Computers in Biology and Medicine 184 (2025) 109389

Contents lists available at ScienceDirect

Computers in Biology and Medicine

journal homepage: www.elsevier.com/locate/compbiomed

A hybrid healthy diet recommender system based on machine learning techniques

Sara Sweidan [a,b,*], S.S. Askar [c], Mohamed Abouhawwash [d,e], Elsayed Badr [f,g]

[a] Department of Artificial Intelligence, Faculty of Computer and Artificial Intelligence, Benha University, Benha, Egypt
[b] Faculty of Computer Science and Engineering, New Mansoura University, New Mansoura, Egypt
[c] Department of Statistics and Operations Research, College of Science, King Saud University, P.O. Box 2455, Riyadh, 11451, Saudi Arabia
[d] Department of Computational Mathematics, Science and Engineering (CMSE), College of Engineering, Michigan State University, East Lansing, MI, 48824, USA
[e] Department of Mathematics, Faculty of Science, Mansoura University, Mansoura, 35516, Egypt
[f] Department of Scientific Computing, Faculty of Computer and Artificial Intelligence, Benha University, Benha, Egypt
[g] Department of Information Systems, College of Information Technology, Misr University for Science & Technology, Giza, Egypt

ARTICLE INFO

Keywords:
Calorie
Machine learning
Regression
Obesity treatment
Expert system

ABSTRACT

Obesity is a chronic disease correlated with numerous risk factors that not only negatively affect all body functions but also increase the chances of developing chronic diseases and the associated morbidity and mortality rates. This study proposes a novel system that bridges the gap between healthcare providers and patients by offering both parties some tools for navigating the intricacies of dietary planning. In this system, machine learning techniques are used to determine the required calories before starting an obesity treatment. A hybrid precision model with minimal parameters is also developed to estimate the appropriate number of calories for losing weight and to formulate a healthy diet plan. A real dataset of 15 anthropometric measurements is analyzed using SVR, LR, and DTR regression models, and all the data are preprocessed before analysis to enhance model performance. Results show that the required calories can be estimated with a high correlation ($R = 0.985$) from independent measurements. The proposed model also calculates the healthy daily percentages of fats, proteins, and carbohydrates based on a knowledge base of medical rules and functions, thus facilitating the sequential treatment of obese patients. In sum, this study applies different models to design a practical, cost-effective approach for accurately determining the required calories and formulating valuable diet plans for obesity treatment and management.

1. Introduction

Obesity is characterized by the abnormal or excessive accumulation of body fat, which can negatively affect health (WHO 2000). The physical accumulation of fat and adipocyte dysfunction can lead to metabolic alterations and increased susceptibility to chronic illnesses followed by cancer [1]. The body mass index (BMI), a measure derived from an individual's weight and height, is frequently employed to diagnose obesity among adults [1]. As efforts to promote healthy weight management, nutritious dietary habits, and adult obesity prevention gain traction, accurate calorie estimation in food items has become increasingly crucial. The primary driver of obesity stems from an imbalance between the caloric intake from one's diet and the energy expended through the body's metabolic processes [2]. Excessive calorie consumption can produce detrimental effects and increase an individual's susceptibility to various diseases, including cancer [1]. Nutritionists advocate for maintaining a balanced caloric intake, which contributes to one's overall well-being regardless of his/her weight status [2]. There has also been a surge in technological advancements in response to these concerns [3].

Obesity is a chronic and multidimensional health concern that necessitates a comprehensive and multidisciplinary approach spanning several domains, such as geriatrics, fitness, sports medicine, and psychiatry. However, the existing definitions of obesity often fail to capture its intricate nature. Empirical evidence suggests that naturally reversing obesity poses significant challenges and is associated with elevated mortality risks [4]. While genetic and gene-regulatory factors play a role, obesity primarily stems from lifestyle choices, such as excessive caloric intake and inadequate physical activity [5], hence underscoring the importance of obesity prevention and intervention across all life

* Corresponding author. Department of Artificial Intelligence, Faculty of Computer and Artificial Intelligence, Benha University, Benha, Egypt.
E-mail address: sweidan_ds@fci.bu.edu.eg (S. Sweidan).

https://doi.org/10.1016/j.compbiomed.2024.109389
Received 31 May 2024; Received in revised form 14 October 2024; Accepted 7 November 2024
Available online 20 November 2024
0010-4825/

stages even though such efforts may be arduous. A cure for the onset of obesity is yet to be defined, and an individual's past weight history does not reliably predict his/her future tendencies for weight gain [4]. The "100 Million Health" survey conducted in Egypt in 2019, which involved 49.7 million adult Egyptians aged 18 years and above, revealed that the prevalence of obesity in the country stood at 39.8 % (BMI $\geq$30 kg/m2). This condition was most prevalent among adult females, with 49.5 % of Egyptian adult women classified as obese compared with 29.5 % of their male counterparts [6].

This study aims to develop different machine learning models that can help improve calorie predictions for obesity treatment. These models can identify patterns and relationships that may be overlooked by human experts, thus reducing bias in calorie estimations and leading to qualified nutrition. Maintaining a nutritious diet routine plays a key part in mitigating the risk of developing significant health issues, such as cardiovascular disorders, strokes, cancers, and diabetes. BMI offers a straightforward and accessible metric for estimating body fat levels. However, BMI calculations only account for height and weight data and disregard other vital aspects, such as body composition, muscle mass distribution, and fat storage patterns [4]. Due to this limitation, people with substantial muscle bulk or high bone density may exhibit an elevated BMI despite having an acceptable adipose tissue proportion.

The proposed system motivates patients to address their obesity by promoting their interest in nutritious daily practices in a new fashion. Unlike existing methods, this system gathers and analyzes data that capture complex relationships between various factors and required calories while considering multiple obesity-related factors by using non-linear models, such as support vector regression (SVR) and decision tree regression (DTR). In other words, this system offers an improved strategy for solving obesity without compromising one's health.

In addition to standard measures, such as weight, height, and BMI, the proposed system incorporates nuanced measurements, such as muscle mass, body fat percentage, visceral fat, basal metabolic rate (BMR), and total body water (TBW), to provide a comprehensive view of an individual's metabolic profile and to offer personalized health recommendations. Unlike conventional obesity management methods, the proposed system adopts an inclusive methodology.

The key contributions of this research are summarized as follows:

- Developed a hybrid approach that combines data-driven methods with expert knowledge-based systems.
- Trained a precise calorie calculation model on a real dataset to estimate the required calories for weight loss.
- Applied data visualization and preprocessing techniques improve the accuracy and efficiency of the hybrid model.
- Applied different machine-learning techniques to improve the confidence of the regression model.
- Ensured the validity of the rule-based system by medical expert test verifies its results.
- Generated personalized nutritional recommendations based on individual characteristics, such as metabolic differences and body composition, to promote healthy eating habits.

The rest of this paper is organized as follows. Studies related to the proposed model are summarized in Section 2. The methodology of the hybrid model is presented along with a description of the dataset and the model components in Section 3. The visualization process and an evaluation of the proposed model are presented in Section 4. The conclusions and recommendations for future work are presented in Section 5.

2. Literature review

To describe the goal of this study and ensure that readers can derive value from our work, we explore the state-of-the-art related to our research domain and review the existing literature. This section presents a detailed examination of previous studies that have developed effective methodological approaches and strategies for obesity diagnosis.

Lingren et al. [7] developed an accurate phenotypic model for childhood obesity diagnosis. They employed machine learning techniques to obtain unique feature set identifiers from the Unified Medical Language System, ICD-9 codes, and RxNorm codes extracted from electronic health records (EHRs). Rios-Julian et al. [8] identified overweight and obese children in Guerrero, Mexico and found that a feature set that encompasses various anthropometric variables and skinfold thickness measurements can effectively diagnose obesity. Both of these studies utilized EHR data and anthropometric measures, such as BMI, waist circumference, and skinfold thickness. Fergus et al. [9] classified different physical activities, such as drawing, free play, jogging, and walking, and used artificial neural networks, accelerometer data from ActiGraph devices, and trained observer recordings to classify children's activities accurately. Overall, these studies [7–9] demonstrate the use of diverse data sources and machine learning models to diagnose obesity and classify physical activity among children, which falls outside our target community's scope.

