Acadlore Transactions on AI and Machine Learning
https://www.acadlore.com/journals/ATAIML

Body Composition-Based Prediction of Obesity Risk in Saudi Adults Using Explainable Machine Learning

Ayesha Siddiqha Mukthar[1,2,3*], Sayeda Faatin Alvi[4], N. Nithiyanandam[5], S. Nikkath Bushra[5], Keshav Kaushik[6], Wedyan Mohammed Alwashmi[1,2,3]

[1] King Abdullah International Medical Research Center, 31982 AlAhsa, Saudi Arabia
[2] King Saud bin Abdulaziz University for Health Sciences, 31982 AlAhsa, Saudi Arabia
[3] Ministry of National Guards Health Affairs, 31982 AlAhsa, Saudi Arabia
[4] Faculty of Science, Radboud University, 6525 AJ Nijmegen, Netherlands
[5] Department of Computing Technologies, SRM Institute of Science and Technology, 603203 Kattankulathur, India
[6] Center for Cyber Security and Cryptology, Sharda University, 201310 Greater Noida, India

[*] Correspondence: Ayesha Siddiqha Mukthar (ayeshamukthar17@gmail.com)

Received: 07-02-2026 **Revised:** 08-05-2026 **Accepted:** 08-10-2026

Citation: A. S. Mukthar, S. F. Alvi, N. Nithiyanandam, S. N. Bushra, K. Kaushik, and W. M. Alwashmi, "Body composition-based prediction of obesity risk in Saudi adults using explainable machine learning," *Acadlore Trans. Mach. Learn.*, vol. 5, no. 3, pp. 244–261, 2026. https://doi.org/10.56578/ataiml050304.

© 2026 by the author(s). Published by Acadlore Publishing Services Limited, Hong Kong. This article is available for free download and can be reused and cited, provided that the original published version is credited, under the CC BY 4.0 license.

Abstract: Obesity represents a substantial public health burden in Saudi Arabia, yet the predictive contribution of anthropometric and lifestyle characteristics beyond body mass index remains insufficiently characterized. An interpretable machine-learning framework was developed to classify obesity among Saudi adults using anthropometric, demographic, health, and lifestyle variables and to assess whether predictive performance was retained after excluding variables directly related to the body mass index-defined outcome. Of 294 survey responses, 279 were retained after consent and data-completeness criteria were applied. Numerical variables were median-imputed, categorical variables were mode-imputed and one-hot encoded, and obesity was defined as a body mass index $\geq$30 kg/m^2. Random forest performance was evaluated using stratified five-fold cross-validation with fixed tuned hyperparameters. Sensitivity analyses excluded body mass index alone and body mass index, height, and weight simultaneously. Model interpretability was assessed using Shapley additive explanations. With body mass index included, all evaluated performance metrics reached 1.000 ± 0.000, reflecting target leakage because body mass index directly defined the outcome. After body mass index exclusion, accuracy was 0.911 ± 0.038, recall 0.624 ± 0.168, F1-score 0.701 ± 0.125, and area under the receiver operating characteristic curve 0.972 ± 0.023. After simultaneous exclusion of body mass index, height, and weight, accuracy decreased to 0.842 ± 0.023 and area under the receiver operating characteristic curve to 0.815 ± 0.056. Waist circumference, hip circumference, waist-to-hip ratio, age, and selected health and lifestyle characteristics retained predictive information, although sensitivity to obesity decreased substantially after removal of the defining anthropometric variables. Shapley additive explanation analyses clarified feature contributions to individual predictions. These findings demonstrate that the exceptional performance of the complete model was predominantly attributable to target leakage. Complementary characteristics retained meaningful discriminatory information, but reduced sensitivity warrants cautious interpretation. External validation in larger, representative cohorts with independently measured anthropometric data is required before clinical or population-level screening applications are considered.

Keywords: Obesity prediction; Body composition; Machine learning; Explainable artificial intelligence; Saudi Arabia; Public health analytics

1 Introduction

Due to complex genetic, environmental, behavioral, and socioeconomic linkages, obesity is a global public-health issue. Over the past few decades, overweight and obesity rates in the Kingdom of Saudi Arabia have skyrocketed, making it one of the highest in the Middle East and North Africa [1]. Obesity is a prominent risk factor for cardiometabolic illnesses, type 2 diabetes, certain malignancies, and impaired quality of life, and it strains healthcare systems and public-health resources [2]. Given the size and multifaceted nature of the disease, new technologies to identify high-risk individuals for targeted prevention and early intervention are needed urgently. Body mass index is widely used to screen for obesity, but it does not distinguish between fat and lean mass or fat distribution (e.g., visceral versus subcutaneous adiposity), which are important metabolic risk factors [3]. Body composition analysis, using bioelectrical impedance analysis, dual-energy X-ray absorptiometry, and anthropometric indices, provides more granular measures (fat mass, lean mass, percent body fat, and visceral adipose estimates) that better reflect biological substrates of obesity-related risk [4]. Body composition measurements predict cardiometabolic outcomes better than body mass index alone in numerous groups, including Middle Eastern cohorts [5].

Meanwhile, machine learning has enabled the use of high-dimensional body composition data, demographic, lifestyle, and clinical characteristics to predict obesity and its comorbidities. Supervised learning techniques like logistic regression, decision trees, ensemble approaches, and neural networks can capture nonlinear patterns and interactions that typical statistical models overlook, improving health-related prediction [6]. Model interpretability and population specificity hinder clinical and public-health machine learning implementation. Black-box models may be effective, but doctors and policymakers struggle to trust and act on them without clear explanations [7]. Due to genetics, lifestyle, and body composition standards, models trained in one community may not generalize to another. Saudi individuals' region-specific anthropometric and metabolic profiles require unique modeling [8].

Explainable artificial intelligence may bridge predicted accuracy with clinical interpretability. Feature importance scoring, Shapley additive explanations, local interpretable model-agnostic explanations, and rule extraction can reveal which body composition features most predict obesity risk and population-level patterns useful for prevention [9]. Explainable artificial intelligence may improve clinical interpretability and predicted accuracy. Features that indicate obesity risk and population-level patterns for prevention may be identified using feature significance scores, Shapley additive explanations, local interpretable model-agnostic explanations, and rule extraction [10]. Shapley additive explanations provide game-theoretic, additive explanations founded on Shapley values to ensure consistent and locally correct interpretation of each feature's influence [11], whereas local interpretable model-agnostic explanations provide case-level interpretability by approximating complicated model behavior with simpler surrogate models [12]. Clinical risk-prediction models are more transparent and behaviorally valid using explainable artificial intelligence approaches. Explainable artificial intelligence in machine learning processes promotes transparency and allows domain experts to assess model behavior, discover biases, and produce clinically actionable insights, which are necessary for real-world implementation.

Despite these methodological breakthroughs, few studies have used precise body composition profiling, contemporary machine learning, and explainable artificial intelligence to predict obesity risk in Saudis. Regional research has mostly utilized body mass index or basic anthropometrics, traditional statistical modeling, or no model interpretability [13]. This presents a potential to design a culturally and scientifically informed prediction system that uses body composition measures, accommodates local demographic features, and offers explainable outcomes for clinical and public health stakeholders in the Kingdom of Saudi Arabia.

This work tackles this gap by developing and assessing machine-learning models trained on body composition and supplementary variables from Saudi people, using integrated explainable artificial intelligence analysis to identify obesity risk factors. The goals are to (i) determine if body composition features outperform body mass index and basic demographics in predictive performance, (ii) compare several machine learning algorithms for discrimination, calibration, and robustness, and (iii) use explainable artificial intelligence methods to generate transparent, clinically interpretable explanations at the cohort and individual levels. The work focuses on Saudi nationals and embeds explainability into the modeling process to create accurate and actionable tools for clinicians, public-health practitioners, and patients to improve personalized risk assessment and support targeted interventions to combat regional obesity.

2 Literature Review

Numerous studies show that demographic, socioeconomic, lifestyle, and behavioral variables greatly impact obesity risk and body mass index remains a powerful predictor of obesity risk across varied groups [14]. Age and gender also affect obesity. Middle Eastern research shows that older adults and women are disproportionately affected by obesity, reflecting physiological changes, cultural norms, and lifestyle patterns [15]. Another factor is socioeconomic status, with lower-income groups having greater obesity rates due to restricted availability to healthful foods, physical exercise, and energy-dense diets [16]. High-calorie foods, inactivity, and sedentary behavior increase obesity risk [17]. These studies show obesity is impacted by biological, behavioral, and environmental factors. The complexity of obesity factors has made machine learning popular for predictive modeling in healthcare. Machine learning approaches can detect complicated connections and hidden patterns in multidimensional datasets, making them ideal for obesity prediction [18]. Decision trees, random forests, and logistic regression models show that demographic, lifestyle, and socioeconomic characteristics improve prediction performance. Early childhood obesity research has used machine learning-based obesity prediction. These studies suggest that sophisticated algorithms trained on longitudinal lifestyle and dietary data can detect children at high risk of obesity before clinical signs appear, enabling early treatments [19].

