



Predicting risk of obesity and meal planning to reduce the obese in adulthood using artificial intelligence

Rajdeep Kaur¹ · Rakesh Kumar¹ · Meenu Gupta¹ ¹

Received: 1 September 2022 / Accepted: 29 September 2022 / Published online: 12 October 2022
© The Author(s), under exclusive licence to Springer Science+Business Media, LLC, part of Springer Nature 2022

Abstract

Background An unhealthy diet or excessive amount of food intake creates obesity issues in human beings that further may cause several diseases such as Polycystic Ovary Syndrome (PCOS), Cardiovascular disease, Diabetes, Cancers, etc. Obesity is a major risk factor for PCOS, which is a common disease in women and is significantly correlated with weight gain.

Methods This study is providing a one-step solution for predicting the risk of obesity using different Machine Learning (ML) algorithms such as Gradient Boosting (GB), Bagging meta-estimator (BME), XG Boost (XGB), Random Forest (RF), Support Vector Machine (SVM), and K Nearest Neighbour (KNN). A dataset is collected from the UCI ML repository having features of physical description and eating habits of individuals to train the proposed model.

Results The model has been experimented with different training and testing data ratios such as (90:10, 80:20, 70:30,60:40). At a data ratio of 90:10, the GB classifier achieved the highest accuracy i.e., 98.11%. Further, at the 80:20 ratio, the GB and XGB provide the same result i.e., 97.87%. For the 70:30 data ratio, XGB achieves the highest accuracy i.e., 97.79%. Further, the Nearest Neighbour (NN) learning method is applied to meal planning to overcome obesity.

Conclusion This method predicts the meal which includes breakfast, morning snacks, lunch, evening snacks, and dinner for the individual as per caloric and macronutrient requirements. The proposed research work can be used by practitioners to check obesity levels and to suggest meals to reduce the obese in adulthood.

Keywords Obesity · Meal planning · ML · Ensemble learning · Bagging · Boosting

Introduction

Health issues are the major challenge faced by human beings around the world. A high number of people are suffering from non-communicable diseases such as diabetes, heart attack, kidney diseases, cancers, etc. According to the World Health Organization [1], an epidemic of overweight and obesity is sweeping the globe, and linked to several non-communicable diseases. Obesity has also been associated with the development of insulin resistance, which is linked to a weakened immune system. An outbreak of COVID-19 was encountered at the end of the year 2019, when the whole world suffered from this disease and the weak immune system was the major cause behind it [2].

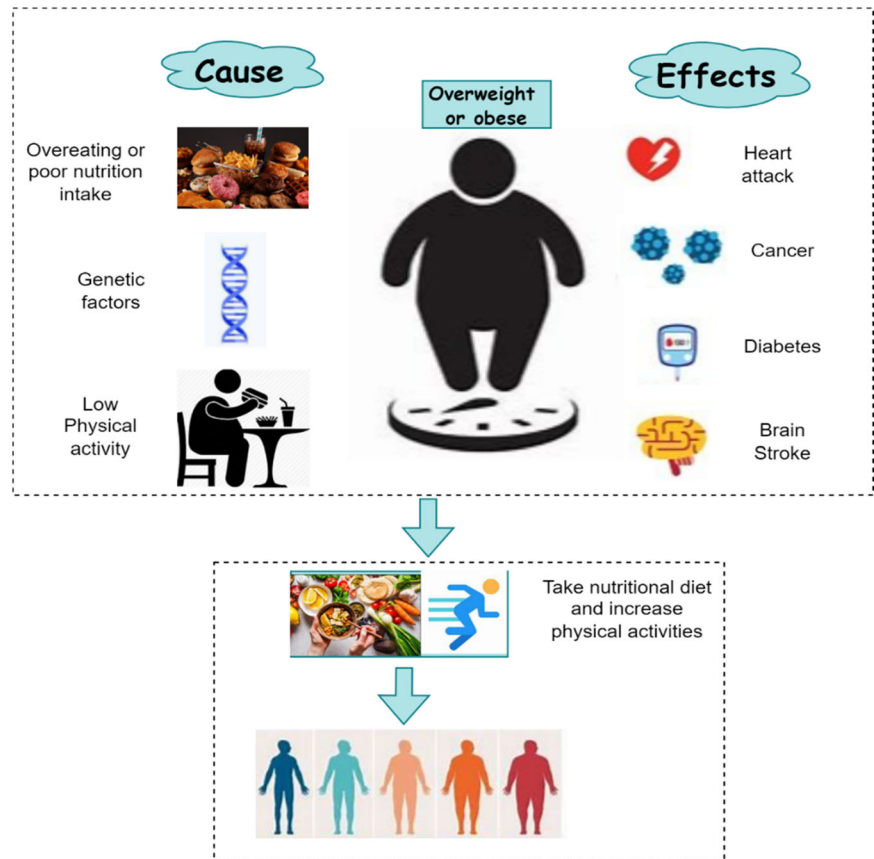
Obesity is also linked with infertility and PCOS in women [3]. Furthermore, the majority of PCOS-afflicted women are obese as they are more prone to weight gain. In obesity, a large amount of fat is accumulated in an individual's different parts of the body and increases the risk of various diseases which varies as per age and gender of humans [4]. Obesity has become a major concern across the world in recent times because it is strongly linked with several other chronic disorders [5] as shown in Fig. 1.

Obesity prevention is difficult because it requires changes in physical activity and dietary habits. To anticipate obesity at an early stage, academicians and healthcare practitioners are using a large quantity of datasets related to obesity from various sources such as electronic medical health records, insurance data, smartphone apps, etc. This data can be examined to generate deeper insights to prevent and cure obesity in the early phase [6]. The most successful weight-loss therapies are dietary, physical activity, and behavior modifications [7]. As the prevalence of overweight and obesity increases, the demand for computational tools also increases for the prediction of obesity to assist obese

✉ Meenu Gupta
gupta.meenu5@gmail.com

¹ Department of Computer Science & Engineering, Chandigarh University, Chandigarh, Punjab, India

Fig. 1 Obesity cause, effects, and prevention



people in taking care of daily meal planning. Furthermore, a healthy diet with proper nutrient intake is an important factor to maintain weight as per body mass index (BMI).

