European Journal of Clinical Nutrition www.nature.com/ejcn

ARTICLE OPEN

Check for updates

Predicting non-responders to lifestyle intervention in prediabetes: a machine learning approach

Andrea Foppiani[1,2]✉, Ramona De Amicis[1,3], Alessandro Leone[1,2], Federica Sileo[1,2], Sara Paola Mambrini[1,4], Francesca Menichetti[1], Giorgia Pozzi[1], Simona Bertoli[1,3] and Alberto Battezzati[1,2]

BACKGROUND: The clinical care process for people with prediabetes starts with lifestyle intervention, often escalating to more intense treatment due to the low success rate of the first-line intervention. Clinicians lack clear guidelines on which patients would benefit from early treatment with more intensive therapeutic options, so we aimed to develop an algorithm to early identify non-responders to lifestyle intervention for prediabetes.
METHOD: Several statistical and machine learning algorithms were screened with internal cross-validation on the basis of accuracy and discrimination ability to correctly classify patients that would fail to normalize fasting glycemia within one year of being prescribed a lifestyle intervention, solely based on the first examination measurements.
RESULT: Of the many screened algorithm, only a random forest model performed with sufficient accuracy to exceed the historical failure rate of patients within our center, with an accuracy of 0.689 (CI 0.669, 0.710) and an AUROC of 0.687 (CI 0.673, 0.701).
CONCLUSIONS: This study showcases the ability of machine learning models to provide useful insight in clinical practice leveraging knowledge contained in routinely collected data.

European Journal of Clinical Nutrition; https://doi.org/10.1038/s41430-024-01495-9

INTRODUCTION

People with prediabetes have abnormal glucose metabolism while not meeting the criteria for diabetes [1]. As they are often characterized by abdominal obesity, screening is recommended for all adults with a body mass index indicative of overweight or obesity and with one or more risk factors. Diagnosis is made with fasting glucose levels between 100 and 125 mg/dL and/or and 2 h glucose levels between 140 to 199 mg/dL during a 75 g oral glucose tolerance test [2].

After diagnosis, therapy focuses on normalization of fasting glycemia through body weight management, physical activity, and/or hypoglycemic medications. There are no clear guidelines between choosing lifestyle behavior change or pharmacological interventions and the intensity of treatment is often escalated once a more conservative approach has proven to be not conclusive [3].

Big data and artificial intelligence may provide insight in situations where guidelines lack a clear course of action. Leveraging data and outcomes collected in everyday clinical practice with traditional or more recent tools may inform the clinician of the probability of success of alternative treatments based on site- and population-specific historical success rate. Also, combining the formalized knowledge contained in guidelines with "learnt context-specific knowledge" may constitute a promising strategy to deal with transparency and explainability issues arising with the use of new artificial intelligence algorithms [4].

Considering the lack of clear guidelines and the high rate of failure of more conservative approaches, we aimed to develop an algorithm to early identify non-responders to lifestyle intervention for prediabetes.

METHODS

Sources of data and participants

The database used for developing the predictive algorithm was the International Center for the Assessment of Nutritional Status (ICANS, University of Milan, Milan, Italy) database, which contains data of a large ongoing open-cohort nutritional study. As part of the protocol of the study, all patients at baseline receive a full nutritional assessment, they are prescribed a lifestyle intervention and eventually also a pharmacological intervention, and a follow-up examination is scheduled. At follow-up, a more limited number of parameters are routinely collected to evaluate changes in weight, body composition, and laboratory exams. For the development of the algorithm, all patients with prediabetes enrolled between 2009 and the beginning of 2019 were included. The complete database include 18.973 baseline observations, and a total of 45.148 follow-up observations. In this study we have included a total of 59 variables from the database.

