NTU Journal of Engineering and Technology (2025) 4 (2) : 95-104
DOI: https://doi.org/10.56286/ntujet.v4i2

IRAQI Academic Scientific Journals

P-ISSN: 2788-9971 E-ISSN: 2788-998X
NTU Journal of Engineering and Technology
Available online at: https://journals.ntu.edu.iq/index.php/NTU-JET/index

Predicting Obesity Levels Based on Lifestyle and Activity Patterns

Iman Nozad Mahmood[1], Swash Sami Mohammed[2]
[1]Kirkuk General Education Directorate, Kirkuk, Iraq, [2]Technical Engineering College – Kirkuk, Northern Technical University, Mosul, Iraq.
imanmahmood821@gmail.com, swash.sami23@ntu.edu.iq

Article Informations

Received: 27-02- 2025,
Revised: 21-03- 2025,
Accepted: 24-03-2025,
Published online: 23-06-2025

Corresponding author:
Name: Iman Nozad Mahmood
Affiliation : Kirkuk General Education Directorate, Kirkuk, Iraq
Email: imanmahmood821@gmail.com

Key Words:
Machine Learning,
Obesity Prediction,
Gradient Boosting,
Predictive Analytics,
Hyperparameter Tuning.

ABSTRACT

Obesity is a growing global health concern driven by genetic, behavioral, and environmental factors. Machine learning (ML) offers potential for predicting and classifying obesity, however data accessibility and model scalability present challenges. This study evaluates various machine learning algorithms for obesity prediction, including Random Forest (RF), K-Nearest Neighbors (KNN), Gradient Boosting, and Support Vector Machines (SVM). The dataset comprises 1,610 individuals, considering health, behavioral, and demographic characteristics. The aforementioned metrics were employed to evaluate the model performances; accuracy, precision, recall, and F1-score. Of interest, Logistic Regression had the lowest accuracy score (76.39%), while Gradient Boosting had the highest (88.82%). Similarly, Gradient Boosting performed well with the other metrics, reinforcing its valid conclusions for obesity classification. There is substantial potential for machine learning, as demonstrated in this study, as it will enable the early detection of obesity, and provide intervention prospects for health management professionals.

1. Introduction

Obesity has emerged as a critical public health issue worldwide, contributing significantly to the burden of chronic diseases and associated healthcare costs [1]. According to the World Health Organization (WHO), global obesity rates have nearly tripled since 1975, with over 650 million adults classified as obese in 2022 [2]. This alarming trend underscores the urgent need for effective strategies to address obesity and its associated health risks. In recent decades, the prevalence of obesity has steadily increased, affecting populations across diverse age groups, genders, and socioeconomic backgrounds [3]. There are many converging factors that contribute to this rise in obesity such as sedentary lifestyles, unhealthy eating and urbanization which may restrict physical activity [4]. There are severe health risks associated with obesity, which significantly increases the likelihood of developing an extensive list of diseases including Type 2 diabetes, cardiovascular disease, hypertension and several cancers [5, 6]. Early identification and receptor of individuals at risk of obesity will be important to improve public health outcomes and minimize the burden on the health care systems [7]. Despite the need for early detection of obesity, current obesity measures target Body Mass Index (BMI) almost exclusively, which whilst being a common measure is limited as it may not capture the complexity of obesity [8, 9]. BMI represents weight in relation to height, but likely neglects individual behaviours, demographics, and other contributory complexity [10]. The intention in this study is to assess ML algorithms for obesity classification, and identify the best performing model to offset the shortcomings of BMI alone by also including behaviour and demographic datasets. In this way, BMI-based assessment may miss critical aspects of individuals leading to less accurate obesity risk prediction. The technology related to ML has advanced and now provides an interesting path to more complex obesity classification [11]. Given the ability for machine learning algorithms to analyse complex datasets incorporating behavioural, physical, and demographic characteristics, machine learning based algorithms provide opportunities to increase the accuracy of obesity predictions, and better provide personalized risk assessments [12]. The dataset in this study consists of details for 1,610 subjects covering a variety of demographic, physical, and lifestyle factors including age, sex, height, dietary patterns, levels of physical activity and family obesity background [13]. The dataset includes this variety of factors that connotates several different components of obesity related risks, thereby allowing for greater complexity in analysis than traditional obesity assessments. Each individual in the dataset is classified into one of four obesity classes namely, underweight, normal weight, overweight, and obese, which provides a multi- class definition suitable for classification algorithms. To better prepare our analysis of the dataset, several pre-processing steps were followed including normalizing the data and encoding categorical data, before utilizing machine learning algorithms.

The primary purpose of this work is to use machine learning models that classify obesity state based on the collection of complex variables and overcome the limitations of traditional methods of evaluating obesity status [14, 15]. The study proposes that ensemble machine learning models (i.e., Gradient Boosting and Random Forest) will outperform simpler model methods (e.g. Logistic Regression) at classifying obesity. The study explores and evaluates different varying ML algorithms systematically to identify the best model to classify obesity status. The intended outcome is to advance healthcare informatics by providing a data driven approach to determining obesity status risk that can lead to more accurate and individualized interventions.