We now discuss the role of artificial intelligence in forecasting obesity among adults. In their systematic review, Simmonds et al. [10] explored the relationship between BMI and the metrics for assessing childhood obesity and their capacity to forecast obesity among adults. Their findings reinforced the notion that obesity during adolescence represents a significant public health issue as it often persists into adulthood. Siqueira et al. [11] recently examined the possible association between obesity and COVID-19. They examined various factors, such as hospitalization rates, diagnosis and recovery outcomes, and mortality rates, and unveiled a substantial link between BMI values surpassing 30 kg/m^2 and adverse COVID-19 outcomes. These findings underscore the importance of recognizing obesity as a risk factor for unfavorable prognoses in individuals afflicted with COVID-19. Singh and Tawfik [12] used data from the Millennium Cohort Study to forecast the BMI of teenagers based on their previous BMI values. Several regression methods, including artificial neural network models, were also assessed, yielding promising results with a prediction accuracy exceeding 90 %. Selya and Anschutz [13] used machine learning approaches, namely, support vector machines (SVM) and neural networks, along with activity variables to predict obesity status and achieved a 61 % accuracy. Uçar et al. [14] determined body fat percentage (BFP) using a hybrid machine learning approach with minimal parameters and a real dataset consisting of 13 anthropometric measurements. They extracted a feature set using feature selection algorithms, such as PCA, and developed hybrid regression models: decision tree (DT), multilayered feed-forward neural network, and SVM. Results show that BFP can be predicted with a correlation value of R = 0.79. Sun et al. [15] applied regression models to model obesity rate among adults at the local scale. Ferenci and Kovács [16] developed regression models using anthropometric measurements and laboratory data to predict BFP and used machine learning techniques to compute the required caloric intake.

Lee and Chun [17] proposed a methodology based on machine learning techniques for investigating those complex factors contributing to overweight and obesity among Korean adolescents. Their model achieved an accuracy of 0.8403 using the XGBoost algorithm. However, they could not determine the relationship between the input features and the outcome, and the limited measurement parameters in their dataset indicate that some potential risk factors may have been missed, thus affecting their obesity prediction outcomes. Calderón-Díaz et al. [18] proposed a methodology for classifying young Chileans with varying weight statuses based on 13 biochemical variables and achieved promising results. However, some limitations or gaps may be present in the correlation between bilirubin and other related conditions. Mahapatra and Singh [19] predicted the causes of obesity in India by using supervised learning approaches and collecting data on various obesity-related parameters from 550 Indians. However, some clinical parameters have a limited capability for predicting obesity. Anisat et al. [20] proposed an obesity prediction methodology based on machine learning techniques. The XGBoost classifier achieved a promising accuracy of 99.05 %, followed by SVM learning. However, the generalizability of their methodology to other healthcare settings warrants further research.

Based on the above review, we have innovated a hybrid model that can help physicians in treating obesity. Traditional methods for predicting burned calories typically involve formulas and general guidelines based on an individual's height, weight, activity level, age, and weight loss goals [21–24]. While these methods can provide a rough estimate, they often lack personalization and may not consider individual variations in metabolism, body composition, and lifestyle. Consequently, their nutrition recommendations may lack precision and fail to contribute to individuals' weight loss progression.

Using regression machine learning models for calorie estimation offers several benefits [16]. For instance, these algorithmic techniques can analyze extensive datasets while accounting for a broad spectrum of factors, hence providing a personalized approach for calorie estimation that is meticulously tailored to individuals' unique characteristics and specific requirements.

Fig. 1 illustrates the research workflow and describes how machine learning algorithms estimate the required calories to lose weight based on the available anthropometric measurements. We used raw datasets, preprocessing techniques, and state-of-the-art supervised machine learning algorithms to develop a system for accurately predicting obesity risk. We preprocessed all our data to enhance the system's precision. Through rigorous analysis and evaluation, we identified the most effective algorithms for calorie estimation based on the measurements applicable to the related domain.

Our proposed system significantly contributes to obesity prediction and calorie estimation and distinguishes itself from prior research in several key aspects. First, we adopt a hybrid artificial-intelligence-based model to accurately estimate the calories needed to achieve a specific weight loss rate based on machine learning techniques.

Second, we propose a rule-based system that calculates the daily nutrition quantities the body needs. This recommendation system aims to minimize the calculations needed to estimate daily nutrition intake. This system can also adapt to data patterns that may not be immediately apparent to reduce complexity and potential errors while enhancing efficiency.

Third, our model uses sophisticated data preprocessing techniques to selectively incorporate only the most pertinent features from the available anthropometric measurements. This strategic approach enhances prediction accuracy and reduces the overall data collection and analysis cost, thereby rendering our model practical and economical.

Lastly, our proposed recommendation model for determining the optimal distribution of nutrition from carbohydrates, proteins, and fats is grounded in rule-based reasoning. This unique methodology generates intuitive and interpretable recommendations that allow people to make informed dietary choices that are aligned with their weight loss objectives. Evaluation results from medical experts reveal that this model can be generalized across different populations. These pioneering elements make our study a significant advancement in obesity prediction and calorie estimation.

3. Proposed methodology

This section illustrates the methodology adopted in our proposed calorie prediction/nutrition recommendation system. This system considers various patient features, including gender, age, weight, height, muscle composition, BFP, BMI, visceral fat, TBW, BMR, fitness score, and level of physical activity, to determine their total daily caloric needs for weight loss. A real dataset of 1000 individual medical records were utilized in the learning model. Several hybrid learning models were also developed, such as linear regression (LR), whose simple structure provides interpretable results for the linear relationship between features and target output [25], SVR, which can handle the linear and nonlinear relationships among the features, output, and outliers [26], and DTR, a single and hybrid model that has innate interpretability [27] and can capture complex decision boundaries and feature interactions. These models were also compared to determine which of them achieves the highest accuracy in predicting the needed nutrients based on the user's measurements. The predicted required calorie regression value was then used to formulate a healthy diet plan through the recommender model, which follows a set of rules to calculate the percentage of fats, proteins, and carbohydrates needed in daily meals. Fig. 1 presents *Dietitian*, the proposed hybrid model for formulating healthy diet plans.

The first model is an obesity prediction module that predicts the needed calories to lose weight, while the second model is a recommender module, an expert system that infers the healthy daily quantity of nutrients, such as fats, proteins, and carbohydrates, based on the regressed estimated calories.

Dietitian aims to help obese patients develop a balanced, healthy diet plan to lose weight and improve their health status. *Dietitian* also helps physicians calculate the required calories for their patients to lose weight based on machine learning models. In sum, this model plans a healthy daily diet based on an expert system that determines the body's needs for nutrition (distributed as proteins, fat, and carbohydrates) to support its internal fat-burning system. However, the expert system does not specifically tailor recommendations for weight loss or obesity control based on the user's preferences.

3.1. Dataset description

Our model employs a real dataset from Mansoura University Hospital

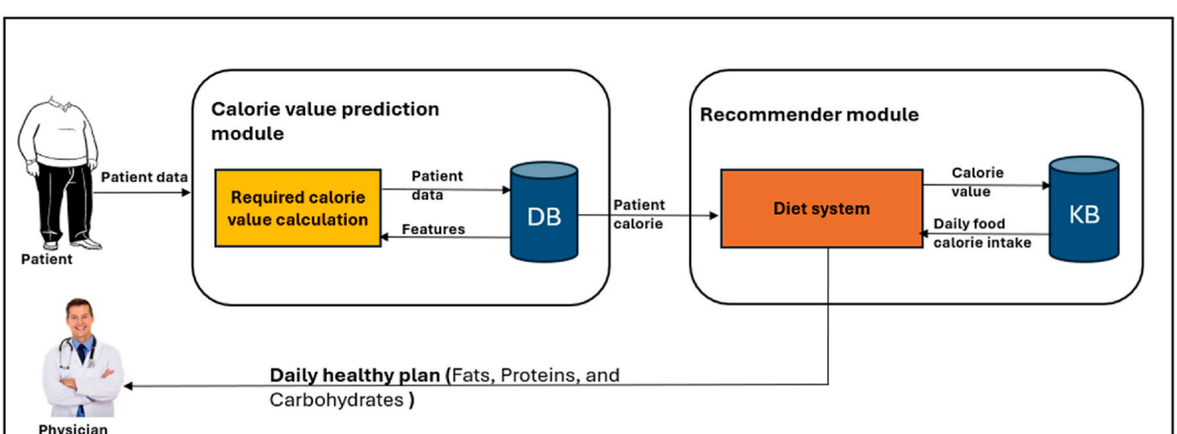

Fig. 1. Proposed framework for the hybrid model (Dietitian).

