Current Research in Food Science 7 (2023) 100544

Contents lists available at ScienceDirect

Current Research in Food Science

journal homepage: www.sciencedirect.com/journal/current-research-in-food-science

Short Communication

Using machine learning models to predict the quality of plant-based foods

Christabel Tachie [a], Nii Adjetey Tawiah [b], Alberta N.A. Aryee [a,*]

[a] Delaware State University, College of Agriculture, Science and Technology, Department of Human Ecology (Food Science and Biotechnology Program), 1200 N DuPont Highway, Dover, DE, 19901, USA

[b] Delaware State University, College of Humanities, Education and Social Sciences, Department of Sociology and Criminal Justice, 1200 N DuPont Highway, Dover, DE, 19901, USA

ARTICLE INFO

Handling Editor: Professor A.G. Marangoni

Keywords:
Nutri-Score
Plant-based foods
Machine learning
FNDDS data
Data exploration

ABSTRACT

Plant-based foods (PBFs) are considered healthy, especially, minimally processed whole foods, fruits, whole grains, and legumes while highly processed PBFs maybe less nutritious. Educating consumers on nurient quality will help to guide their choices. This study was aimed at estimating and predicting the nutrient quality of PBFs using their Nutri-Score and micronutrient content. The NHANES (2017–2020) data shows the output for foods consumed in the US and their nutrient composition based on a 24-h recall. Though the Nutri-Score label has been used to discriminate food quality, it still needs to be implemented in most countries. It computes mostly macronutrients with less consideration for micronutrients which also contributes to product quality. ML methods used in this study combine the Nutri-Score grade and micronutrient content in predicting food quality. The FNDDS data of PBFs for 2017–2020 were split into training (n = 300) and testing (n = 74) datasets. Eight ML models were used to predict the Nutri-Score and the Nutri-Score grade of PBFs. Random forest (RF) and light gradient boost model (LightGBM) performed best with accuracy and coefficient of determination (R^2) scores of 0.88 and 0.96, respectively, while DT had the least scores in predicting the Nutri-Score grade (0.81) and Nutri-Score (0.93). These results suggest that ML can be effectively leveraged to predict PBFs quality.

1. Introduction

Generally, plant-based foods (PBFs) are rich in nutrients such as fiber, sources of antioxidants and low in saturated fats and cholesterol compared to animal-source foods reducing the risk of developing cardiovascular diseases (Kahleova et al., 2017). The negative impact associated with the consumption of animal-source foods have driven consumers to accept PBFs (Tachie et al., 2023). For instance, lactose intolerance and allergenic reactions, the risk of contamination by pathogenic microbes, and the association of animal protein with increased risk of cancer have raised the demand for plant-based alternatives (Vanga and Raghavan, 2018). However, the quality of PBFs can vary depending on factors such as nutrient composition and content, processing, and preparation methods (Tachie et al., 2023). For instance, highly processed PBFs like veggie burgers and other meat alternatives may contain added salt, sugar, and synthetic preservatives reducing their overall nutritional value (Pratt, 2020; Tachie et al., 2023). Conversely, PBFs like fresh fruits, vegetables, and whole grains tend to be more nutritious.

Conventional methods, such as the the back-of-package labels (BOPLs), commonly contain numerical information of the amount of energy and specific nutrients that describe the nutritional makeup of the product. The nutrition facts panel on BOPLs was aimed at aiding consumers in making wiser food selections. However, consumers, particularly those in vulnerable communities, find the nutrition data panel on the back of packages challenging to read and comprehend, and often overlook. The labeling formats can be broadly classified as nutrient-specific and summary (Ducrot et al., 2022; Hawley et al., 2013). The latter provides a general assessment of the product's nutritional quality and is more straightforward for consumers to understand and utilize. The Nutri-Score falls under this category; it assigns a score to the nutritional value of foods using a five-color graded scale that is paired with letters to make it easier to understand (from A for "better nutritional quality" to E for "lower nutritional quality") (Sante Publique, 2022). Nutri-Score is a tool proposed to promote healthier food selection and simplify nutritional information for consumers (Huybrechts et al., 2018). The algorithm allocates positive points (0–10) for nutrients and values of concern, including energy (kJ), total sugars (g), saturated fatty acids (g), and sodium (mg), while negative points (0–5) are assigned to healthy foods and nutrients including fruits/vegetables/pulses/nuts

* Corresponding author.
E-mail address: aaryee@desu.edu (A.N.A. Aryee).