The structure of the paper is as follows: Section 1 introduces the challenges of predicting obesity levels based on lifestyle and activity patterns, along with a review of related studies. Section 2 describes the dataset and details the methodology applied in this research, including data pre-processing, feature engineering, and the selection and tuning of machine learning models. Section 3 showcases the results obtained, discussing key performance metrics and comparing the effectiveness of different models. Section 4 concludes the paper with final remarks and suggestions for future research.

2. Literature Review

The reviewed studies explored the use of ML in obesity prediction and management, highlighting advancements and limitations in this domain. For instance, Chatterjee, et al. [16] examined the effectiveness of eHealth interventions aimed at promoting healthier lifestyles, noting challenges related to data accessibility and variability, which affected model reliability and scalability. However, the study acknowledged the potential for positive health impacts, positioning eHealth as a valuable tool in preventive care. The primary limitation lay in logistical and ethical barriers, suggesting that these factors must be addressed for effective implementation. Similarly, Cheng, et al. [17] used recurrent neural network (RNN) models to make use of electronic health records (EHR) to predict childhood obesity. While these RNN models demonstrated good accuracy, using combination methods like RNN or deep learning will be highly data-dependent; it will not be effective in the absence of substantial datasets when using these machine-learning (ML) techniques. The study showed RNN models exhibited accuracy in predicting obesity, assuming the breastfeeding data was available from there were a sufficient number of clinical encounters. While these RNN models showed promise and value in clinical healthcare research; possible future models will likely require a long-term application of ML models. In a related study, Cheng, et al. [18] focused specifically on models incorporating long short-term memory (LSTM), which predicted obesity in the pediatric population and concluded requiring repeated measures of clinical input data increased performances. The authors highlighted significant potential associated with using ML models in early health interventions, but were largely limited by data availability; examining the quality of these data and the number of records will always be a factor affecting the predicted reliability. Moreover, Koklu and Sulak [19] compared a number of ML models and found random forests (Rf) performed best and their random forest model achieved an 87.82% accuracy rating for classifying obesity; they also found SVM models performed worse. This demonstrates the possible benefit of using random forests, however using other ML methods will likely require a substantial amount of optimization and a balanced dataset across the sample population to improve predicted generalization. This is a serious consideration with respect to dataset structure, and random forests were able to manage more complexity in health data better than the other models. Furthermore, Pang, et al. [20] investigated specifically a data mining process as part of the exploratory research process and noted some data treatment processes were improved prediction in terms of speed and accuracy after a feature selection approach. This made feature selection a very important consideration, particularly in managing very large datasets in health research. However, the researchers still acknowledge the need for a fairly large amount of data to get maximum predictive capabilities from their model, suggesting a system's predictive performance will also be very contingent on the size of the dataset it is provided to analyse. Likewise, Pereira, et al. [21] discussed a similar approach when reminding the reader that variable selection would improve the models efficiency and reduce processing time while using data mining techniques and ML techniques were still supportive of a rapid assessments of health indicators. The authors also offered the caveat that simplicity in variable selection (as an additional efficiency consideration) would risk losing a degree of thoroughness from the interpretation of prediction because it is data driven; as they pointed out, it is possible to a degree to oversimplify variable selection on models to increase predictive accuracy. While these studies showcase the potential for the ML process to provide good accuracy and efficiency rates, they all would suggest that continuing to examine access and be consistent with comprehensive datasets as a predominant consideration will support the chances that the results may be complete with and build reliability and generalizability across multiple populations.

3. Method

The use of several machine learning approaches to categorize obesity levels based on behavioral, physical, and demographic characteristics is examined in this study. The methodology adheres to a systematic framework that includes feature engineering, model selection, hyperparameter tuning, model training, data preparation, and performance assessment. Each stage is carefully implemented to ensure the models perform optimally in addressing this multi-class classification task.

A. Dataset Description

The dataset consists of 1,610 samples obtained from Kaggle [13], ensuring a diverse and representative sample of obesity-related factors. Each record represents an individual and containing 14 input features along with a target variable indicating obesity status. The input features encompass:

1. Demographic Features: These include variables such as Sex, Age, and Height, representing basic physical and gender-related information.
2. Behavioral and Lifestyle Attributes: These include factors like Consumption of Fast Food, Physical Exercise, and Smoking, which reflect the habits and behaviors influencing obesity.
3. Obesity Status (Target Variable): The target variable categorizes individuals into different obesity classes, such as underweight, normal weight, overweight, and obese.

Since the proposed dataset is large, Figures 1–4 highlight key trends, while other parts of the dataset are omitted for clarity.