in Egypt. Data from the EHR of 1000 diagnosed obese patients were collected, with each record comprising 15 features. Our dataset included the demographics and anthropometric measurements of patients aged from 12 to 67 years, with 65 % females and 35 % males. This dataset also contained the most relevant clinical features for the diagnosis. Table 1 provides a comprehensive overview of the dataset and its features. Data collection and representation are vital for performing quantitative risk analysis and formulating effective management plans. We used a set of 12 body measurements (gender, age, weight, height, muscle, fat, BMI, TBW, visceral fat, BMR, fitness score, and physical activity) as input features and the calorie feature is the output. The normal scales of these features are listed in Table 1. We chose these measurements due to their potential effects on calorie prediction. Specifically, gender affects calorie needs due to differences in body composition, with males requiring more calories than females [28]. Younger individuals often require more calories for growth [28]. While weight and height are considered primary factors in determining calorie needs, BMI alone is insufficient due to its inability to distinguish between muscle and fat [4]. Higher muscle mass generally leads to higher calorie needs by increasing the number of calories burned [28], whereas a higher body fat percentage may indicate lower calorie needs. TBW indicates an individual's hydration status, which may influence his/her body processes [28]. Visceral fat refers to fat stored around organs, with high levels affecting calorie needs [29]. BMR denotes the number of calories burned at rest, which is important in estimating daily calorie needs [29]. Fitness scores and physical activity can influence one's calorie needs, especially at higher levels [29]. Machine learning models can identify the complex interactions among these variables that traditional methods cannot observe.

3.2. Data preprocessing

The dataset was structured as a data matrix, with each row representing a patient case and each column representing predictor features. The accuracy of diagnosis outcomes heavily relies on the quality of the data. Therefore, preprocessing the dataset is crucial to address issues commonly found in medical data, such as noise, missing data, and redundancy.

3.2.1. Anonymization, UoM, and normalization

To ensure privacy, we conducted anonymization to eliminate patient-identifying data, such as names, addresses, and ID numbers. In Table 1, we selected a unified unit of measurement (UoM) for each feature and converted all feature values into this standard unit using a standard scaler. We managed the variations in these UoMs by using union units for each. We then integrated our data to minimize redundancy and inconsistency in our final dataset. We adopted a manual approach for the schema integration, where we assigned an attribute ID to each patient.

We also interviewed some nutritionists to determine the system requirements. The expertise of these nutritionists significantly contributed to the data collection and description in all development processes.

3.2.2. Data encoding

Our dataset includes categorical features, such as gender and physical activity, which were converted into numeric values to enable the machine learning models to operate effectively on the data. Before encoding, we transformed the non-numeric data into a numeric format. Table 2 presents the process of discretizing continuous numerical variables into categorical features, which enhances computational efficiency and promotes compatibility across different analytical platforms. We asked a medical domain expert to strategically categorize each anthropometric parameter as vector dimensions to reflect its unique distribution patterns while considering gender-specific and age-related variations in normal ranges and clinical significance. Each feature is categorized by multiple components. Such as, muscle mass has four categories (low, normal, high, very high); fat percentage has four categories (underfat, healthy, overfat, obesity); BMI has five categories (underweight, normal, overweight, obese, Ex obese); TBW has three categories (low, normal, high); visceral fat has categories (normal, high, very high); and BMR is segmented into low, average, high along with gender and age group. This vectorized approach enables precise comparisons and analysis across demographic groups.

We applied data encoding to streamline data retrieval and preprocessing for machine learning regression tasks. This approach also enhances the model's performance and interpretability while facilitating an efficient data integration from diverse sources.

3.2.3. Missing values and feature selection

Missing or unknown values are common in practical medical data. For instance, certain anthropometric measurements may be unavailable due to limitations in healthcare facilities, judgment by physicians, or cost considerations. These missing data should be addressed during the model learning process. In our study, all patient cases had less than 25 % missing values, and any features in the dataset with 25 % or more missing values were excluded from the analysis. Among the 15 features mentioned in Table 1, two features, namely, residence and job, were excluded due to their high amount of missing data. We employed the hot deck imputation method to handle these missing values, where we replaced each missing value with an observed value of a similar unit within the same dataset ("donor"). This approach aids in preserving the integrity and comprehensiveness of the data and in ensuring accurate and dependable analytical outcomes [30]. We used the heterogeneous Euclidean-overlap metric (HEOM) as our similarity criterion [30]. The HEOM distance between two sample cases is computed as follows:

$d(x_i, x_j) = \sqrt{\sum_{z=1}^{n} d_z(x_{iz}, x_{jz})^2}$, where $d(x_i, x_j)$ is the distance be-

Table 1
Obesity related features, (data type: N= Numerical, C=Categorical, O=Ordinal).

Feature type	name	Data type	UoM	Normal range	Min-Mean-Max	F. No.
Demographic Data	Residence	Qualitative/C	–	{rural, urban}	–	1
	Job	Qualitative/C	–	{teacher, engineer}	–	2
	Gender	Qualitative/C	Female/Male	0,1	–	3
	Age	Qualitative/N	–	16–67	16, 35.43, 67	4
Anthropometric Measurements	Weight	Quantitative/N	Kg	46.9–164.9	46.9, 99.13, 164.9	5
	Height	Quantitative/N	Cm	149–185.3	149, 165, 185.3	6
	Muscle	Quantitative/N	Kg	12–87.73	12, 50.14, 87.73	7
	Fat	Quantitative/N	cm^2	0.51–77.17	0.51, 44.78, 77.17	8
	BMI	Quantitative/N	kg/m^2	12–52	12, 36.16, 52	9
	TBW	Quantitative/N	L	12–63.4	12, 39.46, 63.4	10
	Visceral fat	Quantitative/N	cm^2	1–32.9	1, 16.12, 32.9	11
	BMR	Quantitative/N	Cal/day	12–2708	12,1629.7, 2708	12
	Fitness score	Quantitative/N	–	12–65	12, 39.49, 65	13
	Physical activity	Qualitative/O	Active/moderate/slightly/sedentary	1,2,3,4	–	14
Decision	Calories	Quantitative/N	Kcal	1000–1774	1000, 1405, 1774	15

Table 2
Feature distribution sets.

Feature	Distribution sets					
Muscle	Gender	Age	Low	Normal	High	very high
	Female	18–40	<24.4	24.4–30.2	30.3–35.2	≥35.3
		41–60	<24.2	24.2–30.3	30.4–35.3	≥35.4
		61–80	<24.0	24.0–29.8	29.9–34.8	≥34.9
	Male	18–40	<33.4	33.4–39.4	39.5–44.1	≥44.2
		41–60	<33.2	33.2–39.2	39.3–43.9	≥44.0
		61–80	<33.0	33.0–38.7	38.8–43.4	≥43.5
Fat	**Gender**	**age**	**under fat**	**healthy**	**Overfat**	**Obesity**
	Female	18	1–16	17–30	31–35	36–50
		19	1–18	19–31	32–36	37–50
		20–39	1–20	21–32	33–38	39–50
		40–59	1–22	23–33	34–39	40–50
		60	1–23	24–35	36–41	42–50
	Male	18	1–9	10–19	20–23	24–50
		19	1–8	9–19	20–23	24–50
		20–39	1–7	8–19	20–24	25–50
		40–59	1–10	11–21	22–27	28–50
		60	1–12	13–21	22–29	30–50
BMI	**Gender**	**Underweight**	**Normal**	**Overweight**	**Obese**	**Ex obese**
	Female	<18.5	18.5–24.9	25–29.9	30–34.9	35<
	Male	<18.5	18.5–24.9	25–29.9	30–34.9	35<
TBW	**Gender**	**Low**	**Normal**	**High**		
	Female	40 44	45–59	60–70		
	Male	40–49	50–64	65–70		
Visceral fat	**Gender**	**Normal**	**High**	**Very high**		
	Female	1–9	10–14	15–30		
	Male	1–9	10–14	15–30		
BMR	**Gender**	**Age**	**Low**	**Average**	**High**	
	Female	16–19	1200–1500	1501–1800	1801–2100	
		20–29	1300–1600	1601–1900	1901–2200	
		30–39	1200–1500	1501–1800	1801–2100	
		40–49	1100–1400	1401–1700	1701–2000	
		50–59	1000–1300	1301–1600	1601–1900	
		60<	900–1200	1201–1500	1501–1800	
	Male	16–19	1400–1700	1701–2000	2001–2300	
		20–29	1500–1800	1801–2100	2101–2400	
		30–39	1400–1700	1701–2000	2001–2300	
		40–49	1300–1600	1601–1900	1901–2200	
		50–59	1200–1500	1501–1800	1801–2100	
		60<	1100–1400	1401–1700	1701–2000	

tween the x_i and x_j of two sample cases on the n *th* attribute.