https://doi.org/10.1016/j.crfs.2023.100544
Received 17 May 2023; Accepted 27 June 2023
Available online 28 June 2023

fiber (g) and proteins (g). The sum of positive points (0 to +40 points) and negative points (0 to −15 points) is computed, yielding a global score ranging from −15 for the healthiest foods to 40 for less healthy foods (Du et al., 2019). Nutri-Score was established in France, considering the requirements of vulnerable communities regarding nutritional information to lessen social inequalities in health and nutrition, and has been adopted by Germany, Belgium, the Netherlands, Spain, and Switzerland (Ducrot et al., 2022). In Germany, the Nutri-Score algorithm assisted in differentiating the nutritional quality of foods at various levels while avoiding dichotomous thinking of foods in 'healthy' and 'less healthy' categories that promote the contention that foods are either 'all good' or 'all bad.' It informs consumers of the unique nutritional qualities of foods and serves as a guide to enable consumers to make healthier decisions at the point of purchase (Szabo De Edelenyi et al., 2019).

In the last two decades, the growth of artificial intelligence (AI) and machine learning (ML) has enabled researchers to solve real-world issues and reach unbiased conclusions. Supervised ML techniques such as support vector regressor, multilayer perceptron, K-nearest neighbor (KNN) and ensemble ML methods have been used to predict the nutrient content (carbohydrates, fats, proteins, and dietary fiber) of labeled foods from databases (Davies et al., 2021; Ma et al., 2021; Tachie et al., 2023). However, despite the importance of PBF as part of a healthy diet, ML techniques have not yet been applied to predict its quality. The Nutri-Score is calculated based on the macronutrients such as carbohydrates, fats, and proteins and does not consider micronutrients from the food. Hence the need to improve the quality prediction by combining both Nutri-Score and micronutrients. This study aims to predict the quality of foods and beverages substituted with plant-based ingredients using the Nutri-Score and ML techniques.

2. Materials and method

The methods and the ML algorithms used for predicting the quality of PBFs are described in this section, and a summary of the process is provided in Fig. 1.

2.1. Data source

Data from the U.S. National Health and Nutrition Examination Survey (NHANES) cycles from 2017 to 2019 were examined. NHANES combines interviews and physical examinations in several surveys on various demographic groups and subjects to evaluate the health and nutritional status of adults and children in the United States (Yao et al., 2021). Dietary data of years 2017 to 2020 were obtained from the Food and Nutrient Database for Dietary Studies (FNDDS) (https://www.ars.usda.gov/northeast-area/beltsville-md-bhnrc/beltsville-human-nutritionresearch-center/food-surveys-research-group/docs/fndds-download-databases/) based on a 24-h recall of the PBFs usually consumed by people in the US. The PBF data from the two years were merged and split into 80% for training (300) and 20% for testing (74). The FNDDS data contains each food product's description, ingredient list, macro- and micronutrients, and unique identifiers. The current investigation examined the quality of PBFs and ingredients based on the Nutri-Score (grade) and Nutri-Score (Sante Publique, 2022) as the dependent variables. The Nutri-Score calculation included components such as saturated fatty acids (SFA), energy (kJ), protein (g) sodium (g) and dietary fiber (g). These were used to grade the quality of the PBF from A - E (highest to least quality) based on a calculated Nutri-Score ranging between −15 and 40. The PBF categories used were legumes, seeds, nuts, processed bean products, milk substitutes, cereals and grains.

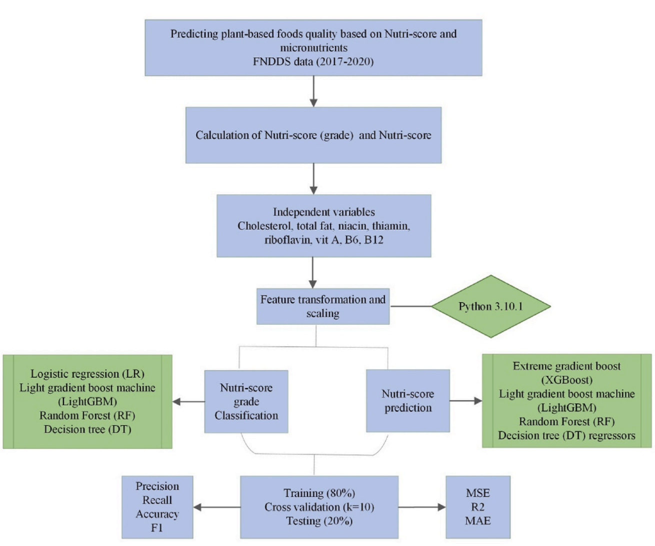

Fig. 1. Overview of methodology and machine learning pipeline.