The chart given in Fig. 1., indicates that individuals classified as Normal and Overweight are more likely to not have a family history of obesity. In contrast, individuals classified as Obese show a higher association with having a family history of obesity, suggesting that genetic and environmental factors may play a significant role in weight regulation. From a predictive modeling perspective, this makes family history a valuable feature for training an AI model, as it introduces a potential genetic component that could help differentiate between obesity classes. Including this feature in a predictive model could improve classification accuracy, especially in distinguishing between Normal and Obese categories where genetic predisposition is a major factor.

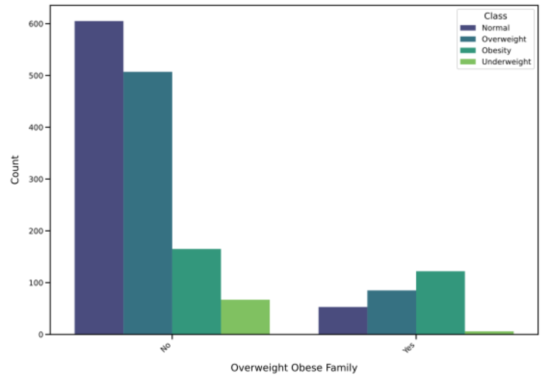

Fig. 1. Overweight Obese Family.

Furthermore, regular consumption of fast food is clearly more common among individuals classified as Obese compared to those classified as Normal or Overweight (see Fig. 2). Fast food consumption reflects dietary habits and nutritional quality, making it a strong predictive feature for an AI model. An AI model could learn to recognize patterns of high fast food consumption and predict increased obesity risk, potentially enabling targeted dietary interventions. Including this feature in model training would likely enhance the model's ability to distinguish between Overweight and Obese categories.

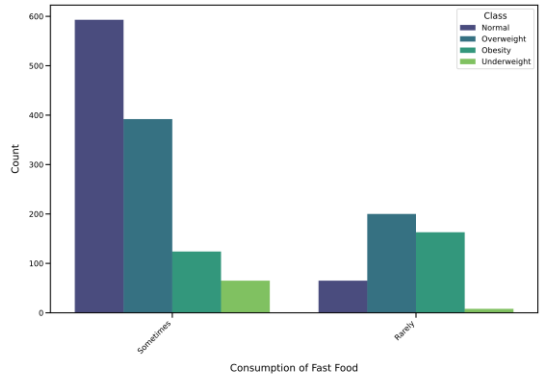

Fig. 2. Consumption of Fast Food.

Furthermore, the graph depicted in Figure 3, indicates that individuals classified as Normal and Overweight are likely to eat three main meals a day, and individuals classified as Obese are likely to eat >3 meals per day, suggesting that meal frequency is a behavioral pattern that influences body weight. The use of number of meals per day as an input feature may help improve the predictive performance of an AI model so that it recognizes behaviors and patterns in eating behavior and caloric intake that may be associated with weight gain. The model could indicate the ways in which variability in meal frequency and size influenced the transition from Normal to Overweight and the transition from Overweight to Obese.

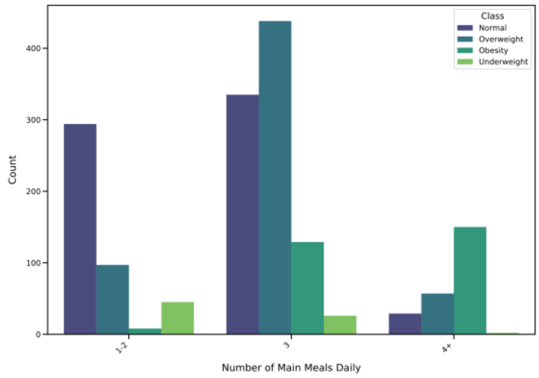

Fig. 3. Number of Main Meals Daily.

Last but not least, the chart depicted in Fig. 4, shows a clear distinction in eating behavior between obesity classes. Individuals classified as Normal and Overweight tend to consume food between meals sometimes or usually, whereas those classified as Obese have a higher tendency to eat between meals always or sometimes. This suggests that snacking behavior could be a major contributing factor to weight gain and obesity. Regular snacking increases total daily calorie intake and may lead to poor dietary habits, such as consuming high-sugar or high-fat snacks. Including food intake between meals as an input feature in an AI model could significantly enhance its predictive power by identifying patterns in eating frequency and timing that are linked to obesity. The model could potentially differentiate between individuals with controlled meal patterns and those who engage in frequent snacking, improving classification accuracy for the Obese category.

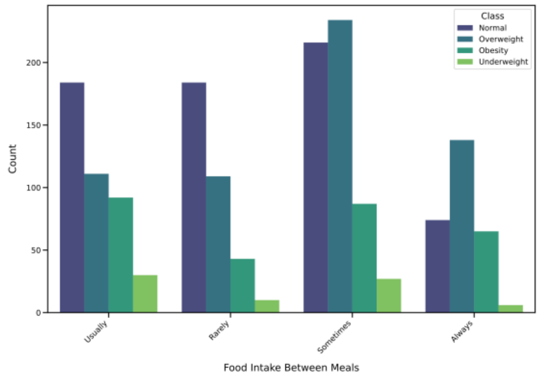

Fig. 4. Food Intake Between Meals.