In the computational era, a practitioner has a lot of different ways to get early predictions related to health issues. Data science and machine learning (ML) approaches play an important role in developing models for the early prediction of several diseases. In this research, bagging (gradient boosting (GB) and Bagging meta-estimator (BME)), boosting (XGBoost (XGB) and Random Forest (RF)), and other ML techniques (support vector machine (SVM) and K nearest neighbor (KNN)) are used to predict obesity level in humans. Dataset is collected from the UCI ML repository that consists of attributes such as physical description and eating habits [8]. Furthermore, a meal planning dataset is collected from open-source websites to predict the meal according to individual nutritional requirements [9].

This study is further divided into sections as follows: the section “Related work” concludes the review on obesity prediction using ML techniques, weight management to overcome obesity and related disorders, and nutritional analysis and meal prediction systems. The section “Material and methods” presents the proposed methodology for obesity prediction and meal planning in which different ML classifiers are used for prediction. The

results and explanation of the proposed methodology are presented in the section “Results analysis”. Finally, in the section “Conclusion and future scope”, the research work is concluded.

Related work

This section summarizes the studies on obesity prediction using ML techniques and the role of meal planning to overcome obesity. It is divided into three different phases: Obesity prediction using ML techniques, Weight management to overcome obesity and related disorders, and Nutritional analysis and meal prediction models. The section “Obesity prediction using ML techniques” concludes the research papers that predict obesity using ML techniques.

Obesity prediction using ML techniques

Pang et al. [10] discussed an ML approach to predict obesity levels in childhood at an early age. This research may help clinicians automatically identify children with a higher risk of obesity and suggest early intervention. The model is designed for children with ages greater than 2 years and less

than 7 years. XGBoost outperformed as compared to other models with 0.81 (0.001) AUC. Ferdowsy et al. [11] presented a study on ML models for obesity prediction. Data were collected from different people from different age groups including both types of people obese and non-obese. Furthermore, the authors used nine different ML algorithms to generate obesity predictions and compared them using performance metrics (precision, recall, etc.). In comparison to the other classifiers, the logistic regression technique had the best accuracy of 97.09. Singh and Tawfik [12] discussed an ML model to predict the likelihood of young people becoming overweight (or obese). Childhood BMI levels from the ages of 3, 5, 7, and 11 were used to identify whether a teenager would become overweight or obese at the age of 14. The prediction accuracy issue due to the imbalance of data was improved using Synthetic Minority Oversampling Technique. Different ML classifiers tested on balanced and imbalanced data for obesity prediction. Rahman et al. [13] designed a sensor-based wearable device to track physical activity to manage obesity using ML boosting algorithms such as GB, XGB, cat boosting, light GB classifiers, etc. Furthermore, the collected dataset of physical activity using 30 individuals was used to analyze the burnt calories. ML boosting algorithms classify physical activities and achieve 90% accuracy. Palechor and de la Hoz Manotas [8] presented a dataset to predict obesity levels that contains 17 different attributes. The data are labeled with different obesity classes such as underweight, overweight, and obese (Type-I, Type-II, and Type-III). The authors mentioned in the paper that this dataset may be used to create intelligent ML models for assessing an individual's obesity level. Dugan et al. [14] presented an ML model to predict obesity in childhood. The authors presented the six ML models Naïve Bayes, ID3, J48, RF, Random Tree, and Bayes to predict obesity levels in early childhood age using the dataset CHICA. The accuracy and sensitivity achieved with ID3 model 85 and 89%, respectively.

The above-mentioned analysis summarizes supervised and unsupervised algorithms used to identify whether the individual is obese or not. The next section contains a comprehensive review of several researchers' opinions on weight control in order to overcome obesity and obesity-related diseases.

Weight management to overcome obesity and related disorders

Ma et al. [15] presented that obesity is linked to the development of PCOS and has a significant impact on metabolic abnormalities in PCOS patients. The study was used to describe and evaluate medical methods for diagnosing, assessing, and treating obesity in PCOS patients. Barber et al. [16] presented a detailed review of

the major features that contribute to the obesity-PCOS relationship. The authors also discussed the significant relationship between weight gain and obesity as risk factors for PCOS. A healthy lifestyle and weight loss is the most promising management method for obese and PCOS women. Smethers and Rolls [17] presented a study on obesity disease that can be managed by weight reduction. Weight reduction can be achieved by a reduced energy dietary pattern. Clinical studies have shown that decreasing energy density helps people lose weight. This study presented the outline of dietary objectives and suggestions for weight loss that may be utilized to assist patients in developing sustainable and satisfying low-energy-dense eating habits.

The above study concludes weight management which can be managed by taking proper nutrients in a meal and may help to overcome obesity or obesity-related disorders. The next section summarizes the research papers on nutrient intake and diet management systems.

Nutritional analysis and meal prediction systems

Mustafa et al. [18] proposed a meal recommender system named iDietScore™ that recommends a meal to athletes and individuals as per their profile, age, training cycle, food preference, and sports category. In the early stage of system design, interviews were conducted with sports nutrition or dieticians to understand the macronutrient requirement of athletes. A set of rules were designed which was then integrated with the meal plan database. This rule-based expert system was designed to match a user's meal requirement with a meal plan in a database and finally recommend a meal to the user. Salloum and Tekli [19] designed a novel framework for nutritional meal recommendations. Three modules were developed such as “weight assessment and suggestion”, “a caloric intake and exercise recommendation”, and “a progress evaluation” modules respectively that allows you to make further caloric intake and activity modifications. It includes fuzzy logic in the above-mentioned modules. The performance of the framework was evaluated using 50 patient profiles and 11 nutrition expert reviewers. Sefa-Yeboah et al. [20] designed a genetic algorithm-based AI application for detecting a user's energy balance and estimating the calorie intake required to fulfill daily calorie demands for obesity management. The algorithm recommends appropriate meals based on the users' input information on favored foods, cholesterol levels, diabetic status, and amount of physical activity. The model was able to estimate both glycaemic and non-glycaemic meals depending on the user's health as well as the macro-and micronutrient needs. Shen et al. [21] proposed an ML model for food image recognition to estimate the nutrients in food to maintain a healthy diet

which helps to control obesity. An experiment was evaluated on the basis of various food classes to achieve higher classification accuracy. CNN model was designed to fine-tune the Inception V3 and V4 model using the food image dataset FOOD-101. The results were improved using augmentation and multi-crop techniques. Kumar et al. [22] proposed a multilayer perceptron model for food type classification and to estimate the calorie values in food. The model recommended a list of food to manage the obesity problem, and the disorders related to the obesity problem respectively. The multilayer perceptron model predicts the calorie values that were close to the actual values and outperformed when compared with SVM.