Adolescent and young adult obesity prediction has improved. Singh and Tawfik [20] used seven algorithms—support vector machine, multilayer perceptron, k-nearest neighbors, pruned trees, random forests, and bagging—to predict teenage obesity, showing algorithmic efficiency variability and the benefits of ensemble-based approaches. In other works, Jindal et al. [21] showed that R-based ensemble learning and Python-driven interfaces may predict obesity risk. Alghnam et al. [22] tested nine machine learning algorithms for obesity prediction in a broad population of over 1,100 people and found logistic regression to be the most accurate (97.09%), exceeding gradient boosting. Cervantes and Palacio [23], who researched young people in Colombia, Mexico, and Peru, again found that machine learning approaches, including k-means clustering, decision trees, and support vector machines, can categorize obesity risk in varied populations. Lifestyle and demographic data are fundamental to many machine learning-based obesity prediction algorithms. Hammond et al. [24] used a gradient boosting method to predict childhood obesity using electronic health records and population-level data on sleep, nutrition, exercise, and socioeconomic variables. Studies have also demonstrated that early-life characteristics can contribute to the prediction of obesity in young adulthood. Hochner et al. [25] used data from the Jerusalem Perinatal and developed logistic regression models incorporating genetic, sociodemographic, and perinatal characteristics to predict young-adult obesity, with the addition of early-life variables improving model discrimination. These studies indicate that machine learning may find obesity factors by analyzing variables that interact in nonlinear, complicated ways that standard statistical approaches miss.

Recent studies show that machine learning models using anthropometric and body composition parameters can improve obesity prediction. Ferdowsy et al. [26] and Singh and Tawfik [20] showed how machine learning can analyze demographics, lifestyle, body mass index, and body fat distribution data to predict risk more accurately. Most research has been undertaken outside the Middle East, and few have included complex body composition metrics like visceral fat percentage, waist-to-hip ratio, and lean mass distribution in machine learning frameworks. Despite high obesity rates in Saudi Arabia, there is little research on how body fat distribution, lifestyle behaviors, and demographic characteristics interact in prediction models for Saudis [27]. This difference is crucial because population-specific physiological and cultural factors may significantly affect obesity trends and determinants.

Global studies have shown the value of machine learning for predicting obesity using demographic, lifestyle, and basic anthropometric data, but few have integrated comprehensive body composition metrics—including fat distribution indicators—into Saudi-specific machine learning-based obesity prediction models. In Saudi Arabia, most research uses body mass index or broad lifestyle indicators without addressing nonlinear relationships between factors such as visceral fat levels, socioeconomic inequalities, and culturally influenced diets. Current models are limited for clinical and public-health decision-making since no substantial research has used machine learning and explainable artificial intelligence to assess obesity risk variables in Saudi populations. This work develops an interpretable, high-accuracy machine learning model with precise body composition, lifestyle, and demographic characteristics to improve obesity prediction and preventive measures for Saudis.

3 Methodology

This cross-sectional analytical study examined associations among anthropometric measures, lifestyle behaviors, demographic characteristics, and obesity status in adults surveyed in Saudi Arabia. A total of 500 individuals were invited and 294 responses were submitted. Of these, 279 respondents provided consent and completed the questionnaire and were included in the analysis. The sample was therefore a convenience sample rather than a probability sample. The cross-sectional design supports classification and association analysis but does not establish temporal prediction or causality. Saudi citizens aged 18–65 who could provide informed consent and complete the survey properly were eligible. Pregnant or nursing women, those with Cushing's syndrome or chronic corticosteroid usage, and those with physical impairments that prohibited good anthropometric measures were excluded. All participants gave electronic informed consent, and the King Abdullah International Medical Research Centre Institutional Review Board approved.

Using Microsoft Forms, a thorough online survey (available as a supplementary file) collected data confidentially and consistently. The questionnaire collected demographic data such as age, gender, education, marital status, employment, nationality, and family income. Body mass index and waist-to-hip ratio were calculated using self-reported or externally measured height, weight, waist circumference, and hip circumference. A Saudi Food Frequency Questionnaire measured dietary patterns, the International Physical Activity Questionnaire–Short Form measured physical activity, a validated sleep quality scale measured sleep duration and quality, and smoking and alcohol consumption histories were recorded. Diabetes, hypertension, cardiovascular disease, and obesity by family were also assessed.

3.1 Feature Engineering and Data Processing

The feature engineering and preprocessing pipeline followed established machine learning practices for healthcare data to maximize data quality while minimizing bias and information leakage [28]. Prior to model development, data quality checks were performed to identify missing values, inconsistent measurements, and implausible anthropometric records. Height values greater than 10 were interpreted as centimeters, whereas entries below 10 (e.g., 5.2, 5.4, and 5.6) were assumed to represent feet and converted to centimeters. Non-positive waist and hip circumference values were treated as missing. Derived anthropometric indices were subsequently calculated, with body mass index computed as weight (kg) divided by height squared (m^2) and waist-to-hip ratio calculated as waist circumference divided by hip circumference. Consistent with previous obesity prediction studies, categorical behavioral variables representing sleep quality, dietary habits, and physical activity were also derived to better capture lifestyle-related risk factors [29, 30].

To preserve statistical validity during model training, numerical missing values were imputed using the median of the corresponding training fold, while categorical variables were imputed using the most frequent category (mode). Categorical predictors were transformed using one-hot encoding to facilitate their use by the machine learning algorithms. Importantly, all preprocessing operations—including imputation, feature engineering, and categorical encoding—were performed within each cross-validation training fold before being applied to the corresponding validation fold, thereby preventing information leakage and ensuring an unbiased evaluation of model performance [28, 31]. Potential outliers were identified using domain-informed thresholds based on the World Health Organization and Centers for Disease Control and Prevention anthropometric guidelines, ensuring that implausible measurements did not adversely influence model training or evaluation [30, 31].

3.2 Machine Learning Model Development

3.2.1 Evaluated machine learning models

The primary analysis used 29 predictors: age, gender, education, employment, income, height, weight, waist circumference, hip circumference, chronic illness, family history of obesity, physical and mental health scores, blood pressure, cholesterol, glucose, sleep duration, fruit and vegetable intake, processed-food intake, sugary-drink intake, portion size, daily caloric intake, two dietary-pattern variables, exercise hours, smoking, stress score, body mass index, and waist-to-hip ratio. Obesity was defined as a body mass index $\geq$30 kg/m^2. Because body mass index directly defines this outcome, three random forest specifications were evaluated: (i) the complete model, including body mass index; (ii) a body mass index-excluded model; and (iii) a stricter anthropometric sensitivity model, excluding body mass index, height, and weight. The latter two analyses were designed to quantify predictive information beyond the variables used to construct the label.

Five machine-learning models were trained and evaluated for obesity prediction, as described below.

• Logistic regression: It was a supervised machine learning algorithm used as a baseline linear classifier. This model was trained on factors like lifestyle and demographic data to predict the obesity risk. This model distinguished between obese and not obese based on the model's performance such as accuracy, precision, recall, F1-score, area under the receiver operating characteristic curve, and confusion matrix [32]. The model was configured with an L2 penalty and the liblinear solver, while the regularization parameter C was tuned across 0.01, 0.1, 1, 10, and 100, with the maximum number of iterations set to 1000.

• Random forest: An ensemble of 200 decision trees was used with unrestricted depth, max_features = 0.5, min_samples_leaf = 1, class_weight = "balanced", and random_state = 42. These settings were selected from a grid considering 200 or 500 trees; maximum depth of none, 8, or 12; minimum leaf sizes of 1, 2, or 4; and max_features of square root or 0.5, with area under the receiver operating characteristic curve as the tuning criterion. The same fixed settings were used in the sensitivity analyses to ensure a comparable assessment of feature-set effects [33].

• Gradient boosts (extreme gradient boosting): Extreme gradient boosting was an advanced machine learning algorithm based on gradient boosted decision trees [34]. It built many small decision trees and each tree tried to correct the error of the previous tree. It was very fast and accurate in handling mixed data types. It used gradient descent to minimize loss and regularization to reduce overfitting.

• Support vector machine: It was a supervised machine learning algorithm that performed well for small datasets. Both linear and nonlinear classification could be conducted with different kernel functions. Support vector machines achieved high-dimensional separation and favorable prediction accuracy. Obesity prediction constituted a nonlinear classification task, and radial basis function kernels were adopted in this analysis [35].

• Multilayer perceptron: It was a basic neural network model. Multilayer perceptrons handled nonlinear features effectively. The multilayer perceptron captured nonlinear relationships between lifestyle factors and obesity levels. It delivered strong predictive performance for obesity in this study [36].

3.2.2 Hyperparameter optimization

To maximize predictive performance while reducing the risk of overfitting, hyperparameter optimization was performed using a grid search within a stratified five-fold cross-validation framework. Candidate hyperparameter combinations for each machine learning algorithm in Table 1 were systematically evaluated, and the optimal configuration was selected based on the highest mean area under the receiver operating characteristic curve across the validation folds. Stratified cross-validation preserves the class distribution within each fold and is recommended for healthcare datasets with limited sample sizes because it provides more reliable estimates of model stability and internal generalizability than a single train-test split [31, 37].

