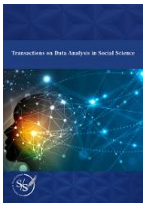
ISSN Online: 2821-1936

Transactions on Data Analysis in Social Science

Journal Homepage: https://transoscience.ir



A Transparent Knowledge-Based Diet Recommender System: Evaluation on NHANES Data and a Test of the Added Value of Machine Learning

R. Kapoor [1], A. Roy [2,*]

[1] Department of Computer Engineering, Faculty of Engineering, University of Delhi, Delhi, India
[2] Department of Computer Engineering, Jadavpur University, Kolkata, India

ARTICLE INFO

Article History:
Received 9 February 2026
Received in revised form
25 April 2026
Accepted 10 May 2026
Available online 9 June 2026

Keywords:
Recommender system;
Diet recommendation;
Knowledge-based system;
Basal metabolic rate;
NHANES; Machine learning; Interpretability

ABSTRACT

Overweight and obesity are major public-health problems, and access to dietary counselling is limited. This paper presents a knowledge-based diet recommender that estimates daily energy requirements from age, sex, height, weight and physical-activity level using validated equations (Mifflin–St Jeor, standard activity factors and a BMI-based adjustment with a safety floor) and recommends one of twelve dietitian-designed diets (1,000–3,500 kcal); users with diabetes or dyslipidaemia are referred to a specialist. The system was evaluated on public data from 4,785 adults in the US National Health and Nutrition Examination Survey (NHANES 2009–2012). Of these adults, 68.9% were overweight or obese and 44.6% met the referral criterion. Replacing the Mifflin–St Jeor equation with the 1918 Harris–Benedict equation left only 46.6% of non-referred adults with the same diet. An earlier prototype, which used ideal instead of actual weight and a single diet for all overweight users, assigned 77.2% of adults to one diet and agreed with the revised method in only 36.4% of the remaining cases. No machine-learning model outperformed the rule that generated the labels: the best, a multilayer perceptron, reached 92.5% accuracy, with errors clustered near energy-range boundaries. Typical self-reporting errors in weight and height changed the recommendation for 24.5% of users. A transparent, testable rule-based engine is therefore preferable to black-box models for this task; machine learning adds value only when labels come from real outcomes such as weight change or independent expert judgement. Code and unit tests are publicly available.

1. INTRODUCTION

Over recent decades, information and communication technology has changed the way people obtain knowledge and services. The volume of data generated and transmitted every day has grown rapidly, and some observers have called this change an “information revolution”. Data mining, the search for hidden patterns and relationships in large volumes of data for description and prediction, is one of the disciplines that have made this change possible [1]. Many everyday decisions, from buying goods to choosing health services, are now made with the help of intelligent systems.

* Corresponding Author: ananya.roy@gmail.com
Department of Computer Engineering, Jadavpur University, Kolkata, India

http://dx.doi.org/10.22034/tdas.2026.256211

© 2026 by the authors. Licensee TDASS, Tehran, Iran. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

Recommender systems are one of the most widely used technologies in this area. Liang et al. [2] regard them as a subset of decision-support systems: information systems that analyse past behaviour and make recommendations for current requests. In other words, they try to suggest the most suitable items (data, information, goods or services) to a user by analysing the user's characteristics and activity, and so help the user reach a goal more quickly in a large volume of information [3]. Recommender systems are usually grouped into three families: collaborative filtering, which relies on similarities between users' behaviour; content-based filtering, which compares item features with the user profile; and knowledge-based systems, which derive recommendations from explicit domain rules and knowledge [3–5]. In health applications, where a wrong recommendation can cause harm and behavioural data are scarce, the third family is particularly important.

Health and nutrition are among the most important fields of application. A sedentary, mechanised lifestyle has brought overweight and obesity to a large part of the population. The World Health Organization regards obesity as a major public-health challenge associated with cardiovascular disease, type 2 diabetes and some cancers [16]; in the data analysed in this paper, 68.9% of adults are overweight or obese. Demand for dietary counselling has increased, but dietitians are not always available, consultations are costly and repeated visits are time-consuming. An intelligent nutrition system can help people assess their situation more easily, cheaply and continuously [4], and it can prepare baseline information for the dietitian before a consultation. Its users fall into two groups: individuals looking for an eating plan suited to their circumstances, and dietitians and nutrition centres that can use the system as a decision-support tool [5].

A health recommender differs fundamentally from a film or product recommender, because a wrong recommendation may harm the user. Three properties are therefore essential: (1) scientific correctness, that is, the use of valid and current equations to estimate energy requirements; (2) transparency, that is, the ability to explain why a particular diet is recommended; and (3) safety, that is, the identification of users whose clinical condition lies outside the system's competence and their referral to a specialist [20, 21, 30]. Many studies apply machine-learning and deep-learning models to every problem. Such models are useful only when the relationship between input and output is unknown and has to be learned from data. If the output (the recommended diet) is constructed from the inputs by an explicit rule, a learning model can at best approximate that rule, and its accuracy measures nothing more than its ability to reproduce the formula.

An earlier prototype of the present system, developed by the authors, followed this pattern. It used 1,415 records with seven features compiled from internet sources; the label of each record (one of 13 diets) was set with the Harris–Benedict equation and a table of energy ranges; and three algorithms (non-linear regression, a multilayer perceptron and a convolutional neural network) were trained to predict this label. The prototype had no reproducible data source, contained several calculation errors (for example, ideal weight was substituted for actual weight in the basal-metabolism equation, and a factor of 1.9 was used for "high" activity), and assigned a single "near-normal" diet 13 to all users with diabetes, dyslipidaemia or overweight. This paper replaces the prototype with a knowledge-based system and tests each of these design choices on real data.

The contributions of this paper are:

The design and implementation of a knowledge-based diet-recommendation engine with validated equations (Mifflin–St Jeor [14], standard activity factors [19] and an energy adjustment based on guidelines for the management of overweight [25]), together with a safety rule that refers users with diabetes or dyslipidaemia to a specialist, and unit tests that check every equation against hand-calculated values.

An evaluation of the system on real, public data from the US National Health and Nutrition Examination Survey (NHANES) [17, 18] covering 4,785 adults, with the distribution of recommended diets in the population.

A quantitative assessment of the effect of the choice of basal-metabolism equation and of the calculation errors of the earlier prototype on the system's decisions.

An explicit test of whether machine-learning models, including neural networks, add value in this problem, and an analysis of where their errors occur.

An analysis of the sensitivity of the recommendations to measurement error in self-reported inputs, which is unavoidable in online systems [26].

Section 2 reviews previous work; Section 3 describes the proposed model and its equations; Section 4 presents the data and experimental design; Section 5 reports the results; Section 6 discusses the findings, the limitations and a protocol for future clinical evaluation; and Section 7 concludes.

2. RELATED WORK

Computer-based nutrition counselling has a history of several decades, and most systems belong to the family of expert and knowledge-based systems. Ramachandran et al. [7] presented one of the first diet programmes, the Nutrition Diet Programme (NDP), developed to help rural populations without access to a dietitian or physician. Another expert system, NCMM, supported menu planning and the management of eating habits. Lee et al. [6] developed a type-2 fuzzy ontology that assessed the nutritional suitability of meals and recommended personalised diets for people with diabetes.

Chen et al. [8] presented a nutrition-diagnosis expert system based on the Nutrition Care Process and Model endorsed by the American Dietetic Association; it integrated the diagnostic knowledge of several dietitians into a rule-based system. Krbez and Shaout [9] proposed a fuzzy nutrition system and compared fuzzy and crisp approaches on the basis of user feedback, arguing that the decision to eat or not eat a food is not inherently crisp. Heinonen et al. [10] presented a fuzzy expert system for a nutritional-guidance application that acquired expert knowledge through tables of variables and fuzzy rules and suggested food-intake quantities. Iwendi et al. [11] proposed a machine-learning diet recommendation system for Internet of Medical Things settings that assigned patients to suitable diet categories on the basis of their health attributes; in terms of problem structure (assigning a user to one of several predefined diets) it is the closest work to the present study.

