Received: 8 August 2025 | Revised: 25 December 2025 | Accepted: 30 January 2026 | Published online: 11 February 2026

RESEARCH ARTICLE

Artificial Intelligence and Applications
2026, Vol. 00(00) 1–10
DOI: 10.47852/bonviewAIA62027134

BON VIEW PUBLISHING

A Multi-class Obesity Risk Prediction Using Machine Learning and Explainable Artificial Intelligence

Tarequl Hasan Sakib[1] **and Mahfuzulhoq Chowdhury**[1,*]

[1]*CSE Department, Chittagong University of Engineering and Technology, Bangladesh*

Abstract: Because of the intricate relationships between dietary practices, physical characteristics, and lifestyle, weight-related health issues have grown to be a global concern. The shortcomings of current machine learning (ML) techniques include inefficient feature selection, imbalanced datasets, a binary classification focus, decreased accuracy, and inadequate hyperparameter tweaking. This paper uses a clinically validated dataset of 1,638 patients from Bangladeshi healthcare facilities to provide a complete framework for ML-based multiclass obesity risk prediction in order to fill these gaps. The proposed approach combines 5-fold cross-validation with GridSearchCV for systematic hyperparameter tuning and ensemble feature selection. The Synthetic Minority Over-sampling Technique was used just on the training set to address class imbalance and guarantee balanced learning across the seven weight categories. The proposed XGBoost outperformed the other ML algorithms that were assessed for obesity risk prediction due to their high accuracy score of 95.4%. According to the results, the proposed method outperformed previous works by at least 5.18% in accuracy increase and 10.30% in F1-score gain. Furthermore, explainable Artificial Intelligence methods based on SHAP offer insights into the decision-making process through feature contributions.

Keywords: obesity risk prediction, ensemble feature selection, hyperparameter tuning, XGBoost, explainable AI (XAI)

1. Introduction

Obesity has been a major worldwide health concern throughout the last 30 years. Over 2.5 billion people are overweight, up from 25% in 1990, and over one billion people globally suffer from obesity as of 2022, making up over 16% of the adult population [1, 2]. Obesity rates have tripled or quadrupled in certain countries, and this growth has increased among children and adolescents. According to projections, by 2055, the disease related to obesity could increase by an additional 30.7% [3]. Around 9% of individuals worldwide suffer from undernutrition, which is more common in some South Asian countries and at the other end of the BMI spectrum [3, 4]. Weakened immunity, osteoporosis, and compromised reproductive health are intimately associated with this illness [4, 5]. Bangladesh is a striking example of the “double burden” of malnutrition, when states such as underweight and overweight coexist. A study of 2,690 urban schoolchildren and adolescents in Dhaka city found that 19% were underweight, 15%–17% overweight, and 6% obese [6]. Models that take into account all body mass index (BMI) categories and their dynamic interactions are necessary to address the complicated health demands created by such concurrent changes [7–11]. The number of people with obesity increased from 10% to 24% within 2004 and 2014 [8]. Although BMI is a commonly used classification tool, it defines categories: normal weight level (18.5 to below 25), overweight level (25 to below 30), underweight level (<18.5), and obesity level I–III (≥30) [11]. No appropriate machine learning (ML) model for predicting obesity was included in some previous publications [10, 11, 12]. These oversimplified models have poor accuracy and inadequate handling of imbalances, and they frequently ignore important components like strict feature selection and interpretability. Limited geographic datasets were used in the obesity prediction research that was previously conducted [13–21]. Across a range of demographic groups, they mostly concentrated on binary classification schemes and the uneven application of thorough lifestyle evaluation procedures. Although existing studies include demographic, physical, and behavioral characteristics, they frequently overlook important lifestyle drivers like thorough eating habits, modes of mobility, and family history aspects because they lack systematic techniques for ensemble feature selection and relevance ranking. Black-box models predominate without sufficient interpretability frameworks, restricting clinical applicability and healthcare provider trust. Additionally, the majority of ML-based existing works [13–21] concentrate on binary obesity classification (obese vs. non-obese) rather than comprehensive weight spectrum analysis encompassing underweight to severe obesity categories. They did not use any ensemble feature selection techniques, advanced data preprocessing, class imbalance handling, or hyperparameter tuning techniques. The prediction accuracy of the existing works is poor. Additionally, they didn’t create any web or mobile applications for predicting the health risks associated with obesity. This paper presents a ML based obesity risk prediction method by using an XGBoost classifier. This paper investigates different underweight, normal weight, obese, and overweight categories. The proposed work used ensemble feature selection approaches to offer higher prediction results. While SHAP-based explainable Artificial Intelligence (AI) algorithms highlight the contributions of individual features, the Synthetic Minority Over-sampling Technique (SMOTE) is utilized to address the imbalance seen in lower-frequency classes.

The key contributions of this paper are illustrated as follows:

1) A comprehensive dataset of 1,638 people was gathered from healthcare facilities in Bangladesh, encompassing a variety of real-world lifestyles across a range of demographic groupings. The dataset includes 16 physical and lifestyle characteristics that have been verified by three medical experts.

*Corresponding author: Mahfuzulhoq Chowdhury, CSE Department, Chittagong University of Engineering and Technology, Bangladesh. Email: mahfuz@cuet.ac.bd

2) This paper presents a brand-new five-method ensemble feature selection architecture for suitable feature selection that combines Borda Count Aggregation, Recursive Feature Elimination (RFE), SHapley Additive exPlanations (SHAP), Analysis of Variance (ANOVA) *F*-Test, and Chi-Square Test.
3) The proposed ML-based framework enables comprehensive classification across seven distinct weight categories. This work selects the best ML model by evaluating the accuracy and F1-score of the examined ML models.
4) Additionally, advanced preprocessing and hyperparameter tuning methods were used in the proposed model. Using SHAP analysis, explainable AI integration offers both local and global interpretability. React Native was used to create a useful mobile application for assessing obesity risk in real time.

A synopsis of relevant research is provided in Section 2. The proposed methodology for classifying obesity risk is explained in Section 3. The comparison results are displayed in Section 4. The work's conclusion, which summarizes its main conclusions, is found in Section 5.

2. Literature Review

The literature on predicting the risk of obesity is highlighted in this section. Parra et al. [12] employed a binary classification approach to predict middle-aged persons in Spain who were overweight or obese. Both normal and obese subjects were evenly distributed throughout the dataset. After selecting evolutionary features using a genetic algorithm, researchers tested 10 ML techniques. They achieved 79 percent accuracy by employing the random forest (RF) algorithm. Although the study performed well, it was constrained by its binary categorization method, which did not differentiate between various levels of obesity severity. Jeon et al. [13] examined age-specific metabolic risk variables for obesity prediction. The RF showed better accuracy in its work, while the MLP model had the highest AUC scores.