Table 1. Hyperparameter settings and optimization strategy for the evaluated machine learning models

Model	Main Hyperparameters	Optimization Strategy
Logistic regression	Penalty = L2; Solver = liblinear; C = {0.01, 0.1, 1, 10, 100}; Maximum iterations = 1000	Grid search with stratified five-fold cross-validation; best model selected based on mean ROC-AUC.
Random forest	Number of trees (n_estimators) = {100, 200, 300}; Maximum depth (max_depth) = {None, 10, 20}; Minimum samples split (min_samples_split) = {2, 5, 10}; Minimum samples leaf (min_samples_leaf) = {1, 2, 4}; Bootstrap = True	Grid search with stratified five-fold cross-validation; optimal parameters selected using mean ROC-AUC.
Gradient boosting (XGBoost)	Number of estimators = {100, 200}; Learning rate = {0.01, 0.05, 0.10}; Maximum depth = {3, 5, 7}; Subsample = {0.8, 1.0}; Column sample (colsample_bytree) = {0.8, 1.0}	Grid search with stratified five-fold cross-validation using mean ROC-AUC as the optimization criterion.
Support vector machine	Kernel = RBF; C = {0.1, 1, 10, 100}; Gamma = {scale, 0.01, 0.1, 1}	Grid search with stratified five-fold cross-validation; best parameter combination selected using mean ROC-AUC.
Multilayer perceptron	Hidden layer sizes = {(50), (100), (100, 50)}; Activation = ReLU; Solver = Adam; Learning rate = Adaptive; Maximum iterations = 500; Alpha = {0.0001, 0.001, 0.01}	Grid search with stratified five-fold cross-validation; model selected based on mean ROC-AUC.

Note: ROC-AUC = receiver operating characteristic-area under the curve; RBF = radial basis function; ReLU = rectified linear unit.

To prevent information leakage, all preprocessing operations—including missing-value imputation, derivation of anthropometric indices, and categorical variable encoding—were performed independently within each training fold before being applied to the corresponding validation fold. This approach ensures that information from the validation data does not influence model training and is considered good practice for developing reproducible machine learning models in clinical research [28, 37]. The optimization process followed standard recommendations for supervised machine learning in healthcare, balancing predictive performance with model robustness and interpretability [15, 37].

3.3 Model Evaluation and Interpretation

Performance was assessed using accuracy, precision, recall, F1-score, and area under the receiver operating characteristic curve [37, 38]. To quantify model stability, stratified five-fold cross-validation with shuffling and random_state = 42 was used. Mean performance and the sample standard deviation across the five held-out folds are reported. This fold-wise analysis was prioritized because the modest sample size and low number of obese cases can produce unstable single-split estimates. Shapley additive explanations and local interpretable model-agnostic explanations were used for global and case-level feature attribution. Both methods provided strong interpretability, supporting recent medical artificial intelligence guidelines for multi-method explanation frameworks [39]. Two representative case explanations were also examined to show how combinations of anthropometric, health, and lifestyle variables affected individual predictions.

3.4 Statistical Software and Analysis

All statistical analyses and machine learning procedures were performed using Python 3.13.5 with scikit-learn 1.8.0, NumPy, statsmodels, and Matplotlib. Data preprocessing and model development were implemented using scikit-learn Pipeline and ColumnTransformer objects. Random seeds were fixed at 42 to ensure reproducibility. The complete analysis was conducted using the final analytical dataset comprising 279 participants who provided informed consent and completed the study questionnaire. Figure 1 presents the overall system architecture.

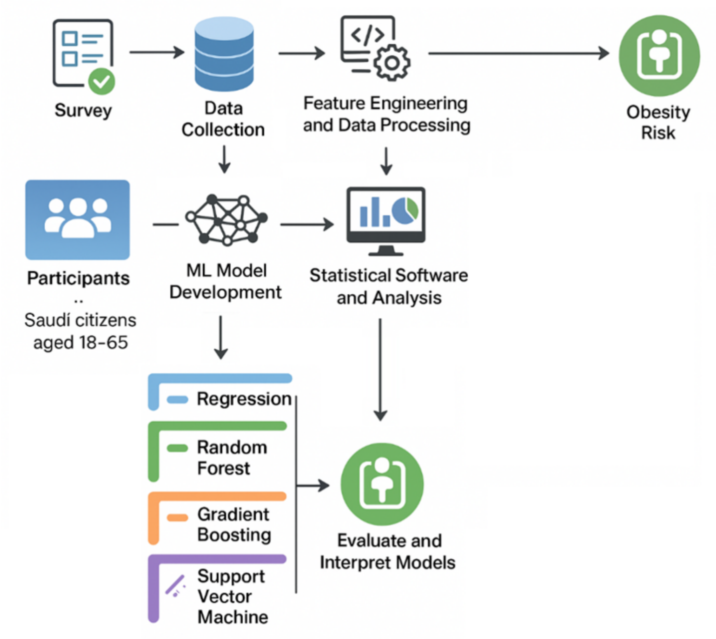


Figure 1. System architecture design
Note: ML = Machine Learning.

4 Results

This section presents the full findings of the machine learning framework developed to predict obesity risk among Saudi individuals. Drawing on survey-collected data, the analysis objectively measured body composition metrics, processed through a pipeline of classical machine learning algorithms and explainable artificial intelligence methods.

4.1 Experimental Setup and Data Partitioning

To obtain a robust estimate of model performance and minimize overfitting, model stability was evaluated using stratified five-fold cross-validation rather than relying solely on a single train-validation-test split. Stratified sampling preserved the proportion of obese and non-obese participants in each fold, ensuring representative evaluation across the relatively small cross-sectional dataset. During each iteration, one-fold was reserved for testing while the remaining folds were used for model training, with all preprocessing steps—including imputation, encoding, and feature scaling where applicable—performed exclusively on the training data before being applied to the corresponding test fold. This procedure prevented information leakage and provided an unbiased assessment of model performance [28, 31].

The predictive performance of each algorithm was summarized using the mean and standard deviation of the evaluation metrics across the five folds, providing a more informative measure of internal model stability than a single train-test split. Hyperparameter optimization was performed within the cross-validation framework using grid-search-based tuning, thereby reducing the risk of overfitting while making efficient use of the available data. Stratified cross-validation is widely recommended for machine learning studies involving relatively small clinical datasets because it provides more reliable estimates of model generalizability and predictive robustness than a single hold-out validation approach [31, 37].

4.2 Data Preprocessing and Cleaning

The spreadsheet contained 294 submitted responses. Fourteen respondents declined consent in the bilingual form, and one declined in the English-only form, leaving 279 consented and completed records. Height values expressed in feet were converted to centimeters. Non-positive waist and hip values were treated as missing and imputed within the modeling pipeline. Obesity prevalence was 48/279 (17.2%), indicating class imbalance; stratification and class weighting were therefore used [28]. Obesity risk was operationalized using two standard anthropometric indices: body mass index and waist-to-hip ratio. In line with the World Health Organization criteria for obesity classification, a body mass index $\geq$30 kg/m^2 serves as the primary obesity threshold [30]. As shown in Figure 2, the body mass index distribution was predominantly concentrated in the normal-to-overweight range, with values clustering between 18 and 30 and a modal peak around 22–24 kg/m^2. A rightward skew extending toward a body mass index of 60 indicated the presence of extreme obesity in a minority of participants, a pattern consistent with

Saudi population health data reporting that obesity prevalence varies substantially across age groups and regions within the Kingdom [19, 27].

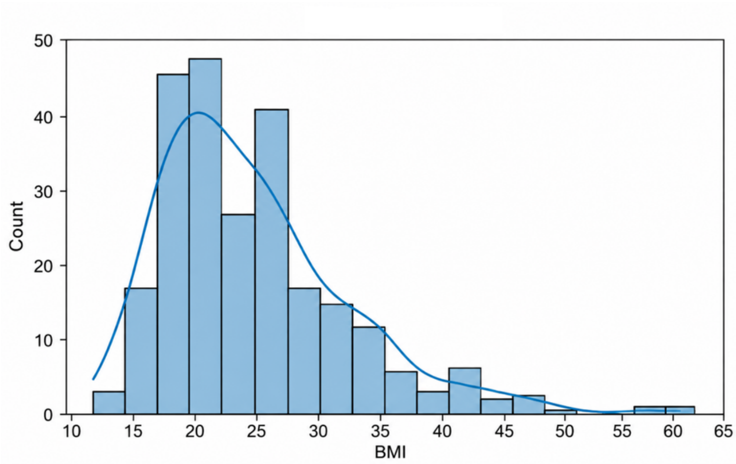


Figure 2. Exploring body mass index distribution

Note: BMI = Body Mass Index.

This definition creates a direct mathematical relationship between the label and body mass index and also an indirect relationship with height and weight. Consequently, the complete model should be interpreted as reproducing body mass index-based classification rather than independently forecasting future obesity. The body mass index-excluded and stricter sensitivity analyses were therefore essential for assessing information supplied by the remaining variables.

Figure 3 reveals height irregularities at values of 5.2, 5.4, and 5.6 feet alongside extreme weight entries exceeding 130 kg and 164 kg, which likely reflect data-entry inconsistencies rather than genuinely exceptional physiological profiles. These anomalies justified the preprocessing step and reinforced the need for filtering before modeling. Class balance across the five folds was maintained through stratified k-fold cross-validation (n = 5), a particularly important safeguard given the inherent class imbalance common in health survey datasets [33].