Patients included in this study were self-referring patients seeking a weight loss program, mainly resident in Milan or nearby cities, with a new or recent diagnosis of prediabetes. Eligibility criteria were: age ≥18 years; not pregnant and not nursing; no condition severely limiting movements and physical activity; no severe cardiovascular, neurological, endocrine, or psychiatric disorder; prescribed only a lifestyle intervention. The lifestyle

[1]International Center for the Assessment of Nutritional Status and the Development of Dietary Intervention Strategies (ICANS-DIS), Department of Food, Environmental and Nutritional Sciences (DeFENS), University of Milan, 20133 Milan, Italy. [2]IRCCS Istituto Auxologico Italiano, Clinical Nutrition Unit, Department of Endocrine and Metabolic Medicine, 20100 Milan, Italy. [3]IRCCS Istituto Auxologico Italiano, Obesity Unit and Laboratory of Nutrition and Obesity Research, Department of Endocrine and Metabolic Diseases, 20145 Milan, Italy. [4]IRCCS Istituto Auxologico Italiano, Laboratory of Metabolic Research, San Giuseppe Hospital, 28824 Piancavallo, Italy. ✉email: andrea.foppiani@unimi.it

Received: 3 April 2024 Revised: 7 August 2024 Accepted: 12 August 2024
Published online: 23 October 2024

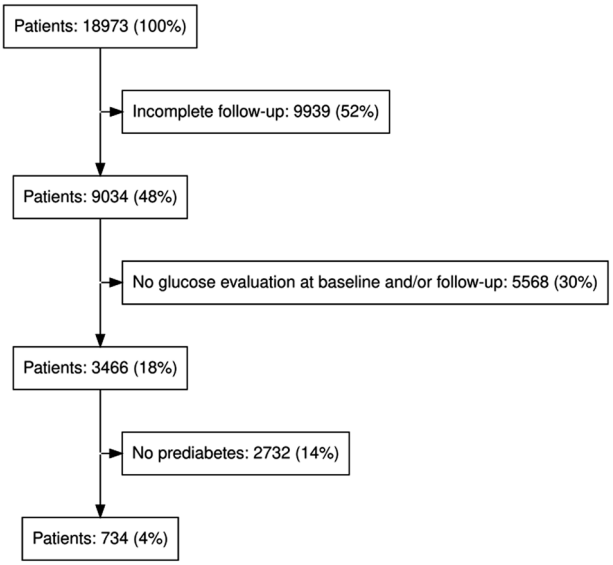

Fig. 1 Flow diagram for study participants.

intervention consisted of an hypocaloric omnivorous diet, with macro- and micronutrient levels set according to the Italian recommended daily allowances [5], and with a Mediterranean pattern; physical activity recommendation were also provided according to the WHO physical activity guidelines [6].

The study complied with the principles established by the Declaration of Helsinki, and written informed consent was obtained by each subject. The ethical committee of the University of Milan (n. 6/2019) approved the study procedures.

Outcome and predictors

The outcome was normalization of glycemia within 1 year of starting the lifestyle intervention (dichotomous, fasting glucose <100 mg/dL).

A total of 59 predictor variables were used in the analysis:

- demographic data: age, sex, education, occupation, marital status
- anthropometry: height, weight, arm length, arm circumference, wrist circumference, waist circumference, biceps skinfold, triceps skinfold, subscapular skinfold, suprailiac skinfold, arm muscle area, arm fat area, body density, fat mass, fat free mass
- bioimpedance analysis: intracellular water, extracellular water
- abdominal ultrasound: sternum subcutaneous adipose tissue, sternum visceral adipose tissue, abdomen subcutaneous adipose tissue, abdomen visceral adipose tissue
- indirect calorimetry: oxygen consumption, carbon dioxide production, respiratory quotient, resting energy expenditure
- medical history: family status, menstruation, pregnancies, diet status, diet history, physical activity, smoking, pharmacological treatments, clinical signs, weight history
- vital signs: heart rate, systolic pressure, diastolic_pressure
- blood and urine exams: white blood cell count, red blood cell count, hemoglobin, mean corpuscular volume, glucose, total cholesterol, HDL cholesterol, LDL cholesterol, triglycerides, glutamic-pyruvic transaminase, glutamic-oxaloacetic transaminase, gamma-glutamyl transferase, thyroid stimulating hormone, creatinine, uric acid, urea

Statistical and machine learning analysis methods

All eligible patients at time of study were included, determining the sample size (no a priori calculations were made).