As per the above analysis, there were several techniques available to identify whether the person is obese or not at an early stage. Obesity leads to several diseases and disorders. It is difficult to manage obesity but weight management can treat obesity by improving lifestyle or habits (eating and physical activities). The objective of this study is to employ a wide range of ML techniques to determine whether a person is obese. It also helps them to follow a healthy meal that includes the proper nutrients as per the individual requirement.

Material and methods

In this section, various supervised algorithms such as GB, XGB, RF, BME, SVM, and KNN are applied for the prediction of obesity. Furthermore, the unsupervised nearest neighbor learning method is used to predict the nutritional meal as per the nutrient requirements of the individual.

Dataset used

In the proposed work, two different datasets are used for the prediction of obesity and required meals. The first dataset (refer Table 1) is used for the obesity prediction in which physical description and eating habit features are used to make the predictions of obesity. The second dataset is used for the meal recommendation in which a one-day meal is further divided into the sub-meals such as breakfast, morning snack, lunch, evening snacks, and dinner [9].

Dataset for obesity prediction

The dataset used for the evaluation of the different ML models is collected from the UCI ML repository [8]. The collected dataset consists of 16 attributes with one output variable and contains 2111 records. A data frame is created by considering 13 attributes (out of 16 attributes) to train and evaluate ML models. These attributes include the physical description features (age, height, weight, gender, family history of overweight) and eating habit features

Table 1 Obesity dataset with attributes description and values

Attribute (obesity dataset)	Values
Gender	{Male, Female}
Age	Numeric value in years {14, 61}
Height	Numeric value in meters {1.45, 1.98}
Weight	Numeric value in Kg {39, 143}
Family history	{Yes, No}
Intake of high-caloric food	{Yes, No}
Intake of vegetables	Count in numeric values
Count of main meals	Count in numeric values
Intake of food between meals	{No, Sometimes, Always, Frequently}
Daily water intake	Numeric value
Intake of alcohol	{No, Sometimes, Always, Frequently}
Obesity level	Normal weight, Overweight-1, Overweight-II, Obesity-Type-I, Obesity-Type-I, Obesity-Type-I, Underweight

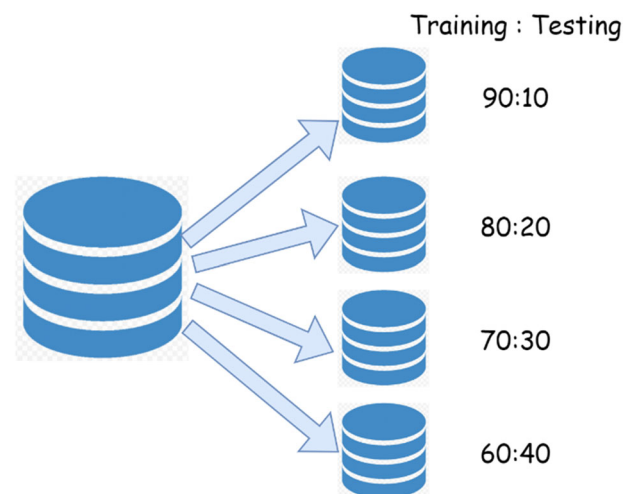


Fig. 2 Division of data for training and testing (different train test split ratios to evaluate the model)

(frequency of main meals, frequency of consumption of high-calorie meals, frequency of consumption of vegetables, smoke, consumption of alcohol, etc.). All the data are labeled with a target feature (obesity level) that is of categorical type as described in Table 1.

The ML models are evaluated on four different data ratios of training and testing (90:10, 80:20, 70:30, and 60:40) as shown in Fig. 2.

Meal planning dataset

The meal planning dataset is collected from an open-source website [9] that consists of 93 meal plans. Each meal is

Table 2 Sample meal plan for 1151 calories

1151 calories that include 152 g protein, 16 g fat, 86 g carbohydrates	
Breakfast (calories: 296)	Vegan pancakes (2 serving), blackberries (1/4 cup), 1 peanut butter (1 tbsp)
Morning snack (calories: 150)	Edamame pods (3/4 cup)
Lunch (calories: 245)	White bean and avocado toast (1 serving), sliced cucumber (1 cup)
Evening snacks (calories: 30)	Small plum (1 serving)
Dinner (calories: 499)	Falafel salad (1 serving)

further categorized into the five sub-meals such as breakfast, morning snacks, lunch, evening snacks, and dinner of a day as shown in Table 2. Each meal contains total amount of calories and macronutrients (protein, fat, carbohydrate). The meal planning dataset includes a wide range of nutritional foods that are beneficial to human health.

The proposed work focuses to predict the obesity levels in adulthood by using ML algorithms. The features related to the physical description and eating habits are considered to train the models. The obesity dataset is pre-processed, and features related to the physical description and eating habits are considered to train the models. Furthermore, the calorie requirements of the individual are calculated from the basic information such as age, height, and weight using BMR, and as per the calorie requirement, a meal with macronutrients detail is suggested to the user.

Proposed methodology

Few steps are used to pre-process the collected dataset such as removing unusual data, imbalanced classes, assessing the correlation level between attributes, etc. In this dataset, many of the feature values are categorical and converted into continuous variables as ML models allow the use of numerical data only (using Eq. (1)). After this, data transformation is performed using StandardScaler to scale the data in the range of 0–1.

Algorithm

Step 1: Clean and pre-process (unusual data) the obesity dataset D .

Step 2: Convert the categorical variables ('Gender', 'family_history_with_overweight', 'FAVC', 'CAEC', 'SMOKE', 'CALC', 'NOBeyesdad') into continuous values.

Step 3: Use StandardScaler to transform the variables into a standardized format.

$$Z = \frac{x - u}{s} \quad (1)$$

Where x represents the training sample, u represents the mean of the training samples, and s represents the standard deviation.

Step 4: Repeat Steps 1–3 for each step 5, 6, and 7, respectively.