Recent reviews give a broader picture of healthy-food recommendation. Tran et al. [21] classify recommender systems in the healthy-food domain into food, recipe, menu and group recommenders and show that many menu recommenders are knowledge- or constraint-based because they must satisfy explicit nutritional constraints. Trattner and Elsweiler [20] list the main challenges of food recommenders: the lack of valid labelled data, the difficulty of evaluating health outcomes, and the conflict between user preference and health. A common point of these reviews is that, in health, evaluation by "label-prediction accuracy" is insufficient; the clinical correctness and safety of the recommendations must be assessed.

Almost all of these systems rest on an estimate of daily energy requirements. The first widely used equation for basal metabolic rate was published by Harris and Benedict [12] in 1918; Roza and Shizgal [13] re-estimated its coefficients in 1984; and Mifflin et al. [14] derived a new equation in 1990 from 498 healthy normal-weight and obese individuals. The systematic review of Frankenfield et al. [15] found that the Mifflin–St Jeor equation gives the most accurate estimates in normal-weight and obese adults (predictions within ±10% of measured values in about 70–80% of individuals) and that the Harris–Benedict equation overestimates basal metabolism, especially in obese people; the American Dietetic Association therefore recommends the Mifflin–St Jeor equation for adults. Multiplying the basal metabolic rate by the physical activity level (PAL) is the standard way to estimate total daily energy expenditure, and reference values are given in the joint FAO/WHO/UNU report [19].

Two gaps remain in comparable systems. First, many have been evaluated on data whose source cannot be reproduced or whose labels were produced by the same rule that the system implements. Second, the effect of the choice of energy equation and of input errors on the final decision has rarely been measured. The present study addresses both gaps.

3. PROPOSED MODEL

3.1. Design principles

Nutrition matched to an individual's body provides the energy needed to maintain tissues and organ function, and a persistent deviation in either direction has adverse health consequences. The system is intended to recommend an eating plan for healthy adults, that is, people without a specific disease who want to know their body status and energy needs. It is a health recommender system (HRS): using the individual's profile, which is part of a personal

health record (PHR), it selects one plan from a set of predesigned eating plans. Because the selection rules are explicit nutritional knowledge rather than a pattern to be learned from data, the system is knowledge-based. Three principles guided its design:

Correctness: every calculation rests on a published, validated equation and is checked by a unit test against a hand-calculated value.

Transparency: the output is not only a diet number; the basal metabolic rate, total energy expenditure, energy adjustment and the rule that led to the diet are shown to the user.

Safety: the system knows the limits of its competence. Users with diabetes or dyslipidaemia are referred to a dietitian instead of receiving an automatic diet, and the target energy never falls below a safety floor.

Figure 1 shows the workflow. The inputs are age, sex, height, weight, physical-activity level and the answers to two questions: "Do you have diabetes?" and "Do you have dyslipidaemia?". After the inputs are checked (for example, height between 120 and 230 cm and age 18 years or more), the safety rule is applied; the energy requirement is then estimated and adjusted to the user's goal, and finally the diet corresponding to the energy range is selected and presented together with an explanation of the calculations.

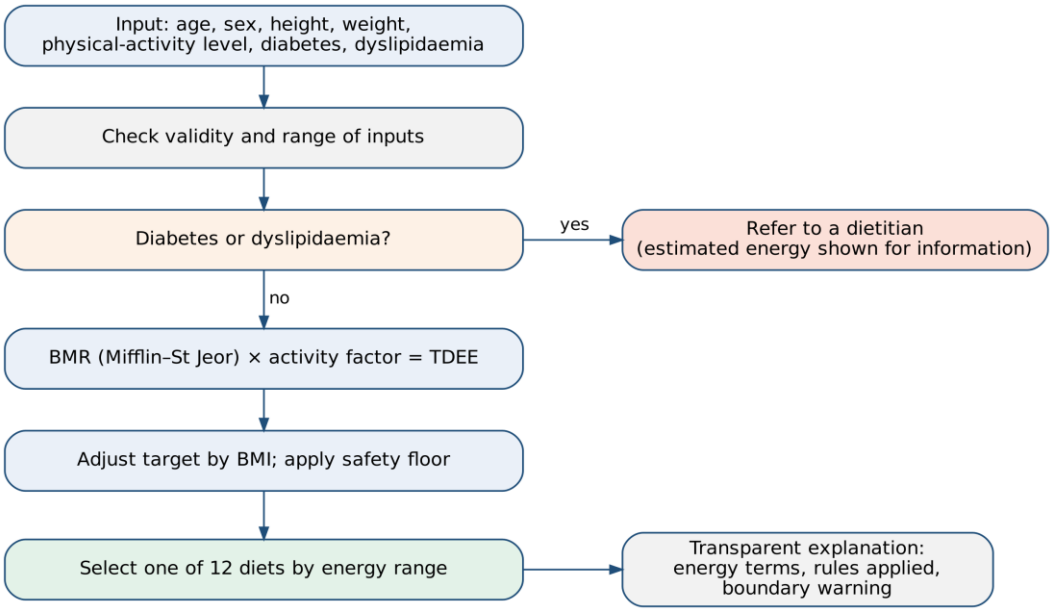


Fig. 1. Workflow of the proposed recommender system.

3.2. Body mass index

The body mass index (BMI) is the ratio of weight to the square of height:

$$BMI = \frac{w}{(h/100)^2} \tag{1}$$

where w is weight in kilograms and h is height in centimetres. BMI does not distinguish between fat mass and lean mass and is therefore not a complete measure of body fat, but it is the standard screening index. The WHO classification for adults [16] is: underweight (BMI < 18.5), normal (18.5 to < 25), overweight (25 to < 30) and obese (≥ 30). In the proposed system BMI is used only to set the direction of the energy adjustment (weight loss, maintenance or gain).

3.3. Basal metabolic rate

The basal metabolic rate (BMR) is the energy needed to maintain vital functions at rest. It is not the energy a person should consume each day; total daily energy expenditure is obtained by multiplying it by the activity level (Section 3.4). The primary equation is that of Mifflin–St Jeor [14]:

$$BMR_{MSJ} = 10\,w + 6.25\,h - 5\,a + s, \qquad s = \begin{cases} +5 & \text{men} \\ -161 & \text{women} \end{cases} \tag{2}$$

where a is age in years. For comparison, two forms of the Harris–Benedict equation were implemented: the original 1918 form [12], which the earlier prototype used with rounded coefficients,

$$BMR_{HB18} = \begin{cases} 66.473 + 13.7516\,w + 5.0033\,h - 6.755\,a & \text{men} \\ 655.0955 + 9.5634\,w + 1.8496\,h - 4.6756\,a & \text{women} \end{cases} \tag{3}$$

and the revised form of Roza and Shizgal [13]:

$$BMR_{HBR} = \begin{cases} 88.362 + 13.397\,w + 4.799\,h - 5.677\,a & \text{men} \\ 447.593 + 9.247\,w + 3.098\,h - 4.330\,a & \text{women} \end{cases} \tag{4}$$

The Mifflin–St Jeor equation was chosen as the primary equation because of the systematic-review evidence [15] that it is more accurate in contemporary populations with a high prevalence of obesity. Actual body weight is used in all equations. Substituting an "ideal weight" (the desirable BMI for the age group multiplied by the square of height), as the earlier prototype did, is inconsistent with equations whose coefficients were fitted on actual weight; it underestimates the requirements of obese people and overestimates those of underweight people (Section 5.4). Weight-loss or weight-gain goals are applied explicitly through the energy adjustment (Section 3.5).