Understanding the nature of the association among variables is crucial for generating precise interpretations and predictions across various domains, including statistics, data analysis, and machine learning [31]. We used the Spearman correlation coefficient as our feature selection algorithm to assess the strength and directionality of the relationships among our variables [32]. These coefficients can offer valuable insights into the intensity and directionality of the relationship among the variables present in a dataset. Fig. 2 presents a correlation matrix that offers a compact and visual representation of the dependencies among our variables.

The above matrix facilitates the identification of the highly correlated variables that may be redundant or multicollinear, which can adversely impact the performance of our model. For instance, the strong positive correlations among weight, muscle, fat, BMI, visceral fat, and BMR may suggest that heavier patients typically have higher values for these related variables. This matrix also reveals the nature of the relationship (linear/nonlinear) between the variables and the target, hence informing our selection of appropriate models.

Fig. 2 shows that gender negatively correlates with most of the other variables, thereby suggesting that males tend to have higher values for these variables than females. Meanwhile, age has a weak positive correlation with most variables, indicating a slight tendency for these variables to increase with age. TBW negatively correlates with fat, BMI, and visceral fat, which aligns with the expectation that individuals with higher fat levels tend to have lower TBW. Calorie shows moderate to strong positive correlations with weight, muscle, fat, BMI, visceral fat, and BMR, thereby suggesting that individuals with higher values for these variables tend to require more calories.

Apart from highlighting the importance of each variable, the above correlation matrix also plays an important role in distinguishing linear and nonlinear associations when analyzing data and building predictive models.

The linear association model assumes a straight-line relationship between the predictor and the response variable. By contrast, nonlinear regression models or machine learning algorithms are used when the relationship is nonlinear [25]. Fig. 3 visualizes the linear associations between calorie and various anthropometric and body composition variables. The positive slopes of the regression lines indicate that as the values of the independent variables (muscle, fat, BMI, TBW, and visceral fat) increase, the predicted calorie requirements also increase.

Fig. 3 presents the measured degree of linear association between the measurement features and calories, which is non-zero. We observe some positive linear relationships, such as higher muscle mass being associated with higher calorie requirements. Individuals with higher fat content also tend to require more calories, those with higher BMI values require more calories, and those with higher visceral fat levels have higher calorie needs. We also observe a negative linear relationship between some variables. For instance, those individuals with higher TBW tend to have lower calorie requirements. Some features show a weak or nonlinear association with the target output as reflected in their correlation coefficient (R) being close to zero, and such relationship cannot be accurately represented by a straight line. For example, age, BMR, and calories may exhibit a nonlinear association with one another.

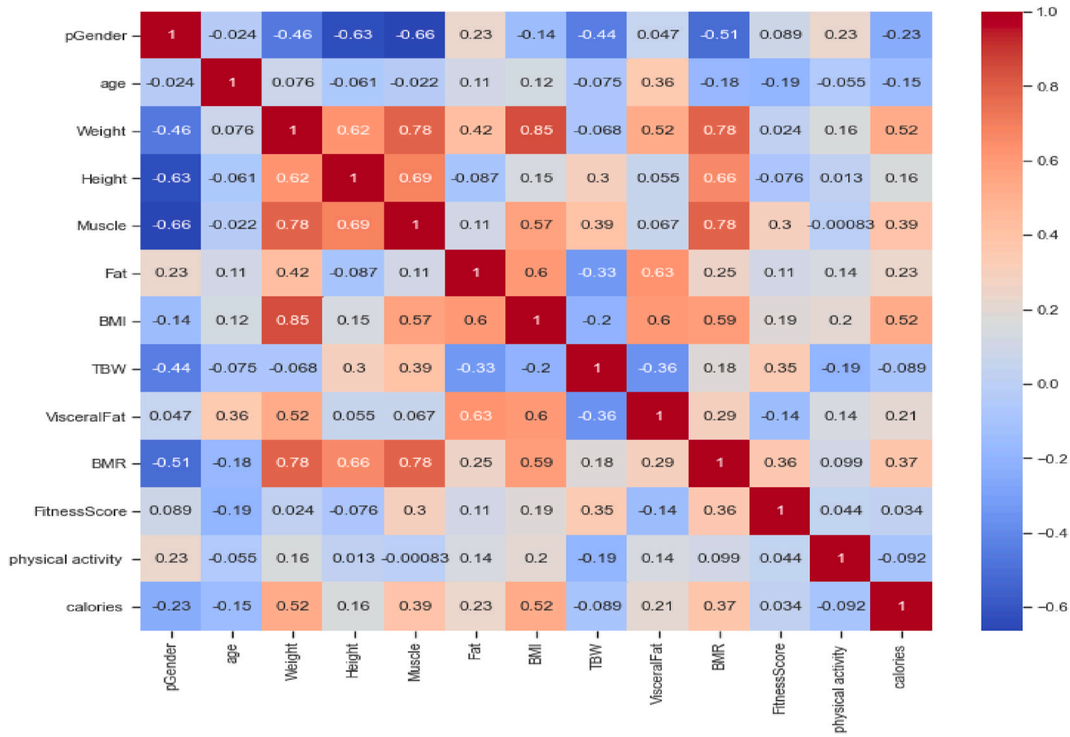

Fig. 2. Measurement features correlation matrix.

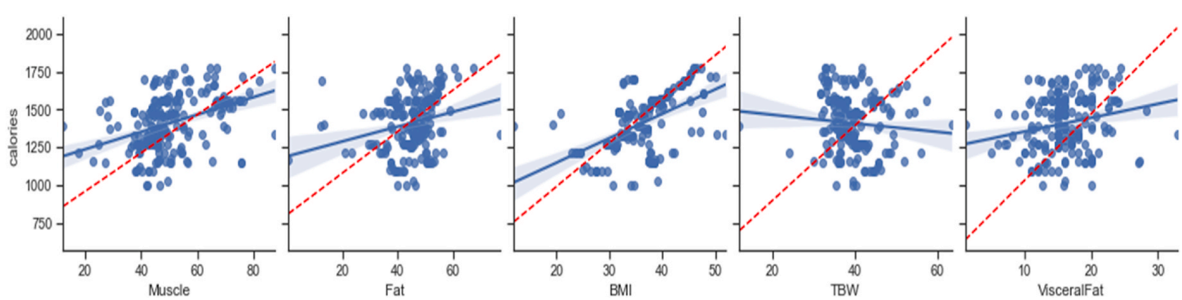

Fig. 3. Linear association between features and target output.

3.3. *Required calorie value prediction module*

We employed SVR [26], DTR [27], and LR [25] as machine learning algorithms. Each of these algorithms was utilized as an independent predictive model and implemented in hybrid structures (Table 3).

We selected these algorithms for several reasons. First, SVR, DTR, and LR are known for their high accuracy rates and short learning time [33,34], hence attracting wide usage in the literature [35,36]. We constructed a hybrid model in the form of ensemble learning, which generates improved predictions by leveraging the strength of multiple models. Due to its simple interpretation and effectiveness, we used the average method to calculate the regression output of this hybrid model. For example, in the case of SVR and DT hybrid models, the decision-making process should consider the consensus of one or both models. If SVR generates an output of 20 while DT produces an output of 10 for a specific data point, then the SVR + DT hybrid model would yield an output of $(20 + 10)/2 = 15$. We calculated and adopted the average output as the output of the new hybrid model. After processing all data points, we recalculated the performance evaluation criteria using the reference values. We divided the dataset comprising 1000 data points for the machine learning module into 70 % for training and 30 % for testing.

Table 3
Individual and hybrid models.

	Individual models		Hybrid models
1	SVR	4	SVR + DTR
2	DTR	5	SVR + LR
3	LR	6	SVR + LR + DTR

3.3.1. *SVR regression*

SVR operates based on the principles of an SVM with some minor distinctions. This algorithm aims to find a curve that best fits the given data points. However, as a regression algorithm, SVR seeks to fit the curve within a hyperplane, allowing it to pass over the dataset. Support vectors are vital in determining the closest match between the data points and the function used to represent them. The choice of a sigmoid Kernel function, its parameters, and any necessary regularization is essential for the SVR algorithm to produce accurate results [26].

3.3.2. *DTR regression*

DTR is a supervised machine learning algorithm for predicting continuous numerical values. This algorithm is based on a hierarchical structure comprising roots, branches, nodes, and leaves. When constructing the tree, each attribute is linked to a node, while the branches allow information to flow throughout the tree. Decisions are made at each node, and the ultimate decision is reached at the final leaf node [27]. DTR has been widely used in the literature due to its simplicity, interpretability, and ability to handle nonlinear relationships in the data.