Step 5: Bagging approach (Bagging meta-estimator, Random Forest algorithms used in proposed work)

$\{d_1, d_2, d_3 \dots d_n\}$ = Data samples generated from dataset D using replacement method

$\{c_1, c_2, c_3 \dots c_n\}$ = initialize weak learners for $i = 1$ to n

where n is the number of weak learners

$$p_i = c_i(d_i) \quad (2)$$

Fit a weak learner model c_i on data sample d_i to generate the prediction p_i

Final_{pred} = aggregation of output of weak learners $\{p_1, p_2, p_3 \dots p_n\}$

Step 6: Boosting (GB, XGB algorithms used in proposed work)

for $i = 1$ to n

where n is the number of weak learners

d_i = sub-dataset generated from dataset D

c_i = initialize weak learners

$$p_i = c_i(d_i) \quad (3)$$

Fit a weak learner model c_i on data sample d_i

e_i = evaluate the error of p_i

d_i = adjust the next data sample(D_i, e_i)

Final_{pred} = combined output of classifier predictions $\{p_1, p_2, p_3 \dots p_n\}$

Step 7: Other ML classifier (SVM, KNN algorithms used in proposed work)

D = original dataset

C = initialize classifier

Fig. 3 Bagging approach to generate the final prediction (multiple datasets trained on multiple datasets to generate the final prediction)

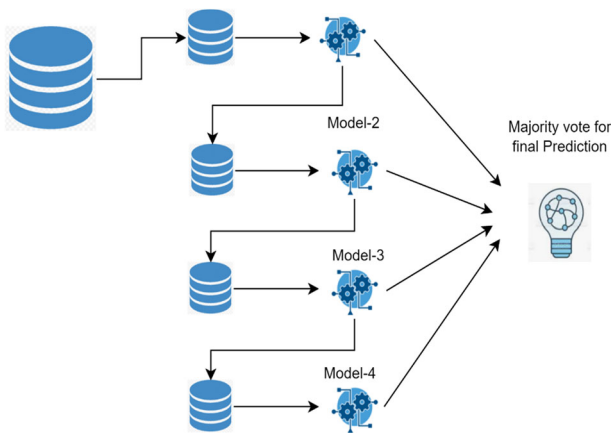
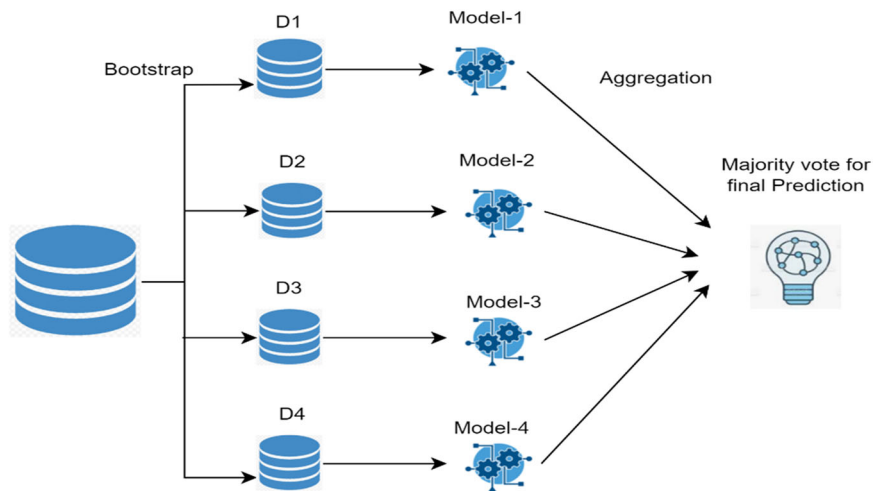


Fig. 4 Boosting algorithms to generate the final prediction

$Final_{Pred} = C(D)$ fit a classifier Con dataset D

Step 8: Evaluate the ML models performance using performance metrics (Precision, Recall, F-1 score, and Accuracy)

Bagging, boosting, and ML models are applied to the clean dataset. Bagging which is also called bootstrap aggregation is a popular ensemble approach that uses the bootstrap method to construct various training sets as shown in Fig. 3. These training sets are created by randomly selected samples from the dataset. Every model with the training dataset generates a prediction or output using Eq. (2) which is then aggregated to make the final prediction. Boosting used multiple weak learners to generate a strong learner for final prediction using Eq. (3). It generates the weak learners sequentially to improve the performance of the model as shown in Fig. 4. Six models (GB, BME, XGB, RF, SVM, and KNN) are applied in the proposed work to predict obesity levels as shown in Fig. 5.

The obesity prediction model’s parameter settings are shown in Fig. 6. GB is a robust ensemble ML approach that integrates many weak base learning models (learners)

into a robust prediction model. The parameter n estimators represent the number of trees or samples in the model, and the final result is derived from the combined predictions of the trees. Furthermore, XGB algorithm is one of the most popular ML techniques for structured or tabular data. It is similar to the GB algorithm and consists of several hyperparameters which can be fine-tuned. In the algorithm, $n_estimators$ parameter is set as 200, and all the other parameters are not changed and assigned a default value. Next, the RF generates decision trees based on the number of dataset subsamples and enhances accuracy by controlling overfitting. The parameter bootstrap is set to “True” to train the classifier on a subsample rather than the complete dataset. A total of 95.7% accuracy of the RF model has been achieved using these parameters such as $n_estimators = 200$, $criterion = 'entropy'$, $random_state = 0$. In addition to this, BC is an ensemble meta-estimator that used a base classifier on the sub-datasets and then aggregates their predictions to generate a final prediction. The default classifier is set as a decision tree which achieves better results as compared to the SVM as a base estimator. Next, SVM is a supervised learning technique that uses a separating hyperplane to make fine classifications. In other words, the method returns an optimal separation hyperplane that classifies new samples using trained labeled data. KNN is also applied to find out KNNs to classify the objects, where the value of k is 3.