For algorithms requiring complete data, we imputed missing data in the pre-processing phase using k-nearest neighbors imputation (Gower's distance, number of neighbors = 5).

Maximum predictive strength was sought through optimization of the correct classification fraction (CCF) and the receiver operating characteristic area under the curve (AUROC). Between accuracy and discrimination ability, accuracy was selected as the most relevant metric in the clinical settings (ie. maximization of the CCF).

Several statistical and machine learning models were compared using 10-fold cross-validation resampling. For models requiring tuning parameters, a grid made of several combinations of tuning parameters was tested via 10-fold cross-validation.

Prior to model selection, per-model preprocessing steps were defined in order to guarantee the best predictive ability for the specific model. To capture uncertainty about non-deterministic data manipulation, all preprocessing steps were repeated in each cross-validation fold.

Principal component analysis (PCA) was employed as an optional preprocessing step aimed to reduce the dimensionality of the dataset. In these cases, PCA was used to transform the set of predictors in a reduced number of predictors designed to capture the maximum amount of information in the original variables. A potential benefit of this approach, other than the dimensionality reduction, is the production of statistically independent predictors that can ameliorate the problem of inter-variables correlations in the dataset.

The following models were evaluated :

- *logistic regression*
- *linear discriminant analysis*
- *quadratic discriminant analysis*
- *naive Bayes*, tuned for kernel smoothness, and Laplace correction
- *K-nearest neighbour*, tuned for number of nearest neighbors, and distance weighting function, Minkowski distance order
- *ridge regression and LASSO*, tuned for the amount of regularization, and the proportion of LASSO penalty
- *decision trees*, tuned for tree depth, minimal node size, and cost-complexity parameter
- *bagged trees*, tuned for the cost/complexity parameter used by CART models, the maximum depth of a tree, the minimum number of data points in a node that are required for the node to be split further, and a cost value to assign to the class corresponding to the first factor level
- *random forest*, tuned for number randomly selected predictors, number of trees, and minimal node size
- *boosted trees*, tuned for tree depth, the number trees, the learning rate, the number randomly selected predictors, the minimal node size, the minimum loss reduction, the proportion observations sampled, and the number iterations before stopping
- *linear support vector machine*, tuned for cost, and insensitivity margin
- *single layer neural network*, tuned for the number of hidden units, the amount of regularization, and the number of epochs

Sensitivity, specificity, positive predictive value, and negative predictive value, were calculated for the best model as: Sensitivity $=TP/(TP+FN)$, Specificity $=TN/(TN+FP)$, positive predictive value $=TP/(TP+FP)$, negative predictive value $=TN/(TN+FN)$, where FN, false negative; FP, false positive; TN, true negative; TP, true positive.

All statistical analysis was performed with R 4.1.1 [7]. Model preprocessing, tuning, resampling, and fitting were performed with the addition of the Tidymodels package to R (for algorithm-specific packages see the appendix).

RESULTS

Participants

A total of 734 patients were selected for this study, Fig. 1 reports a complete flow diagram for study participants.

Table 1 reports patients characteristics in the overall sample, and by sex. The historical fraction of patient glycemia within 1 year of starting the lifestyle intervention in this sample was 0.68.

Model development and screening

Model screening results are shown in Fig. 2. Machine learning models based on decision trees (in particular boosted trees and random forest) were the best performing models, producing models with both relatively high CCF and AUROC. PCA was generally not useful in improving the predictive ability of these models. Comparing the accuracy results with the historical probability of event, random forest models were the only ones able to exceed a naive classifier that assumes that all patients would experience the event.

Model performance

The model reached an accuracy of 0.689 (CI 0.669, 0.710) and an AUROC of 0.687 (CI 0.673, 0.701). The sensitivity and specificity

Table 1. Patients characteristics at baseline.

