together to make the obesity labeling system more open and easy to use. By using LIME and SHAP, the system gives users understandable information about model results, which helps them figure out how lifestyle, dietary, and physiological factors affect obesity levels. LIME gives reasons for each prediction by looking at how each feature affects all predictions, while SHAP gives reasons for all predictions by measuring how each feature affects them all.

As a simple web framework, Flask links the backend for machine learning to a user interface that can be interacted with. It lets users guess and see XAI results in real time, which makes it easier for them to understand how complex models make decisions. This combination makes AI both understandable and easy to use, so it can make accurate predictions and clear, user-friendly decision support for obesity risk analysis.

4. EXPERIMENTAL RESULTS

Accuracy: How well a test can tell the difference between sick and healthy people is called its accuracy. To get an idea of how accurate a test is, we should figure out what percentage of cases are true positives and true negatives. In terms of math, this can be written as

$$Accuracy = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{FP} + \text{TN} + \text{FN}} (10)$$

Precision: Precision is the percentage of correctly classified cases or samples compared to those that were correctly classified as positives. So, here is the method to figure out the precision:

$$Precision = \frac{\text{True Positive}}{\text{True Positive } + \text{False Positive}} (11)$$

Recall: In machine learning, recall is a metric that shows how well a model can find all the important instances of a certain class. It shows how well a model captures instances of a certain class. It is calculated by dividing the number of correctly predicted positive observations by the total number of real positives.

$$Recall = \frac{\text{TP}}{\text{TP } + \text{ FN}} (12)$$

F1-Score: The F1 score is a way to rate the correctness of a machine learning model. It takes a model's accuracy and recall scores and adds them together. The accuracy metric counts how many times, across the whole dataset, a model made a correct guess.

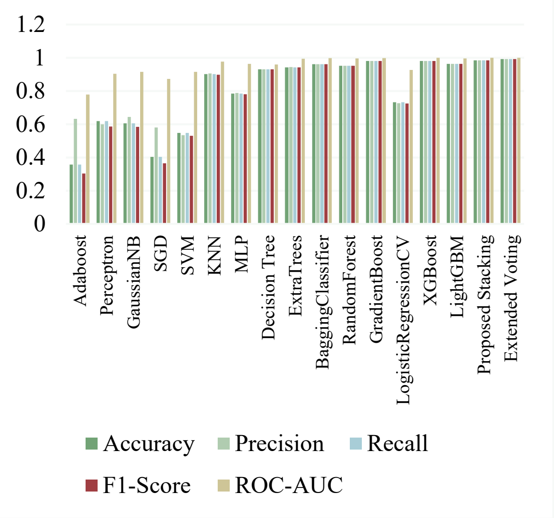
$$F1\ Score = 2 * \frac{Recall\ X\ Precision}{Recall + Precision} * 100 (13)$$

Table.1 Performance Evaluation Table

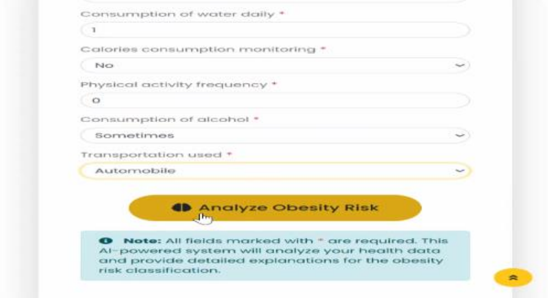
ML Model	Accuracy	Precision	Recall	F1-Score	ROC-AUC
Adaboost	0.358	0.633	0.358	0.304	0.779
Perceptron	0.618	0.600	0.618	0.587	0.903
GaussianNB	0.606	0.643	0.606	0.584	0.915
SGD	0.404	0.581	0.404	0.365	0.872
SVM	0.547	0.535	0.547	0.531	0.915
KNN	0.902	0.906	0.902	0.898	0.977
MLP	0.785	0.788	0.785	0.780	0.964
Decision Tree	0.931	0.931	0.931	0.930	0.960
ExtraTrees	0.943	0.944	0.943	0.943	0.995
BaggingClassifier	0.961	0.962	0.961	0.961	0.998
RandomForest	0.951	0.952	0.951	0.951	0.997
GradientBoost	0.980	0.980	0.980	0.980	0.998
LogisticRegressionCV	0.732	0.726	0.732	0.725	0.927
XGBoost	0.980	0.980	0.980	0.980	0.999
LightGBM	0.963	0.964	0.963	0.963	0.996
Proposed Stacking	0.984	0.984	0.984	0.984	0.999
Extended Voting	**0.993**	**0.993**	**0.993**	**0.993**	**1.000**