3.3.3. LR regression

LR is used to predict the value of a dependent variable (calories) based on a given set of independent variables (input features). This algorithm establishes a linear relationship between the input and output using the following function: $Y = \theta_0 + \theta_1 X$. The primary goal of this algorithm is to find the optimal line that best predicts the value of Y for a given value of X. The model achieves the best regression fit line by determining the most suitable values for θ_i [25].

3.4. Recommender module

The recommender module is an expert system that is based on specialized medical and mathematical functions to mimic physicians' decision-making processes. This module uses mathematical functions and algorithms to calculate the recommended healthy fats, proteins, and carbohydrates that the human body needs every day.

We used equations (6)–(8) to calculate protein, carbohydrates, and fat, respectively, when the current weight value is outside the minimum and maximum values obtained using the required calorie value predicted by the machine model and equation (2). We asked a medical expert to document and revise all previous equations, which were used by the proposed system to generate personalized recommendations for calorie intake and for determining the optimal balance of protein, carbohydrates, and fat in the user's diet based on his/her characteristics and the burned calories estimated from the machine learning model.

This recommender module incorporates a medical domain knowledge base that comprises medical rules, facts, and mathematical functions (Table 4), hence offering an expert medical understanding of nutritional needs and health guidelines. This approach allows the system to account for individual variations in metabolism and body composition, adapt its recommendations based on the user's health status (inferred from his/her in-body measurements), provide scientifically grounded advice that incorporates clinical expertise, and offer specific, actionable dietary guidelines in terms of total calories and macronutrient balance.

Table 4 presents a set of mathematical functions employed in the proposed system to calculate various quantities related to body weight, ideal weight ranges, and recommended daily protein, carbohydrates, and fat intake. Equation (1) calculates the ideal body weight (IBW) based on an individual's height. Specifically, 100 is subtracted from the height in centimeters, and then the optimal weight for that specific height is estimated. Equations (2) and (3) determine the adjusted body weight (ABW) for males and females, respectively, while factoring in the individual's current weight and the IBW obtained from equation (1). The ABW for males and females is derived by taking 32 % and 38 % of the difference between their current weight and IBW and adding the difference to their IBW, respectively. Equations (4) and (5) define the lower and upper boundaries of the ideal weight range, respectively. The minimum value is obtained by subtracting 30 % of the IBW from the IBW itself, while the maximum value is calculated by adding 30 % of the IBW to the IBW itself. Equation (6) calculates the recommended daily protein intake when the individual's current weight falls within the ideal weight range defined by equations (4) and (5). This equation multiplies the IBW by 1.5, multiplies the result by 4 (assuming 4 calories per gram of protein), and divides the result by the predicted calorie requirement from the machine learning model. The quotient is then multiplied by 100 to obtain the percentage of protein needed daily. Equation (7) calculates the recommended daily carbohydrate intake based on the predicted calorie requirement from the machine learning model. This equation assumes that 60 % of the total daily calories consumed should come from carbohydrates. Equation (8) calculates the recommended daily fat intake based on the predicted calorie requirement from the machine learning model and the protein percentage calculated in equation (6). This equation initially subtracts the sum of the protein percentage and 60 % (representing the carbohydrate percentage) from 100 % and then multiplies the remaining percentage by the predicted calorie requirement.

Table 4
Used mathematical equations & their function.

Function	Mathematical equation	Eq. no.
calculate IBW (Ideal Body Weight) by using the height entered by the user [4]	IBW= Height-100	(1)
calculate ABW (Adjusted Body Weight) for males by using the Weight entered by user and eq. (1) [37]	ABW = [(Current Weight-IBW)* 0.32]+IBW	(2)
calculate ABW (Adjusted Body Weight) for females by using Weight entered by user and eq. (1) [37]	ABW = [(Current Weight-IBW)*0.38]+IBW	(3)
calculate the minimum value of the ideal zone [37]	Minimum value = IBW-(30/100 * IBW)	(4)
calculate the maximum value of the ideal zone [37]	Maximum value = IBW+(30/100 *IBW)	(5)
calculate protein [37] when the value of current weight is greater than the minimum value and less than the maximum value using calorie predicted by machine model and eq. (1)	Protein = (((IBW*1.5)*4)/calorie)*100	(6)
calculate carbohydrates using calories predicted by machine model [38]	CHO = (60/100)*Calorie	(7)
calculating fat [38] using predicted calorie value by machine learning model and eq. (6)	Fat = ((100 - (Protein +60))/100) * Calorie	(8)

4. Performance evaluation criteria

We implemented our performance criteria as follows. First, we used 70 % of the current dataset to train the machine learning algorithms and the remaining 30 % to test the performance of the model. Second, we calculated the performance criteria by comparing the model's predicted responses (y_i) with the actual responses (t_i). Third, we used six performance evaluation criteria to assess the six regression models. These criteria include root mean square error (RMSE), mean absolute error (MAE), standard error (SH), correlation coefficient (R), explanatory coefficient (R^2), and mean squared error (MSE). Table 5 presents the performance evaluation results.

The RMSE represents the square root of MSE ($RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n} e_i^2}$), MAE measures the variation between the actual and predicted values ($MAE = \frac{1}{n}\sum_{i=1}^{n} |t_i - y_i|$), and SH is a standard deviation whose reduction corresponds to an increase in system reliability ($SH = \sqrt{\frac{\sum_{i=1}^{n} e_i^2}{n-2}}$). The Pearson correlation coefficient (r) is computed as =

Table 5
Machine learning models performance evaluation.

Model	Performance evaluation criteria					
	RMSE	MAE	SH	R	R^2	MSE
SVR	178.44	151.84	0.128	0.2215	0.0491	31842.4
DTR	30.97	3.28	0.022	**0.9856**	0.9714	959.2
LR	140.5	110.13	0.100	0.64	0.41	19740.74
SVR + DTR	91.15	77.56	0.065	0.8671	0.7519	8308.79
SVR + LR	150.05	127.80	0.107	0.5724	0.3277	22514.198
SVR + DTR + LR	100.97	86.297	0.072	0.834	0.6956	10194.55

$\frac{\sum(x_i-\bar{x})(y_i-\bar{y})}{\sqrt{\sum(x_i-\bar{x})^2\sum(y_i-\bar{y})^2}}$, R^2 defines the share of change by the regression model, and MSE is defined as the average of the squares of errors ($MSE = \frac{1}{n}\sum_{i=1}^{n} e_i^2$) [32].

We then calculated the required calories by using a real dataset with a high accuracy rate and six machine learning methods and then determined the performance evaluation criteria for each model. These models may differ in their performance for the same dataset (Table 5), and such difference can positively and negatively describe the performance of single and hybrid models. Fig. 4 summarizes the comparison results, where (a) presents the results for RMSE and MAE, (b) presents the results for MSE, and (c) presents the results for SH, R, and RR. The DTR model performs exceptionally well in predicting the required calories based on the given dataset. Such performance may be ascribed to its ability to capture nonlinear relationships and its robustness to outliers. The hybrid models, particularly the SVR + DTR model, also achieve promising results, thereby suggesting that combining different models can improve prediction accuracy.

Evaluating the recommender system is critical to determining its accuracy, reliability, and overall performance. The evaluation procedure tests the system's ability to accurately diagnose and recommend appropriate intakes of fats, proteins, and carbohydrates based on the estimated required calories and level of physical activity. Given our lack of a benchmark model, we only compare the performance of the proposed recommender system with the decisions made by human experts to ensure consistency and validity. The system's rules must be continuously evaluated and validated to identify and address any shortcomings or inconsistencies and consequently improve its effectiveness. We used Cohen's Kappa metric to evaluate the consistency between the recommendations of the proposed system and those of human experts as follows [32]:

$$k = \frac{P(A) - P(E)}{1 - P(E)} \tag{9}$$

where P(A) is the observed agreement between the two recommendations, and P(E) is the expected agreement achieved by the recommender system. The recommender system obtained an average Kappa value of 0.8, indicating nearly perfect agreement.

We also validated the performance of our system using real-world data to guarantee its adaptability and generalizability across a range of medical scenarios. Medical rules in knowledge expert systems can be improved and adjusted via thorough reviews, making these systems valuable tools for assisting healthcare providers and enhancing patient outcomes.