2.2. Data pre-processing

Data was modeled using the python software version (3.10.1). The data was imported into the software after loading library packages such as NumPy, matplotlib, and pandas versions 1.22.4, 1.4.4, and 3.6.3, respectively. The exploratory data analysis involved extracting relevant features using the variable inflation factor (VIF), and independent variables with VIF (>7) eliminated to minimize multicollinearity. The selected explanatory variables included the Nutri-Score, total fats, cholesterol, and micronutrients such as B12, K, B6, A, D, thiamin, riboflavin, folate, and niacin, while the Nutri-Score grade (A-E) and Nutri-Score (−8 to 24) were the response variables. The data distribution was visualized using the box and whisker plot and heatmap.

2.3. Data transformation

The data had no missing values and duplicates were deleted. The data were scaled using the standard scaler function and transformed into dummy variables using the label encoding function for features with categorical values such as food description and Nutri-Score (grade). The scaled data was used for modeling.

2.4. Machine learning development - model training and testing

Supervised ML classification models: decision tree (DT), random forest (RF), light gradient boost machine (LightGBM), and logistic regression (LR) were imported and trained using the leave-one-out cross-validation method with 10 splits (k). The data was split into 10, and nine parts were used for training and one for testing. The average accuracy scores from each training were used as the final prediction accuracy. The models were evaluated using accuracy, weighted micro-precision, recall, and F1 score.

RF, DT, LightGBM, and XGBoost regressor models were used to predict the exact Nutri-Score. The cross-validation method was similarly used for training; however, these models were evaluated based on the coefficient of determination (R^2), mean squared error (MSE), and mean absolute error (MAE) to measure how much a model deviates from the actual values in its predictions. The calculation for metrics has been described (Ananey-Obiri and Sarku, 2020)

The models were saved using the pickle function and used to classify the PBFs based on the Nutri-Score grade in the testing datasets. The saved models were re-loaded. The models were also evaluated using the same metrics as the training sample to assess their ability to generalize trends to new data.

$$Accuracy = \frac{True\ Positives + False\ Negatives}{Total\ number\ of\ samples}$$

$$Precision = \frac{True\ Positives}{True\ Positives + False\ Positves}$$

$$Recall = \frac{True\ Positives}{True\ Positives + False\ Negatives}$$

$$MAE = \frac{1}{N}\sum_{i=1}^{N}(|\widehat{y}_i - y_i|)$$

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \widehat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \overline{y})^2}$$

Where n and N represent the number of the samples, y_i is the measured value of the ith sample, $\widehat{y}_i$ is the predicted value of the ith sample, and $\overline{y}$ is the average value of all the y_i.

DT is a hierarchical structure that recursively partitions the data based on the feature values to make predictions. At each node, the algorithm selects the best feature and split point to partition the data, using a criterion like information gain. The process continues until a stopping criterion is met, such as maximum tree depth, minimum samples per leaf, or no further improvement in the splitting criterion. The final prediction is made by traversing the tree from the root node to a leaf node based on the input feature values (Agarwal, 2014).

RF is an ensemble method that integrates numerous decision trees and finds average predictions from all trees to improve overall accuracy and reduce overfitting (Khan et al., 2022). RF produces great prediction accuracy on big heterogeneous data for both quantitative and qualitative elements. The algorithm creates multiple bootstrapped samples from the original dataset with replacement and trains a decision tree for each sample, selecting a random subset of features for each split. The final prediction is obtained by majority voting for classification tasks or averaging for regression tasks (Qian et al., 2022).

LR is a popular machine learning algorithm used for binary classification problems, where the goal is to predict one of two possible outcomes for a given input (Engelhardt et al., 2014). It can also be extended to handle multi-class classification problems through one-vs-rest or multinomial logistic regression techniques.

LightGBM is a gradient-boosting framework that focuses on efficiency and scalability. It constructs an ensemble of weak decision trees stage-wise, additively. However, LightGBM uses a leaf-wise growth strategy, which can lead to faster convergence and better accuracy. Additionally, it employs techniques like Gradient-based One-Side Sampling (GOSS) and Exclusive Feature Bundling (EFB) to reduce the number of data instances and features used during training, improving training speed and memory usage (Mondal, 2022).