3.4. Physical-activity level and total energy expenditure

Total daily energy expenditure (TDEE) is the product of the basal metabolic rate and the activity factor:

$$TDEE = BMR \times PAL \tag{5}$$

The system asks the user for one of four activity levels and uses the standard factors in Table 1. The earlier prototype used a factor of 1.9 for the fourth level ("high activity, e.g. exercising two or three times a day"); in the usual classification this factor corresponds to "extremely active" (heavy physical work or professional training) and overestimates the requirements of people described as "very active" by about 10%. Its factor of 1.37 for the second level was also replaced by the usual value of 1.375.

Table 1. Physical-activity levels, PAL factors and their assignment from the NHANES questionnaire.

Level	Description	Prototype factor	Factor used	Criterion in NHANES	Sample share (%)
1	Sedentary: no regular exercise	1.2	1.2	No moderate/vigorous sport and active on fewer than 3 days a week	41.2
2	Lightly active: light exercise 1–3 days a week	1.37	1.375	Moderate/vigorous sport with fewer than 3 active days, or no sport but 3 or more active days	32.1
3	Moderately active: exercise 3–5 days a week	1.55	1.55	Moderate/vigorous sport and 3–5 active days	20.1
4	Very active: exercise 6–7 days a week	1.9	1.725	Moderate/vigorous sport and 6–7 active days	6.6

3.5. Energy adjustment and safety floor

TDEE is the energy needed to maintain current weight. For people with overweight or obesity, the guideline for the management of overweight in adults [25] recommends a daily reduction of 500–750 kcal and an intake usually of 1,200–1,500 kcal/day for women and 1,500–1,800 kcal/day for men. The target energy was therefore defined as

$$E = \min\big(3500,\ \max(E_{\min},\ TDEE + \Delta)\big), \qquad \Delta = \begin{cases} -500 & BMI \geq 25 \\ +300 & BMI < 18.5 \\ 0 & \text{otherwise} \end{cases} \tag{6}$$

where the safety floor E_{min} is 1,200 kcal for women and 1,500 kcal for men. The reduction of 500 kcal is the lower end of the recommended range and was deliberately chosen to be conservative; the increase of 300 kcal for underweight people is a common starting value that should be adjusted with weight monitoring. The ceiling of 3,500 kcal is the upper limit of the designed diets.

3.6. Diet selection and safety rule

Twelve diets were designed with the help of a dietitian on the basis of energy ranges, which are given in Table 4 (Section 5). Diet k is assigned to a user whose target energy is at least the lower bound and below the upper bound of the k-th range; the first bound is 1,000 kcal and the last is 3,500 kcal.

The earlier prototype had a thirteenth, “near-normal” diet for “people with diabetes, dyslipidaemia and overweight”. This design is problematic in two respects. First, dietary treatment of diabetes and dyslipidaemia is individual and depends on medication (especially insulin and glucose-lowering drugs that carry a risk of hypoglycaemia), glycaemic control, kidney function and blood-lipid levels [23, 24]; a single diet for all of these people, whose energy requirements range from 1,200 to more than 3,000 kcal, cannot be appropriate. Second, grouping “overweight” with these diseases, given the high prevalence of overweight, deprives a large share of users of a diet matched to their energy needs (Section 5.4). Diet 13 was therefore removed and replaced by the following rule:

If the user has diabetes or dyslipidaemia, the system does not recommend a diet; it displays the estimated energy requirement for information and refers the user to a dietitian or physician.

Users with overweight or obesity but without these diseases receive, like everyone else, the diet that matches their adjusted target energy. In the NHANES data, diabetes was defined by the answer to the question on a physician’s diagnosis, and dyslipidaemia by laboratory values: total cholesterol ≥ 6.22 mmol/L (240 mg/dL) or HDL cholesterol below 1.03 mmol/L in men and 1.29 mmol/L in women (40 and 50 mg/dL), following the NCEP ATP III thresholds [24]. In online use, both items are asked of the user.

3.7. Explainable output

For each user, the system reports, in addition to the diet number, the BMI and its category, the BMR, the activity factor and TDEE, the adjustment applied and its reason, whether the safety floor was activated, and the distance of the target energy from the nearest diet boundary. The last item matters in practice: if the target energy lies less than about 50 kcal from a boundary, a small error in weight or activity level can change the diet (Section 5.6), and the system issues the warning “the adjacent diet is also suitable”. This kind of explanation is not readily available from black-box models [30].

3.8. Implementation and unit tests

The system was implemented in Python with NumPy and pandas. All functions are vectorised, and the whole NHANES sample is processed in a fraction of a second. Unit tests compare the three BMR equations with hand-calculated values for reference cases (for example, a 30-year-old man of 180 cm and 80 kg has a BMR of 1,780 kcal by Eq. (2)) and check the BMI category bounds, the diet-range bounds, the safety floor and the referral rule. All tests passed.

4. DATA AND EXPERIMENTAL DESIGN

4.1. Data source

The earlier prototype used 1,415 records with seven features compiled from "various internet databases", including the UCI repository and the US Food Environment Atlas. Neither source provides a dataset with these features at the individual level (the Food Environment Atlas contains county-level data), and the dataset identifier, merging procedure and exclusion criteria were not documented, so those data cannot be reproduced. The present study therefore used data from the US National Health and Nutrition Examination Survey (NHANES) [17]. NHANES is conducted continuously by the National Center for Health Statistics (NCHS); height and weight are measured in a standardised examination, blood lipids are determined in the laboratory, and physical activity and medical history are recorded with structured questionnaires. The version used is the ready-made NHANES package [18], which contains 10,000 records from the 2009–2010 and 2011–2012 cycles and is freely available without registration. It contains the same seven features as the prototype data (age, sex, height, weight, activity level, diabetes and blood lipids), but the values are measured.

4.2. Data preparation

The records in the file were created by resampling with the survey weights, so an individual may appear more than once; the 10,000 records contained 6,779 unique identifiers, and duplicates were removed. Individuals aged 18 years or older were then selected (4,835), because the energy equations and adult BMI thresholds are not valid for children and adolescents. Individuals with missing height, weight or diabetes status (50) were excluded, giving a final sample of 4,785 (2,554 from 2009–2010 and 2,231 from 2011–2012). Cholesterol was missing for 5.8% of individuals; for them dyslipidaemia was treated as "unknown" and the referral rule was applied on the basis of diabetes alone. This version of the data has no variable for current pregnancy; in real use, pregnant and lactating women must be excluded from the scope of the system (Section 6.2). Activity level was derived from two questionnaire variables: participation in moderate or vigorous sports or recreational activities (yes/no) and the number of physically active days in a typical week (Table 1).

Table 2 describes the final sample. Age ranges from 18 to 80 years (higher ages are top-coded at 80 in NHANES), height from 134.5 to 200.4 cm and weight from 37.0 to 230.7 kg; these ranges are wider than those of the prototype data (weight 43–130 kg) and include people with severe obesity.

Table 2. Characteristics of the NHANES 2009–2012 adult sample after preparation.

Characteristic	Value
Number of adults	4,785
Men (%)	48.7
Age (years), mean ± SD	46.3 ± 17.9
BMI (kg/m²), mean ± SD	28.8 ± 6.8
Underweight / normal / overweight / obese (%)	1.9 / 29.3 / 33.6 / 35.3
Diabetes (%)	11.2
Total cholesterol ≥ 6.22 mmol/L (%)	11.4
Low HDL cholesterol (%)	29.7
Diabetes or dyslipidaemia, i.e. referred (%)	44.6

4.3. Research questions and experiments

Five experiments were designed to answer five questions:

Experiment 1, population behaviour: which diets does the system recommend and in what proportions, and for what share of people are the safety rule and the energy floor activated?