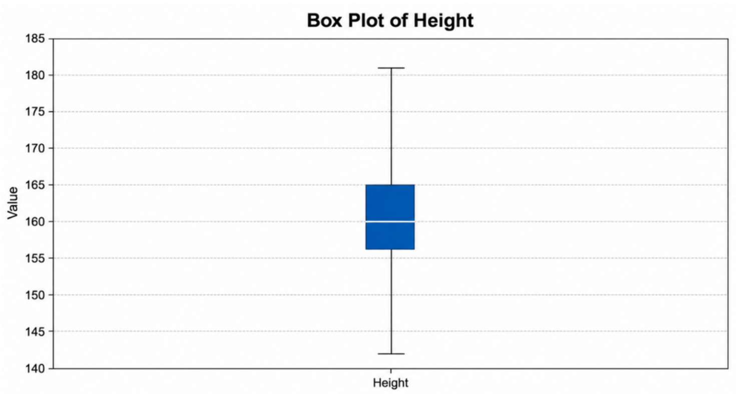


Figure 3. Outlier in height and weight measurement identification

4.3 Classifier Performance Comparison

Machine learning classifiers were evaluated: logistic regression, random forest, gradient boosting, support vector machine, and a multilayer perceptron neural network. Table 2 consolidates their performance across accuracy, precision, recall, F1-score, and area under the receiver operating characteristic curve—the five metrics most commonly reported in obesity prediction literature for a balanced view of classifier behavior under class imbalance [15, 37].

The random forest as tree-based and logistic regression as a linear classifier emerged as the strongest performers, achieving accuracy figures of approximately 95.9% and 96.6%, respectively, alongside area under the receiver operating characteristic curve values are 0.956 and 0.901. These results are notable because they indicate near-perfect separation between obese and non-obese individuals on the test set. A random forest's balanced precision of 1.00 alongside a recall of 0.80 reflects a conservative but clinically sensible profile: the model rarely produces false positives (important when avoiding unnecessary clinical alarms), though it misses roughly one in five at-risk individuals—a trade-off that would require contextual calibration in a real screening scenario.

The comparatively weak performance of gradient boosting (85%), support vector machine (83%), and multilayer perceptron (84%) was not unexpected. Support vector machines are known to struggle with high-dimensional or non-linearly separable feature spaces without careful kernel selection and hyperparameter tuning [35], while multilayer perceptron classifiers are prone to instability and precision collapse when training data is imbalanced [36]. Gradient boosting's underperformance relative to random forest echoes findings in prior obesity risk studies where ensemble tree methods demonstrated superior stability on relatively small, heterogeneous health datasets [34, 40]. Logistic regression, while conceptually the simplest model, performed surprisingly well (area under the curve $\approx$ 0.90), likely because the dominant features—body mass index, weight, and waist circumference—carry a strong linear signal, even if the full dataset contains nonlinear interaction terms [41].

Table 2. Comparative performance of machine learning classifiers

Algorithm	Accuracy	Precision	Recall	F1-Score	ROC-AUC	Summary Assessment
Logistic regression	∼95.9%	0.89	0.60	0.714	∼0.901	Decent baseline; limited with complex non-linear patterns
Random forest	∼96.6%	1.00	0.80	0.889	∼0.956	Very strong; robust across folds; interpretable via feature importance
Gradient boosting	∼85%	0.77	0.54	0.630	∼0.856	Underperformed expectations; marginal discrimination ability
Support vector machine	∼83%	0.45	0.34	0.212	∼0.667	Poor result; likely a non-linear or high-dimensional challenge without tuning
Multilayer perceptron (neural network)	∼84%	0.57	0.45	0.191	∼0.776	Undefined precision warnings; underperformed tree-based models

Note: ROC-AUC = receiver operating characteristic-area under the curve.

Five-fold cross-validation in Table 3 showed that the complete random forest model achieved 1.000 ± 0.000 for every reported metric. This perfect separation is not evidence of broad clinical generalizability; it reflects inclusion of body mass index, the variable used to define the outcome. Removing body mass index reduced accuracy to 0.911 ± 0.038 and recall to 0.624 ± 0.168, although the area under the receiver operating characteristic curve remained high at 0.972 ± 0.023. Excluding body mass index, height, and weight further reduced accuracy to 0.842 ± 0.023 and area under the receiver operating characteristic curve to 0.815 ± 0.056, while recall fell to 0.333 ± 0.082. These findings demonstrate both residual predictive information and substantial dependence on the defining anthropometric variables.

The complete-model results were identical across folds because body mass index encoded the outcome threshold. The body mass index-excluded model was less stable, particularly for recall and precision, showing that the limited number of obese observations affected sensitivity estimates. The stricter model retained moderate discrimination but identified only about one-third of obese participants, which would be inadequate for direct screening use without recalibration and external validation. Accordingly, the high original accuracy should not be interpreted as superiority over other algorithms or evidence of readiness for deployment. It instead provides an upper-bound classification result under a feature set containing the label-defining measurement. The ablation analyses offer a more relevant estimate of how much information is available from other body-composition, health, and lifestyle factors.

Table 3. Five-fold cross-validation for the random forest model

Feature Set	Accuracy	Precision	Recall	F1-Score	ROC-AUC	Interpretation
Complete model, including BMI	1.000 ± 0.000	1.000 ± 0.000	1.000 ± 0.000	1.000 ± 0.000	1.000 ± 0.000	BMI directly defines the outcome; the result reflects label reconstruction.
BMI excluded	0.911 ± 0.038	0.822 ± 0.133	0.624 ± 0.168	0.701 ± 0.125	0.972 ± 0.023	Other variables retain discrimination, but sensitivity is lower and variable.
BMI, height, and weight excluded	0.842 ± 0.023	0.583 ± 0.128	0.333 ± 0.082	0.419 ± 0.088	0.815 ± 0.056	Moderate discrimination; insufficient sensitivity for direct screening.
Validation design	Stratified 5-fold cross-validation	Mean ± standard deviation	Random state 42	Class-weighted random forest	n = 279	48 obese and 231 non-obese participants.

Note: BMI = body mass index; ROC-AUC = receiver operating characteristic-area under the curve.

4.4 Feature Importance Analysis

After identifying random forest as the best-performing model, the study then interrogated which variables were driving predictions. Table 4 presents the normalized importance scores derived from the ensemble's internal feature-weighting mechanism.

Table 4. Feature importance scores (random forest)

Feature	Importance Score
Body mass index	0.364
Weight	0.224
Waist circumference	0.062
Age	0.062
Hip circumference	0.056
Waist-hip ratio	0.038
Height	0.034
Family history of obesity (yes)	0.021
Chronic illness (e.g., diabetes, cardiovascular disease) (yes)	0.016
Employment status (unemployed)	0.013

The feature-importance analysis confirmed that body mass index and weight dominated the complete model. This is expected because obesity status was defined from body mass index, while weight is a direct component of the body mass index calculation. Their importance therefore represents target-definition dependence rather than a novel biological discovery. In the body mass index-excluded analysis, waist and hip measures, waist-to-hip ratio, age, physical-health score, and selected lifestyle variables contributed to prediction, indicating that central adiposity and broader health patterns provided information beyond body mass index alone. The practical value of explainable artificial intelligence in this setting is not simply to restate that larger body size is associated with obesity. Rather, explainable artificial intelligence exposes why the complete model appears exceptionally accurate, distinguishes label-defining from supplementary predictors, and identifies which variables continue to influence classification when body mass index is removed. This helps prevent an inflated interpretation of model performance and supports clinically meaningful sensitivity analysis.

These rankings are not merely a modeling artefact—they align closely with clinical reasoning in obesity risk screening, where body composition indices are typically assessed first before behavioral and socioeconomic determinants are explored [30, 32]. Nevertheless, the non-zero contribution of employment status is noteworthy in the Saudi context, where labor market participation intersects with physical activity patterns and dietary behaviors in ways that are population-specific [42].

4.5 Explainable Artificial Intelligence Interpretations: Shapley Additive Explanations and Local Interpretable Model-Agnostic Explanations

Beyond aggregate performance metrics, a major contribution of this study is the use of explainable artificial intelligence to provide both global and individual-level interpretations of the machine learning models. Two complementary explainable artificial intelligence techniques were employed: Shapley additive explanations quantify the global contribution of each predictor across the dataset, and local interpretable model-agnostic explanations explain individual predictions at the participant level [11, 12]. Together, these methods improve transparency and facilitate understanding of how the model reaches its predictions, an important consideration for clinical decision-support systems [43].

The Shapley additive explanations summary plot shown in Figure 4 illustrates the distribution of feature contributions across all observations, with color representing the magnitude of each feature value. In the complete random forest model, body mass index and weight produced the largest Shapley additive explanations values, confirming that the classifier relied predominantly on the variables used to define the obesity outcome. This finding complements the sensitivity analyses and demonstrates the dependence of the complete model on label-defining anthropometric variables. When body mass index was excluded from the model, the relative importance shifted towards waist circumference, hip circumference, waist-to-hip ratio, age, physical health, and lifestyle-related variables. This change demonstrates that the principal value of the explainable artificial intelligence analysis extends beyond identifying important predictors; it reveals potential target leakage, illustrates how feature rankings change following predictor ablation, and quantifies the incremental predictive contribution of non-body mass index variables [10].