Calderón-Díaz et al. [14] used multiple ML models for obesity prediction. The dataset included lipid profile and biochemical information on 21 variables, including VLDL cholesterol, bilirubin, glycemia, and total cholesterol. With 100 iterations and 80% cross-validation, the model's accuracy was 87.5%. The study's shortcomings included the removal of finer-grained obesity stages and the lack of lifestyle-related indicators, which may have limited the comprehensiveness of obesity risk modeling even if it demonstrated high performance. Lim et al. [15] took into account mother characteristics (e.g., BMI and self-esteem) as well as lifestyle choices (e.g., eating habits and physical activity). It demonstrated the predictive power of family and environmental effects with an accuracy of 74%. An ML model utilizing the Cat Boost algorithm was created by Lin et al. [16] to forecast the risk of obesity in persons who are overweight. The model made use of a number of physiological and physical characteristics, such as gender, systolic blood pressure, and waist and hip size. With an AUC of 0.87 on the test dataset, it performed well and showed good predictive power. The study was limited, though, in that it only employed the Cat Boost model, making it unable to compare it to other methods. Furthermore, the study excluded underweight participants and only examined overweight and obese people, which limited the prediction framework's comprehensiveness. To categorize adults as overweight/obese or normal weight, Gutiérrez-Gallego et al. [17] suggested a cascaded ensemble ML model. There were 1,179 Madrid residents in the dataset. The model's efficacy in binary classification was demonstrated by its 79% overall accuracy, 84% precision, and 89% recall. Using RISKESDAS 2018 data from over 618,898 Indonesian individuals, a large-scale binary classification study by Thamrin et al. [18] found that 21.8% of them were obese. The best result was demonstrated by logistic regression, which had an accuracy of about 72%. The study's limitation is its binary result (obese vs. non-obese), which ignores granularity in obesity levels and underweight classification, even with the huge dataset and fair model performances. The goal of the study of Singh and Tawfik [19] was to use ML for youth overweight and obesity prediction. Eleven thousand one hundred ten samples from three classes—normal (8,160), overweight (2,126), and obese (824)—were included in the collection. Despite the use of several algorithms, the models' accuracy was often above 80%. Recall for the overweight class, however, was just 62%, indicating issues with unbalanced data that the study failed to sufficiently address. Rousset et al. [20] used behavioral and personal data to create a neural network model that predicts weight status. Additionally, the study's relevance was limited because underweight people were not taken into consideration. Dirik [21] presented an RF algorithm to predict obesity. The model's remarkable accuracy of 95.78% was attained. Although this indicates significant predictive capacity, the study ignored the underweight category in favor of focusing primarily on differentiating between overweight and obese people. The clinical efficacy of thorough weight health assessments is diminished by this lack of class granularity, which also ignores the entire range of dietary health hazards. In the articles of Dutta et al. [22], Görmez et al. [23], and Helforoush and Sayyad [24], the multiclass obesity prediction problem was not examined. Adaboost and MLP algorithms were employed in the article of Devi et al. [25] to predict schoolchildren's obesity. The issues of handling imbalances and multiclass were not addressed. Big data analytics was used by Vemulapalli et al. [26] to anticipate obesity trends. They did offer comparison results for several ML techniques. DL techniques were employed in the study of Jeong et al. [27] to predict teenage obesity. Additionally, they failed to provide an appropriate feature selection scheme and data validation. The article by Nalini et al. [28] used body fat percentage data to predict overweight. Titu et al. [29] discussed the negative effects of advertisements for unhealthy foods using the Explainable AI approach. The decision tree algorithm, habit, and health data were used in the paper of Bahrin [30] to forecast the type of body (healthy or sick). Garcia-d'Urso et al. [31] used an ML algorithm to predict an individual's total cholesterol level. It is evident from the previous discussion that the majority of works concentrate on binary classification. Their obesity risk prediction accuracy results are also not proper. They did not use any appropriate feature selection, explainable AI-based feature impact analysis, hyperparameter tweaking, class imbalance handling, or advanced preprocessing strategies.

In order to overcome these drawbacks, this study presents an obesity risk prediction dataset that includes standardized lifestyle assessment procedures, integrated behavioral and physical profiling, and rigorous ensemble feature engineering that combines Borda Count Aggregation, ANOVA *F*-Test, RFE, SHAP analysis, and Chi-Square. The objective of this work is to close the gap between algorithmic performance and practical clinical decision-making in preventive healthcare and customized obesity risk management by creating interpretable models with explainable AI approaches and thorough seven-category weight classification.

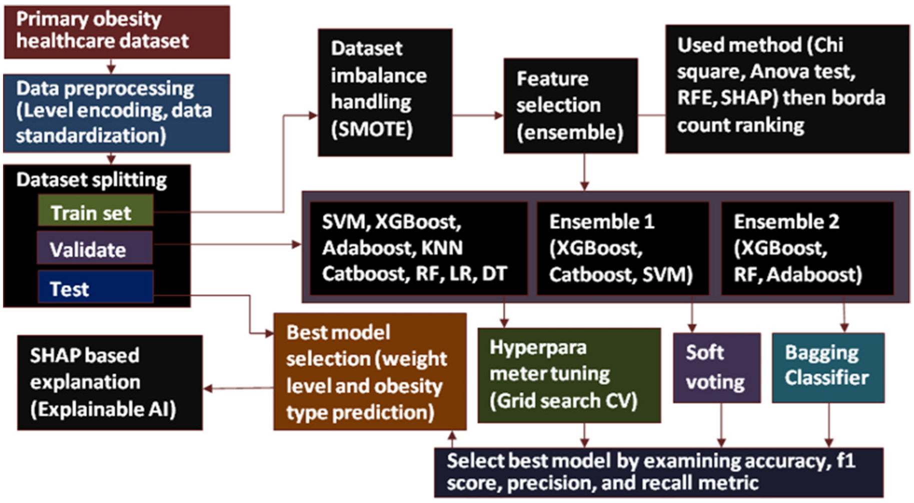
3. Proposed ML Framework for Obesity Risk Prediction

This section presents the overall methodology for ML-based obesity risk prediction with detail discussion.