Characteristic	N	Overall, N = 734[a]	Responder, N = 235[a]	Non responder, N = 499[a]
Age (years)	734	59 (51, 66)	58 (49, 65)	59 (53, 67)
Sex	734			
Female		376 (51%)	118 (50%)	258 (52%)
Male		358 (49%)	117 (50%)	241 (48%)
Married	734	492 (67%)	155 (66%)	337 (68%)
Working	734	368 (50%)	125 (53%)	243 (49%)
With higher degree	734	232 (32%)	78 (33%)	154 (31%)
In menopause	376	253 (67%)	72 (61%)	181 (70%)
Previous diets	734	412 (56%)	130 (55%)	282 (57%)
Physically active	734	210 (29%)	74 (31%)	136 (27%)
Smoker	734	386 (53%)	127 (54%)	259 (52%)
Familiarity with diabetes	734	216 (29%)	67 (29%)	149 (30%)
Weight (kg)	733	91 (80, 104)	89 (79, 103)	92 (80, 104)
Height (cm)	733	167 (159, 174)	166 (160, 175)	167 (159, 173)
Body mass index (kg/m²)	733	32.7 (29.3, 36.6)	32.4 (28.4, 35.7)	32.8 (29.6, 37.2)
Waist circumference (cm)	733	110 (101, 120)	108 (98, 119)	110 (102, 120)
Arm muscle area (cm²)	715	59 (47, 69)	58 (46, 69)	59 (48, 69)
Body fat fraction of total body weight, as %	633	41.3 (36.7, 45.1)	40.7 (36.2, 44.4)	41.6 (37.1, 45.5)
Total body water (L)	628	40 (34, 49)	41 (34, 49)	40 (34, 48)
Extra-cellular water (L)	628	15.5 (12.8, 18.5)	15.6 (12.9, 18.6)	15.4 (12.8, 18.4)
Intra-cellular water (L)	628	24.9 (20.6, 30.1)	24.9 (20.7, 30.4)	24.9 (20.6, 29.8)
Abdomen subcutaneous fat thickness (cm)	609	2.52 (1.81, 3.34)	2.63 (1.90, 3.29)	2.43 (1.80, 3.35)
Abdomen visceral fat thickness (cm)	604	8.47 (6.25, 10.33)	7.93 (5.41, 10.01)	8.63 (6.44, 10.39)
Respiratory quotient	715	0.83 (0.79, 0.87)	0.83 (0.79, 0.88)	0.83 (0.79, 0.87)
Resting energy expenditure (kcal/die)	715	1623 (1410, 1845)	1591 (1402, 1849)	1630 (1419, 1838)
Glucose (mg/dL)	734	120 (114, 132)	116 (112, 122)	124 (116, 140)
Total cholesterol (mg/dL)	731	214 (186, 241)	214 (190, 240)	214 (184, 241)
HDL cholesterol (mg/dL)	729	50 (43, 61)	53 (45, 64)	49 (41, 59)
LDL cholesterol (mg/dL)	680	140 (116, 166)	138 (118, 166)	141 (116, 167)
Triglycerides (mg/dL)	730	125 (92, 173)	115 (85, 155)	132 (98, 181)
Aspartate transaminase (U/L)	722	21 (18, 26)	21 (18, 27)	22 (17, 26)
Alanine transaminase (U/L)	723	26 (19, 37)	26 (18, 38)	26 (19, 36)
Gamma-glutamyltransferase (U/L)	697	27 (19, 44)	27 (19, 43)	27 (19, 44)
Thyroid stimulating hormone (mUI/L)	700	1.89 (1.31, 2.71)	1.91 (1.36, 2.79)	1.88 (1.30, 2.68)

[a]Median (IQR); n (%).

values for the best model were 0.877 (CI 0.851, 0.903) and 0.28 (CI 0.203, 0.357), while the positive and negative predictive values of the best model were 0.721 (CI 0.7, 0.743) and 0.514 (CI 0.433, 0.596). Variable importance analysis is presented in Figure Fig. 3 and Fig. 4.

DISCUSSION

In this paper we aimed to identify the highest predictive algorithm for early detection of non-responders to lifestyle intervention among patients with prediabetes. After screening several statistical and machine learning algorithms, we found that a machine learning algorithm, and specifically a random forest model, performed best in internal validation.