B. Data Pre-Processing

Ensuring that the dataset is properly prepared for machine learning models requires effective data pretreatment. This process involves steps such as normalization, encoding categorical features, and handling the multi-class nature of the target variable.

1. Normalization Using Min-Max Scaling

Min-max scaling is applied to numerical features such as Age and Height [22]. This technique rescales the features to a standard range, typically between 0 and 1. The transformation is defined in Eq. (1).

$$X_{Scaled} = \frac{X - X_{min}}{X_{max} - X_{min}} \quad (1)$$

Where, X is the original feature value, X_{min} and X_{max}_max are the minimum and maximum values of the feature in the dataset, and X_{Scaled} is the rescaled value of X, which is now between 0 and 1. By scaling the features, the magnitude of the feature values are not biased towards models like SVM and KNN that are impacted by the scale of the features.

2. One-Hot Encoding of Categorical Features

One-hot encoding is the transformation of categorical variable data into distinguishable binary columns; Categorical variables in this dataset included Sex (with two levels) and Transportation Type (with five levels), this encoding ensures that proper representation is accomplished without any ordinal implications. One-hot encoding prepares our categorical data for use in ML, we also checked our data set for missing data and outliers. Any instances of missing data were filled using imputation; extreme outliers were either transformed or removed to ensure they do not influence model predictions.

3. Conversion to Categorical Format (Target Variable)

The target variable, which represents different obesity classes, is transformed into a categorical format for multi-class classification. This transformation ensures that the model treats each obesity class as a separate category rather than a continuous variable. In multi-class classification jobs, where the objective is to categorize occurrences into distinct groups instead of forecasting a continuous result, this is very crucial.

C. Data Splitting

After the pre-processing stage, the data partitioning occurs. The models (that were implemented) are trained with around 80% of the data, and the remaining 20% will assess the performance and evaluation of the models. A data partition allows for models to be evaluated on a set of data that they have never encountered, and thus a more objective evaluation of how the models generalize to new data.

D. Feature Engineering

Feature engineering provides a benefit to the dataset in the way it configures the input variables to be in the best format for model training. In this research, although it is clear that no additional features were generated from the existing data, one can see that normalization and encoding are important procedures. They assure that the models will interpret the data correctly. For example, the normalized feature values of Height and Age produce sage feature scales that do not wield disproportionate weight on the models. Additionally, categorical features that are one-hot encoded avert the assumption of ordinality between categories.

E. Model Selection

The study explores five machine learning models for the classification task, each representing different approaches to multi-class classification:

1. Random Forest is an ensemble learning technique creates a number of decision trees, each of which casts a vote on the final classification. This model is good at capturing intricate relationships between features and is resistant to overfitting.
2. Support Vector Machine (SVM) divides classes by building a hyperplane in a high-dimensional feature space. To capture non-linear correlations between features, a radial basis function (RBF) kernel is employed in this work.
3. 3. K-Nearest Neighbors (KNN) is a distance-based model that tends to classify a sample based upon the majority class of the k nearest neighbors in the feature space. Thus, KNN allows us to predict a class based upon all of its neighbor's class which can be representative of the neighbors' class. KNN works based on the idea that samples having similar feature values are likely having the same class.
4. The logistic (S-shaped) function is used in the linear classification model of logistic regression to estimate class probabilities. Logistic regression is the method of choice when the relationship between input features and target variable can be approximated as linear for either binary or multi-class classification problems.
5. The Gradient Boosting Classifier is an ensemble method that builds decision trees iteratively, thereby correcting errors from previous trees. Gradient boosting classifiers are generally successful at improving predictions through good overall summary prediction accuracy, minimizing the total residual error, and are a mature and appropriate option for various tasks. The selection of these particular methods arises from their success in multi-class classification, and their ability to accommodate the linear and non-linear relationships usually seen with obesity-related data.

F. Hyperparameter Tuning

Optimizing hyperparameters is a valuable step in enhancing each of the models' performance. Hyperparameter tuning is the process of choosing a good set of hyperparameters for your model based upon cross-validation. For this study, five-fold cross validation is used to determine the best hyperparameters for each model; the dataset is divided into five segments. The model is fit on four segments and validated on the fifth, and this process is repeated 5 times; this improves the model's generalization capabilities. The hyperparameters were chosen based upon performance measures such as accuracy and F1-score, which achieved a balance between precision and recall for classification tasks. The hyperparameters optimized for each model were:

1. Random Forest: Number of trees ($n_{estimators}$), maximum depth of trees (Max_{depth}), minimum samples required to split a node ($Min_{samples\ Split}$), and minimum samples required to be at a leaf node ($Min_{samples\ leaf}$) [23].
2. SVM: Regularization parameter (C), kernel type (kernel), and kernel coefficient (gamma) [24].
3. KNN: Number of neighbors ($n_{\text{neighbors}}$), weighting function (weights), and the algorithm used to compute nearest neighbors ($Algorithm$) [25].
4. Logistic Regression: Regularization strength (C), regularization type (penalty), and the optimization solver (solver) [26].
5. Gradient Boosting: Number of boosting stages ($n_{estimators}$), learning rate ($Learning_{rate}$), and the maximum depth of each tree (Max_{depth}) [27].