3.1. System design

The proposed methodology model for obesity risk prediction is displayed in Figure 1. This methodical process consists of five separate stages, each of which adds crucial elements to the finished prediction system. The approach starts with thorough data gathering using standardized questionnaires given to participants in different

Figure 1
Proposed system methodology for multi-class obesity risk prediction

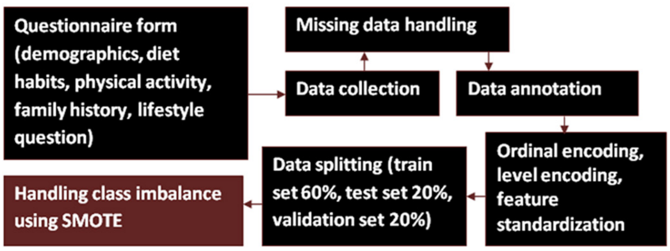


Bangladeshi healthcare facilities. Demographic data, food habits, physical work, lifestyle patterns, family, and personal health issues are all covered in the questionnaire. As seen in the first stage of our approach, a thorough preprocessing pipeline after data collection guarantees data quality and consistency. This covers dealing with missing values and carrying out the necessary data conversions. To confirm the obesity categories in accordance with accepted clinical recommendations, the dataset is subsequently annotated by qualified medical professionals. The framework then moves on to feature selection, employing a thorough ensemble approach that combines five different techniques: Borda Count voting for consensus-based aggregation, ANOVA *F*-Test, RFE, SHAP Global Importance, and Chi-Square Test. This multi-method strategy minimizes selection bias while ensuring robust feature detection, as illustrated in Figure 1. RF, XGBoost, Support Vector Machine (SVM), and K-Nearest Neighbor (KNN) are among the ML algorithms that are methodically implemented as the technique moves on to model training and evaluation. The framework uses cross-validation approaches to optimize hyperparameters and integrates evaluation of individual models and ensemble methods. The latter section uses explainable AI methods, specifically SHAP analysis, to offer clear insights into how the model makes decisions.

3.2. Dataset collection, dataset preparation

The thorough dataset preparation pathway for the study on obesity risk prediction is shown in Figure 2. We created a thorough survey using Google Forms in order to create a solid and clinically useful dataset for classifying obesity risk. In order to create the questionnaire, items from well-known research studies on obesity and lifestyle assessment were analyzed and integrated. We developed a thorough questionnaire by methodically examining overlapping and redundant items through a thorough evaluation of studies on lifestyle factors and obesity prediction, drawing on the validated study findings by Dutta et al. [22], Görmez et al. [23], and Helforoush and Sayyad [24].

Figure 2
Data preparation procedure



The study incorporates key lifestyle characteristics that are known to affect the risk of obesity, such as eating patterns, physical work issues, health and disease information, and pertinent demographic information. Medical experts were consulted during the assessment and improvement of the questionnaire to guarantee that it was clear and appropriate for the Bangladeshi population. The Google Form was disseminated throughout Bangladesh through clinics, community health centers, and medical facilities. All participants gave their informed consent in accordance with stringent ethical requirements, and the questionnaire was completed remotely or during clinical appointments. One thousand six hundred thirty-eight people in all took part, offering thorough information on all the features that were mentioned. Many aspects of lifestyle, physical traits, and behavioral patterns that may affect the risk of obesity are included in the extensive feature set gathered by our questionnaire. The collected dataset section is displayed in Tables 1 and 2. To guarantee dataset quality and clinical accuracy, a methodical annotation and validation procedure was put in place after the data gathering phase. In order to validate obesity diagnoses, the annotation process involves working with trained healthcare specialists, such as general practitioners and nutritionists, who examined each participant's data. In order to guarantee uniformity and dependability, disagreements among reviewers were settled by a consensus procedure that included group deliberation and majority voting. The expert's comments are taken into account to avoid bias, sample representativeness, and demographic balance in the final dataset. Table 3 provides an illustration of the dataset annotation procedure by experts. Table 3 also shows how weight groups are given based on BMI thresholds, and expert assessment provides an illustration of this classification procedure. The obesity risk dataset gathered for this study includes a thorough sample of 1,638 people who span a wide range of weight categories, from underweight to severe obesity.

3.3. Data preprocessing and feature engineering

To change the raw data into a proper format for ML-based prediction work, the gathered dataset underwent methodical preprocessing. To

Table 1
Portion of the collected dataset (first part)

Row	Gender	Age	Height	Weight	Family history obesity	High-calorie food intake	Vegetable consumption frequency	Daily meal count
1	F	21	1.62	64	Yes	No	2	3
2	F	21	1.52	56	Yes	No	3	3
3	M	23	1.80	77	Yes	No	2	3
4	M	27	1.80	87	No	No	3	3
5	M	22	1.78	89.8	No	No	2	1

Table 2
Portion of collected dataset (remaining part)

Snacks eat frequency	Smoking status	Daily water consumption	Calorie intake tracking	Physical activity frequency	Screen time duration	Beverage consume frequency	Mode of transportation
Sometimes	No	2	No	Low	1	No	Public
Sometimes	Yes	3	Yes	High	0	Sometimes	Motor bike
Sometimes	No	2	No	Medium	1	Frequently	Public
Sometimes	No	2	No	Medium	0	Sometimes	Walking
Sometimes	No	2	No	Low	0	Sometimes	Motor bike

Table 3
Annotation process of the dataset

Target	Annotator 1	Annotator 2	Annotator 3	Weight	Height	Weight category
Normal weight	Insufficient weight	Normal weight	Normal weight	45	1.65	Insufficient weight
Overweight Level I	Overweight Level I	Overweight Level I	Normal weight	65	1.70	Normal weight
Insufficient weight	Insufficient weight	Insufficient weight	Normal weight	72	1.75	Normal weight
Obesity Type I	Overweight Level I	Obesity Type I	Obesity type I	91	1.68	Obesity Type I
Obesity Type I	Overweight Level II	Obesity Type I	Obesity type I	96	1.60	Obesity Type II

guarantee the best possible model performance, this preprocessing stage concentrated on feature scaling, categorical variable encoding, and data quality verification. Following the analysis, we eliminated the data that had null values. Using label encoding techniques, categorical features were methodically converted into numerical representations. Although allowing ML algorithms to efficiently process the data, this encoding approach maintained the significant correlations within categorical categories. Label encoding was used for lifestyle-related categorical variables based on patterns of intensity or frequency.

A systematic numerical mapping that maintains the clinical progression from underweight to severe obesity conditions was used to encode the target variable representing weight categories. To guarantee consistent contributions throughout model training, continuous numerical variables were standardized. In order to prevent variables with greater scales from controlling the learning process, this scaling strategy converts all numerical features to have a zero mean and unit variance.