Meal prediction as per the calorie requirement of the users

Nutritional management and weight reduction are essential components of treating overweight and obesity in individuals. In the proposed work, a model is designed to assist dieticians or practitioners to predict meal plans for individuals as per their calorie requirements to

Fig. 5 Obesity prediction ML model workflow

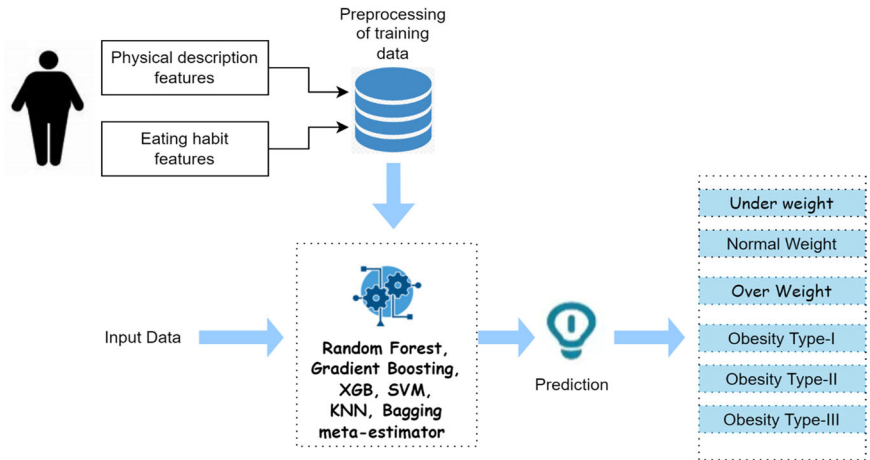


Fig. 6 Models parameter for obesity prediction

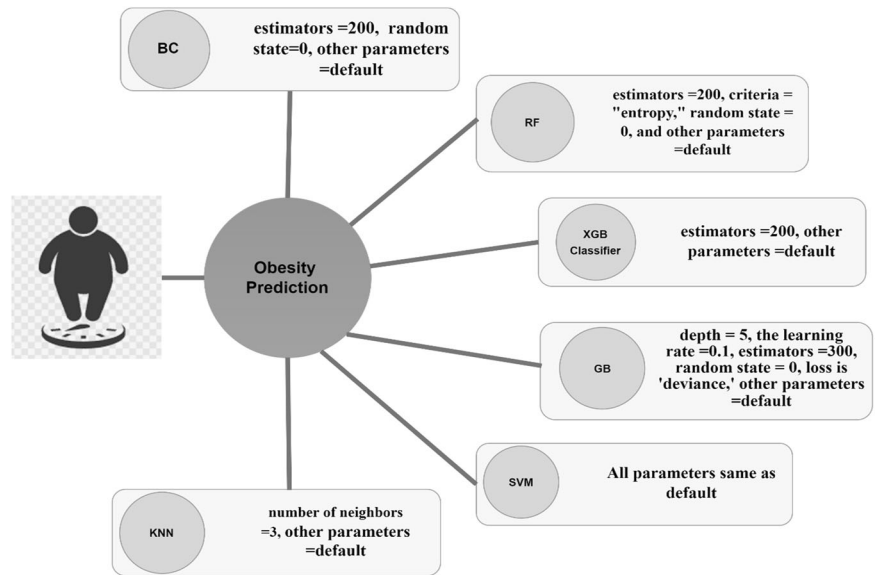
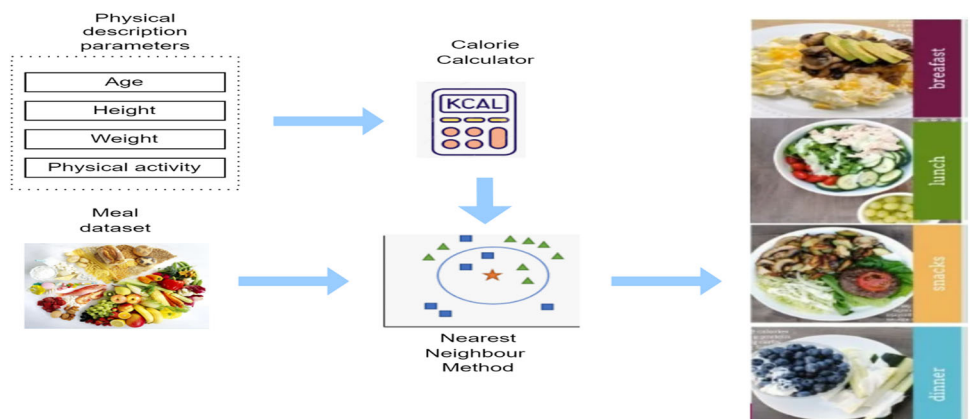


Fig. 7 A machine learning model to predict meals as per the caloric requirements



maintain a healthy weight to overcome obesity. The model discussed in the section “Proposed methodology” is used to predict individual weight status such as underweight, normal weight, overweight or obese. An

initial set of characteristics, such as height, weight, and age, are obtained from the user’s input to compute the user’s daily calorie intake requirement. The calorie requirements for the males and females are different that

Table 3 Performance evaluation metrics and comparison of different classifiers

Algorithm	Accuracy (%)	Weighted avg precision (%)	Weighted avg recall (%)	Weighted avg F-1 score (%)
RF	95.74	96	96	96
Bagging classifier	95.90	96	96	96
XGB classifier	97.79	98	98	98
GB classifier	97.16	97	97	97
SVM	87.7	88	88	88
KNN	82.3	82	83	82

Bold values shows better performance as compared to other models

are calculated using the Harries–Benedict equation as mentioned in Eqs. (4) and (5) [23].

$$\text{Total}_{\text{Cal}}(\text{Women}) = 655 + 9.563 \times \text{wt} + 1.8496 \times \text{ht} - 4.6756 \times \text{age} \tag{4}$$

$$\text{Total}_{\text{Cal}}(\text{Men}) = 66.47 + 13.75 \times \text{wt} + 5.003 \times \text{ht} - 6.7 \times \text{age} \tag{5}$$

As per the calorie requirement, a meal plan with five sub-meals (breakfast, morning snacks, lunch, evening snacks, and dinner) is suggested to the individual using the Near-estNeighbors learning method as shown in Fig. 7. It is an unsupervised method to compute the neighbors, where k is the number of neighbors. If a user required 1200 calories per day then this method will return, k number of meals (approximately 1200 calories). The detail of the parameters set for the experiment is the number of neighbors 3 (n_neighbors = 3), the algorithm to search for the neighbors = ball_tree, parameter metric is set as Euclidean, and all other parameters of the search method are set the same as by default.

In the proposed research work, obesity levels are predicted using six ML algorithms and a meal is suggested to the individual as per their calorie requirements. Results of all the obesity prediction and meal planning models are discussed in the section “Performance analysis metrics”.