For Random Forest, the number of trees ($n_{estimators}$) was tested in the range of 50 to 200, with a step size of 50, while the maximum depth (Max_{depth}) was tested in the range of 10 to 50. For Gradient Boosting, the learning rate ($Learning_{rate}$) was tested in the range of 0.1 to 0.5, and the number of boosting stages ($n_{estimators}$) was tested in the range of 50 to 150. For SVM, the regularization parameter (C) was tested in the range of 0.1 to 10, and the kernel coefficient (gamma) was tested with values of 'scale' and 'auto'. For KNN, the number of neighbors ($n\_neighbors$) was tested in the range of 3 to 10, and the weighting function ($weights$) was tested with 'uniform' and 'distance' options. For Logistic Regression, the regularization strength (C) was tested in the range of 0.1 to 10, and the regularization type ($penalty$) was tested with $'l1'$ and $'l2'$ norms.

G. Model Training

The chosen models are trained on the training dataset using the best hyperparameters identified through cross-validation. During training, the models adjust their internal parameters to minimize the loss function, which quantifies the discrepancy between predicted and actual class labels. This process enables each model to identify patterns in the data associated with various obesity categories.

H. Model Evaluation

The evaluation of the models' effectiveness in classifying obesity status based on performance measures on the test dataset are done [22]. In this case the performance measures are F1-score, recall, accuracy and precision respectively. Accuracy is the measure of the proportion correct, including true positives and true negatives, of total instances. The accuracy is calculated as shown in Eq. (2).

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{2}$$

True Positives are denoted by TP, True Negatives by TN, False Positives by FP, and False Negatives by FN. Precision measures the accuracy of positive classifications by dividing the number of genuine positive predictions by the total number of positive predictions produced by the model. It is calculated using Eq. (3):

$$Precision = \frac{TP}{TP + FP} \tag{3}$$

Recall quantifies the percentage of real positive cases that the model successfully detects; it is also known as sensitivity or the true positive rate. It is computed using Eq. (4) and highlights the model's capacity to identify positive instances.

$$Recall = \frac{TP}{TP + FN} \tag{4}$$

The F1-score is a balanced statistic that takes into account both false positives and false negatives. It is computed as the harmonic mean of accuracy and recall. Because of this, it is especially helpful when there is a class disparity. The expression given in Eq. (5) is used to determine the F1-score.

$$F1\_Score = 2 \times \frac{Precision \times Recall}{Precision + Recall} \tag{5}$$

Assessing each model's performance using these metrics offers a thorough insight into their strengths and limitations in accurately classifying obesity status.

4. Training and Results

A pre-processed obesity dataset was used to model the machine learning algorithms, including Random Forest, SVM, KNN, Logistic Regression, and Gradient Boosting, after splitting the dataset into a 20% test set (322 instances) and 80% training set (1,288 instances).

To search through hyperparameter tweaks and find an optimal configuration for each model, a grid search with five-fold cross validation was conducted. Importantly, this involved searching different ranges of important hyperparameters with the goal of generating the best performance while limiting overfitting. The use of five-fold cross validation to each model further ensured that each model generalized across subsets of data from the training set.

Hyperparameters for the Random Forest model included a prescribed number of trees (n_estimators), the minimum number of samples required at a leaf, maximum depth, and minimum samples required to split a node. The best configuration used a maximum depth of 30, one sample for a leaf and 100 decision trees - trying to balance model complexity with overfitting.

The SVM used an RBF kernel to deal with non-linear interaction. The C parameter and gamma were optimized using grid tuning, the best configuration observed was C = 10 and gamma = scale, effectively separating classes in high-dimensional space.

For the KNN model, the distance measure, number of neighbours (n_neighbors), and ball tree were optimized to maximize computing time efficiency through memory use. The best configuration used 3 neighbours with distance-based weights, which allowed the model to detect local patterns per class in the dataset.

Similar to KNN model hyperparameter tuning, the Logistic Regression tuning largely focused on the regularization parameter (C). The configuration that performed best was regarded to have a value of 10 as shown above in the hyperparameter tuning process, with L2 regularization which supported a strong trade-off between fit to the model and generalization. This ultimately served as a baseline for other, more complex models.