Continuous variables such as age, height, weight, daily water intake, and screen time duration were all subjected to this standardization process. The scaling procedure guarantees that gradient descent optimization methods and distance-based algorithms can function efficiently without favoring features with varying distributional properties or greater numerical ranges by nature. With percentages ranging from 5.9% to 27.6% across categories, the sample exhibits significant class imbalance characteristics. In order to guarantee balanced representation throughout training procedures, this imbalance pattern calls for specific handling strategies during model construction, especially the application of synthetic minority oversampling approaches.

The class distribution on the training set before and after SMOTE application is shown in Table 4. Effective pattern recognition across all weight categories is made possible by the balanced training set, which also maintains the natural class distribution in the test and validation sets for accurate performance assessment.

The Chi-Square Test, ANOVA *F*-Test, RFE, SHAP, and ensemble aggregation were among the competitive feature selection techniques

Table 4
Class distribution before and after SMOTE apply

Weight category	Before SMOTE	After SMOTE
Normal weight	271	271
Overweight Level I	185	271
Obesity Type I	130	271
Overweight Level II	133	271
Insufficient weight	136	271
Obesity Type II	70	271
Obesity Type III	58	271

Note: SMOTE = Synthetic Minority Over-sampling Technique.

used to select the best features for training. We used the Borda count aggregation technique to aggregate the outcomes of several feature selection techniques in order to guarantee stable and trustworthy feature selection. Borda count gathers the ranks from every technique and generates a final rating based on their combined agreement rather than depending on just one. Every approach (e.g., Chi-Square, ANOVA, RFE, and SHAP) ranks each characteristic separately. The final Borda score is calculated by adding the ranks for each feature. Higher priority is given to features with lower overall scores.

Table 5
Feature importance score

Features	Chi-square	ANOVA	RFE	SHAP
Weight	1,247	89	0.28	0.44
High-calorie food intake	734	65	0.07	0.27
Physical activity frequency	658	58	0.08	0.29
Height	892	71	0.14	0.31
Vegetable consumption frequency	356	44	0.04	0.21
Snacks eating frequency	467	48	0.08	0.24
Age	589	52	0.03	0.10
Beverage consumption frequency	298	39	0.09	0.14
Family history of obesity	289	41	0.04	0.12
Screen time duration	225	36	0.03	0.06
Mode transportation	198	28	0.05	0.03
Daily water consumption	41	16	0.03	0.02
Daily meal count	164	32	0.06	0.21
Gender	87	19	0.05	0.03
Calorie intake tracking	76	22	0.08	0.02
Smoking status	52	25	0.02	0.01

Note: ANOVA = Analysis of Variance, RFE = Recursive Feature Elimination, SHAP = SHapley Additive exPlanations.

Table 6
Feature importance ranking

Features	Chi-square rank	ANOVA rank	RFE rank	SHAP rank	Ensemble rank
Weight	1	1	1	1	1
Height	2	3	2	2	2
High calorie food intake	3	2	5	4	3
Physical activity frequency	4	4	3	3	4
Snacks eating frequency	9	6	4	5	5
Age	5	5	7	10	6
Beverage consumption frequency	8	8	12	8	7
Vegetable consumption frequency	7	7	9	6	8
Family history obesity	6	11	8	7	9
Screen time duration	11	9	11	9	10
Mode transportation	10	12	10	12	11
Daily meal count	12	10	6	11	12
Gender	13	15	13	13	13
Calorie intake tracking	16	13	15	14	14

Note: ANOVA = Analysis of Variance, RFE = Recursive Feature Elimination, SHAP = SHapley Additive exPlanations.

For the selected ML model, the features with the highest rankings are chosen. Only the chosen features are then used to validate the model's performance. Tables 5 and 6, respectively, provide thorough documentation of the resulting feature subset as well as specific relevance scores and ranks. We carried out a systematic examination across various feature subset sizes to ascertain the ideal amount of features to keep. We assessed model performance at different feature subset sizes using RF as the basic classifier because features cannot be removed at random. We determined the saturation point at which additional feature removal either no longer significantly improves or starts to damage model performance by tracking the accuracy, F1-score, precision, recall, and Receiver Operating Characteristic-Area Under Curve (ROC-AUC) throughout these iterations. The top nine features are selected. A balanced distribution approach is used in the dataset splitting strategy: A total of 983 samples, or 60% of the data, were used for model training, which included both algorithm fitting and initial parameter learning. Twenty percent (328 samples) was put aside as a validation set to track model performance throughout training and inform choices about hyperparameter adjustment. The test set containing the remaining 20% (327 samples) was kept for the last, objective assessment of the model's generalization ability. In our methodology, the validation set has two functions: it facilitates hyperparameter adjustment through cross-validation techniques and tracks training progress to avoid overfitting.

3.4. Model selection, hyperparameter tuning, and SHAP-based model explain ability

This paper used the Grid Search CV hyperparameter tuning strategy to maximize the performance of the proposed ML models regarding obesity risk category prediction, along with k cross-validation ($k = 5$). Based on the highest weighted F1-score, the optimal hyperparameters were chosen. Table 7 shows the whole hyperparameter

Table 7
Hyperparameter grid for ML model

Model	Parameters	Values
XGBoost	N_estimators, max depth, learning rate	100, 5, .05
Random forest	N_estimators, max depth, min samples split, min samples leaf, bootstrap	100, 20, 10, 4, true
Decision tree	Max depth, min samples split, min samples leaf, criterion	20, 10, 4, gini
Logistic regression	C, penalty, solver, max iter	1, l2, lbfgs, 200
Catboost	Iterations, depth, learning rate	200, 8, .05
SVM	C, kernel, gamma	1, linear, auto
KNN	N_neighbors, weight, metric	9, uniform, minkowski
Adaboost	N_ estimator, learning rate, estimator	100, .05, DT classifier (max depth = 3)

Note: KNN = K-Nearest Neighbor, ML = machine learning, SVM = Support Vector Machine.

grid for each of the ML models that were studied (XGBoost, RF, DT, CatBoost, AdaBoost, LR, KNN, and SVM). We used sophisticated ensemble approaches to take advantage of the combined power of several algorithms, going beyond the optimization of individual ML models.

In order to investigate different elements of model performance and robustness in obesity risk prediction, two separate ensemble approaches were used: the Bagging Classifier (XGBoost, RF, and AdaBoost) and the Soft Voting Classifier (XGBoost, CatBoost, and SVM). This study looked into eight individual ML models and two ensemble models to determine which one was better for the obesity risk prediction task. A thorough comparison of several ML algorithms was carried out after the most informative features were chosen by incorporating the proposed ensemble feature selection technique (Table 8). This study assesses the performance of the ML models using a number of significant criteria, including accuracy value and F1-score value. The ratio of all correct predictions to all instances is used to calculate accuracy for multiple classes.