Performance analysis metrics

Performance metrics such as precision, recall, F-1 score, and accuracy are derived from the confusion matrix and computed using Eqs. (6)–(8). A confusion matrix is a matrix that allows analyzing the performance of each class by computing the true positive, true negative, false positive, and false negative of the prediction model, and it helps in the evaluation of the quality of the classification model.

$$\text{Accuracy} = \frac{TP + TN}{TP + FN + TN + FP} \tag{6}$$

Table 4 Data split ratio effect on the percentage

ML classification model	90–10%	80–20%	70–30%	60–40%
RF	96.23%	95.74%	95.74%	95.5%
Bagging classifier	95.75%	95.98%	95.9%	95.03%
XGB classifier	97.64%	97.87%	97.79%	96.45%
GB classifier	98.11%	97.87%	97.16%	96.33%
SVM	88.21%	88.18%	87.7%	87.34%
KNN	83.49%	82.03%	82.33%	81.42%

Bold values shows better performance as compared to other models

$$\text{Percision} = \frac{TP}{TP + FP} \tag{7}$$

$$\text{Recall} = \frac{TP}{TP + FN} \tag{8}$$

Results analysis

The comparisons of the ML models’ performance are discussed in this section. The models are evaluated on the obesity dataset with different train test ratios.

Classification models performance analysis

The models used in this proposed work are evaluated on the obesity dataset using the performance metrics (precision, recall, and F-1 score). All models’ accuracy is very close to each other as shown in Table 3, but the GB and XGB classifiers models are outperformed with the difference in the accuracy of 0.16%. SVM and KNN algorithms perform poorly when compared to the other algorithms. The difference between the best models (GB and XGB) and the worst model (KNN) is approximately 15% in terms of accuracy. Performance metrics are shown in Table 4. These are evaluated on training and testing data for 70 and 30%, respectively, and visualization of the model performance is presented in Fig. 8.

Fig. 8 Comparison of different models using performance evaluation metrics (precision, recall, F-1 score)

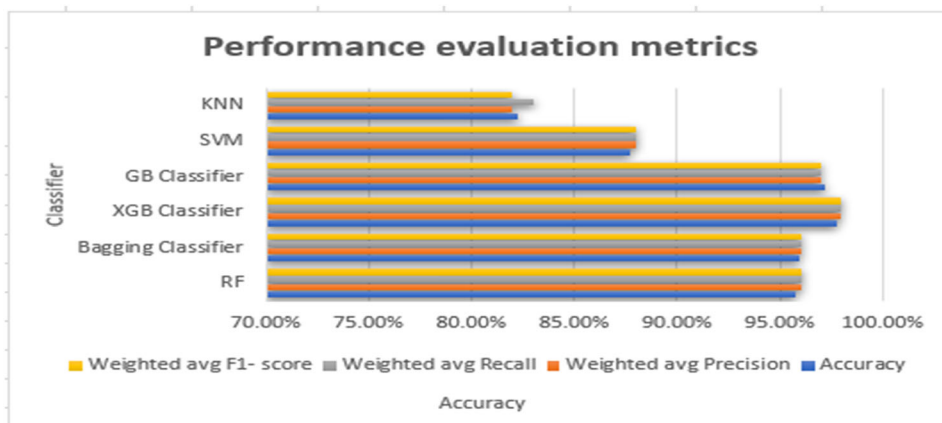


Figure 9 shows the graphical representation of the confusion matrix of six ML models with 70 and 30% training and testing dataset ratio, respectively. The confusion matrix X-axis and Y-axis show the predicted label and observed label, respectively.

In the first experiment, the data ratio for training and testing is 90 and 10%, respectively. The GB classifier achieves the highest accuracy 98.11% (GB ~ XGB > RF ~ BME > SVM > KNN). In the second experiment, the data ratio for training and testing is 80 and 20%, respectively. In this experiment GB and XGB both classifiers achieve the best accuracy GB = XGB > RF ~ BME > SVM > KNN. In the third experiment, the data ratio for training and testing is 70 and 30%, respectively. In this experiment XGB classifiers achieve the best accuracy XGB ~ GB > RF ~ BME > SVM > KNN. In the fourth experiment, the data ratio for training and testing is 60 and 40%, respectively. In this experiment again XGB classifiers achieve the best accuracy XGB ~ GB > RF ~ BME > SVM > KNN. In all the experiments with different data ratios, GB and XGB outperform compared to the other classifiers.

Furthermore, XGB and GB accuracy changes concerning the different train test data splits are shown in Table 4. GB achieved the highest accuracy at a 90:10 data ratio and XGB achieved the highest accuracy at 70:30. XGB and GB classifier detailed classification reports to predict the output variable are shown in Table 5 that include precision, recall, F-1 score, and support performance metrics. Obesity-Type-I to Type-III output class label prediction is highest as compared to the output classes in terms of precision, recall, and F-1 score. Furthermore, F-1 score of GB and XGB algorithm for different output variable classes is shown in Fig. 10.

Discussion and challenges

The bagging, boosting, and other ML techniques are applied for the classification of obesity levels based on features

(physical description attributes and eating habit attributes). These techniques with all the parameter values are discussed in the section “Proposed methodology”. Obese people are more prone to developing diseases and it can be overcome using weight reduction. Weight can be managed by taking healthy meals as per macronutrient requirements that vary from person to person and by doing physical activities to burn extra calories. A nutritional meal plan for the individuals as per their caloric requirements is discussed in the section “Meal prediction as per calorie requirement of the users”. The major challenges to designing a meal planning system are to understand the nutritional requirements and food preferences of individuals as their culture, season, and existing health conditions. Furthermore, there is no open-source meal dataset with a large number of records available to make the model more robust to per best of our knowledge.

Conclusion and future scope

People who are obese have a higher risk of developing some serious diseases diabetes, hypertension, stroke, cancer, etc. It is essential to diagnose and predict such diseases in their initial stages to prevent them from developing in their late stages. ML and ensemble learning approaches are effective in the field of diagnosis prediction. In this study, GB, BME, XGB, RF, SVM, and KNN models are discussed for obesity prediction. All the models are trained on a dataset that contains the physical description attributes (age, height, weight, etc.) and eating habit attributes (number of meals, etc.). The models are trained on different data ratios and evaluated using performance metrics (accuracy, precision, recall, F-1 score). GB classifiers achieve the highest accuracy that is 98.11% when models are trained on data ratio (90:10). GB and XGB models performed equally with an accuracy of 97.87% when evaluated on data ratio (80:20), whereas the XGB model achieve the highest