Christodoulou et al. [8] compared performance of traditional (logistic regression) methods to machine learning for clinical prediction models and found no evidence of superior performance of machine learning over logistic regression. In this study logistic regression ranked 6th and 4th in the accuracy and discrimination ranking respectively. Moreover, the best model obtained an accuracy of 0.689 (CI 0.669, 0.710), which exceeded the historical rate of failure within our center. This is a fundamental and unique characteristic of the best model, and a necessary condition for its usefulness in clinical practice. An algorithm not exceeding the historical rate of failure would not beat a naive classifier that would assume that all patients would experience the event, and as such would not be more accurate then assigning by the default all patients to a more intensive treatment scheme. Only an algorithm that is more accurate than this naive classifier has the capability of improving the clinical success rate while conserving resources.

The best model obtained an AUROC of 0.687 (CI 0.673, 0.701), and while this shows only a moderate discrimination ability, the majority of non-responders were correctly detected and may have been treated more intensively from the beginning, possibly lowering the probability of failure to normalize fasting glucose. The algorithm also shows a tendency to include almost half of true responders among non-responders. While this may not be

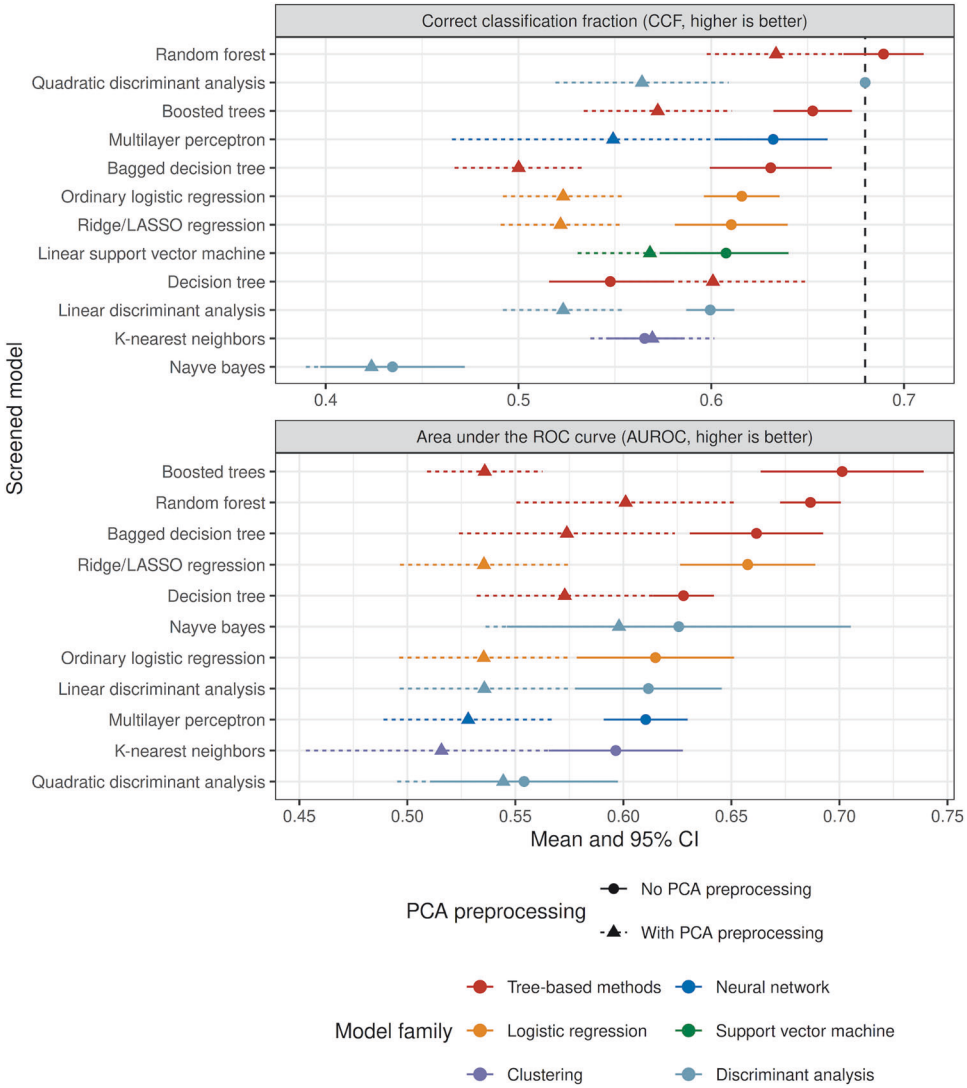

Fig. 2 Comparison of accuracy (correct classification fraction, CCF) and discrimination ability (area under the ROC curve, ROC AUC) of statistical and machine learning models in the prediction of failure to improve fasting glycemia. For each model and metric, mean and confidence bounds (95% confidence) across resamples were computed. For each model, an alternative with and without principal component analysis (PCA, unsupervised learning) is shown. The historical rate of patients experiencing the event is marked with a vertical dashed line.

clinically an issue, as in the context of prediabetes more intensive treatments are usually safe for the patient, it may represent an issue from an economic perspective, as resources are allocated where they are probably not needed. Considering that true responders are the minority of overall patients, the slight increase in effort for false non-responders may be worthwhile.