The ratio of true positives to the total number of positives (true and false positives) is used to determine the precision for a specific class. The true positive rate, which provides the ratio of real positives to both true positives and false negatives, is what defines recall. The harmonic mean of recall and precision is the F1-score. The total multi-class ROC-AUC is often calculated by averaging these per-class AUC scores, commonly via macro or weighted averaging. For multi-class issues, this work initially determined the precision and recall values for each class. The final macro-averaged precision and recall scores are then obtained by averaging the accuracy and recall across all classes. The XGBoost model emerged as a suitable ML model for the obesity risk prediction task (Table 8) with the highest F1-score value (95.3%) and highest accuracy value (95.4%). Out of the seven weight categories in the dataset, XGBoost offered the most accurate and balanced classification performance. For both common and minority classes, the model continuously produced good precision and recall values, demonstrating its capacity to differentiate between various obesity risk levels while preserving clinical dependability. The model's accurate categorization across all weight categories is demonstrated by the confusion matrix in Figure 3 illustrates that the proposed XGBoost model is most suitable for the obesity risk prediction task. Figure 4 illustrates the SHAP-based explanation of how important each feature regarding obesity prediction process. A discrete hierarchy of feature relevance is revealed by the investigation, with lifestyle and demographic factors playing different roles. With a mean SHAP value of almost 0.20, weight is the most significant predictor, which is consistent with its primary function in calculating BMI and classifying obesity in general. With a mean SHAP value of 0.17, height comes in second place, emphasizing how crucial anthropometric measurements are in determining weight category. With a mean SHAP value of 0.13, High Caloric Food Intake had the highest predictive power among lifestyle factors, highlighting the clear link between dietary practices and the risk of obesity. With a value of 0.11, Physical Activity Frequency comes in at number four, highlighting the protective effect of consistent exercise against weight-related health problems.

Table 8
Performance comparison among ML model

Model	Accuracy	Precision	Recall	F1-score	ROC-AUC
XGBoost	0.954	0.944	0.963	0.953	0.981
Random forest	0.942	0.923	0.962	0.947	0.971
CatBoost	0.942	0.965	0.927	0.945	0.974
SVM	0.931	0.918	0.953	0.933	0.971
AdaBoost	0.926	0.908	0.941	0.924	0.967
Logistic regression	0.918	0.891	0.932	0.913	0.956
Decision tree	0.845	0.813	0.872	0.841	0.897
KNN	0.812	0.788	0.847	0.815	0.873
Ensemble 1 (XGBoost, CatBoost, SVM)	0.941	0.933	0.943	0.938	0.971
Ensemble 2 (XGBoost, random forest, AdaBoost)	0.943	0.939	0.931	0.934	0.973

Note: KNN = *K*-Nearest Neighbor, ML = machine learning, ROC-AUC = Receiver Operating Characteristic - Area Under Curve, SVM = Support Vector Machine.

3.5. Mobile app development

The architecture of the mobile application adheres to an organized user journey that places a high value on usability while upholding scientific rigor in data collecting. Figure 5 shows the flowchart for our suggested mobile application.

Through several stages of data collecting, the process flow illustrates the application's systematic approach to obesity risk prediction. Following authentication, users proceed through the evaluation of their food habits, physical activity, and basic information entry (demographics and physical measures). After finishing, the system uses contributing factor analysis to generate weight category forecasts, and users can store their results in the Firebase database for further usage and progress monitoring. React Native and Firebase technologies are used in the development of the application. As shown in Figure 6(a), the program starts with a thorough welcome screen that acts as the