Experiment 2, choice of equation: how do the energy estimate and the recommended diet change if the Harris–Benedict equation of 1918 or its revised form is used instead of the Mifflin–St Jeor equation? To isolate the effect of the equation, the energy difference was computed on TDEE without adjustment, and the agreement of diets after adjustment; agreement is reported as percentage agreement and quadratically weighted kappa [29].

Experiment 3, errors of the earlier prototype: the prototype's calculation was reconstructed exactly (the Harris–Benedict equation with its rounded coefficients, ideal instead of actual weight, activity factors 1.2/1.37/1.55/1.9 and diet 13 for people with diabetes, dyslipidaemia or overweight) and compared with the revised method. Because the table of desirable BMI by age in the prototype documentation was illegible, the common table of desirable BMI by age was used (midpoints of the ranges from 19–24 kg/m² at ages 19–24 to 24–29 kg/m² at age 65 and over) [27].

Experiment 4, value of machine learning: do machine-learning models, given the same inputs and the labels produced by the system's rule, reach an accuracy greater than, or at least equal to, that of the rule itself? This repeats the design of the prototype and reports its result.

Experiment 5, sensitivity to input error: what share of recommendations changes under realistic self-reporting errors in weight, height and activity level?

4.4. Machine-learning models

In Experiment 4 the input of each model consists of the seven features (age, sex, height, weight, activity level, diabetes, dyslipidaemia) and the output is a 13-class label (diet 1–12 or referral). Four models were implemented with scikit-learn [22]: (1) multinomial logistic regression with standardised features, as the linear baseline; (2) non-linear (cubic polynomial) regression that predicts the target energy from the five anthropometric and activity features and then places it in the table of ranges, with the referral rule applied from the two clinical features (the counterpart of the prototype's "non-linear regression"); (3) histogram-based gradient boosting with default settings; and (4) a multilayer perceptron with two hidden layers of 64 neurons, ReLU activation and the Adam optimiser (the counterpart of the prototype's multilayer perceptron). A convolutional neural network (CNN) was not implemented: CNNs are designed for data with spatial or sequential structure (images, signals, text), and the seven tabular features of this problem have no such structure, so convolution reduces to a multilayer network with weight constraints. Moreover, as shown below, even the multilayer network cannot outperform the rule, so a more complex model would not change the conclusion.

Evaluation used stratified five-fold cross-validation [28]; for comparison with the prototype design, a random 70/30 split is also reported. The metrics are accuracy (share of exactly correct recommendations), macro-F1 (mean F1 over all classes, which is sensitive to small classes), ±1-diet accuracy (share of recommendations at most one diet away from the correct one) and the number of referred individuals to whom the model assigned a diet. To examine the effect of data size, a learning curve was computed with 100 to 3,349 training samples on a fixed test set.

4.5. Simulation of input error

In Experiment 5 the inputs were perturbed with realistic errors, and the change of the recommended diet relative to the diet computed from the measured values was recorded. The scenarios were random error in height (SD 2 cm), random error in weight (SD 2 and 5 kg), systematic under-reporting of weight by 2 kg, a shift of activity level to an adjacent level with probability 20%, and combinations of these. The "typical self-report" scenario (weight under-reported by 2 kg on average with SD 2 kg, and height SD 1.5 cm) was based on the systematic review of Connor Gorber et al. [26]. Each scenario was repeated 200 times. In addition, sensitivity to the definition of activity level was assessed with an alternative mapping (level 4 from five active days, level 3 from two active days, and everyone without sport at level 1).

5. RESULTS

5.1. A worked example

Before the population results, the operation of the system is illustrated with an example. Consider a 35-year-old woman, 165 cm tall, weighing 78 kg, who does light exercise two days a week (level 2) and has neither diabetes nor dyslipidaemia. Her BMI is $78 / 1.65^2 = 28.7$ kg/m² (overweight). Her BMR by Eq. (2) is $780 + 1031 - 175 - 161 =$

1,475 kcal, and her TDEE is 1,475 × 1.375 = 2,028 kcal. Because she is overweight, 500 kcal is subtracted, giving a target of 1,528 kcal, which is above the floor of 1,200 kcal. This value lies in the range 1,425 to < 1,575 kcal, so diet 4 is recommended. The target is 47 kcal from the upper bound of the range, less than the 50-kcal threshold, so the system warns that diet 5 is also suitable. With the 1918 Harris–Benedict equation instead of Eq. (2), the target would be 1,621 kcal and diet 5 would be recommended. The example illustrates the phenomenon that Experiment 2 measures at the population level.

5.2. Behaviour of the system in the population (Experiment 1)

Figure 2 and Tables 3 and 4 show the results of running the system on the 4,785 NHANES adults. The first important finding is the high share of referrals: 44.6% of individuals have diabetes (11.2%) or dyslipidaemia, and the system does not give them an automatic diet. Much of this group is due to low HDL cholesterol (29.7% of the whole sample). The figure reflects the epidemiology of the adult population rather than a design flaw: a system for "healthy people" that honestly respects the limits of its competence answers about half of adults directly and acts as a screening and referral tool for the other half. If the referral criterion were limited to diabetes and high total cholesterol, the share would fall to 21.4%. The referral threshold should be decided with a dietitian, and the system provides it as an explicit parameter.

Table 3. Summary of the behaviour of the revised system in the NHANES sample.

Indicator	Value
Referred (%)	44.6
Assigned one of diets 1–12 (%)	55.4
Target energy, mean ± SD (kcal/day), all adults	1,808 ± 446
Safety floor activated (%)	15.2
Women assigned diet 2 (floor of 1,200 kcal) (% of women)	16.2

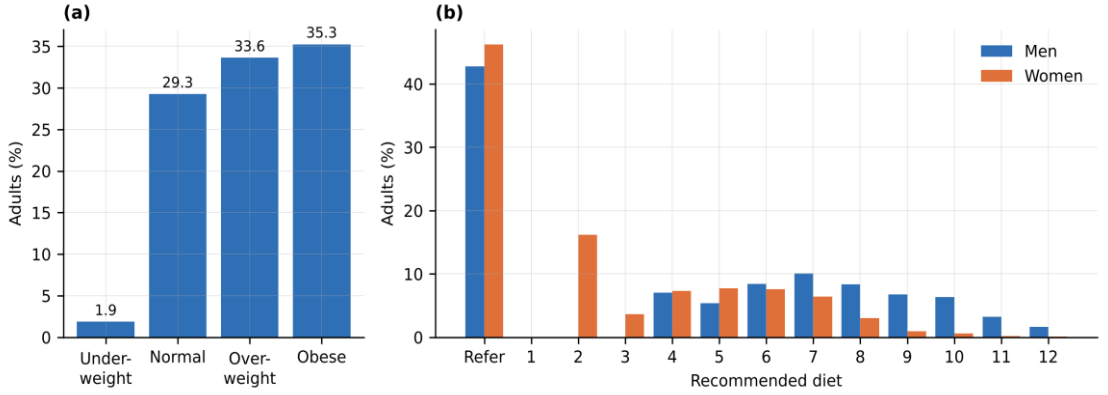


Fig. 2. (a) Distribution of BMI categories in the sample; (b) distribution of the diets recommended by the revised system, by sex. The "Refer" bar shows individuals with diabetes or dyslipidaemia.

Table 4. Energy ranges of the twelve diets and distribution of the diets recommended by the revised system in the NHANES sample (n = 4,785).