Accuracy, Precision, Recall, F1-Score, and ROC-AUC measures are used in Table.1 to show how well the model works. The Extended Voting model does the best overall, with the most accurate classification and the most reliable predictions.

Fig.3 Comparison Graph



The results shown in Fig.3 show how well different machine learning models do at classifying obesity risk. The Extended Voting model does the best in terms of accuracy, precision, recall, F1-score, and ROC-AUC.


Consumption of water daily *
1
Calories consumption monitoring *
No
Physical activity frequency *
0
Consumption of alcohol *
Sometimes
Transportation used *
Automobile

Note: All fields marked with * are required. This AI-powered system will analyze your health data and provide detailed explanations for the obesity risk classification.

Fig.4 Enter Input Data

In Fig.4, the input form lets users give personal and lifestyle data, which makes it easier to classify the risk of obesity based on the data given.

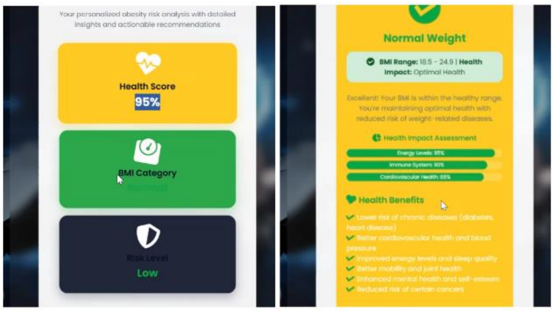
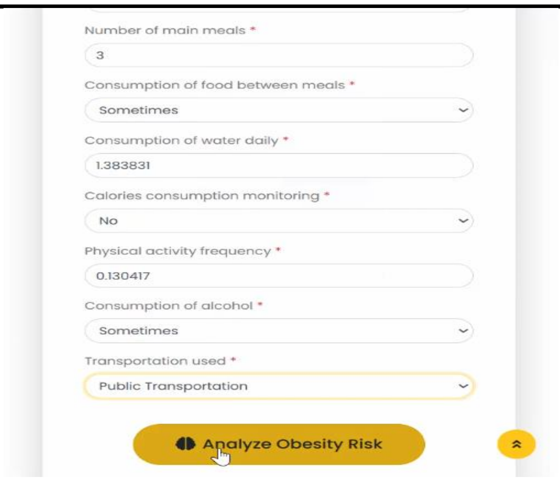

Your personalized obesity risk analysis with detailed insights and actionable recommendations
Health Score
95%
BMI Category
Risk Level
Low
Normal Weight
BMI Range: 18.5 - 24.9 | Health Impact: Optimal Health
Excellent! Your BMI is within the healthy range. You're maintaining optimal health with reduced risk of weight-related diseases.
Health Impact Assessment
Energy Levels: 95%
Immune System: 90%
Cardiovascular Health: 88%
Health Benefits
- Lower risk of chronic diseases (diabetes, heart disease)
- Better cardiovascular health and blood pressure
- Improved energy levels and sleep quality
- Better mobility and joint health
- Enhanced mental health and self-esteem
- Reduced risk of certain cancers

Fig.5 Predicted Results

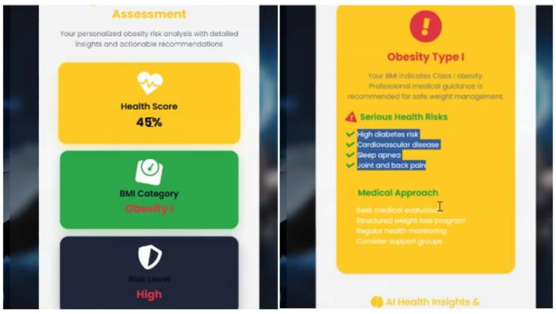
Based on the user's personal and lifestyle data shown in Fig.5, the output screenshot shows that the projected risk of obesity is Normal Weight.