This study showcases the capabilities of machine learning models when used in everyday clinical practice. It is worth reiterating that the accuracy considerations made above single out machine learning models as the only ones that may prove useful in this context. On the other hand, the necessary increase in accuracy comes at the cost of explainability of the results. Machine learning models are notorious for working as a "black box", generating outputs without explaining how it arrived at those outputs. In this case, the best model was a machine learning model of the "decision trees and tree ensembles" family, that are among the best candidates to provide some explanation of their internal working, giving at least the possibility to quantify the importance of each variable included in the model. Here "importance" is a measure of improvement (eg. in accuracy) when the variable is included in calculation of the output. While such a measure of "importance" can provide some insight in which variables may explain most of the prediction, a direct inference is not possible due to the nature of the study. The fact that biochemical parameters represented the most important domain in the variable importance ranking may reflect the importance of phenotyping patients with prediabetes beyond glucose tolerance status. Indeed the blood lipid profile and liver enzymes were among the most important variables and may have signaled disease progression and co-morbidities to the model.

The clinician role remains paramount in the deployment of these algorithms in clinical practice, as they should be viewed as one of the tools that the clinician uses to make the decisions that actually form the therapeutic plan, and not as decision makers themself. On the other hand, these algorithms permit to leverage the data collected in everyday practice, synthesizing context-specific expertise accrued in years of clinical practice.

The study has the following limitations. The algorithm uses a set of predictors that are specific to the clinical practice of the ICANS center and this limits the generalizability of the algorithm. Different interventions were not evaluated. The algorithm does