Diet	Energy range (kcal/day)	All (%)	Men (%)	Women (%)
1	1,000 to < 1,200	0.0	0.0	0.0
2	1,200 to < 1,350	8.3	0.0	16.2
3	1,350 to < 1,425	1.9	0.0	3.7

4	1,425 to < 1,575	7.2	7.0	7.3
5	1,575 to < 1,725	6.6	5.4	7.7
6	1,725 to < 1,900	8.0	8.4	7.6
7	1,900 to < 2,100	8.2	10.1	6.4
8	2,100 to < 2,275	5.6	8.4	3.0
9	2,275 to < 2,425	3.8	6.7	0.9
10	2,425 to < 2,650	3.4	6.3	0.6
11	2,650 to < 2,900	1.7	3.3	0.2
12	2,900 to 3,500	0.8	1.6	0.1
Referral	Diabetes or dyslipidaemia	44.6	42.8	46.3

Among non-referred individuals the distribution of diets is wide and differs clearly between the sexes (Fig. 2b). Diet 1 (1,000–1,200 kcal) is never recommended, because the safety floors of 1,200 kcal for women and 1,500 kcal for men do not allow lower intakes; for the same reason diets 2 and 3 are not recommended to men. Diet 2 is the most frequent diet for women (16.2% of women), because for many sedentary overweight women the 500-kcal reduction brings the target down to the floor. This has two implications for diet design: diet 1 has no use in this population except under clinical supervision, and diet 2 should be designed with particular care for micronutrient adequacy because it has the most users. Diets 11 and 12 are rarely used and are mostly assigned to young, active men.

5.3. Effect of the choice of basal-metabolism equation (Experiment 2)

Table 5 and Figure 3 show that the choice of equation has a systematic and substantial effect. Both forms of the Harris–Benedict equation estimate TDEE higher than the Mifflin–St Jeor equation, by 122 kcal on average for the 1918 form and 115 kcal for the revised form. The difference increases with BMI, from about 88 kcal in normal-weight people to about 160 kcal in obese people (Fig. 3a). This is the pattern that Frankenfield et al. [15] reported from direct calorimetry: the Harris–Benedict equation overestimates the metabolism of obese people.

Table 5. Effect of the choice of basal-metabolism equation on energy and recommended diet; reference: Mifflin–St Jeor. Diet agreement is computed for non-referred adults.

Indicator	Harris–Benedict 1918	Revised Harris–Benedict (1984)
Mean difference from Mifflin–St Jeor, kcal (± SD)	122 ± 82	115 ± 72
Mean difference, men / women	126 / 117	132 / 99
Mean difference, normal weight	88	86
Mean difference, obese	160	144
Difference > 200 kcal (%)	12.6	9.8
Same diet (%)	46.6	49.0
At most one diet apart (%)	94.3	95.8
Two or more diets apart (%)	5.7	4.2
Quadratically weighted kappa	0.950	0.955

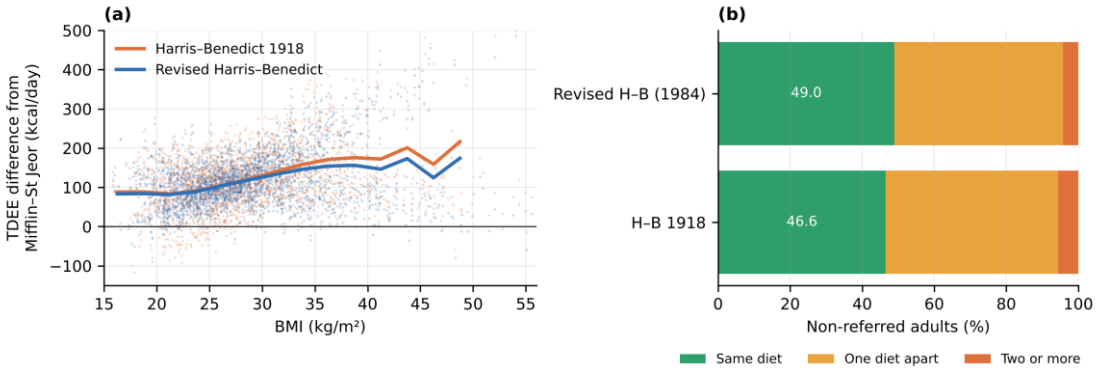


Fig. 3. Effect of the choice of basal-metabolism equation: (a) difference between the TDEE estimated with the Harris–Benedict equations and with the Mifflin–St Jeor equation, by BMI (points: a random subsample of 2,500 individuals; lines: means in BMI bins); (b) agreement of the recommended diet with the diet obtained with the Mifflin–St Jeor equation.

The consequences for the system's decision are large. Because the diet ranges are narrow (75 to 225 kcal, most of them 150–200 kcal), a difference of about 100 kcal is enough to change a person's diet. With the 1918 Harris–Benedict equation only 46.6% of non-referred individuals receive the same diet as with the Mifflin–St Jeor equation, and 5.7% move by two or more diets. The high weighted kappa (about 0.95) shows that the order of individuals is almost preserved, that is, the two equations agree on who needs more energy, but they disagree on the exact diet of more than half of the individuals. An important consequence is that the "accuracy" of any system that derives its labels from one of these equations depends on the choice of equation and cannot be interpreted as a measure of clinical correctness; the only way to arbitrate between equations is direct measurement of energy expenditure (indirect calorimetry) or follow-up of weight change.

5.4. Effect of the calculation errors of the earlier prototype (Experiment 3)

Table 6 and Figure 4 compare the calculation of the earlier prototype with the revised method. Three main differences were found.

First, diet 13 covers the majority of the population. Because its criterion included "overweight", 77.2% of adults are sent to this single diet, and 32.7% of all individuals fall into it only because their BMI is ≥ 25, without any disease. In other words, the prototype made no personalised decision for more than three quarters of the population and gave the same diet to everyone from a sedentary 60-year-old woman to an athletic 25-year-old man.

Second, ideal weight reverses the direction of the adjustment. Substituting ideal for actual weight reduces the energy requirement by 510 kcal on average in obese people and by 152 kcal in overweight people and increases it by 234 kcal in underweight people. In direction this resembles a goal adjustment, but its size depends on the degree of overweight and it has no safety floor: in people with severe obesity (BMI ≥ 40) it creates a mean deficit of about 870 kcal, beyond the guideline recommendation [25], while it creates almost no deficit in people with mild overweight. As a result, the prototype's target energy in the overweight group is on average 415 kcal higher than that of the revised method (Fig. 4).

Third, the factor of 1.9 for the fourth level overestimates the requirements of the 6.6% of individuals at this level by about 10%.

Overall, the target energies of the two methods differ by 266 kcal on average (absolute value), and by more than 300 kcal in 40.7% of individuals. Even among the 1,089 individuals to whom the prototype gives a diet other than 13, only 36.4% of recommendations agree with the revised method.

Table 6. Comparison of the calculation of the earlier prototype (Harris–Benedict with rounded coefficients, ideal weight, activity factors 1.2/1.37/1.55/1.9 and diet 13) with the revised method; energy in kcal/day.