Finally, the Gradient Boosting model hyperparameters include number of boosting stages (n_estimators), learning rate, and maximum depth of a tree were optimized. The best performance achieved was for a configuration of 100 boosting stages, a learning rate of 0.5, and a maximum depth of 7 as the model could update iteratively through the boosting process, contributing to less learning prompt through the optimization iterations.

The models were trained with their optimal hyperparameters, and were then used for evaluation against the test set. Models were assessed based on accuracy, precision, recall, and F1 score, all of which are summarized in Table 1 for the best scores achieved for each model.

Table 1. Frequencies of wbc total and differential count.

Model	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)
Random Forest	84.16	84.37	84.16	84.21
SVM	81.98	81.95	81.98	81.95
K-Nearest Neighbors	77.02	77.03	77.02	76.9
Logistic Regression	76.39	76.2	76.39	76.32
Gradient Boosting	88.82	89.04	88.82	88.85

The notable gradient boosting model had the best performance after, among the models described, it had the best accuracy of 88.82% along with the best precision of 89.04%, Recall of 88.82% and F1-score of 88.85%. In high-dimensional spaces with many features, the average performance of models can prove difficult as features can present complex relations or patterns, and gradient boosting models can also iteratively correct misclassification using the tree ensemble as well as promote the attention to more challenging records. Based on the iterative process of gradient boosting, the model reduced both the amount of false positives and false negatives in predictions as it focused on learning from the earlier iterations. Additionally, the nature of gradient boosting allows for "stacked" models of weak learners to improve the performance of predicting obesity increasing its overall performance output.

In terms of next best performance, the random forest model had a good accuracy performance of 84.16% (accuracies were comparable along with precision and recall scores being comparable). Random Forest models provide attainment of robustness as they take multiple decision trees, but they fell short to gradient boosting as the random forest models have no previous trees to adjust random sampling with and are random related to sequence; random forest models while they produce many trees emphasize on comparison, so the order of trees was not relevant to predicting obesity status which when giving many factors contributed to complexity behind model use use. In some cases, they still chose random forest as a preferred mode due at least as a relatively balancing between interpretability of predictions, to robustness of prediction, to reducing overfitting considerably better than alternatives outlined.

The SVM had relatively good performance (accuracy 81.98%). The SVM effectively created simple decision boundaries to distinguish between obesity classes. The SVM by applying a RBF kernel where the SVM utilized a non-linear decision boundary to accommodate for feature interactions, was found to produce reasonable performance compared to the competing models. SVM was simply ranked below the performance of gradient boosting and random forest because of the amount of complexity with the dataset and models capabilities limited with a fixed kernel function.

In contrast, KNN and logistic regression were the least performing models with the accuracy results of 77.02% and 76.39% of predictive accuracy, respectively. KNN possibly accomplished with lower accuracy than expected as the model attempts to classify using local patterning of the data. Alternatively, the KNN requires distance metrics which may be insufficient in capturing the appropriate global structure of the data when incorporating higher variability and considering the aforementioned increased dimensionality of the dataset.

Again, Logistic Regression was a simpler linear model and as a linear model, therefore should have been speculated tendency to underperform compared to the complex dataset where it did not have appropriate capabilities to represent non-linear relationships among features. Although, specified as baseline, logistic regression predictors were found to perform poorly comparatively in predictive accuracy, precision, recall, F1-score, etc. along with the other models gradient boosting and random forest provided the capacity to more adequately describe interactions among features in their core structures as predictive models.

From the results, we found that models that used ensemble methods, mainly gradient boosting, were the only suitable approaches available for the multi-class classification problem of predicting obesity based on a number of demographic, physical, and behavioral factors about obesity. However, there were some recognition about limitations of the study. For one, there was no guaranteeing that the dataset was widely representative of populations globally thus engaging in the need for carefully self-reported lifestyle measures could evoke biasing in data responses overall. In future recommended research recommendations we suggest embarking with the models mentioned but larger and more diverse data points beyond 30 participations and also would be useful to explore multi-modal datasets, (possibly increases in genomic and accelerometry to immurement acceptable accuracy measures while still accepting the limitations of relatively larger global population use and addresses) but can also connect as inherently shared nature to collecting multi-modal datatypes along quit an undertaking for recognizing biases and limitations regarding contributing interaction with datasets. (moreover this also deals with potential issues with validated use and appeal of datasets with the more covert moves of activity engagement- in possible real-time predictive system- we imagine to engage datasets with wearables could recommend early prompts to intervene based on the predictive states not addressing classification).

Although the findings indicate that machine learning models can successfully classify obesity, it is valuable to recognize sources of variability and uncertainty. While the diversity of this dataset is certainly a strength, diversity in demographic and lifestyle aspects of the individuals in the dataset may also contribute to variability in inputs and results. Specific models such as KNN, can be sensitive, meaning the local data it receives will affect whether it generalizes correctly. SVM relies heavily on its kernel (this is where it deviates from being a strictly generalized model), and Random Forest relies on randomness so many variable components of how results were achieved may contribute variability in the results. Future studies may address these limitations by utilizing more diverse and larger datasets, as well as utilize more sophisticated techniques such as cross-validation and ensemble to acknowledge uncertainty.