5. Discussion

We proposed a dynamic approach for predicting calorie needs before obesity treatment based on anthropometric measurements. Unlike traditional static recommendations, our system can adapt to changing body compositions and thus provide evolving recommendations as the health of each patient progresses. We comprehensively evaluated the performance of multiple machine learning models in predicting calorie needs as presented in Table 5. Our proposed system improves the accuracy of its predictions by incorporating a wide range of body composition data as presented in Table 1.

We used SVR, DTR, and LR and assessed the efficacy of these models based on their MSE, R, SH, MAE, and RMSE. We observed that the DTR model outperformed the other models in terms of RMSE, MAE, SH, R, R^2, and MSE, thereby highlighting its impressive predictive capabilities and accuracy. This model also obtained the lowest RMSE (30.97) and MAE (3.28) values, which highlights its better regression accuracy compared with the other models. The high R (0.985) of this model also underscores a strong correlation between its predicted and observed outcomes, while its R^2 (0.971) addresses the large variance in the data. The high MSE value of DTR further supports the excellent performance of this model in explaining a significant percentage of the variance in the actual values.

When comparing the hybrid models, we found that the SVR + DTR hybrid model yielded promising results with relatively low RMSE (91.15), MAE (77.56), and SH (0.065) and a high correlation coefficient (R = 0.8671). To determine the statistical significance of the performance differences among DTR, SVR, and LR, we conducted a Friedman test was conducted and obtained $Q(2) = 5.000$, $p = 0.0150$, which suggests significant differences among these models with a confidence level of 98.50 %. Our post-hoc analysis based on the Wilcoxon signed-rank test revealed significant differences between the performances of DTR and SVR ($p = 0.012$) and of SVR and LR ($p = 0.005$). However, the difference between DTR and LR was insignificant ($p = 0.045$) after applying a Bonferroni correction ($\alpha = 0.0167$). Thus, DTR and SVR as well as SVR and LR demonstrated significant performance differences, while DTR and LR did not.

Our dataset contains noisy or skewed data points. The DTR model is less sensitive to outliers than the other regression models, making it particularly suitable for the nature of our developed medical dataset. Furthermore, Fig. 2 shows a non-linear association between some predictors and the target output, and the DTR model can capture these nonlinear relationships more effectively than the other models.

Anthropometric measurements in many studies [10–15] rely on basic parameters, such as age, BMI, and height, to estimate BFP. By contrast, our proposed system incorporates a wide range of body composition data, including BFP, muscle mass, and visceral fat. The proposed model offers a more nuanced understanding of the needed energy for everyone, as relying only on anthropometric measurements may oversimplify the complex factors that influence metabolic requirements. By using machine learning models, our proposed system can capture complex, nonlinear relationships between various factors and

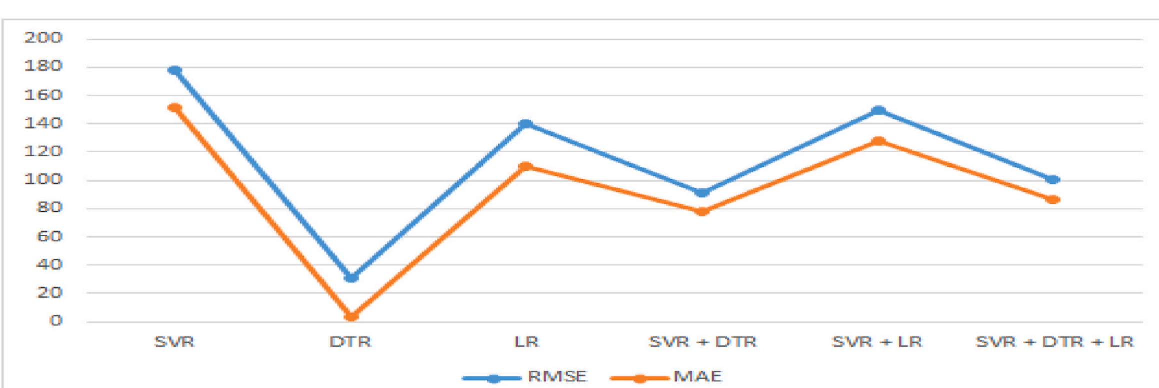

Fig. 4a. comparison between measure terms (RMSE, MAE).

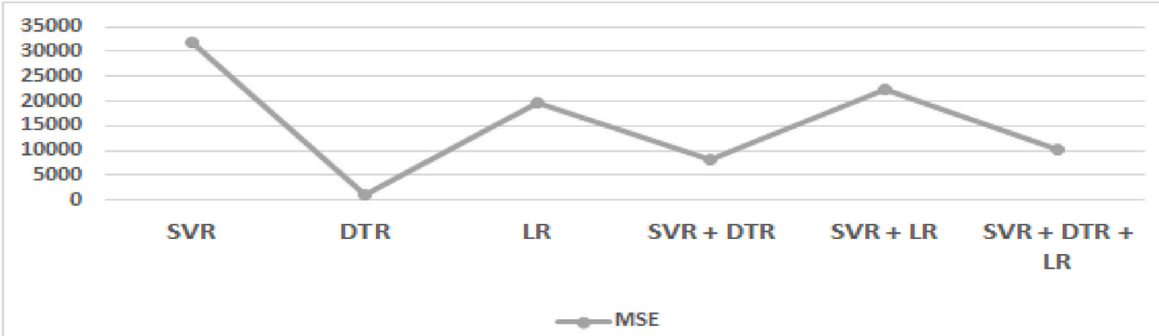

Fig. 4b. MSE results for different models.

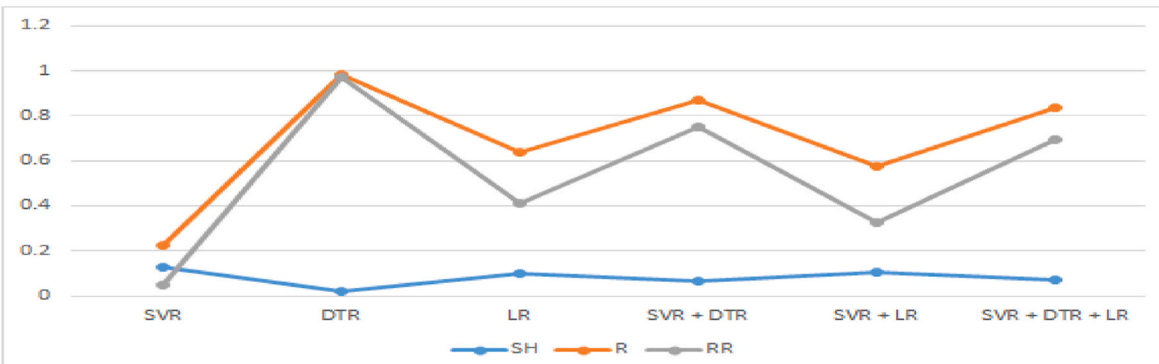

Fig. 4c. A comparison between measure terms (SH, R, RR).

calorie requirements, thereby generating highly nuanced predictions. Many traditional approaches provide static recommendations that do not adapt to an individual's changing conditions. Meanwhile, our proposed system can adjust its calculations based on whether an individual's weight falls within certain ranges, hence offering highly dynamic recommendations. Our system also uses current body composition data and adaptable calculations to offer evolving recommendations as a patient progresses in his/her treatment.

Our main objective was to evaluate the performance of different ML models in predicting the calorie needs of obese patients. We found that ML models obtained excellent regression outputs and generated solid estimations for the proposed recommender system. Table 6 summarizes the literature that applied ML models to predict the required number of calories burned for weight loss.

Current studies [21–24] have several limitations that the proposed model aims to address. Existing solutions focus on narrow prediction (calories or body fat percentage) based on limited dataset and simplified input parameters. They fail to provide comprehensive health assessments and do not offer dietary advice. They do not account for individual differences in health conditions or dietary preferences. The proposed model improves upon using these by using a more diverse dataset of 1000 obese patients aged 16–65 and incorporating 15 anthropometric measurements for a more comprehensive analysis.

We developed our medical decision-making model based on DTR and LR given their better interpretability compared with complex models, such as neural networks. We evaluated our model on a sample of 1000 obese patients. Simpler models may be less prone to overfitting than highly complex models, such as neural networks and RF. Furthermore, DTR can handle numerical and categorical data without extensive processing. SVR and LR can also efficiently handle high-dimensional data, which might be relevant given the 15 anthropometric measurements.