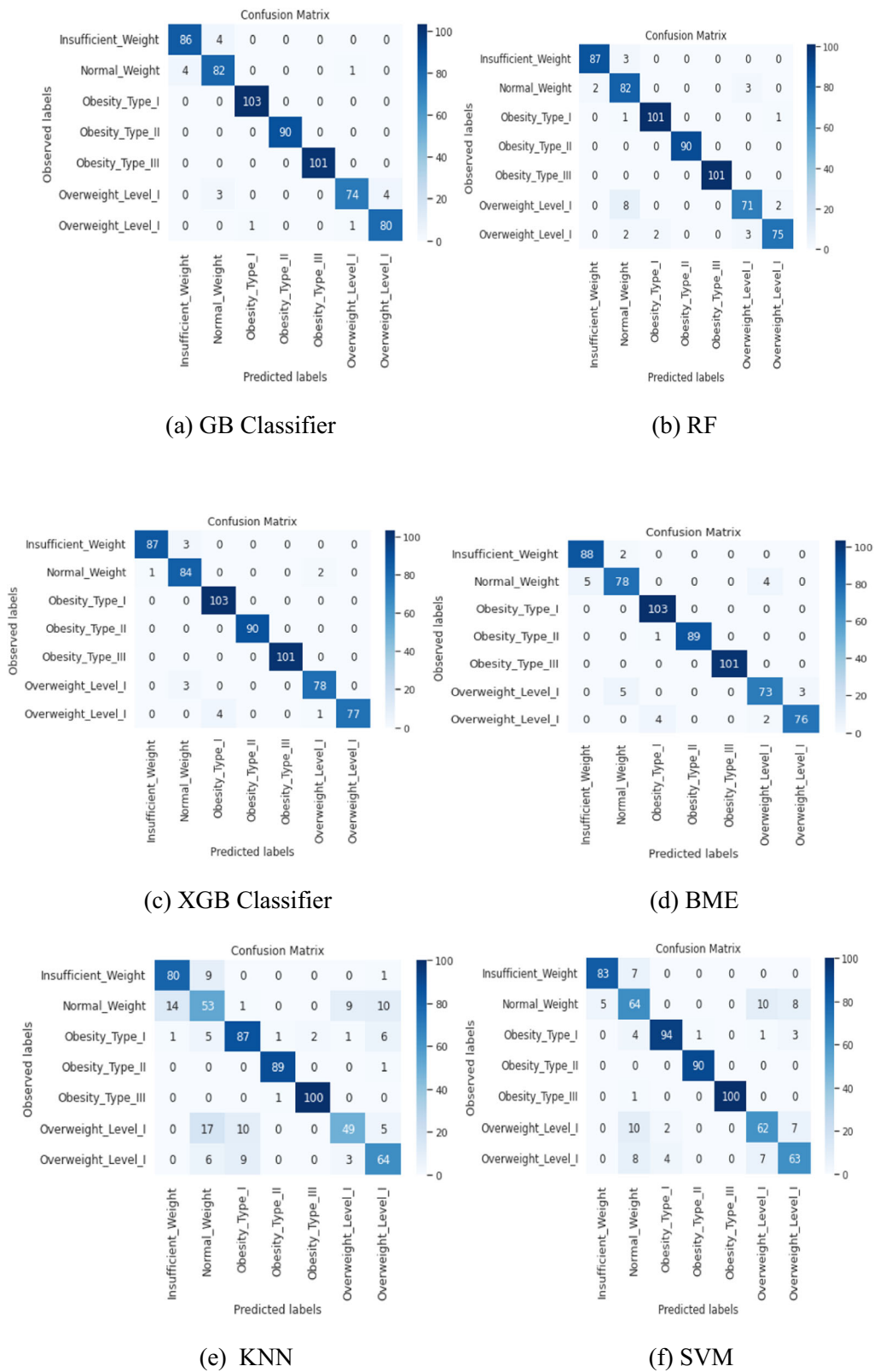
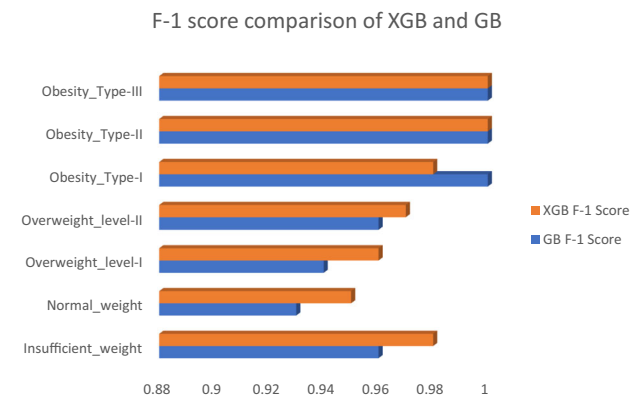


Fig. 9 Confusion matrices of classifiers. **a** GB, **b** RF, **c** XGB, **d** BME, **e** KNN, and **f** SVM

Table 5 Classification report of best-performed algorithms XGB and GB classifier

Output class label	{Precision, Recall, F-1 score}
Insufficient_weight _{XGB}	{0.99, 0.97, 0.98}
Insufficient_weight _{GB}	{0.96, 0.96, 0.96}
Normal_weight _{XGB}	{0.93, 0.97, 0.95}
Normal_weight _{GB}	{0.92, 0.94, 0.93}
Overweight_level-I _{XGB}	{0.96, 0.96, 0.96}
Overweight_level-I _{GB}	{0.97, 0.91, 0.94}
Overweight_level-II _{XGB}	{1.00, 0.94, 0.97}
Overweight_level-II _{GB}	{0.95, 0.98, 0.96}
Obesity_Type-I _{XGB}	{0.96, 1.00, 0.98}
Obesity_Type-I _{GB}	{0.99, 1.00, 1.00}
Obesity_Type-II _{XGB}	{1.00, 1.00, 1.00}
Obesity_Type-II _{GB}	{1.00, 1.00, 1.00}
Obesity_Type-III _{XGB}	{1.00, 1.00, 1.00}
Obesity_Type-III _{GB}	{1.00, 1.00, 1.00}

**Fig. 10** F-1 score of GB and XGB algorithm for different output variable classes

accuracy 97.79% when trained on data ratio (70:30). Furthermore, weight management through healthy eating and increasing physical activity are two common therapies to overcome obesity or overweight. A meal planning dataset is created to suggest a nutritional meal for obese individuals. One meal is further divided into five sub-meals breakfast, morning snacks, lunch, evening snacks, and dinner. Individual calorie requirements are calculated using the Harries–Benedict equation and a meal is recommended as per the calorie requirements. To identify the meal that is as per the calorie requirements of the individual, the nearest neighbor algorithm is applied with $k = 3$ (3 neighbor meals as per calorie requirement). Practitioners may use this model to identify those who are at risk of obesity and help them to improve their food and lifestyle habits.