Gradient Boosting was the best performing model of the ones included in this study achieving close to 88.82% classification accuracy. This is a function of its boosting approach that iteratively improves its predictions by correcting for mistakes made on earlier rounds, leading to more precise results. Random Forest classification accuracy was also strong showing close to 84.16% classification accuracy. Random Forest relies on an ensemble learning method that makes use of an aggregation of several decision trees but does not include any prior sequential correction like the boosting method. There is the possibility that a generalized SVM could be a reasonable performer, however, it operates with restrictions associated with fixed kernel function that may limit its ability to adapt to the correct structures within varied data structures. Meanwhile, KNN and Logistic Regression are good for easier classification tasks, but struggled with the complexity of this dataset, as it is reasonable to suggest the relationships are likely complex, and perhaps even non-linear.

5. Conclusions

This research demonstrates how well ensemble approaches, in particular, Random Forest and Gradient Boosting, classify obesity according to physical, behavioral, and demographic characteristics. With the best accuracy of 88.82% and impressive precision, recall, and F1-scores, gradient boosting proved its capacity to identify intricate, non-linear patterns in the data. Random Forest did well too, although its accuracy was somewhat lower since it relied on bagging instead of sequential boosting.

The SVM achieved competitive performance with an accuracy of 81.98% when using an RBF kernel; however, because this kernel does not change in relation to the data, some adaptability to the complex structure of the data was limited. KNN achieved an accuracy of 77.02% and Logistic Regression performed the worst with 76.39%, largely because they are less equipped to model the complex, non-linear relationships in the data, respectively.

Building upon this work for future studies could involve refining feature selection and feature engineering procedures and selecting features that provided the strongest association to obesity.

Additionally, incorporating all relevant multi-modal data such as genetic and lifestyle tracking information could improve accuracy and potential programs for assisted real-time obesity management. Although modest, the present findings provide generalizable implications in practice for clinical practice and public health. In clinical practice, the Gradient Boosting model could easily be adapted into an electronic health record (EHR) system so healthcare providers might be able use the model in real time to evaluate an individual's risk of developing obesity. This type of assessment would enable clinicians to identify at-risk people earlier, and to provide individualized interventions as necessary (e.g., individualized diet plans, individualized exercise plans). In public health contexts, the Gradient Boosting model may inform the design of obesity prevention programs that target populations already at increased risk of obesity. There is also the option to integrate the model into continuously-distributed wearables and monitor a person's lifestyle variables, such as exercise and diet behaviors, to provide more serious real-time preventative healthcare. If used effectively, machine-learning could provide a method for early detection of obesity and relieve the healthcare system from some of the burden of obesity-related diseases and the costs associated. Although less tractable due to multicollinearity, we could harness novel features using multicollinearity-reducing approaches such as recursive feature elimination (RFE) or domain-specific transformations to interpret and offer more effective models that can influence practice. However, data quality imbalance may still plague the models despite the data preparation. Data augmentation options such as SMOTE might help balance our classes and create more robust models, especially for classes capable.

Deep learning models may have more to contribute to obesity classification than we currently know because of their potential capacity to learn complex structures without human intervention, perhaps making prediction accuracy improvements possible through the use of Convolutional Neural Networks (CNNs) or Recurrent Neural Networks (RNNs). In addition, and despite the complexity, we would want to enhance our ability to make interpretable decisions about which features are most important to practice by using available model-interpretation technologies, such as SHAP or LIME, to explore our most influential decision variables and their relationships to our prediction model in health care contexts. In the long term, integrating multi-modal data sources, including genomic, dietary, and activity data, could pave the way for a comprehensive obesity prediction model. Developing real-time obesity prediction systems may facilitate early intervention and preventive healthcare strategies. These advancements represent promising directions for improving obesity classification and contributing to public health initiatives.

Acknowledgments
The authors would like to express their sincere gratitude to the Kirkuk General Education Directorate, Kirkuk, Iraq, for their support and resources that contributed to this research.

Competing Interests
The authors declare no conflict of interests.

References

[1] M. Safaei, E. A. Sundararajan, M. Driss, W. Boulila, and A. Shapi'i, "A systematic literature review on obesity: Understanding the causes & consequences of obesity and reviewing various machine learning approaches used to predict obesity," Computers in Biology and Medicine, vol. 136, p. 104754, 2021.

[2] D. Lakdawalla and T. Philipson, "The growth of obesity and technological change," Economics & Human Biology, vol. 7, pp. 283–293, 2009.

[3] K. W. DeGregory et al., "A review of machine learning in obesity," Obesity Reviews, vol. 19, pp. 668–685, 2018.