Indicator	Value
Assigned diet 13 by the prototype (%)	77.2
— of whom only because BMI ≥ 25 (no diabetes or dyslipidaemia) (% of all)	32.7
Mean energy difference (prototype minus revised)	231
Mean absolute energy difference	266
Difference > 300 kcal (%)	40.7
Mean difference, normal / overweight / obese	150 / 415 / 135
Diet agreement among adults given diets 1–12 by the prototype (n = 1,089) (%)	36.4

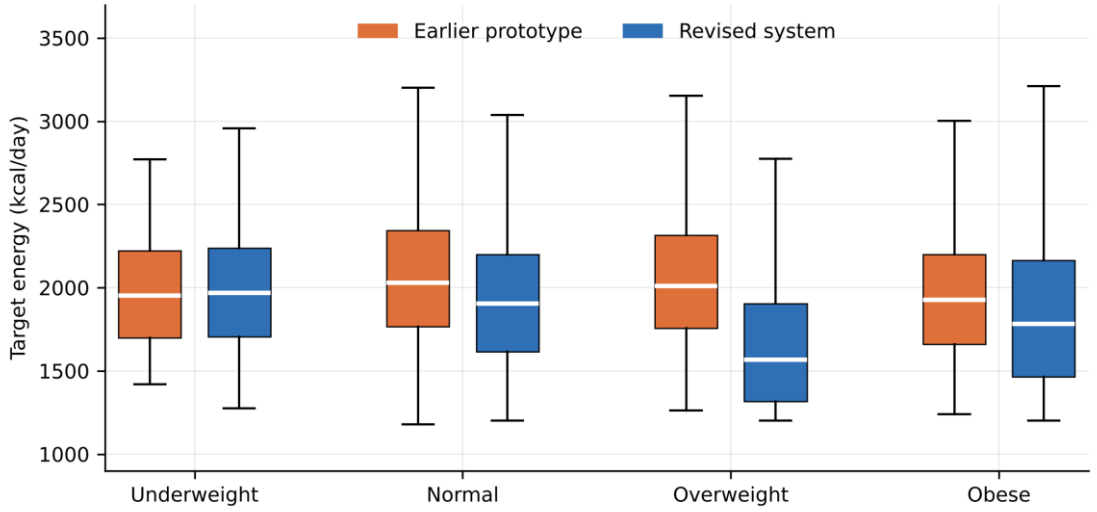


Fig. 4. Distribution of target energy with the earlier prototype and the revised method, by BMI category (boxes: quartiles; white line: median; whiskers: 1.5 × IQR).

5.5. Does machine learning add value? (Experiment 4)

Table 7 and Figure 5 show the results of training machine-learning models to reproduce the recommended diet. By definition, the system's transparent rule has an accuracy of 100%, because the labels were generated by this rule. The question is whether a learning model can reach it.

Table 7. Reproduction of the rule-based recommendation by machine-learning models (mean ± SD over the five stratified cross-validation folds; %). ±1-diet accuracy is computed for non-referred individuals only.

Model	Accuracy (5-fold)	Macro-F1 (5-fold)	±1-diet accuracy	Accuracy (70/30 split)
Multinomial logistic regression	67.3 ± 0.8	41.1 ± 2.1	79.0	66.5
Polynomial regression (degree 3)	70.6 ± 1.1	48.8 ± 2.4	93.4	70.7
Gradient boosting (histogram)	80.3 ± 1.3	63.4 ± 2.8	91.3	78.8
Multilayer perceptron (64–64)	92.5 ± 0.8	85.3 ± 2.0	97.1	90.6
Transparent rule (proposed system)	100	100	100	100

The best model, the multilayer perceptron, reaches an accuracy of 92.5% in cross-validation and 90.6% on the 70/30 split. This looks high at first sight, but it means that for 7.5% of individuals (361 of 4,785) the model recommends a different diet from the one determined by the nutritional rule. This error is not compensated anywhere: in no case is the model "better" than the rule, because the rule is itself the reference. Simpler models are much weaker. Logistic regression reaches only 67.3%, because the input–label relationship involves a product (weight

times activity factor), a threshold (BMI ≥ 25) and truncation (the safety floor), which linear boundaries cannot represent; it also assigned a diet to 7 individuals who should have been referred, and the multilayer perceptron to one. Polynomial regression, although it has the structure "predict energy, then bin", reaches 70.6%, because a smooth polynomial cannot reproduce the 500-kcal jump at BMI = 25 and the truncation at the safety floor; its ±1-diet accuracy (93.4%) shows that its errors are mostly to an adjacent diet.

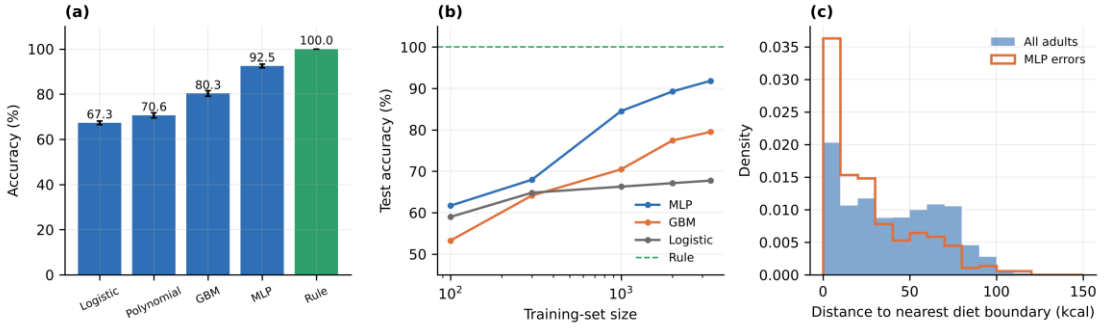


Fig. 5. (a) Accuracy of reproducing the recommended diet (five-fold cross-validation; error bars: SD over folds); (b) learning curve on a fixed test set (dashed line: rule); (c) distribution of the distance between the target energy and the nearest range boundary for all non-referred adults and for those whose diet the multilayer perceptron predicted incorrectly.

The learning curve (Fig. 5b) shows that the accuracy of the multilayer perceptron rises from 61.7% with 100 training samples to 91.9% with 3,349 samples and has not yet levelled off. With more data it would come closer to 100%, but at best it would reach the rule that was available from the start. With the data size of the prototype (about 1,000 training samples, 70% of 1,415 records), the expected accuracy is about 85%. Figure 5c shows where the errors occur: the median distance between the target energy and the nearest boundary is 19 kcal for individuals the model misclassified, against 38 kcal for all individuals. In other words, the model approximates the sharp boundaries of the table with "soft" boundaries and errs near them. These are precisely the individuals for whom the rule-based system issues the "adjacent diet" warning (Section 3.7), with the difference that the learning model is unaware of its error.

The conclusion of this experiment is clear: when the labels are a known deterministic function of the inputs, machine learning does not solve an unsolved problem. At best the learning model approximates the rule; it adds error, removes transparency and may violate the safety rule. The accuracy reported for such a model measures only its ability to approximate the formula, not the quality of the nutritional recommendation. A claim that a convolutional network would be superior in this setting is therefore not supported.

5.6. Sensitivity to input error (Experiment 5)

In online use, weight, height and activity level are reported by the users themselves and contain errors [26]. Table 8 and Figure 6 show that these errors have a large effect on the system's decision. A random error of 2 kg in weight changes the diet of 16.1% of individuals, and systematic under-reporting of 2 kg, which is common in self-reports, changes it for 20.9%. Under the typical self-report pattern, the diet of 24.5% of individuals changes, and if errors in the estimate of activity level are added, the share reaches 34.7%. The definition of activity level also matters: with an alternative and equally defensible mapping of the questionnaire answers to the four levels, the diet of 18.9% of individuals changes.

Table 8. Sensitivity of the recommended diet to measurement error in the inputs (200 Monte Carlo repetitions; non-referred individuals only).