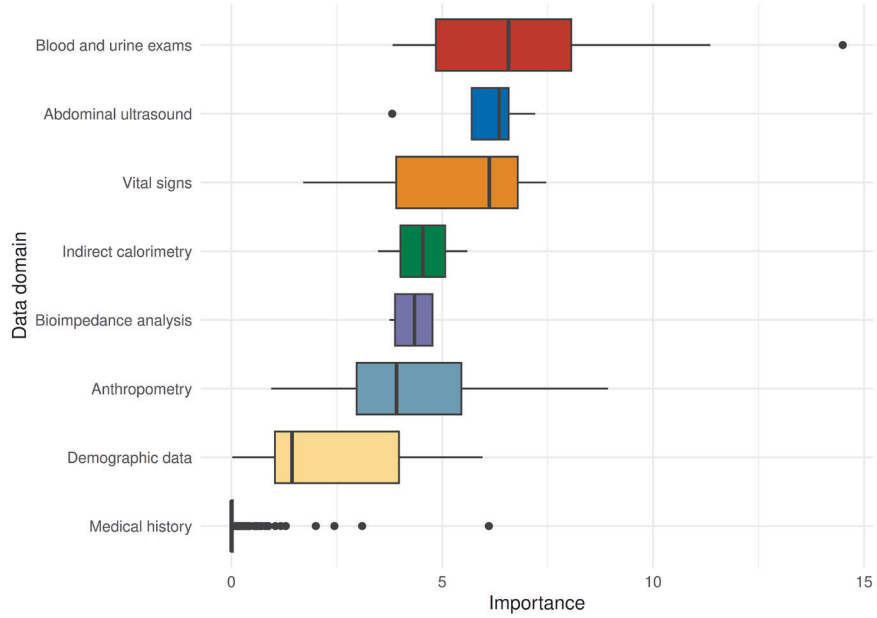

Fig. 3 Relative importance of predictor variables by data domain.

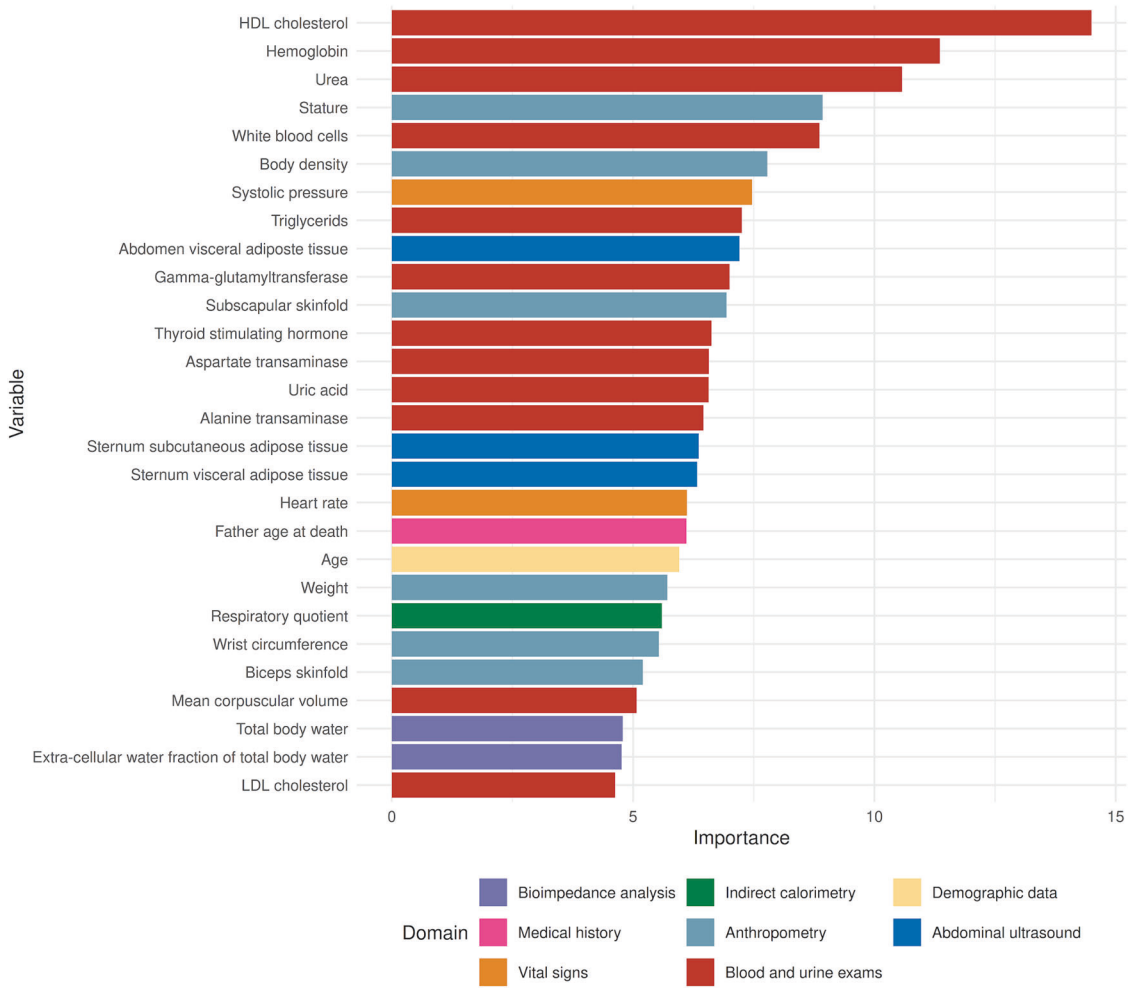

Fig. 4 Relative importance of predictor variables (showing only predictors with an importance values >95th percentile).

not take into account loss to follow-up as an outcome. Self-selection bias could have influenced the demographic characteristics and the probability of success in our sample.