[4] C. Suresh et al., "Obesity prediction based on daily lifestyle habits and other factors using different machine learning algorithms," in Proc. 2nd Int. Conf. Adv. Comput. Eng. Commun. Syst. (ICACECS), 2022, pp. 397–407.

[5] O. A. Mahmood, A. S. Yousif, and A. Adam, "Brief review for multi-class brain tumor diseases schemes using machine learning techniques," Al-Kitab J. Pure Sci., vol. 8, pp. 94–108, 2024.

[6] S. Aggarwal and K. Pandey, "Early identification of PCOS with commonly known diseases: Obesity, diabetes, high blood pressure and heart disease using machine learning techniques," Expert Syst. Appl., vol. 217, p. 119532, 2023.

[7] Q. Wang et al., "Predicting risk of overweight or obesity in Chinese preschool-aged children using artificial intelligence techniques," Endocrine, vol. 77, pp. 63–72, 2022.

[8] C. A. C. Montañez et al., "Machine learning approaches for the prediction of obesity using publicly available genetic profiles," in Proc. Int. Joint Conf. Neural Netw. (IJCNN), 2017, pp. 2743–2750.

[9] H. M. Essa and A. M. Murshid, "Optimizing image processing with CNNs through transfer learning: Survey," Al-Kitab J. Pure Sci., vol. 7, pp. 57–68, 2023.

[10] 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.

[11] S. Mohan, C. Thirumalai, and G. Srivastava, "Effective heart disease prediction using hybrid machine learning techniques," IEEE Access, vol. 7, pp. 81542–81554, 2019.

[12] R. C. Cervantes and U. M. Palacio, "Estimation of obesity levels based on computational intelligence," Informatics Med. Unlocked, vol. 21, p. 100472, 2020.

[13] D. N. Koklu, "Obesity Dataset," Kaggle. [Online]. Available:

https://www.kaggle.com/datasets/suleymansulak/obesity-dataset

[14] S. A. Ahmed, H. Desa, and A.-S. T. Hussain, "Performance improvements using deep learning based object-identification," Al-Kitab J. Pure Sci., vol. 6, pp. 1–13, 2022.

[15] A. Prentice, A. Black, W. Coward, and T. Cole, "Energy expenditure in overweight and obese adults in affluent societies: An analysis of 319 doubly-labelled water measurements," Eur. J. Clin. Nutr., vol. 50, pp. 93–97, 1996.

[16] 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, p. 2734, 2020.

[17] X. Cheng et al., "Does physical activity predict obesity—A machine learning and statistical method-based analysis," Int. J. Environ. Res. Public Health, vol. 18, p. 3966, 2021.

[18] E. R. Cheng, R. Steinhardt, and Z. Ben Miled, "Predicting childhood obesity using machine learning: Practical considerations," BioMedInformatics, vol. 2, pp. 184–203, 2022.

[19] N. Koklu and S. A. Sulak, "Using artificial intelligence techniques for the analysis of obesity status according to the individuals' social and physical activities," Sinop Univ. J. Sci., vol. 9, pp. 217–239, 2024.

[20] X. Pang, C. B. Forrest, F. Lê-Scherban, and A. J. Masino, "Prediction of early childhood obesity with machine learning and electronic health record data," Int. J. Med. Inform., vol. 150, p. 104454, 2021.

[21] N. C. Pereira, J. D'souza, P. Rana, and S. Solaskar, "Obesity related disease prediction from healthcare communities using machine learning," in Proc. 10th Int. Conf. Comput., Commun. Netw. Technol. (ICCCNT), 2019, pp. 1–7.

[22] S. S. Mohammed and H. G. Clarke, "Conditional image-to-image translation generative adversarial network (cGAN) for fabric defect data augmentation," Neural Comput. Appl., vol. 36, pp. 20231–20244, 2024.

[23] A. Parmar, R. Katariya, and V. Patel, "A review on random forest: An ensemble classifier," in Int. Conf. Intell. Data Commun. Technol. Internet Things (ICICI), 2019, pp. 758–763.

[24] A. Kurani, P. Doshi, A. Vakharia, and M. Shah, "A comprehensive comparative study of artificial neural network (ANN) and support vector machines (SVM) on stock forecasting," Ann. Data Sci., vol. 10, pp. 183–208, 2023.

[25] R. K. Halder, M. N. Uddin, M. A. Uddin, S. Aryal, and A. Khraisat, "Enhancing K-nearest neighbor algorithm: A comprehensive review and performance analysis of modifications," J. Big Data, vol. 11, p. 113, 2024.

[26] Q. Li, S. Zhao, S. Zhao, and J. Wen, "Logistic Regression Matching Pursuit algorithm for text classification," Knowl.-Based Syst., vol. 277, p. 110761, 2023.

[27] M. Esmaeili-Falak and R. S. Benemaran, "Ensemble extreme gradient boosting based models to predict the bearing capacity of micropile group," Appl. Ocean Res., vol. 151, p. 104149, 2024.