Figure 3
Confusion matrix for the proposed XGBoost model

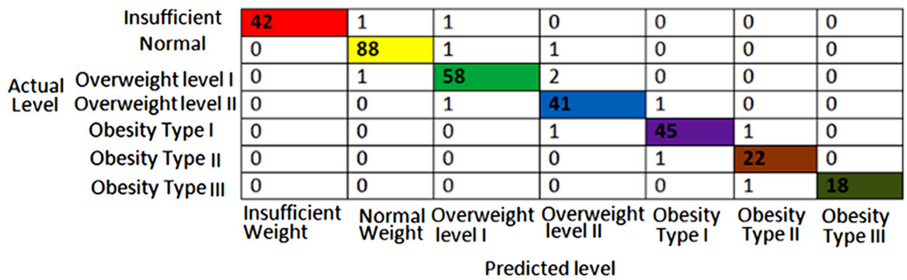


Figure 4
Global interpretation using SHAP

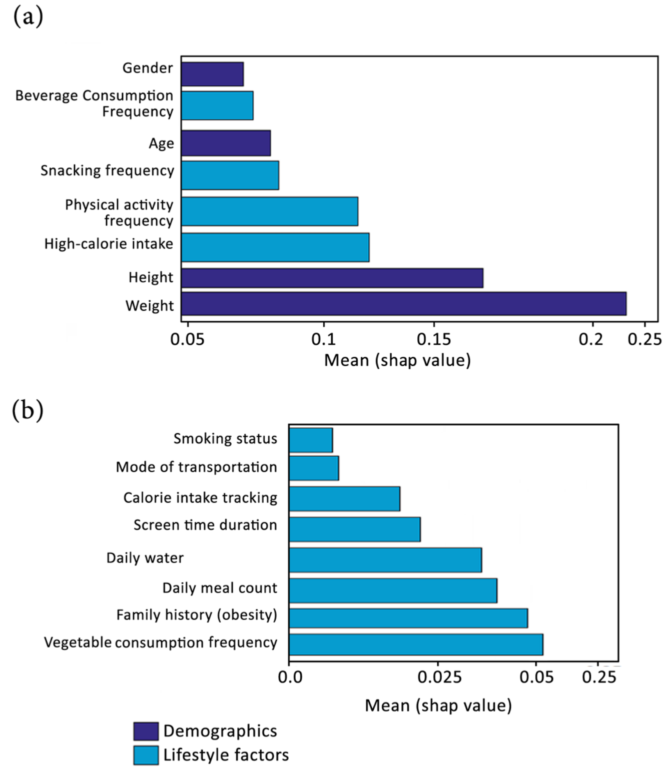


Figure 5
Mobile app flowchart model

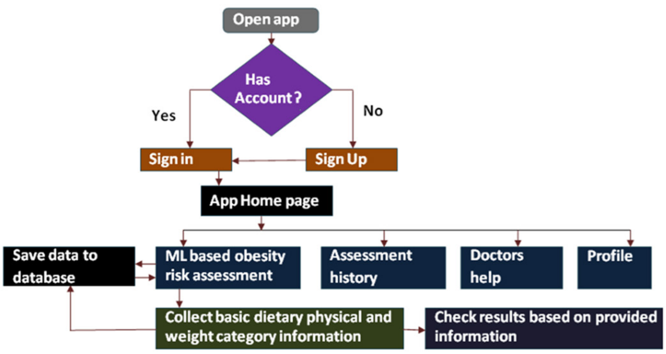
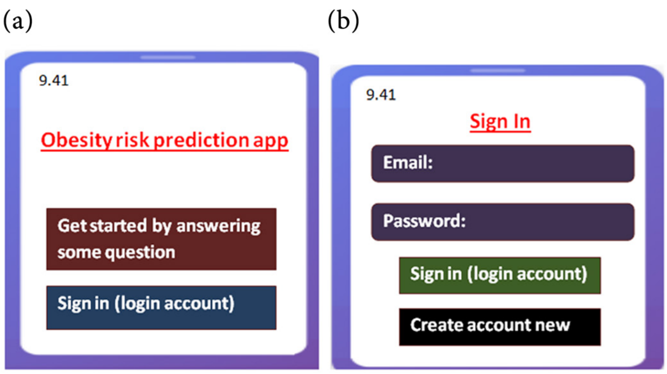


Figure 6
Mobile app welcome screen and sign in screen



main gateway to the health prediction system. The application's user-centric design approach is evident in the welcome screen layout, which has simple navigation elements and clear typography. With options for both new user registration and existing user sign-in, the authentication system offers safe user account management (see Figure 6(b)). Figure 7 displays the Basic Information and Family History Data Collection Screens. As seen in Figure 8, the last phases of data collection focus on specific lifestyle aspects such as nutritional consumption habits, frequency of physical activity, screen time, and preferred modes of transportation. The smooth transition between ML methodology and mobile application functionality is exemplified by the prediction processing phase. User input is converted into standardized features via the system's implementation of real-time data processing capabilities (Figure 9). Figure 10 shows how the application's user-friendly interface ranks lifestyle factors according to their impact on the prediction outcome, providing a full study of contributing factors. An advanced application of explainable AI concepts in mobile health applications is the factor analysis interface. Complex longitudinal

Figure 7
Basic information and family history info of the proposed app

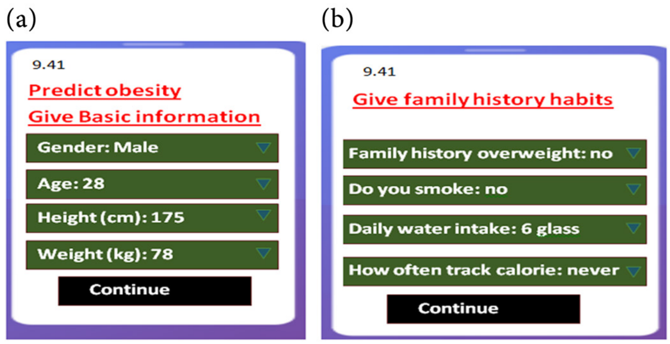


Figure 8
Diet and physical activity information screen of the mobile app

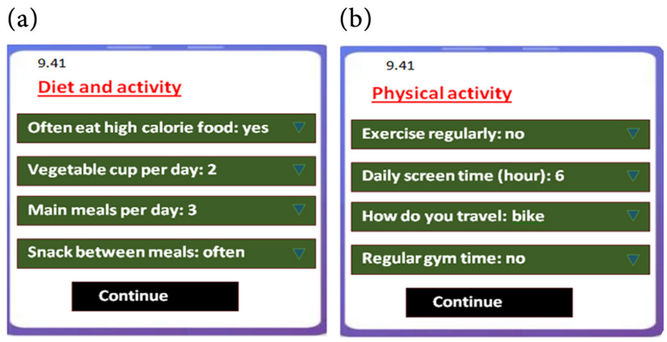


Figure 9
Data analysis and result display page of the mobile app

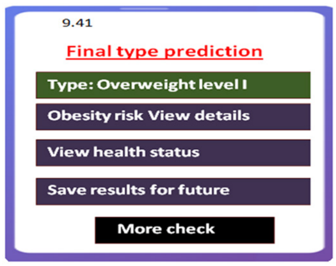


Figure 10
Main contributing factors and result saving interface of the mobile app

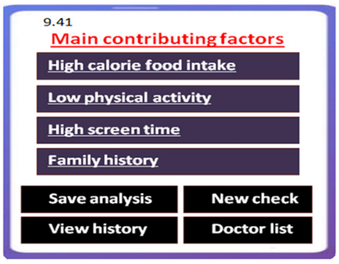


Figure 11
Progress tracking and assessment history interface of the mobile app

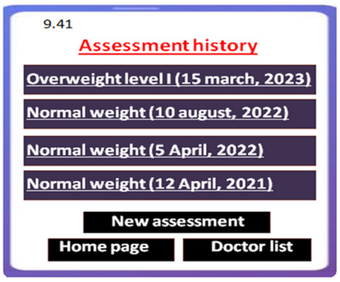


Table 9
Comparison results (proposed method vs. existing method)

Reference name	Accuracy	Precision	Recall	F1-score	Used method	ROC-AUC
Lim et al. [15]	79.6%	78.1%	80.7%	78.8%	Logistic regression	.864
Lin et al. [16]	89.2%	87.8%	85.1%	86.4%	Cat Boost	0.894
Gutiérrez-Gallego et al. [17]	90.7%	88.3%	86.4%	85.8%	Random forest	.903
Thamrin et al [18]	81.3%	83.7%	84.9%	84.2%	Decision tree	0.823
Our Proposed Model	95.4%	94.4%	96.3%	95.3%	XGBoost	0.981

Note: ROC-AUC = Receiver Operating Characteristic - Area Under Curve.

health data is effectively converted into easily comprehensible visual representations by the progress tracking interface, which encourages users to stay engaged (Figure 11).

4. Evaluation Results

Table 9 shows how well our proposed model performs in contrast to important literature works by investigating metrics, such as accuracy value along with F1-score gain. The performance of the proposed ML-based obesity prediction system was compared with different existing and important literary works (e.g., Lim et al. [15], Lin et al. [16], Gutiérrez-Gallego et al. [17], and Thamrin et al. [18]). All compared models are evaluated under the same dataset used by the proposed model.