Input-error scenario	Diet changed (%, mean ± SD)
Random height error (SD 2 cm)	11.4 ± 0.5
Random weight error (SD 2 kg)	16.1 ± 0.6
Random weight error (SD 5 kg)	34.9 ± 0.8
Systematic weight under-reporting (2 kg)	20.9
Shift of activity level to an adjacent level (20%)	14.4 ± 0.6
Typical self-report (weight −2 ± 2 kg, height ± 1.5 cm)	24.5 ± 0.6
Typical self-report + 20% activity error	34.7 ± 0.8
Alternative activity-level mapping	18.9

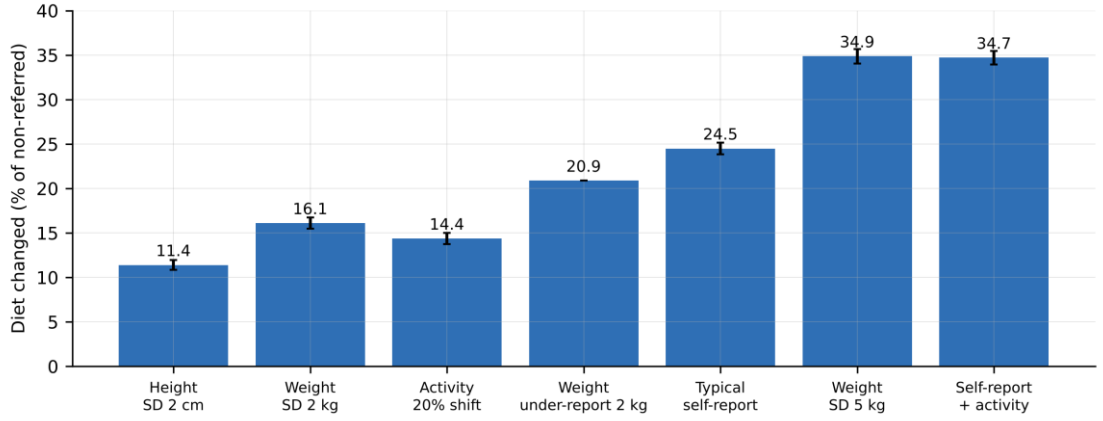


Fig. 6. Percentage of non-referred adults whose recommended diet changes under the input-error scenarios (mean ± SD over 200 repetitions; the systematic under-reporting scenario is deterministic).

The reason for this sensitivity is the narrow width of the diet ranges: one kilogram of weight changes the BMR in Eq. (2) by 10 kcal and, after multiplication by the activity factor, the TDEE by 12–17 kcal, and a shift of one activity level changes it by 11–15%. The analysis shows that the nominal precision of the recommended diet exceeds the precision of the inputs. We therefore recommend that diet designers either widen the ranges and reduce the number of diets, or design adjacent diets so that switching between them is harmless and the system offers both options near the boundaries.

6. DISCUSSION

6.1. Main findings

This study showed that, for the problem of choosing one of several predesigned diets on the basis of energy requirements, a transparent rule-based engine is superior to a machine-learning model: it is exact, explainable, needs no training data and guarantees the safety rules. This agrees with Rudin's argument [30] that in high-stakes decisions a black-box model should not replace an interpretable model that performs adequately. The real uncertainty of the system lies elsewhere: in the choice of energy equation (Experiment 2), in input errors (Experiment 5) and in the correctness of the diets and thresholds themselves, which can only be assessed clinically. None of these uncertainties is reduced by adding a learning model.

Machine learning has a legitimate place in this field, but in a different problem: when the label is derived from a real outcome. For example, if weight change after three months on a diet were recorded for a group of users, a model could learn which diet is most effective for whom. Or, if dietitians independently selected a diet for each individual

without seeing the system's output, the rule could be compared with their decisions, or a model could be trained to imitate clinical judgement that does not fit a simple formula. In both cases the data must be generated independently of the system's rule.

6.2. Limitations

No clinical evaluation: this paper assesses the correctness of the calculations and the behaviour of the system, not the effect of the diets on health. The recommendations of the earlier prototype were said to have been checked by a specialist, but the method, number of cases and outcome of that check were not documented; a protocol for a clinical evaluation is therefore proposed for future work (Section 6.3).

Population: the NHANES data come from US adults. The distribution of BMI, the prevalence of diabetes and dyslipidaemia and the pattern of activity differ in the Indian population, and the energy equations may have different accuracy in Asian populations. Evaluation on Indian data, for example from the National Family Health Survey (NFHS), is necessary.

Survey weights: the analyses did not use the survey design weights, so the percentages describe the sample and are not exact population estimates.

Pregnancy and lactation: this version of the data has no variable for current pregnancy. Pregnant and lactating women have different energy requirements and must be excluded from the scope of the system or referred.

Activity level: the mapping of questionnaire answers to four activity levels is the authors' assumption; the sensitivity of the results to it (18.9% of diets change with an alternative mapping) is reported.

Ideal-weight table of the prototype: the table of desirable BMI by age in the prototype documentation was illegible, and the common table was used to reconstruct the prototype's calculation; the exact values of Experiment 3 may therefore differ slightly from the original method, but the qualitative conclusions (the reversed effect of ideal weight and the effect of diet 13) do not depend on this table.

Diet content: the food content of the twelve diets is not evaluated in this paper; the system is responsible only for selecting the correct energy range.

6.3. Proposed protocol for clinical evaluation

The following design is proposed for the clinical evaluation of the system: (1) at least two dietitians, independently and without seeing the system's output, determine the target energy and appropriate diet for a random sample of at least 200 individuals with complete profiles; (2) agreement between the two dietitians and between each dietitian and the system is reported with weighted kappa and, for energy, with Bland–Altman plots; (3) cases of disagreement are reviewed in a joint session and their causes classified (equation, adjustment, referral rule, or a characteristic not in the profile); and (4) in a short follow-up study, weight change and adherence to the recommended diet are recorded. The data from step (4) are exactly the independent data that make training a learning model meaningful.

7. CONCLUSIONS

A knowledge-based diet recommender with validated equations and an explicit safety rule was implemented and evaluated on real data from 4,785 adults in the NHANES survey. The results show that (1) the system recommends a diet matched to energy requirements for 55.4% of adults and refers 44.6% to a specialist because of diabetes or dyslipidaemia; (2) the choice of basal-metabolism equation changes the diet of more than half of the individuals and must be made consciously and on the basis of evidence, the Mifflin–St Jeor equation being the recommended choice; (3) the calculation of the earlier prototype, because of the use of ideal weight, an incorrect activity factor and a single diet for overweight and disease, led to decisions that differed greatly from those of the correct method; (4) machine-learning models, including neural networks, cannot add value in this problem and only approximate the rule with error; and (5) errors in self-reported inputs change the diet of about one quarter of individuals, which the design of the ranges must take into account. The next steps are a clinical evaluation of the system with the proposed protocol,

an evaluation on data from the Indian population, and the collection of outcome data that would make machine-learning methods useful for predicting the effectiveness of diets.

Data and code availability

The NHANES data are freely available from the website of the US National Center for Health Statistics and in the NHANES package [18]. The complete code of the system, the unit tests and the scripts that reproduce all tables and figures are provided as supplementary material (Appendix A).

Ethics

The study used only public, de-identified NHANES data; NHANES protocols were approved by the NCHS Research Ethics Review Board, and no additional approval was required.

Conflict of interest

The authors declare no conflict of interest.

Funding

This research received no specific grant from any funding agency.

Use of AI tools

An AI-based tool was used to translate and edit the language of the manuscript; the authors reviewed the text and take full responsibility for it.

Appendix A. Code Structure

dietrec.py: BMI, the three BMR equations, activity factors, energy adjustment with the safety floor, diet selection and the referral rule.

test_dietrec.py: unit tests against hand-calculated values.

experiments.py: preparation of the NHANES data and Experiments 1–5; all results are stored in results.json.

figs.py: generation of the figures from the stored results.

A full run of the experiments takes about two minutes on an ordinary computer. All random seeds are fixed, and the results are exactly reproducible.