For future work, the physical activity attributes can also be considered for obesity prediction. In addition, more ensemble learning algorithms and hybrid models can be developed to

improve the accuracy of disease prediction at an early stage. For the nutritional management of obese patients, a nutritional meal dataset can be improved by including the distribution of macronutrients (protein, fat, carbohydrate).

Data availability

The dataset for this research work is obtained from two different sources, one from the UCI repository and another from the online source. Data are available at: [https://archive.ics.uci.edu/ml/datasets/Estimation+of+obesity+levels+based+on+eating+habits+and+physical+condition+](https://archive.ics.uci.edu/ml/datasets/Estimation+of+obesity+levels+based+on+eating+habits+and+physical+condition) and <https://www.eatingwell.com/>.

Compliance with ethical standards

Conflict of interest The authors declare no competing interests.

References

1. World Health Organization (WHO), Controlling the global obesity epidemic. <https://www.who.int/activities/controlling-the-global-obesity-epidemic> (2003). Accessed June 2022
2. F. Demeulemeester, K. de Punder, M. van Heijningen, F. van Doesburg, Obesity as a risk factor for severe COVID-19 and complications: a review. *Cells* **10**(4), 933 (2021)
3. H. Cena, L. Chiovato, R.E. Nappi, Obesity, polycystic ovary syndrome, and infertility: a new avenue for GLP-1 receptor agonists. *J. Clin. Endocrinol. Metab.* **105**(8), e2695–e2709 (2020)
4. R.C. Cervantes, U.M. Palacio, Estimation of obesity levels based on computational intelligence. *Inform. Med. Unlocked* **21**, 100472 (2020)
5. E. Rodríguez, E. Rodríguez, L. Nascimento, A.F. da Silva, F.A.S. Marins, Machine learning techniques to predict overweight or obesity. In *IDDM* (2021), pp. 190–204
6. K.W. DeGregory, P. Kuiper, T. DeSilvio, J.D. Pleuss, R. Miller, J.W. Roginski, D.M. Thomas, A review of machine learning in obesity. *Obes. Rev.* **19**(5), 668–685 (2018)
7. S. Pfeifflié, F. Pellegrino, M. Kruseman, C. Pijollet, M. Volery, L. Soguel, S. Bucher Della Torre, Current recommendations for nutritional management of overweight and obesity in children and adolescents: a structured framework. *Nutrients* **11**(2), 362 (2019)
8. F.M. Palechor, A. de la Hoz Manotas, Dataset for estimation of obesity levels based on eating habits and physical condition in individuals from Colombia, Peru and Mexico. *Data Br.* **25**, 104344 (2019). <https://doi.org/10.1016/j.dib.2019.104344>
9. Diet Meal Plan to Lose Weight. <https://www.eatingwell.com/>. Accessed July 2022
10. X. Pang, C.B. Forrest, F. Lê-Scherban, A.J. Masino, Prediction of early childhood obesity with machine learning and electronic health record data. *Int. J. Med. Inform.* **150**, 104454 (2021)
11. F. Ferdowsy, K.S.A. Rahi, M.I. Jabiullah, M.T. Habib, A machine learning approach for obesity risk prediction. *Curr. Res. Behav. Sci.* **2**, 100053 (2021)
12. B. Singh, H. Tawfik, Machine learning approach for the early prediction of the risk of overweight and obesity in young people. In *International Conference on Computational Science* (Springer, Cham, 2020), pp. 523–535
13. S. Rahman, M. Irfan, M. Raza, K. Moyeezullah Ghori, S. Yaqoob, M. Awais, Performance analysis of boosting classifiers in

- recognizing activities of daily living. *Int. J. Environ. Res. Public Health* **17**(3), 1082 (2020)
14. T.M. Dugan, S. Mukhopadhyay, A. Carroll, S. Downs, Machine learning techniques for prediction of early childhood obesity. *Appl. Clin. Inform.* **6**(03), 506–520 (2015)
 15. R. Ma, Y. Zou, W. Wang, Q. Zheng, Y. Feng, H. Dong, A. Sun, Obesity management in polycystic ovary syndrome: disparity in knowledge between obstetrician-gynecologists and reproductive endocrinologists in China. *BMC Endocr. Disord.* **21**(1), 1–10 (2021)
 16. T.M. Barber, P. Hanson, M.O. Weickert, S. Franks, Obesity and polycystic ovary syndrome: implications for pathogenesis and novel management strategies. *Clin. Med. Insights Reprod. Health* **13**, 1179558119874042 (2019)
 17. A.D. Smethers, B.J. Rolls, Dietary management of obesity: cornerstones of healthy eating patterns. *Med. Clin. North Am.* **102**(1), 107–124 (2018)
 18. N. Mustaf, A.H. Abd Rahman, N.S. Sani, M.I. Mohamad, A.Z. Zakaria, A. Ahmad, N.H. Yatiman, R.A. Talib, P.B. Koon, N.S. Safii, iDietScoreTM: meal recommender system for athletes and active individuals. *Int. J. Adv. Comput. Sci. Appl.* **11**(12) (2020)
 19. G. Salloum, J. Tekli, Automated and personalized nutrition health assessment, recommendation, and progress evaluation using fuzzy reasoning. *Int. J. Hum. Computer Stud.* **151**, 102610 (2021)
 20. S.M. Sefa-Yeboah, K. Osei Annor, V.J. Koomson, F.K. Saalia, M. Steiner-Asiedu, G.A. Mills, Development of a mobile application platform for self-management of obesity using artificial intelligence techniques. *Int. J. Telemed. Appl.* **2021**, 6624057 (2021)
 21. Z. Shen, A. Shehzad, S. Chen, H. Sun, J. Liu, Machine learning based approach on food recognition and nutrition estimation. *Procedia Computer Sci.* **174**, 448–453 (2020)
 22. R.D. Kumar, E.G. Julie, Y.H. Robinson, S. Vimal, S. Seo, Recognition of food type and calorie estimation using neural network. *J. Supercomputing* **77**(8), 8172–8193 (2021)
 23. D.C. Frankenfield, E.R. Muth, W.A. Rowe, The Harris-Benedict studies of human basal metabolism: history and limitations. *J. Am. Dietetic Assoc.* **98**(4), 439–445 (1998)

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.