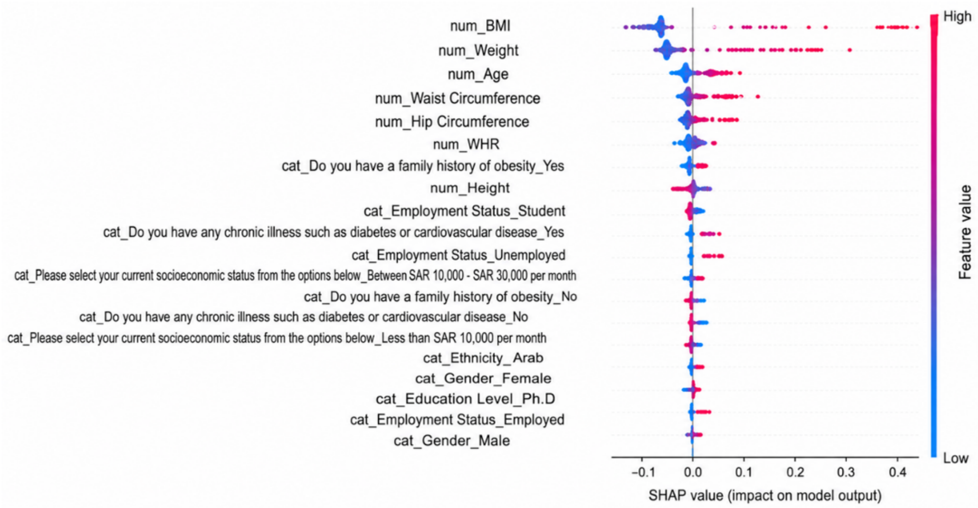


Figure 4. Shapley additive explanations summary plot illustrating global feature contributions to random forest obesity prediction

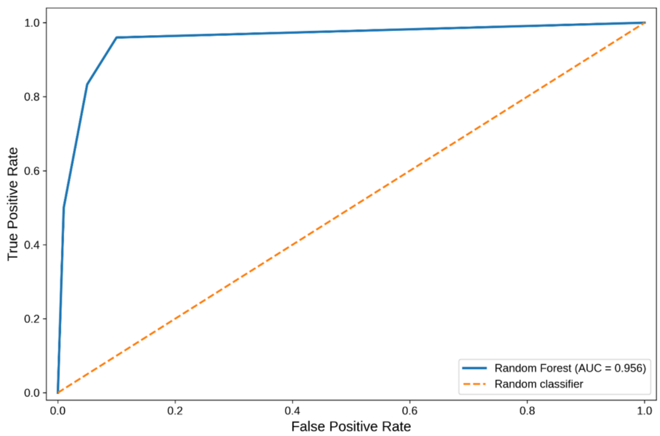
Note: BMI = Body Mass Index; WHR = Waist-to-Hip Ratio; SHAP = SHapley Additive exPlanations.

Local interpretable model-agnostic explanations provided complementary instance-level explanations that illustrate how the model combines multiple variables when classifying individual participants. In Case 1, a 23-year-old woman with a weight of 99 kg, waist circumference of 94 cm, hip circumference of 121 cm, poor physical health, a family history of obesity, and limited weekly physical activity received a body mass index-excluded random forest obesity probability of 0.975. The local explanation indicated that the prediction was driven by the combined influence of anthropometric characteristics, health status, and behavioral factors rather than body mass index itself. In Case 2, a 40-year-old woman with a weight of 69 kg, waist circumference of 38 cm, hip circumference of 40 cm, a waist-to-hip ratio of 0.95, and generally favorable health behaviors received a substantially lower predicted probability of 0.680. This example demonstrates that local predictions arise from the balance of multiple risk-increasing and risk-reducing factors rather than from any single variable in isolation [12].

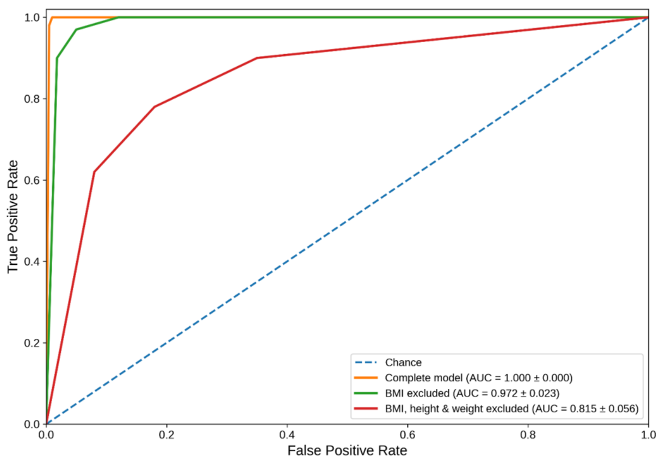
These individual explanations provide information beyond global feature rankings by illustrating how the same model can combine predictors differently for different participants. Such explanations may improve clinician confidence, facilitate personalized risk communication, and support transparent review of borderline cases [2, 44]. However, Shapley additive explanations and local interpretable model-agnostic explanations describe model behavior rather than causal relationships and therefore should not be interpreted as evidence that individual predictors directly cause obesity. Furthermore, explainability methods cannot compensate for limitations related to sample size, self-reported anthropometric measurements, or the absence of external validation. Consequently, the proposed explainable artificial intelligence framework should be regarded as a tool for improving model transparency and hypothesis generation rather than as a substitute for independent clinical validation before implementation in routine healthcare practice [43, 44].

4.6 Receiver Operating Characteristic Curve Analysis

Figure 5a illustrates the receiver operating characteristic curve for the complete random forest model. The curve lies close to the upper-left corner of the receiver operating characteristic space, indicating excellent discrimination between obese and non-obese participants in an individual validation run [31]. However, to provide a more robust estimate of model performance, the study reports fold-wise cross-validation results rather than relying solely on a single train-test split. Across five stratified cross-validation folds, the complete model achieved a mean area under the receiver operating characteristic curve of 1.000 ± 0.000, reflecting the fact that body mass index directly defines the obesity outcome. When body mass index was removed from the predictors, the mean area under the receiver operating characteristic curve decreased to 0.972 ± 0.023, while the stricter model, excluding body mass index, height, and weight, further decreased to 0.815 ± 0.056, as shown in Figure 5b. The progressive decline in discrimination demonstrates that much of the predictive performance of the complete model is driven by label-defining anthropometric variables, whereas the remaining body composition, demographic, and lifestyle features provide more moderate but meaningful discriminatory information.



(a)



(b)

Figure 5. Receiver operating characteristic curve analysis of the random forest models

Note: AUC = area under the curve; BMI = body mass index.

(a) Receiver operating characteristic curve for the complete random forest model based on an individual validation split (area under the curve 0.956); and (b) Comparison of receiver operating characteristic curves for the three random forest models: the complete model, including body mass index; the model excluding body mass index; and the model excluding body mass index, height, and weight. The progressive reduction in area under the receiver operating characteristic curve illustrates the dependence of discrimination on label-defining anthropometric variables while demonstrating the remaining predictive contribution of body composition, demographic, and lifestyle features.

4.7 Physical and Mental Health Correlation

An auxiliary correlation analysis explored the relationship between participants' physical and mental health self-reports. As depicted in Figure 6, a moderate positive correlation (r = 0.4048) was observed: individuals who rated their physical health more favorably also tended to report better mental well-being. While this correlation does not establish causation, it is broadly consistent with a body of evidence linking physical fitness, obesity, and psychological outcomes across diverse populations. In the Saudi context, where obesity rates have risen alongside increasing sedentary lifestyles, this association may reflect a shared lifestyle determinant influencing both domains simultaneously [27].

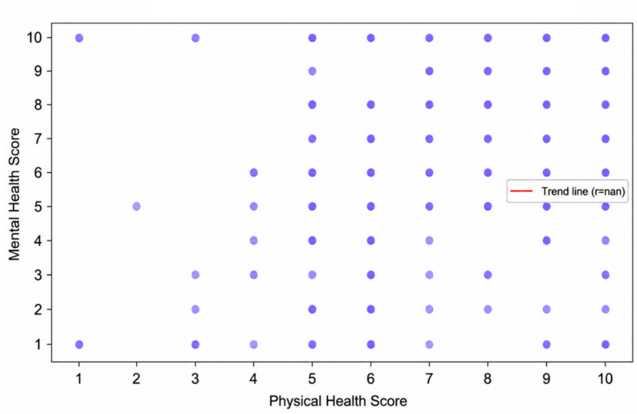


Figure 6. Physical–mental health score correlation

4.8 Blood Pressure Distribution

Figure 7 presents the distribution of blood pressure readings among participants. The majority fell within the normal range, as would be expected in a community survey sample. However, 24 participants recorded occasional high blood pressure readings and 6 reported low blood pressure episodes. Though these counts are relatively small, the co-occurrence of hypertension and overweight or obesity is well established [22], and their presence in this sample—even at modest frequencies—is a reminder that the dataset captures clinically heterogeneous individuals whose comorbidity profiles extend beyond adiposity alone. These blood pressure irregularities were not used as direct predictors but are worth noting as potential confounders in future longitudinal extensions of this work.