REFERENCES

[1] Rajamani, K., & Sheela, D. (2018). Data mining techniques and algorithms in cloud environment. *International Journal of Pure and Applied Mathematics, 119*(7), 599–602.

[2] Liang, T.-P., Lai, H.-J., & Ku, Y.-C. (2006). Personalized content recommendation and user satisfaction: Theoretical synthesis and empirical findings. *Journal of Management Information Systems, 23*(3), 45–70. https://doi.org/10.2753/MIS0742-1222230303

[3] Ricci, F., Rokach, L., & Shapira, B. (2015). Recommender systems: Introduction and challenges. In F. Ricci, L. Rokach, & B. Shapira (Eds.), *Recommender systems handbook* (2nd ed., pp. 1–34). Springer.

https://doi.org/10.1007/978-1-4899-7637-6_1

[4] Adomavicius, G., Sankaranarayanan, R., Sen, S., & Tuzhilin, A. (2005). Incorporating contextual information in recommender systems using a multidimensional approach. *ACM Transactions on Information Systems, 23*(1), 103–145. https://doi.org/10.1145/1055709.1055714

[5] Melville, P., & Sindhwani, V. (2011). Recommender systems. In C. Sammut & G. I. Webb (Eds.), *Encyclopedia of machine learning* (pp. 829–838). Springer. https://doi.org/10.1007/978-0-387-30164-8_705

[6] Lee, C.-S., Wang, M.-H., & Hagras, H. (2010). A type-2 fuzzy ontology and its application to personal diabetic-diet recommendation. *IEEE Transactions on Fuzzy Systems, 18*(2), 374–395.

[7] Ramachandran, S., Singh, H., & Bajaj, K. K. (1992). Nutrition diet programme – An expert system. In *AAAI Fall Symposium Series* (Technical Report FS-92-01). AAAI Press.

[8] Chen, Y., Hsu, C.-Y., Liu, L., & Yang, S. (2012). Constructing a nutrition diagnosis expert system. *Expert Systems with Applications, 39*(2), 2132–2156.

[9] Krbez, J. M., & Shaout, A. (2013). Fuzzy nutrition system. *International Journal of Innovative Research in Computer and Communication Engineering, 1*(7), 1–12.

[10] Heinonen, P., Mannelin, M., Iskala, H., & Juuso, E. (2009). Development of a fuzzy expert system for a nutritional guidance application. In *Proceedings of the Joint 2009 International Fuzzy Systems Association World Congress and 2009 European Society of Fuzzy Logic and Technology Conference (IFSA-EUSFLAT 2009)*. Lisbon, Portugal.

[11] Iwendi, C., Khan, S., Anajemba, J. H., Bashir, A. K., & Noor, F. (2020). Realizing an efficient IoMT-assisted patient diet recommendation system through machine learning model. *IEEE Access, 8*, 28462–28474.

[12] Harris, J. A., & Benedict, F. G. (1918). A biometric study of human basal metabolism. *Proceedings of the National Academy of Sciences, 4*(12), 370–373. https://doi.org/10.1073/pnas.4.12.370

[13] Roza, A. M., & Shizgal, H. M. (1984). The Harris Benedict equation reevaluated: Resting energy requirements and the body cell mass. *American Journal of Clinical Nutrition, 40*(1), 168–182. https://doi.org/10.1093/ajcn/40.1.168

[14] Mifflin, M. D., St Jeor, S. T., Hill, L. A., Scott, B. J., Daugherty, S. A., & Koh, Y. O. (1990). A new predictive equation for resting energy expenditure in healthy individuals. *American Journal of Clinical Nutrition, 51*(2), 241–247. https://doi.org/10.1093/ajcn/51.2.241

[15] Frankenfield, D., Roth-Yousey, L., & Compher, C. (2005). Comparison of predictive equations for resting metabolic rate in healthy nonobese and obese adults: A systematic review. *Journal of the American Dietetic Association, 105*(5), 775–789. https://doi.org/10.1016/j.jada.2005.02.005

[16] World Health Organization. (2000). *Obesity: Preventing and managing the global epidemic* (WHO Technical Report Series No. 894). World Health Organization.

[17] Centers for Disease Control and Prevention, National Center for Health Statistics. (2009–2012). *National Health and Nutrition Examination Survey data, 2009–2010 and 2011–2012* [Data set]. U.S. Department of Health and Human Services.

[18] Pruim, R. (2015). *NHANES: Data from the US National Health and Nutrition Examination Study* (Version 2.1.0) [R package]. Project MOSAIC. https://github.com/ProjectMOSAIC/NHANES

[19] FAO/WHO/UNU. (2004). *Human energy requirements: Report of a joint FAO/WHO/UNU expert consultation* (FAO Food and Nutrition Technical Report Series No. 1). Food and Agriculture Organization of the United Nations.

[20] Trattner, C., & Elsweiler, D. (2017). *Food recommender systems: Important contributions, challenges and future research directions* (arXiv:1711.02760). arXiv. https://doi.org/10.48550/arXiv.1711.02760

[21] Tran, T. N. T., Atas, M., Felfernig, A., & Stettinger, M. (2018). An overview of recommender systems in the healthy food domain. *Journal of Intelligent Information Systems, 50*(3), 501–526. https://doi.org/10.1007/s10844-017-0469-0

[22] Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., & Duchesnay, É. (2011). Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research, 12*, 2825–2830.

[23] American Diabetes Association. (2024). Standards of care in diabetes—2024. *Diabetes Care, 47*(Suppl. 1).

[24] National Cholesterol Education Program (NCEP) Expert Panel on Detection, Evaluation, and Treatment of High Blood Cholesterol in Adults. (2002). Third report of the National Cholesterol Education Program (NCEP) Expert Panel on Detection, Evaluation, and Treatment of High Blood Cholesterol in Adults (Adult Treatment Panel III) final report. *Circulation, 106*(25), 3143–3421. https://doi.org/10.1161/circ.106.25.3143

[25] Jensen, M. D., Ryan, D. H., Apovian, C. M., Ard, J. D., Comuzzie, A. G., Donato, K. A., Hu, F. B., Hubbard, V. S., Jakicic, J. M., Kushner, R. F., Loria, C. M., Millen, B. E., Nonas, C. A., Pi-Sunyer, F. X., Stevens, J., Stevens, V. J., Wadden, T. A., Wolfe, B. M., & Yanovski, S. Z. (2014). 2013 AHA/ACC/TOS guideline for the management of overweight and obesity in adults. *Circulation, 129*(25, Suppl. 2), S102–S138. https://doi.org/10.1161/01.cir.0000437739.71477.ee

[26] Connor Gorber, S., Tremblay, M., Moher, D., & Gorber, B. (2007). A comparison of direct vs. self-report measures for assessing height, weight and body mass index: A systematic review. *Obesity Reviews, 8*(4), 307–326. https://doi.org/10.1111/j.1467-789X.2007.00347.x

[27] National Research Council. (1989). *Diet and health: Implications for reducing chronic disease risk*. National Academy Press.

[28] Kohavi, R. (1995). A study of cross-validation and bootstrap for accuracy estimation and model selection. In *Proceedings of the 14th International Joint Conference on Artificial Intelligence (IJCAI)* (Vol. 2, pp. 1137–1143). Morgan Kaufmann.

[29] Cohen, J. (1968). Weighted kappa: Nominal scale agreement with provision for scaled disagreement or partial credit. *Psychological Bulletin, 70*(4), 213–220. https://doi.org/10.1037/h0026256

[30] Rudin, C. (2019). Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. *Nature Machine Intelligence, 1*, 206–215. https://doi.org/10.1038/s42256-019-0048-x