Interestingly, our proposed XGBoost-based obesity risk prediction framework performs better than most previous research, with an accuracy of 95.4%, F1-score of 95.3%, and ROC-AUC of 0.981, whereas the majority of existing studies show accuracy between 79.6% and 90.7%, F1-scores between 78.8% and 86.4%, and ROC-AUC between 0.823 and 0.894. In both binary and multiclass scenarios, our suggested system performs noticeably better than previous research. The existing works did not use suitable data preprocessing, feature selection, hyperparameter tuning, and cross-validation like the proposed method. Its powerful and generalizable method for real-world obesity risk prediction that includes the capacity to manage imbalanced classes, include diverse behavioral and lifestyle characteristics, and provide interpretable insights via SHAP values.

5. Conclusion

This paper delivers a ML based obesity risk prediction system by taking seven weight categories into account. Prior to using an ensemble feature selection strategy that combined Chi-Square, ANOVA *F*-test, RFE, SHAP, and Borda Count aggregation, the proposed work used label encoding and standardization for preprocessing. SMOTE was used to correct class imbalance on the training set, and GridSearchCV was used to carry out systematic hyperparameter optimization. The best prediction model is chosen for the obesity prediction by investigating 10 ML classifiers perfromance. The proposed XGBoost method outperforms other ML methods by achieving the highest accuracy (95.4%) and F1-score value (95.3%) for the prediction task. The results show that at least 5.18% gain in accuracy and 10.30% gain in F1-score is achieved in the proposed XGBoost-based obesity risk prediction scheme over previous works. Additionally, a SHAP-based feature importance analysis for predicting the risk of obesity was reported in this work. This research also presents the React Native-based obesity risk prediction mobile application, which includes features like lifestyle data input, real-time obesity risk prediction, SHAP-based factor analysis, assessment history monitoring, and user registration. In the future, sophisticated feature engineering utilizing new lifestyle factors and real-time data gathering based on Internet of Things and Deep Learning will also be included for the prediction of obesity risk.

Future research will also look at block chain-based obesity risk data security management systems, FL and AI-based obesity reason determination, Large Language Model-based obesity related problem suggestion generation, among others. The limitations of synthetic oversampling and validate robustness through external or cross-population testing will also be investigated in future. Future research scope would also focus on larger dataset scale, larger geographic scope, reliance on SMOTE, and more precise external validation, among others.

Recommendations

Our research discussed that XGBoost model is suitable for multi-class obesity risk prediction.

Acknowledgement

The authors are grateful to CUET, CSE Faculty for its learning and research facilities.

Ethical Statement

This study did not require formal ethical approval because Chittagong University of Engineering and Technology, Bangladesh does not require IRB/ethics committee approval for this type of survey based on data collection research. This exemption is based on formal ethical policy of CUET authority, issued by Chittagong University of Engineering and Technology.

Conflicts of Interest

The authors declare that they have no conflicts of interest to this work.

Data Availability Statement

The data that support the findings of this study are openly available in GitHub at https://github.com/pymche/Machine-Learning-Obesity-Classification.

Author Contribution Statement

Tarequl Hasan Sakib: Conceptualization, Methodology, Software, Validation, Formal analysis, Investigation, Resources, Data curation. **Mahfuzulhoq Chowdhury:** Conceptualization, Methodology, Validation, Writing – original draft, Writing – review & editing, Visualization, Supervision, Project administration.

References

[1] WHO (2025). Obesity and overweight. Retrieved from: https://www.who.int/news-room/fact-sheets/detail/obesity-and-overweight

[2] WHO (2024). One in eight people are now living with obesity. Retrieved from: https://www.who.int/news/item/01-03-2024-one-in-eight-people-are-now-living-with-obesity

[3] Ng, M., Gakidou, E., Lo, J., Abate, Y. H., Abbafati, C., Abbas, N., ... & Azargoonjahromi, A. (2025). Global, regional, and national prevalence of adult overweight and obesity, 1990–2021, with forecasts to 2050: A forecasting study for the Global Burden of Disease Study 2021. *Lancet Journal*, *405*(10481), 785–812. https://doi.org/10.1016/S0140-6736(25)00355-1

[4] Phelps, N. H., Singleton, R. K., Zhou, B., Heap, R. A., Mishra, A., Bennett, J. E., ... & Barbagallo, C. M. (2024). Worldwide trends in underweight and obesity from 1990 to 2022: A pooled analysis of 3663 population-representative studies with 222 million children, adolescents, and adults. *The Lancet*, *403*(10431), 1027–1050. https://doi.org/10.1016/S0140-6736(23)02750-2

[5] Prithishkumar, I. J., Sappani, M., Ranjan, V., Garg, C., Mani, T., Babu, M., ... & Lakshmanan, J. (2024). Double burden of malnutrition among women of reproductive age: Trends and determinants over the last 15 years in India. *Plos one*, *19*(6), e0304776. https://doi.org/10.1371/journal.pone.0304776

[6] Tariqujjaman, M., Sheikh, S. P., Smith, G., Hasan, A. R., Khatun, F., Kabir, A., ... & Rasheed, S. (2022). Determinants of double burden of malnutrition among school children and adolescents in Urban Dhaka: A multi-level analyses. *Frontiers in Public Health*, *10*, 926571. https://doi.org/10.3389/fpubh.2022.926571

[7] Tanwi, T. S., Chakrabarty, S., & Hasanuzzaman, S. (2019). Double burden of malnutrition among ever-married women in Bangladesh: A pooled analysis. *BMC Women's Health*, *19*(1), 24. https://doi.org/10.1186/s12905-019-0725-2

[8] Rawal, L. B., Biswas, T., Khandker, N. N., Saha, S. R., Bidat Chowdhury, M. M., Khan, A. N. S., ... & Renzaho, A. (2017). Non-communicable disease (NCD) risk factors and diabetes among adults living in slum areas of Dhaka, Bangladesh. *PloS One*, *12*(10), e0184967. https://doi.org/10.1371/journal.pone.0184967

[9] Nowsheen, F., Islam, F., Siddiqueea, Y., Ahsan, M., Pavel, M. A. M., Majumder, T., ... & Zaman, M. M. (2021). Noncommunicable disease risk factors among postgraduate students in Dhaka city, Bangladesh: A multi-centric cross-sectional study. *Journal of Xiangya Medicine*, *6*. https://dx.doi.org/10.21037/jxym-21-29

[10] Azmi, S., Kunnathodi, F., Alotaibi, H. F., Alhazzani, W., Mustafa, M., Ahmad, I., ... & Arafat, A. A. (2025). Harnessing Artificial intelligence in obesity research and management: A comprehensive review. *Diagnostics*, *15*(3), 396. https://doi.org/10.3390/diagnostics15030396

[11] WHO (2024). Body mass index (BMI). Retrieved from: https://www.who.int/data/gho/data/themes/topics/topic-details/GHO/body-mass-index