Number of main meals *
3
Consumption of food between meals *
Sometimes
Consumption of water daily *
1.383831
Calories consumption monitoring *
No
Physical activity frequency *
0.130417
Consumption of alcohol *
Sometimes
Transportation used *
Public Transportation

Fig.6 Enter Input Data

Figure 6 shows an input form that lets users enter personal and lifestyle data. This lets the risk of obesity be correctly analyzed and categorized.



Assessment
Your personalized obesity risk analysis with detailed insights and actionable recommendations
Health Score
45%
BMI Category
Obesity I
Risk Level
High

Obesity Type I
Your BMI indicates Class I obesity. Professional medical guidance is recommended for safe weight management.
Serious Health Risks
- High diabetes risk
- Cardiovascular disease
- Sleep apnea
- Joint and back pain

Medical Approach
Seek medical evaluation
Structured weight loss program
Regular health monitoring
Consider support groups

Fig.7 Predicted Results

The projected risk of obesity, shown as Obesity Risk I in Fig.7, was made from the user's input data.

5. CONCLUSION

In conclusion, the framework we created makes it easy to classify the risk of obesity using the UCI Machine Learning Repository obesity dataset, which includes information about eating habits, physical exercise, and body measurements. Standard Scaler is used for normalization, RFECV and Logistic Regression are used for feature selection, and SMOTENC is used to fix category imbalance. Stratified 5-Fold Cross-Validation is used to validate the model. The implementation tests a wide range of algorithms, including Adaboost, Perceptron, GaussianNB, SGD, SVM, KNN, MLP, Decision Tree, ExtraTrees, Bagging Classifier, Random Forest, Gradient Boosting, Logistic RegressionCV, XGBoost, and LightGBM. It also looks at a proposed Stacking model to combine learners that work well together. Explainability in the core pipeline uses LIME to come up with local, model-agnostic explanations. An extra ensemble approach uses an Extended Voting ensemble along with SHAP-based global attributions and Flask deployment as an extra. The finished system is very good at making predictions; it has a ROC-AUC of 100%, 99.3% accuracy, high precision, recall, and F1-score, and it can classify obesity risk in a way that is clear, scalable, and reliable. It can be used to help doctors and the public make decisions about health. Integration with electronic health data and long-term monitoring in the future can make it easier to make proactive interventions and personalized lifestyle suggestions for a lot of people.

Adding real-time data from wearable tech and mobile health apps can make the system even better by keeping an eye on lifestyle and physiological metrics all the time. Adding more genetic, environmental, and social factors could help make predictions more accurate. When used with cloud-based systems, assessing the risk of obesity can be scaled up and made available to more people. Integration with electronic health records would make it easier to tailor interventions to each person, and longitudinal monitoring would let you keep track of each person's success and make suggestions that are more appropriate for them. When you combine the framework with more advanced explainable AI methods than just LIME and SHAP, you can learn more about risk factors that can help you make better clinical decisions and public health plans.

REFERENCES

[1] Wang, Y., Shi, S., Wei, X., Wu, Y., Shi, Y., & Cai, J. (2025). Application and Analysis of Random Forest and Support Vector Classification in Risk Prediction of Childhood Obesity and Hyperuricemia. Diabetes, Metabolic Syndrome and Obesity, 2221-2233.

[2] Tripathi, A. K., Chauhan, N. R., & Sharma, A. (2025, February). Obesity classification and prognosis using machine learning. In AIP Conference Proceedings (Vol. 3224, No. 1, p. 020050). AIP Publishing LLC.

[3] Leitzmann, M. F., Stein, M. J., Baurecht, H., & Freisling, H. (2025). Excess adiposity and cancer: evaluating a preclinical-clinical obesity framework for risk stratification. EClinicalMedicine, 83.

[4] Li, X., & Asemi, A. (2025). A Novel AI-Driven Expert System for Obesity Diagnosis and Personalised Treatment. CAAI Transactions on Intelligence Technology.