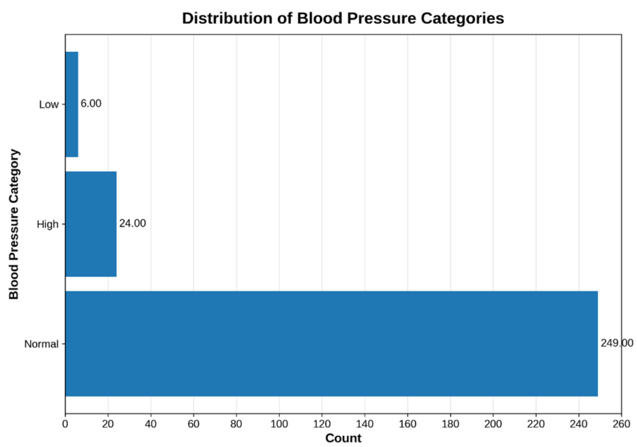


Figure 7. Blood pressure abnormalities

4.9 Comparison With Prior Machine Learning Obesity Prediction Studies

Table 5 compares the proposed work with prior obesity-prediction studies. Because datasets, outcome definitions, feature sets, sample sizes, validation designs, and class distributions differ, accuracy values are not directly interchangeable. Area under the curve, recall, and F1-score should be considered where reported, and studies that include body mass index while defining obesity from body mass index should be interpreted as classification rather than independent prediction.

Table 5. Comparison with related machine learning studies on obesity prediction

Study	Sample Size/Data Source	Best Model	Accuracy	ROC-AUC	Recall/F1	Comparison Note
Proposed complete model	279; cross-sectional survey	Random forest	1.000 ± 0.000	1.000 ± 0.000	1.000/1.000	BMI included and used to define the outcome; not comparable as an independent prediction
Proposed model with BMI excluded	279; cross-sectional survey	Random forest	0.911 ± 0.038	0.972 ± 0.023	0.624/0.701	More relevant estimate of information beyond BMI
Musleh et al. [41]	310; patient records	Logistic regression	0.910	0.910 ± 0.087	0.910/0.087	Different cardiometabolic tasks and clinical data sources
Görmez et al. [45]	500; online survey	Categorical boosting	0.9367	0.993 ± 0.173	0.935/0.149	Different survey, features, and outcome distribution, maintained relatively low SD values
Dutta et al. [40]	800; survey data	Random forest	0.960	Not reported	Not reported	Larger sample and different feature set
Maria et al. [46]	2111; online open data	Random forest	0.948	Not reported	Not reported	Open source, different data sets

Note: BMI = body mass index; ROC-AUC = receiver operating characteristic-area under the curve; SD = standard deviation.

The complete model should not be described as outperforming prior studies because its outcome is directly encoded by body mass index. The body mass index-excluded results provide a fairer assessment: accuracy 0.911 ± 0.038, recall 0.624 ± 0.168, F1-score 0.701 ± 0.125, and area under the receiver operating characteristic curve 0.972 ± 0.023. The stricter model produced lower but more realistic estimates when body mass index, height, and weight were removed. Differences from previous studies may reflect sample size, class prevalence, survey measurement, target definition, and validation design rather than algorithmic superiority. For previous studies, not all additional metrics were available in the cited manuscript text. Therefore, unavailable area under the curve, recall, or F1-score values are indicated as "not reported" rather than inferred. This avoids misleading direct comparisons based on incomplete information. The principal contribution of the present revision is the transparent ablation and stability analysis rather than a claim of highest accuracy. The comparison consequently supports a more balanced conclusion: the proposed framework demonstrates internally useful discrimination, but its generalizability remains uncertain because it was trained on a small convenience sample with self-reported measurements and no external test cohort.

5 Discussion

This study demonstrates that machine learning models integrating anthropometric, lifestyle, demographic, and health-related variables can effectively classify body mass index-defined obesity risk among Saudi adults. Consistent with previous cardiometabolic and obesity prediction studies, ensemble tree-based methods, particularly random forest, achieved the strongest predictive performance, confirming their suitability for modeling complex biological and behavioral datasets [47, 48]. However, the sensitivity analyses performed in response to this study revealed that the excellent performance of the complete random forest model was largely attributable to the inclusion of body mass index, which is mathematically linked to the obesity outcome. The complete model reproduced body mass index-defined obesity with near-perfect discrimination, whereas removing body mass index reduced both predictive accuracy and recall and introduced greater fold-to-fold variability. A stricter analysis, excluding body mass index, height, and weight, further reduced model performance, indicating that although waist and hip measurements, waist-to-hip ratio, age, physical health, and lifestyle variables contain meaningful predictive information, their discriminative ability is considerably more modest. These findings suggest that the complete model should be interpreted primarily as an automated body mass index-classification framework rather than an independent obesity prediction model.

The body mass index-ablation analysis provides important insight into the contribution of non-body mass index variables. While body mass index, weight, and waist circumference remained the strongest contributors, consistent with epidemiological studies highlighting abdominal adiposity and anthropometric indices as major obesity risk factors in Middle Eastern populations [17, 32], the reduced performance of the body mass index-excluded models indicates that self-reported behavioral, demographic, and body composition variables alone cannot reliably identify all obese individuals. The lower recall observed after body mass index removal suggests that a proportion of obese participants cannot be accurately classified using the remaining features, highlighting the continued importance of objective anthropometric measurements for obesity screening while demonstrating that complementary lifestyle and body composition indicators provide additional, although limited, predictive value.

The integration of explainable artificial intelligence strengthened the interpretability of the proposed framework. Shapley additive explanations and local interpretable model-agnostic explanations provided both global and individual-level explanations that are consistent with recent explainable artificial intelligence-enabled clinical prediction studies [43]. Beyond identifying feature importance, the explainability analyses revealed how model behavior changed after body mass index removal, with greater reliance placed on central adiposity measures, age, physical health, and lifestyle-related variables. The individual local interpretable model-agnostic explanations further demonstrated that predictions were driven by combinations of interacting risk factors rather than isolated variables, thereby supporting more transparent and clinically interpretable decision-making [44]. Nevertheless, these explanations should not be interpreted as evidence of causal relationships; Shapley additive explanations and local interpretable model-agnostic explanations describe how the trained model reaches its predictions rather than establishing causal effects.

Overall, the proposed framework compares favorably with previous obesity prediction studies by combining competitive predictive performance with model interpretability (Tables 1–3). However, the results should be interpreted cautiously because the study was based on a relatively small cross-sectional dataset with self-reported measurements and internal validation only. The possibility of target leakage through body mass index emphasizes the importance of sensitivity analyses when outcome-defining variables are included as predictors. Consequently, the present framework should be regarded as an internally validated proof of concept rather than a clinically deployable screening tool. Future studies should evaluate the proposed approach using larger and more geographically diverse cohorts, objective anthropometric measurements, and independent external validation before considering implementation in routine clinical practice or public health programs. Such developments could ultimately support preventive counselling, personalized lifestyle interventions, and resource allocation within the Saudi Vision 2030 Health Sector Transformation Program [49].

5.1 Limitations

Several limitations require emphasis. First, the cross-sectional convenience sample was modest (n = 279) and included only 48 participants meeting the body mass index-based obesity definition, producing uncertainty in recall and precision across folds. Second, anthropometric and lifestyle variables were self-reported, and some height, waist, hip, and caloric-intake entries required harmonization or imputation. Third, body mass index, height, which may introduce reporting bias [50], and weight are mathematically related to the outcome, and including them inflates apparent performance. Fourth, the study used internal cross-validation only; no independent Saudi or non-Saudi cohort was available. Fifth, the sample may not represent the regional, socioeconomic, or demographic distribution of Saudi Arabia. The models should therefore not be described as directly generalizable or clinically deployable. Future work should use objectively measured anthropometry, larger probability-based samples, repeated or nested validation, calibration assessment, subgroup fairness analysis, prospective follow-up, and external validation.

6 Conclusion

This study developed an interpretable machine-learning framework integrating anthropometric, lifestyle, demographic, and health-related variables for body mass index-defined obesity classification among Saudi adults. Tree-based models, particularly random forest, demonstrated excellent predictive performance during internal validation. However, sensitivity analyses showed that this performance was strongly influenced by the inclusion of body mass index, which directly defines the obesity outcome. Removing body mass index reduced the random forest accuracy to 0.911 ± 0.038 and recall to 0.624 ± 0.168, while excluding body mass index, height, and weight further reduced the area under the receiver operating characteristic curve to 0.815 ± 0.056. These findings indicate that although anthropometric variables remain the strongest predictors, waist and hip measurements, waist-to-hip ratio, age, physical health, lifestyle behaviors, sleep quality, and socioeconomic factors provide meaningful complementary information for obesity classification.

The integration of Shapley additive explanations and other explainable artificial intelligence techniques enhanced model transparency by providing both global and individual-level interpretations of prediction outcomes, thereby supporting clinically meaningful understanding of the contributing risk factors. Compared with previous regional studies, the proposed framework demonstrates competitive predictive performance while offering improved interpretability. Nevertheless, the results should be interpreted cautiously because the study represents an internally validated proof of concept based on a relatively small cross-sectional dataset. Larger, geographically and demographically diverse cohorts, objective anthropometric measurements, and external validation are required before clinical or public health implementation. Future research should incorporate longitudinal follow-up, wearable sensor data, and nutritional and metabolic biomarkers, and explore hybrid ensemble, deep-learning, and causal inference approaches to improve robustness, generalizability, and clinical applicability.