Our proposed model, while innovative, has several important limitations; firstly, the sample size of 1000 patients may lead to overfitting, potentially limiting the model's generalization to the broader population. Authors do not account for crucial factors such as disease history, chronic conditions like diabetes, cardiovascular diseases, or liver disorders, which may impact nutritional needs and metabolism. Additionally, the hybrid model lacks some genetic factors and lifestyle features such as activity, stress, and dietary preferences that may influence dietary requirements. Lastly, our dataset includes 15 anthropometric measurements; we did not capture dynamic changes in these measurements over time that may affect dietary interventions.

In our future research, we should address these limitations to enhance the accuracy of the model and overall value in personalized nutrition planning. We will prioritize empowering users to gain more control over their dietary choices by allowing them to select preferred food items from a comprehensive food database. Our system will develop a personalized diet plan for everyone while considering the calculated percentages of fat, protein, and carbohydrates, the specific nutritional requirements of each patient, the number of essential nutrients in different food items, and the patient's history of chronic diseases, such as diabetes. By enhancing our system's adaptability and customizability, we hope to provide a tailored and practical approach to maintaining people's healthy lifestyles and achieving their dietary goals. In addition, we will investigate integrating AI-driven motivational elements to improve dietary plan adherence. Future efforts will also focus on cross-cultural validation, integration with wearable technology, and addressing ethical considerations in AI-driven healthcare solutions.

6. Conclusion

Our "Healthy Life" hybrid model plays a vital role in accurately calculating calorie intake for weight loss. By utilizing machine learning techniques based on 15 anthropometric measurements, our model formulates personalized diet plans for each user. This model also calculates the required intake of fats, proteins, and carbohydrates based on the

Table 6
Empirical analysis of calorie needs prediction studies.

Reference	Outcome	Dataset	Models	Results
Nipas et al. [21]	Burned calorie prediction	15000 patient Parameter (age, height, weight, heart rate, body temp, duration)	Linear regression, XGBoost regressor, Random Forest regression	Random forest (95.77)
Gaikwad et al. [22]	Calorie prediction	15000 patient Parameter (age, height, weight, heart rate, body temp, duration)	Decision Tree, Simple Linear Regression, and Multiple Linear Regression, Random Forest algorithm, Support Vector Machine (SVM), K-NN, and XGBoost regressor	XGBoost (0.99)
Sheng et al. [23]	Calorie prediction	15000 patient Parameter (age, height, weight, heart rate, body temp, duration)	LightGBM, XGBoost, Random Forest, Ridge, Linear, Lasso, and Logistic	LightGBM (MAE 1.27)
kadam et al. [24],	Calorie prediction	15000 patient Parameter (age, height, weight, heart rate, body temp, duration)	Random forest regressor	RMSE (8.3)
Ferenci et al. [16],	Body fat percentage	862 patients Male, aged >18 39 variables	Neural network, Support vector machine, Linear regression	Support vector machine (0.983)
Uçar et al. [14],	Body fat percentage	252 patients 13 anthropometric measurements	Support vector machine, decision tree regression, artificial neural network	DT + SVM (0.802)
Proposed model	Calorie prediction	1000 obese patients aged 16–65 real dataset of 15 anthropometric measurements	Logistic regression, support vector regression, decision tree regression	Decision tree regression (0.985)

estimated calorie needs of each user. This model demonstrates its effectiveness and practical utility with a correlation value of approximately 0.98 and a variance of 0.971. Future research may expand this work by integrating diverse datasets and including individuals with chronic diseases to further refine and validate our model's predictions. Exploring alternative machine learning algorithms and incorporating more advanced techniques may also enhance the calorie estimation accuracy and personalization of our model.

CRediT authorship contribution statement

Sara Sweidan: Writing – original draft, Visualization, Software, Methodology, Formal analysis, Data curation. **S.S. Askar:** Validation, Supervision, Project administration. **Mohamed Abouhawwash:** Supervision. **Elsayed Badr:** Methodology, Formal analysis.

Funding

This project is funded by King Saud University, Riyadh, Saudi Arabia. Researchers Supporting Project number (RSP2025R167), King Saud University, Riyadh, Saudi Arabia.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgment

The authors gratefully acknowledge **Dr. Marwa El-Hindawy** for her dedication to the success of medical research and her essential role in supporting this research. We appreciate her guidance and profound wisdom, and we feel proud to have such a distinguished expert in the medical domain.

Researchers Supporting Project number (RSP2025R167), King Saud University, Riyadh, Saudi Arabia.

References

[1] M. Díaz, L. Castillo, E. Cuevas, A. Salas, M. Jiménez, Detection of variables for the diagnosis of overweight and obesity in young Chileans using machine learning techniques, Procedia Comput. Sci. 220 (2023) 978–983.
[2] M. Mahapatra, K. Singh, Prediction of causes and effects of obesity in India by supervise learning approaches, Obesity Med. 34 (2022).
[3] X. Pang, C. Forrest, F. Scherban, A. Masino, Prediction of early childhood obesity with machine learning and electronic health record data, Int. J. Med. Inf. 150 (2021).
[4] M. Safaei, E. Sundararajan, M. Driss, W. Boulila, I.W. Shapi, A systematic literature review on obesity: understanding the causes & consequences of obesity and reviewing various machine learning approaches used to predict obesity, Comput. Biol. Med. 136 (2021).
[5] M.A. Gomez-Llorente, R. Romero, N. Chueca, A. Martinez-Cañavate, C. Gomez-Llorente, Obesity and asthma: a missing link, Int. J. Mol. Sci. 18 (7) (2017) 1490.
[6] M. Aboulghate, A. Elaghoury, I. Elebrashy, N. Elkafrawy, G. Elshishiney, The burden of obesity in Egypt, Front. Public Health 9 (2021 Aug 27) 718978, https://doi.org/10.3389/fpubh.2021.718978. PMID: 34513789; PMCID: PMC8429929.
[7] T. Lingren, V. Thaker, C. Brady, B. Namjou, S. Kennebeck, J. Bickel, N. Patibandla, Developing an algorithm to detect early childhood obesity in two tertiary pediatric medical centers, Appl. Clin. Inf. 7 (3) (2016) 693–706.
[8] S.P. Choukem, J.N. Tochie, A.T. Sibetcheu, J.R. Nansseu, J.P. Hamilton-Shield, Overweight/obesity and associated cardiovascular risk factors in sub-Saharan African children and adolescents: a scoping review, Int. J. Pediatr. Endocrinol. (2020 Dec; 2020) 1–3.
[9] P. Fergus, A. Hussain, J. Hearty, L. Boddy, A machine learning approach to measure and monitor physical activity in children to help fight overweight and obesity, in: D.S. Huang, K.H. Jo, A. Hussain (Eds.), Intelligent Computing Theories and Methodologies. ICIC, Lecture Notes in Computer Science, vol. 9226, Springer, Cham, 2015, 2015.
[10] M. Simmonds, A. Llewellyn, C. Owen, N. Woolacott, Predicting adult obesity from childhood obesity: a systematic review and meta-analysis, Obes. Rev. 17 (2) (2016) 95–107.
[11] J.V. de Siqueira, L.G. Almeida, B.O. Zica, I.B. Brum, A. Barceló, A.G. de Siqueira Galil, Impact of obesity on hospitalizations and mortality, due to COVID-19: a systematic review, Obes. Res. Clin. Pract. 14 (5) (2020 Sep 1) 398–403.
[12] B. Singh, H. Tawfik, A machine learning approach for predicting weight gain risks in young adults, in: Conference Proceedings of 2019 10th International Conference on Dependable Systems, Services and Technologies, DESSERT, 2019, pp. 231–234, 2019.
[13] A.S. Selya, D. Anshutz, Machine learning for the classification of obesity from dietary and physical activity patterns, Adv. Data Analyt. Health (2018) 77–97.
[14] M. Uçar, Z. Uçar, F. Köksal, N. Daldal, Estimation of body fat percentage using hybrid machine learning algorithms, Measurement 167 (2021).
[15] Y. Sun, S. Wang, X. Sun, Estimating neighbourhood-level prevalence of adult obesity by socio-economic, behavioural and built environment factors in New York City, Publ. Health 186 (2020) 57–62.
[16] T. Ferenci, L. Kovács, Predicting body fat percentage from anthropometric and laboratory measurements using artificial neural networks, Appl. Soft Comput. 67 (2018).
[17] S. Lee, J. Chun, Identification of important features in overweight and obesity among Korean adolescents using machine learning, Child. Youth Serv. Rev. 161 (2024).
[18] M. Calderón-Díaz, L.J. Serey-Castillo, E.A. Vallejos-Cuevas, A. Espinoza, R. Salas, M.A. Macías-Jiménez, Detection of variables for the diagnosis of overweight and obesity in young Chileans using machine learning techniques, Procedia Comput. Sci. 220 (2023) 978–983.
[19] M. Mahapatra, K.K. Singh, Prediction of causes and effects of obesity in India by supervised learning approaches, Obesity Med. 34 (2022) 100436.
[20] Fati Anisat Musa, Federick Duniya Basaky, E.O. Osaghae, Obesity prediction model using machine learning techniques, J. Appl. Artifi. Intellig. 3 (1) (2022) 24–33.
[21] M. Nipas, A.G. Acoba, J.N. Mindoro, M.A.F. Malbog, J.A.B. Susa, J.S. Gulmatico, Burned calories prediction using supervised machine learning: regression