In conclusions we show that machine learning models have the potential to predict non-responders to lifestyle intervention in patients with prediabetes, with an accuracy sufficient to result useful in clinical practice. Validation on new diagnosis and complementary algorithms for different outcomes are warranted.

DATA AVAILABILITY

The datasets generated during and/or analysed during the current study are available from the corresponding author on reasonable request.

REFERENCES

1. ElSayed NA, Aleppo G, Aroda VR, Bannuru RR, Brown FM, Bruemmer D, et al. 3. Prevention or Delay of Diabetes and Associated Comorbidities: standards of Care in Diabetes 2023. Diabetes Care. 2022;46:S41–48. https://doi.org/10.2337/dc23-s003
2. Chadha C, Pittas AG, Lary CW, Knowler WC, Chatterjee R, Phillips LS, et al. Reproducibility of a Prediabetes Classification in a Contemporary Population. Metab Open. 2020;6:100031. https://doi.org/10.1016/j.metop.2020.100031
3. Diabetes Prevention Program Research Group. Reduction in the Incidence of Type 2 Diabetes with Lifestyle Intervention or Metformin. N Engl J Med. 2002;346:393–403. https://doi.org/10.1056/nejmoa012512
4. Montani S, Striani M. Artificial Intelligence in Clinical Decision Support: A Focused Literature Survey. Yearb Med Inf. 2019;28:120–27. https://doi.org/10.1055/s-0039-1677911
5. Società Italiana di Nutrizione Umana. 2014. *LARN: Livelli Di Assunzione Di Riferimento Di Nutrienti Ed Energia Per La Popolazione Italiana*. Milan, MI, Italy: Società Italiana di Comunicazione Scientifica e Sanitaria.
6. World Health Organization. 2020. *WHO Guidelines on Physical Activity and Sedentary Behaviour*. Geneva: World Health Organization.
7. R Core Team. 2023. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. https://www.R-project.org/.
8. Christodoulou E, Ma J, Collins GS, Steyerberg EW, Verbakel JY, Van Calster B. A Systematic Review Shows No Performance Benefit of Machine Learning over Logistic Regression for Clinical Prediction Models. J Clin Epidemiol. 2019;110:12–22. https://doi.org/10.1016/j.jclinepi.2019.02.004

ACKNOWLEDGEMENTS

Project funded under the National Recovery and Resilience Plan (NRRP), Mission 4 Component 2 Investment 1.3 - Call for proposals No. 341 of 15 March 2022 of Italian Ministry of University and Research funded by the European Union – NextGenerationEU; Award Number: Project code PE00000003, Concession Decree No. 1550 of 11 October 2022 adopted by the Italian Ministry of University and Research, CUP D93C22000890001, Project title "ON Foods - Research and innovation network on food and nutrition Sustainability, Safety and Security – Working ON Foods.

AUTHOR CONTRIBUTIONS

AF conceived, drafted, and performed statistical analysis; RDA interpreted results and revised the manuscript; AL interpreted results and revised the manuscript; FS interpreted results and revised the manuscript; SPM acquired data and revised the manuscript; FM acquired data and revised the manuscript; GP acquired data and revised the manuscript; SB conceived and revised the manuscript; AB conceived and revised the manuscript.

COMPETING INTERESTS

The authors declare no competing interests.

ADDITIONAL INFORMATION

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41430-024-01495-9.

Correspondence and requests for materials should be addressed to Andrea Foppiani.

Reprints and permission information is available at http://www.nature.com/reprints

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