Authors Contributions

Conceptualization, A.S.M., S.F.A., S.N.B., and N.N.; methodology, A.S.M., S.F.A., and N.N.; software, A.S.M.; formal analysis, N.N., K.K., and W.M.A.; data curation, S.F.A., K.K., and W.M.A.; writing—original draft preparation, A.S.M.; writing—review and editing, S.F.A., S.N.B., N.N., K.K., and W.M.A.; visualization, S.F.A.; supervision, A.S.M. All authors have read and agreed to the published version of the manuscript.

Informed Consent Statement

This study was conducted in accordance with the ethical principles outlined in the Declaration of Helsinki. The protocol was approved by the Ethics Committee of King Abdullah International Medical Research Center, Al Ahsa Saudi Arabia (Approval No.: IRB/2060/23). Informed consent was obtained from all participants prior to their inclusion in the study and their confidentiality and privacy were strictly protected. Participation was entirely voluntary.

Data Availability

Data will be available from the corresponding author upon reasonable request and with approval from the Ethics Committee. The source code and model configuration files will be made available by the corresponding author upon reasonable request to facilitate reproducibility.

Acknowledgements

The authors would like to thank all those who participated in this study by completing the survey.

Conflicts of Interest

The authors declare no conflicts of interest.

References

[1] M. S. Alkalbi, M. H. Al-Mutairi, A. M. T. Almutairi, S. H. Albeshrey, S. A. Alharthi, N. M. H. Mousa, and K. M. S. Almohammed, "The role of artificial intelligence in predicting and preventing lifestyle-related diseases in rural Saudi Arabia: A systematic review," *J. Med. Life Sci.*, vol. 6, no. 4, pp. 733–739, 2024. https://doi.org/10.21608/jmals.2024.426259

[2] S. Azmi, F. Kunnathodi, H. F. Alotaibi, W. Alhazzani, M. Mustafa, I. Ahmad, R. Anvarbatcha, M. D. Lytras, and A. A. Arafat, "Harnessing artificial intelligence in obesity research and management: A comprehensive review," *Diagnostics*, vol. 15, no. 3, p. 396, 2025. https://doi.org/10.3390/diagnostics15030396

[3] T. Khater, H. Tawfik, and B. Singh, "Explainable artificial intelligence for investigating the effect of lifestyle factors on obesity," *Intell. Syst. Appl.*, vol. 23, p. 200427, 2024. https://doi.org/10.1016/j.iswa.2024.200427

[4] H. G. Lazarjani, S. Zarei, F. Aboutorabi, R. Shalbaf, and A. Shalbaf, "Application of explainable artificial intelligence for predicting and diagnosing neurocognitive disorders using tabular data: A systematic review," *Arch. Comput. Methods Eng.*, pp. 1–45, 2025. https://doi.org/10.1007/s11831-025-10471-w

[5] S. A. Alsareii, M. Awais, A. M. Alamri, M. Y. AlAsmari, M. Irfan, M. Raza, and U. Manzoor, "Machine-learning-enabled obesity level prediction through electronic health records," *Comput. Syst. Sci. Eng.*, vol. 46, no. 3, pp. 3715–3728, 2023. https://doi.org/10.32604/csse.2023.035687

[6] S. R. N. Kalhori, F. Najafi, H. Hasannejadasl, and S. Heydari, "Artificial intelligence-enabled obesity prediction: A systematic review of cohort data analysis," *Int. J. Med. Inform.*, p. 105804, 2025. https://doi.org/10.1016/j.ijmedinf.2025.105804

[7] M. K. Mahadi, R. Rahad, A. Noman, S. Ishrat, and F. Faisal, "Understanding machine learning & its application in obesity estimation by explainable AI," in *2024 International Conference on Inventive Computation Technologies (ICICT)*, 2024, pp. 112–117. https://doi.org/10.1109/ICICT60155.2024.10544438

[8] M. Al-Remawi, R. A. Abdel-Rahem, F. Al-Akayleh, F. Aburub, and A. S. A. Agha, "Transforming obesity care through artificial intelligence: Real-case implementations and personalized solutions," in *2025 1st International Conference on Computational Intelligence Approaches and Applications (ICCIAA)*, 2025, pp. 1–5. https://doi.org/10.1109/ICCIAA65327.2025.11013698

[9] R. Qasrawi, A. Ajab, L. Cheikh Ismail *et al.*, "What drives weight status among female university students? A machine learning analysis of sociodemographic, dietary, and lifestyle determinants," *Front. Nutr.*, vol. 12, p. 1574063, 2025. https://doi.org/10.3389/fnut.2025.1574063

[10] S. M. Ganie, P. K. D. Pramanik, and Z. Zhao, "Lifestyle data-based multiclass obesity prediction with interpretable ensemble models incorporating SHAP and LIME analysis," *Sci. Rep.*, vol. 15, no. 1, p. 36916, 2025. https://doi.org/10.1038/s41598-025-20936-4

[11] L. Wu, "A review of the transition from Shapley values and SHAP values to RGE," *Statistics*, pp. 1–23, 2025. https://doi.org/10.1080/02331888.2025.2487853

[12] M. S. Ahmad and R. A. Canessane, "Explainable AI for public health: Predictive modelling of non-communicable diseases risk using SHAP and LIME," in *2025 3rd International Conference on Sustainable Computing and Data Communication Systems (ICSCDS)*, 2025, pp. 1942–1948. https://doi.org/10.1109/ICSCDS65426.2025.11166917

[13] R. An, J. Shen, and Y. Xiao, "Applications of artificial intelligence to obesity research: Scoping review of methodologies," *J. Med. Internet Res.*, vol. 24, no. 12, p. e40589, 2022. https://doi.org/10.2196/40589

[14] S. Amjad, I. Liaqat, Z. Amjad, S. Fatima, M. Hashmi, W. Khan, and M. Nasir, "Genomic insights: Predicting obesity through AI and machine learning," in *Complementary and Alternative Medicine: Non-Conventional Therapies*, Faisalabad, Pakistan, 2024, pp. 90–98. https://doi.org/10.47278/book.CAM/2024.297

[15] F. M. Delpino, Â. K. Costa, M. C. do Nascimento, H. S. D. Moura, H. G. dos Santos, R. M. Wichmann, A. D. P. Chiavegatto Filho, R. A. Arcêncio, and B. P. Nunes, "Does machine learning have a high performance to predict obesity among adults and older adults? A systematic review and meta-analysis," *Nutr. Metab. Cardiovasc. Dis.*, vol. 34, no. 9, pp. 2034–2045, 2024. https://doi.org/10.1016/j.numecd.2024.05.020

[16] J. Jeon, S. Lee, and C. Oh, "Age-specific risk factors for the prediction of obesity using a machine learning approach," *Front. Public Health*, vol. 10, p. 998782, 2022. https://doi.org/10.3389/fpubh.2022.998782

[17] S. Ghiasi Hafezi, A. Ghasemabadi, N. Soleimani, M. Allahyari, M. Moradi, A. Mansoori, R. Kolahi Ahari, M. Ghamsary, G. Ferns, H. Esmaily *et al.*, "Predictive value of anthropometric indices for incident of dyslipidemia: A large population-based study," *Popul. Health Metr.*, vol. 23, no. 1, p. 48, 2025. https://doi.org/10.1186/s12963-025-00410-z

[18] Y. Hong, X. Chen, L. Wang, F. Zhang, Z. Zeng, and W. Xie, "Machine learning prediction of metabolic dysfunction-associated fatty liver disease risk in American adults using body composition: Explainable analysis based on SHapley Additive eXplanations," *Front. Nutr.*, vol. 12, p. 1616229, 2025. https://doi.org/10.3389/fnut.2025.1616229

[19] H. Wahabi, A. A. Fayed, Z. Shata, S. Esmaeil, R. Alzeidan, E. Saeed, Y. Amer, M. Titi, K. Bahkali, and L. Hneiny, "The impact of age, gender, temporality, and geographical region on the prevalence of obesity and overweight in Saudi Arabia: Scope of evidence," *Healthcare*, vol. 11, no. 8, p. 1143, 2023. https://doi.org/10.3390/healthcare11081143

[20] B. Singh and H. Tawfik, "Machine learning approach for the early prediction of the risk of overweight and obesity in young people," in *Lecture Notes in Computer Science*, 2020, pp. 523–535. https://doi.org/10.1007/978-3-030-50423-6_39

[21] K. Jindal, N. Baliyan, and P. S. Rana, "Obesity prediction using ensemble machine learning approaches," in *Recent Findings in Intelligent Computing Techniques*, Singapore, 2018, pp. 355–362. https://doi.org/10.1007/978-981-10-8636-6_37

[22] S. Alghnam, S. A. Alessy, M. Bosaad, S. Alzahrani, I. I. Al Alwan, A. Alqarni, R. Alshammari, M. Al Dubayee, and M. Alfadhel, "The association between obesity and chronic conditions: Results from a large electronic health records system in Saudi Arabia," *Int. J. Environ. Res. Public Health*, vol. 18, no. 23, p. 12361, 2021. https://doi.org/10.3390/ijerph182312361