[12] Parra, D., Gutiérrez-Gallego, A., Garnica, O., Velasco, J. M., Zekri-Nechar, K., Zamorano-León, J. J., ... & Hidalgo, J. I. (2022). Predicting the risk of overweight and obesity in Madrid—A binary classification approach with evolutionary feature selection. *Applied Sciences*, *12*(16), 8251. https://doi.org/10.3390/app12168251

[13] Jeon, J., Lee, S., & Oh, C. (2023). Age-specific risk factors for the prediction of obesity using a machine learning approach. *Frontiers in Public Health*, *10*, 998782. https://doi.org/10.3389/fpubh.2022.998782

[14] Calderón-Díaz, M., Serey-Castillo, L. J., Vallejos-Cuevas, E. A., Espinoza, A., Salas, R., & Macías-Jiménez, M. A. (2023). Detection of variables for the diagnosis of overweight and obesity in young Chileans using machine learning techniques. *Procedia Computer Science*, *220*, 978–983. https://doi.org/10.1016/j.procs.2023.03.135

[15] Lim, H., Lee, H., & Kim, J. (2023). A prediction model for childhood obesity risk using the machine learning method: A panel study on Korean children. *Scientific Reports*, *13*(1), 10122. https://doi.org/10.1038/s41598-023-37171-4

[16] Lin, W., Shi, S., Huang, H., Wen, J., & Chen, G. (2023). Predicting risk of obesity in overweight adults using interpretable machine learning algorithms. *Frontiers in Endocrinology*, *14*, 1292167. https://doi.org/10.3389/fendo.2023.1292167

[17] Gutiérrez-Gallego, A., Zamorano-León, J. J., Parra-Rodríguez, D., Zekri-Nechar, K., Velasco, J. M., Garnica, Ó., ... & Hidalgo, J. I. (2024). Combination of machine learning techniques to predict overweight/obesity in adults. *Journal of Personalized Medicine*, *14*(8), 816. https://doi.org/10.3390/jpm14080816

[18] Thamrin, S. A., Arsyad, D. S., Kuswanto, H., Lawi, A., & Nasir, S. (2021). Predicting obesity in adults using machine learning techniques: An analysis of Indonesian basic health research 2018. *Frontiers in nutrition*, *8*, 669155. https://doi.org/10.3389/fnut.2021.669155

[19] Singh, B., & Tawfik, H. (2020, June). Machine learning approach for the early prediction of the risk of overweight and obesity in young people. In *International Conference on Computational Science*, 523–535. https://doi.org/10.1007/978-3-030-50423-6_39

[20] Rousset, S., Angelo, A., Hamadouche, T., & Lacomme, P. (2023). Weight status prediction using a neuron network based on individual and behavioral data. *Healthcare*, *11*(8), 1–10. https://doi.org/10.3390/healthcare11081101

[21] Dirik, M. (2023). Application of machine learning techniques for obesity prediction. *Journal of Complexity in Health Sciences*, *6*(2), 16–34. https://doi.org/10.21595/chs.2023.23193

[22] Dutta, R. R., Mukherjee, I., & Chakraborty, C. (2025). Obesity disease risk prediction using machine learning. *International Journal of Data Science and Analytics*, *19*(4), 709–718. https://doi.org/10.1007/s41060-023-00491-9

[23] Görmez, Y., Yagin, F. H., Yagin, B., Aygun, Y., Boke, H., Badicu, G., ... & Aghaei, M. (2025). Prediction of obesity levels based on physical activity and eating habits with a machine learning model integrated with explainable artificial intelligence. *Frontiers in Physiology*, *16*, 1549306. https://doi.org/10.3389/fphys.2025.1549306

[24] Helforoush, Z., & Sayyad, H. (2024). Prediction and classification of obesity risk based on a hybrid metaheuristic machine learning approach. *Frontiers in Big Data*, *7*, 1469981. https://doi.org/10.3389/fdata.2024.1469981

[25] Devi, S., Chavan, R., & Gupta, H. (2023, December). Prediction of obesity among school going children using machine learning algorithms. In *International Conference on Innovative Computing, Intelligent Communication and Smart Electrical Systems*, 1–7. https://doi.org/10.1109/ICSES60034.2023.10465310

[26] Vemulapalli, G., Yalamati, S., Palakurti, N. R., Alam, N., Samayamantri, S., & Whig, P. (2024). Predicting obesity trends using machine learning from big data analytics approach. In *Asia Pacific Conference on Innovation in Technology*, 1–5. https://doi.org/10.1109/APCIT62007.2024.10673429

[27] Jeong, J. H., Lee, I. G., Kim, S. K., Kam, T. E., Lee, S. W., & Lee, E. (2024). DeepHealthNet: Adolescent obesity prediction system based on a deep learning framework. *IEEE Journal of Biomedical and Health Informatics*, *28*(4), 2282–2293. https://doi.org/10.1109/JBHI.2024.3356580

[28] Nalini, N., Nagaraju, S., Jayanthi, R., Ram, C. S., & Pandey, D. (2025). Predict overweight or obesity using body fat percentage with machine learning classification algorithms to enhance accuracy. In *International Conference on Recent Advances in Electrical, Electronics, Ubiquitous Communication, and Computational Intelligence*, 1–5. https://doi.org/10.1109/RAEEUCCI63961.2025.11048274

[29] Titu, M. F. S., Rahman, M. R., Apurba, A. Z., Qayum, M. A., & Khan, R. (2025). Predicting adolescent concern over unhealthy food ads using explainable AI. *IEEE Access*. https://doi.org/10.1109/ACCESS.2025.3588190

[30] Bahrin, U. F., Jantan, H., Sofian, M. A. S. M., Ismail, I. S., & Samsudin, S. H. A. (2022, November). Classifying body type based on eating habits and physical condition using decision tree technique. In *International Visualization, Informatics and Technology Conference*, 95–102. https://doi.org/10.1109/IVIT55443.2022.10033335

[31] Garcia-d'Urso, N., Climent-Pérez, P., Sánchez-Sansegundo, M., Zaragoza-Martí, A., Fuster-Guilló, A., & Azorín-López, J. (2022). A non-invasive approach for total cholesterol level prediction using machine learning. *IEEE Access*, *10*, 58566–58577. https://doi.org/10.1109/ACCESS.2022.3178419=999989

How to Cite: Sakib, T. H., & Chowdhury, M. (2026). A Multi-class Obesity Risk Prediction Using Machine Learning and Explainable Artificial Intelligence. *Artificial Intelligence and Applications*. https://doi.org/10.47852/bonviewAIA62027134