algorithm, in: 2022 Second International Conference on Power, Control and Computing Technologies (ICPC2T), 2022, pp. 1–4. Raipur, India.
[22] V. Gaikwad, M. Khodke, S. Shahare, P. Terkar, R. Talmale, Implementation of XGBoost regression for calories burnt prediction using R, in: M. Singh, V. Tyagi, P. Gupta, J. Flusser, T. Ören (Eds.), Advances in Computing and Data Sciences. ICACDS 2023, Communications in Computer and Information Science, vol. 1848, Springer, Cham, 2023, https://doi.org/10.1007/978-3-031-37940-6_31.
[23] A.T. Jing Sheng, Z. Che Embi, N. Hashim, Comparison of machine learning methods for calories burn prediction, J. Inform. Web Eng. 3 (1) (2024) 182–191, https://doi.org/10.33093/jiwe.2024.3.1.12.
[24] A. Kadam, A. Shrivastava, S.K. Pawar, V.H. Patil, J. Michaelson, A. Singh, Calories burned prediction using machine learning. 2023 6th International Conference on Contemporary Computing and Informatics (IC3I), Gautam Buddha Nagar, India, 2023, pp. 1712–1717, https://doi.org/10.1109/IC3I59117.2023.10397623.
[25] E. Vittinghoff, C. McCulloch, D. Glidden, S. Shiboski, 5 linear and non-linear regression methods in epidemiology and biostatistics, Handb. Stat. 27 (2007) 148–186. Elsevier.
[26] M. Vogt, V. Kecman, Active-set methods for support vector machines, in: Support Vector Machines: Theory and Applications 177, 2005, pp. 133–158. Springer.
[27] M. Gerven, R. Jurgelenaite, B. Taal, T. Heskes, P. Lucas, Predicting carcinoid heart disease with the noisy-threshold classifier, Artif. Intell. Med. 40 (1) (2007) 45–55.
[28] M.M. Maresh, V.A. Beal, A longitudinal survey of nutrition intake, body size, and tissue measurements in healthy subjects during growth on JSTOR, Monogr. Soc. Res. Child Dev. 33 (1970) doi.org/1165811.
[29] E.R. Buskirk, J. Mendez, Energy: caloric requirements, in: R.B. Alfin-Slater, D. Kritchevsky (Eds.), Nutrition and the Adult, Human Nutrition, 3A, Springer, Boston, MA, 1980, https://doi.org/10.1007/978-1-4613-3015-8_2.
[30] S. Sweidan, S. El-Sappagh, H. El-Bakry, et al., A fibrosis diagnosis clinical decision support system using fuzzy knowledge, Arabian J. Sci. Eng. 44 (2019) 3781–3800.
[31] I.H. Sarker, Data science and analytics: an overview from data-driven smart computing, decision-making and applications perspective, SN Comput. Sci. 2 (2021) 377.
[32] R. Alpar, Applied Statistic and Validation - Reliability, Detay Publishing, 2010.
[33] M. Uçar, M. Bozkurt, C. Bilgin, K. Polat, Automatic detection of respiratory arrests in OSA patients using PPG and machine learning techniques, Neural Comput. Appl. 28 (10) (2017), https://doi.org/10.1007/s00521-016-2617-9.
[34] M. Uçar, M. Nour, H. Sindi, K. Polat, The effect of training and testing process on machine learning in biomedical datasets, Math. Probl Eng. 2020 (2020) 1–17.
[35] S. Arielle, Selya, Drake Anshutz, Machine learning for the classification of obesity from dietary and physical activity patterns, Smart Innova. Syst. Technol. 93 (2018) 77–97. Springer.
[36] J. Dunstan, M. Aguirre, M. Bastías, C. Nau, Predicting nationwide obesity from food sales using machine learning, Health Inf. J. 26 (1) (2020) 652–663.
[37] A.R.H. van Zanten, L. Petit, J. De Waele, et al., Very high intact-protein formula successfully provides protein intake according to nutritional recommendations in overweight critically ill patients: a double-blind randomized trial, Crit. Care 22 (2018) 156, https://doi.org/10.1186/s13054-018-2070-5.
[38] E. Hartsook, T. Hershberger, J. Nee, Effects of dietary protein content and ratio of fat to carbohydrate calories on energy metabolism and body composition of growing rats, J. Nutr. 103 (2) (1973) 167–178, https://doi.org/10.1093/jn/103.2.167.

Dr. Sara Sweidan

Assistant Professor of Artificial Intelligence

Artificial Intelligence Department, Faculty of Computer and Artificial Intelligence, Benha University, Egypt

Faculty of Computer Science and Engineering, New Mansoura University, New Mansoura, Egypt

Email: Sweidan_ds@fci.bu.edu.eg

Sara Sweidan received the B.Sc., M.Sc. and PhD. degrees in computer science and Artificial Intelligence from Mansoura University, Mansoura, Egypt, in 2002, 2011 and 2019, respectively. She is an assistant professor of Artificial Intelligence with the department of Artificial Intelligence, faculty of computers and artificial intelligence, Benha University, Benha, Egypt. Her current research interests include machine learning, computer vision, optimization algorithms, data science, quantum computing and biomedical informatics, sentiment analysis, and large language models.

Dr. S.S. Askar

Department of Statistics and Operations Research, College of Science, King Saud University, P.O. Box 2455, Riyadh 11451, Saudi Arabia

Email: saskar@ksu.edu.sa

S. ASKAR received his BSc. Degree in mathematics and the MSc. degree in applied mathematics from Mansoura University, Egypt, in 1998 and 2004, respectively. He got his PhD in Operation Research from Cranfield University in the UK in 2011. He works as an associate Professor at Mansoura University Egypt since 2016. He joined King Saud University in 2012, and till present, he works at the Department of Statistics and Operation Research at King Saud University as a professor. His main interests lie in game theory and its applications that include mathematical economy, dynamical systems and Network analysis.

Dr. Mohamed Abouhawwash

Department of Computational Mathematics, Science and Engineering (CMSE), College of Engineering, Michigan State University, East Lansing, MI 48824, USA;

Department of Mathematics, Faculty of Science, Mansoura University, Mansoura 35516, Egypt Email: abouhaww@msu.edu;saleh1284@mans.edu.eg

MOHAMED ABOUHAWWASH received the B.Sc. and M.Sc. degrees in statistics and computer science from Mansoura University, Mansoura, Egypt, in 2005 and 2011, respectively, and the joint Ph.D. degree in statistics and computer science with the Channel Program between Michigan State University, East Lansing, MI, USA, and Mansoura University, in 2015. In 2018, he was a Visiting Scholar with the Department of Mathematics and Statistics, Faculty of Science, Thompson Rivers University, Kamloops, BC, Canada. He is currently with Michigan State University. He is also an Associate Professor with the Department of Mathematics, Faculty of Science, Mansoura University. His current research interests include evolutionary algorithms, machine learning, image reconstruction, and mathematical optimization. He was a recipient of the Best Master's and Ph.D. Thesis Awards from Mansoura University in 2012 and 2018, respectively.

Dr. Elsayed Badr

Scientific Computing Department, Faculty of Computer and Artificial Intelligence, Benha University, Egypt; Email: badrgraph@gmail.com, alsayed.badr@fci.bu.edu.eg

Elsayed Badr is a full professor of computer science at Faculty of Computers & Artificial Intelligence, Benha University, Egypt. Prof. Dr. Badr received his Ph.D. degree in Parallel Algorithms (mainly in parallel programming applications) in 2006 from the University of Macedonia; Greece. He has worked on several research topics. He has contributed more than 50+ technical papers in the areas of Artificial Intelligence, Machine learning, Data Science, wireless networks, High-Performance Computing, Cloud Computing, graph algorithms, neutrosophic theory and Mathematical Programming in international journals and conferences.