[5] Wang, W., Wang, X., Jiang, Y., Guo, Y., Fu, P., He, W., & Fu, X. (2025). Normal weight obesity, circulating biomarkers and risk of breast cancer: a prospective cohort study and meta-analysis. British Journal of Cancer, 132(2), 203-211.

[6] R. Kaur, R. Kumar, and M. Gupta, “Predicting risk of obesity and meal planning to reduce the obese in adulthood using artificial intelligence,” *Endocrine*, vol. 78, no. 3, pp. 458–469, Oct. 2022, doi: 10.1007/s12020-022-03215-4.

[7] F. M. Palechor and A. D. L. H. Manotas, “Dataset for estimation of obesity levels based on eating habits and physical condition in individuals from Colombia, Peru and Mexico,” *Data Brief*, vol. 25, Aug. 2019, Art. no. 104344.

[8] I. G. S. M. Diayasa, M. Idhom, A. Fauzi, and A. T. Damaliana, “Stacking ensemble methods to predict obesity levels in adults,” in *Proc. IEEE 8th Inf. Technol. Int. Seminar (ITIS)*, Oct. 2022, pp. 339–344.

[9] D. D. Solomon, S. Khan, S. Garg, G. Gupta, A. Almjally, B. I. Alabduallah, H. S. Alsagri, M. M. Ibrahim, and A. M. A. Abdallah, “Hybrid majority voting: Prediction and classification model for obesity,” *Diagnostics*, vol. 13, no. 15, p. 2610, Aug. 2023.

[10] G Loge, T Sunil Kumar Reddy, G Swapna, & G Viswanath. (2025). Interpretable AI for Precision Brain Tumor Prognosis: A Transparent Machine Learning Approach. In International Journal of Health Sciences and Pharmacy (IJHSP) (Vol. 9, Number 1, pp. 180–195). Zenodo. https://doi.org/10.5281/zenodo.15523628

[11] S. A. Thamrin, D. S. Arsyad, H. Kuswanto, A. Lawi, and S. Nasir, “Predicting obesity in adults using machine learning techniques: An analysis of Indonesian basic health research 2018,” *Frontiers Nutrition*, vol. 8, Jun. 2021, doi: 10.3389/fnut.2021.669155.

[12] Z. Zheng and K. Ruggiero, “Using machine learning to predict obesity in high school students,” in *Proc. IEEE Int. Conf. Bioinf. Biomed. (BIBM)*, Nov. 2017, pp. 2132–2138.

[13] I. Machorro-Cano, G. Alor-Hernández, M. A. Paredes-Valverde, U. Ramos-Deonati, J. L. Sánchez-Cervantes, and L. Rodríguez-Mazahua, “PISIoT: A machine learning and IoT-based smart health platform for overweight and obesity control,” *Appl. Sci.*, vol. 9, no. 15, p. 3037, Jul. 2019.

[14 A D Venkatesh, K Bhaskar, G Swapna, & G Viswanath. (2025). Advanced Hybrid Learning Architecture for Precision Cardiovascular Risk Assessment. In International Journal of Health Sciences and Pharmacy (IJHSP) (Vol. 9, Number 1, pp. 50–61). Zenodo. https://doi.org/10.5281/zenodo.15448632

[15] M. J. Raihan, M. A. M. Khan, S. H. Kee, and A. A. Nahid, “Detection of the chronic kidney disease using XGBoost classifier and explaining the influence of the attributes on the model using SHAP,” *Sci. Rep.*, vol. 13, no. 1, p. 6263, Apr. 2023, doi: 10.1038/s41598-023-33525-0.

[16] S. Jahan, K. A. Taher, M. S. Kaiser, M. Mahmud, M. S. Rahman, A. S. M. S. Hosen, and I. Ra, “Explainable AI-based Alzheimer’s prediction and management using multimodal data,” *PLOS ONE*, vol. 18, Nov. 2023, Art. no. 0294253.

[17] D. Dua and C. Graff, “UCI machine learning repository,” Irvine, School Inf. Comput. Sci., Univ. California, Tech. Rep., 2017. [Online]. Available: http://archive.ics.uci.edu/ml

[18] Naresh, M., Gudditti., M., Viswanath, & SunilKumarReddy, M.T. (2014). Distributed Utility-Based Energy Efficient Cooperative Medium Access Control in MANETS.