[23] R. C. Cervantes and U. M. Palacio, "Estimation of obesity levels based on computational intelligence," *Inform. Med. Unlocked*, vol. 21, p. 100472, 2020. https://doi.org/10.1016/j.imu.2020.100472

[24] R. Hammond, R. Athanasiadou, S. Curado, Y. Aphinyanaphongs, C. Abrams, M. J. Messito, R. Gross, M. Katzow, M. Jay, N. Razavian *et al.*, "Predicting childhood obesity using electronic health records and publicly available data," *PLoS ONE*, vol. 14, no. 4, p. e0215571, 2019. https://doi.org/10.1371/journal.pone.0215571

[25] H. Hochner, R. Butterman, I. Margaliot, Y. Friedlander, and M. Linial, "Obesity risk in young adults from the Jerusalem Perinatal Study (JPS): The contribution of polygenic risk and early life exposure," *Int. J. Obes.*, vol. 48, no. 7, pp. 954–963, 2024. https://doi.org/10.1038/s41366-024-01505-7

[26] F. Ferdowsy, K. S. A. Rahi, M. I. Jabiullah, and M. T. Habib, "A machine learning approach for obesity risk prediction," *Curr. Res. Behav. Sci.*, vol. 2, p. 100053, 2021. https://doi.org/10.1016/j.crbeha.2021.100053

[27] N. A. Althumiri, M. H. Basyouni, N. AlMousa, M. F. AlJuwaysim, R. A. Almubark, N. F. BinDhim, Z. Alkhamaali, and S. A. Alqahtani, "Obesity in Saudi Arabia in 2020: Prevalence, distribution, and its current association with various health conditions," *Healthcare*, vol. 9, no. 3, p. 311, 2021. https://doi.org/10.3390/healthcare9030311

[28] P. C. Austin, I. R. White, D. S. Lee, and S. van Buuren, "Missing data in clinical research: A tutorial on multiple imputation," *Can. J. Cardiol.*, vol. 37, no. 9, pp. 1322–1331, 2021. https://doi.org/10.1016/j.cjca.2020.11.010

[29] A. Maulana, R. P. F. Afidh, N. B. Maulydia, G. M. Idroes, and S. Rahimah, "Predicting obesity levels with high accuracy: Insights from a CatBoost machine learning model," *Infolitika J. Data Sci.*, vol. 2, no. 1, pp. 17–27, 2024. https://doi.org/10.60084/ijds.v2i1.195

[30] P. Piqueras, A. Ballester, J. V. Durá-Gil, S. Martinez-Hervas, J. Redón, and J. T. Real, "Anthropometric indicators as a tool for diagnosis of obesity and other health risk factors: A literature review," *Front. Psychol.*, vol. 12, p. 631179, 2021. https://doi.org/10.3389/fpsyg.2021.631179

[31] P. O. Akinwumi, S. Ojo, T. I. Nathaniel, J. Wanliss, O. Karunwi, and M. Sulaiman, "Evaluating machine learning models for stroke prediction based on clinical variables," *Front. Neurol.*, vol. 16, p. 1668420, 2025. https://doi.org/10.3389/fneur.2025.1668420

[32] A. Chatterjee, M. W. Gerdes, and S. G. Martinez, "Identification of risk factors associated with obesity and overweight—A machine learning overview," *Sensors*, vol. 20, no. 9, p. 2734, 2020. https://doi.org/10.3390/s20092734

[33] E. Halabaku and E. Bytyçi, "Overfitting in machine learning: A comparative analysis of decision trees and random forests," *Intell. Autom. Soft Comput.*, vol. 39, no. 6, pp. 987–1006, 2024. https://doi.org/10.32604/iasc.2024.059429

[34] I. L. Cherif and A. Kortebi, "On using extreme gradient boosting (XGBoost) machine learning algorithm for home network traffic classification," in *2019 Wireless Days (WD)*, 2019, pp. 1–6. https://doi.org/10.1109/WD.2019.8734193

[35] S. Ghosh, A. Dasgupta, and A. Swetapadma, "A study on support vector machine based linear and non-linear pattern classification," in *2019 International conference on intelligent sustainable systems (ICISS)*, 2019, pp. 24–28. https://doi.org/10.1109/ISS1.2019.8908018

[36] A. A. Safar, D. M. Salih, and A. M. Murshid, "Pattern recognition using the multi-layer perceptron (MLP) for medical disease: A survey," *Int. J. Nonlinear Anal. Appl.*, vol. 14, no. 1, pp. 1989–1998, 2023. https://doi.org/10.22075/ijnaa.2022.7114

[37] S. E. Abhadiomhen, E. O. Nzeakor, and K. Oyibo, "Health risk assessment using machine learning: Systematic review," *Electronics*, vol. 13, no. 22, p. 4405, 2024. https://doi.org/10.3390/electronics13224405

[38] F. Palmieri, N. F. Akhtar, A. Pané, A. Jiménez, R. P. Olbeyra, J. Viaplana, J. Vidal, A. de Hollanda, P. Gama-Perez, J. C. Jiménez-Chillarón *et al.*, "Machine learning allows robust classification of visceral fat in women with obesity using common laboratory metrics," *Sci. Rep.*, vol. 14, no. 1, p. 17263, 2024. https://doi.org/10.1038/s41598-024-68269-y

[39] M. Taiyeb Khosroshahi, S. Morsali, S. Gharakhanlou, A. Motamedi, S. Hassanbaghlou, H. Vahedi, S. Pedrammehr, H. M. D. Kabir, and A. Jafarizadeh, "Explainable artificial intelligence in neuroimaging of Alzheimer's disease," *Diagnostics*, vol. 15, no. 5, p. 612, 2025. https://doi.org/10.3390/diagnostics15050612

[40] R. R. Dutta, I. Mukherjee, and C. Chakraborty, "Obesity disease risk prediction using machine learning," *Int. J. Data Sci. Anal.*, vol. 19, no. 4, pp. 709–718, 2024. https://doi.org/10.1007/s41060-023-00491-9

[41] D. Musleh, A. Alkhwaja, I. Alkhwaja, M. Alghamdi, H. Abahussain, M. Albugami, F. Alfawaz, S. El-Ashker, and M. Al-Hariri, "Machine learning approaches for predicting risk of cardiometabolic disease among university students," *Big Data Cogn. Comput.*, vol. 8, no. 3, p. 31, 2024. https://doi.org/10.3390/bdcc8030031

[42] R. L. Lee, W. T. Chien, K. Tanida, S. Takeuchi, P. Rutja, S. W. H. Kwok, and P. H. Lee, "The association between demographic characteristics, lifestyle health behaviours, and quality of life among adolescents in Asia Pacific region," *Int. J. Environ. Res. Public Health*, vol. 16, no. 13, p. 2324, 2019. https://doi.org/10.3390/ijerph16132324

[43] H. Hoghooghi Esfahani, S. Toyonaga, and K. Oyibo, "The application of explainable artificial intelligence in the prediction, diagnoses, treatment, and management of chronic diseases: A systematic review," *Digit. Health*, vol. 11, p. 20552076251355669, 2025. https://doi.org/10.1177/20552076251355669

[44] C. Molnar, *Interpretable Machine Learning*, 2022. https://christophm.github.io/interpretable-ml-book/

[45] Y. Görmez, F. H. Yagin, B. Yagin, Y. Aygun, H. Boke, G. Badicu, M. S. De Sousa Fernandes, A. Alkhateeb, M. B. A. Al-Rawi, and M. Aghaei, "Prediction of obesity levels based on physical activity and eating habits with a machine learning model integrated with explainable artificial intelligence," *Front. Physiol.*, vol. 16, p.

1549306, 2025. https://doi.org/10.3389/fphys.2025.1549306

[46] A. S. Maria, R. Sunder, and R. S. Kumar, "Obesity risk prediction using machine learning approach," in *2023 International Conference on Networking and Communications (ICNWC)*, 2023. https://doi.org/10.1109/ICNWC57852.2023.10127434

[47] Z. Helforoush and H. Sayyad, "Prediction and classification of obesity risk based on a hybrid metaheuristic machine learning approach," *Front. Big Data*, vol. 7, p. 1469981, 2024. https://doi.org/10.3389/fdata.2024.1469981

[48] A. S. Sheik Abdullah, V. N. P. Shashank, and D. A. L. Hudson, "Disseminating the risk factors with enhancement in precision medicine using comparative machine learning models for healthcare data," *IEEE Access*, vol. 12, pp. 72 794–72 812, 2024. https://doi.org/10.1109/ACCESS.2024.3400023

[49] Saudi Vision 2030, "Health sector transformation program." https://www.vision2030.gov.sa/en/explore/programs/health-sector-transformation-program

[50] S. Gorber, M. Tremblay, M. Moher, and B. Gorber, "A comparison of direct vs. self-report measures for assessing height, weight and body mass index: A systematic review," *Obes. Rev.*, vol. 8, no. 4, pp. 307–326, 2007. https://doi.org/10.1111/j.1467-789X.2007.00347.x