[19] G. Ke, Q. Meng, T. Finley, T. Wang, W. Chen, W. Ma, Q. Ye, and T. Liu, “LightGBM: A highly efficient gradient boosting decision tree,” in *Proc. 31st Int. Conf. Neural Inf. Process. Syst.*, Red Hook, NY, USA: Curran Associates Inc., Dec. 2017, pp. 3149–3157.

[20] XGboost Vs Lightgbm: Gradient Boosting in the Spotlight. Accessed: Jul. 16, 2024. [Online]. Available: https://dataheadhunters.com/academy/xgboost-vs-lightgbm-gradient-boosting-in-the-spotlight/

[21] L. Breiman, J. H. Friedman, R. A. Olshen, and C. J. Stone, *Classification and Regression Trees*. Monterey, CA, USA: Wadsworth and Brooks, 1984.

[22] M. Azad, I. Chikalov, S. Hussain, and M. Moshkov, “Multi-pruning of decision trees for knowledge representation and classification,” in *Proc. 3rd IAPR Asian Conf. Pattern Recognit. (ACPR)*, Kuala Lumpur, Malaysia, Nov. 2015, pp. 604–608.

[23] Ganesh, B. R. ., B M, P., Prasad K, K. ., Swapna, G., & G, Viswanath. (2025). Data Mining-Driven Multi-Feature Selection for Chronic Disease

Forecasting. Journal of Neonatal Surgery, 14(5S), 108–124. https://doi.org/10.52783/jns.v14.1993

[24] F. Alsolami, M. Azad, I. Chikalov, and M. Moshkov, *Decision and Inhibitory Trees and Rules for Decision Tables With Many-Valued Decisions* (Intelligent Systems Reference Library), vol. 156. Cham, Switzerland: Springer, 2020.

[25] M. Azad, I. Chikalov, and M. Moshkov, "Decision trees for knowledge representation," in *Proc. 28th Int. Workshop Concurrency, Specification Program.* (CS&P), vol. 2571, K. Ropiak, L. Polkowski, and P. Artiemjew, Eds. Olsztyn, Poland: CEUR-WS.Org, Sep. 2019. [Online]. Available: http://ceur-ws.org/Vol-2571/CSP2019_paper_1.pdf

[26] S. Rahman, M. Irfan, M. Raza, K. M. Ghori, S. Yaqoob, and M. Awais, "Performance analysis of boosting classifiers in recognizing activities of daily living," *Int. J. Environ. Res. Public Health*, vol. 17, no. 3, p. 1082, Feb. 2020. [Online]. Available: https://www.mdpi.com/1660-4601/17/3/1082

[27] Kumar, K., Udaya Suriya Rajkumar, D., Viswanath, G., & Mahalakshmi, J. (2024). A Hybrid Particle Swarm Optimization and C4.5 for Network Intrusion Detection and Prevention System. International Journal of Computing, 23(1), 109-115. https://doi.org/10.47839/ijc.23.1.3442

[28] Z. A. Memish, C. El Bcheraoui, M. Tuffaha, M. Robinson, F. Daoud, S. Jaber, S. Mikhitarian, M. Al Saeedi, M. A. AlMazroa, A. H. Mokdad, and A. A. Al Rabeeah, "Obesity and associated factors—Kingdom of Saudi Arabia, 2013," *Preventing Chronic Disease*, vol. 11, p. E174, Oct. 2014, doi: 10.5888/pcd11.140236.

[29] F. A. Hamam, A. S. Eldalo, A. A. Alnofeie, W. Y. Alghamdi, S. S. Almutairi, and F. S. Badyan, "The association of eating habits and lifestyle with overweight and obesity among health sciences students in Taif University, KSA," *J. Taibah Univ. Med. Sci.*, vol. 12, no. 3, pp. 249–260, Jun. 2017.

[30] S. K. Keadle, R. McKinnon, B. I. Graubard, and R. P. Troiano, "Prevalence and trends in physical activity among older adults in the United States: A comparison across three national surveys," *Preventive Med.*, vol. 89, pp. 37–43, Aug. 2016.