3. Results and discussion

3.1. Data characteristics

The data used was PBFs and ingredients including, 7.2% milk substitutes, 2.13% cereals and pseudo-cereal products, 12.3% oatmeal products, 30.5% nuts and seeds, 7.5% processed soy products, 37.1% legumes, and 3.5% other products (chips, dips, and gravies). A total of 198, 53, 70, 51, and 3 PBFs had an A, B, C, D, and E Nutri-Score grades, respectively. This implies that 67.1% (grades A and B) of the PBFs used in this study have high quality ratings.

3.2. Model selection and performance

In total, 374 PBFs were selected and the accuracy, precision, recall, and F1 score for predicting the quality among the models ranged between 0.81- 0.88, 0.84–0.91, 0.81–0.88, and 0.81–0.88, respectively. Box and whisker plots were used to present the distribution of the dataset graphically (Fig. 2). This provides a visual summary of the central tendency, dispersion, overall spread, and potential outliers. Most variables have outliers, and the data range beyond the whiskers after the interquartile range. Some features (niacin, total fat, vitamin B6, K, and E) were positively skewed, indicating that this dataset requires robust models such as ensemble methods that can treat outliers or create suitable models regardless of the data distribution. The data distribution influenced the final model outcome as normally distributed data had higher predictive quality than skewed data (Ma et al., 2021). The distribution of the Nutri-Score was closer to normal, resulting in high prediction values for all models (Table 2).

Table 1 shows the performance metrics of DT, LightGBM, LR, and RF for predicting the quality of PBFs using the Nutri-Score grade. The accuracy, precision, recall, and F1 scores are reported for training and testing datasets. The accuracy score measures the proportion of correct predictions made by the model. In the training dataset, all four models achieved high accuracy scores ranging between 0.82 and 0.99, with 0.99 for LightGBM , followed by DT, RF (0.97), and LR (0.82). However, in the testing dataset, the accuracy scores of all models were minimized, while RF had the highest score of 0.88 showing its ability to generalize the training patterns in predicting new or unseen variables. The precision score is the proportion of true positives (correctly predicted positives) out of all positive predictions. LightGBM achieved the highest precision score (0.99) for training, followed by DT and RF, which were reduced after testing. The testing RF precision score was the highest followed by LR. The recall score measures the proportion of true positives out of all actual positives. RF had the highest recall score for testing though LightGBM had the best training scores. All the other models also had high recall scores in the training dataset but reduced recall in the testing dataset. The F1 score is the average of the precision and recall and gives an overall measure of a model's performance. The RF model performed the best on the testing dataset with the highest accuracy, precision, recall, and F1 score. However, LightGBM performed the best on the training dataset. LR also performed well on the testing dataset, achieving high precision, recall, and F1 scores. These results suggest that the RF is the most suitable model for predicting the Nutri-Score grade of PBFs while LR is also a viable option.

The confusion matrices shown in Fig. 3 enable a more in-depth examination of the classification errors. The diagonal represents the correctly classified samples. The numbers selected for visualization were 41, 10, 14, and 11 for grades A, B, C, and D, respectively. The scores reported for accuracy in Table 1, LR had the lowest score in predicting Nutri-Score grade. This was reflected in the confusion matrix where 3, 3, 2, and 4 samples were misclassified for classes A, B, C, and D, respectively. RF similarly misclassified two samples in grades C and D. However, LightGBM and DT correctly mapped all samples to their respective grades (Fig. 3a).

Table 2 shows the performance metrics of different ML models for predicting the Nutri-Score of PBFs. The R^2 score ranges between 0 and 1 and measures the proportion of variance in the target variable (Nutri-Score) explained by the independent variables (features), with a higher score indicating a better model fit. The MAE measures the absolute average difference between the predicted and actual Nutri-Score values. MSE measured the squared average difference between the predicted and actual values. A lower MAE and MSE indicate better performance. Based on the results, all four models performed well on the training set with R^2 scores ranging from 0.93 to 0.96, showing that the models can explain a large proportion of the variance in the data. However, some of the models' R^2 scores were slightly reduced on the testing set. Table 2 shows that LightGBM had the highest R^2 score (0.95 and 0.96) and the lowest MAE (train = 0.81, test = 1.00) and MSE (train = 1.85, test = 2.13), indicating that it is the most accurate model for explaining the variance in the Nutri-Score and making predictions for the unseen data. On the other hand, the DT model had the least R^2 score (0.93) on both data sets with the highest MSE (5.34) and MAE (1.241) rates, particularly in the testing data indicating possible overfitting.

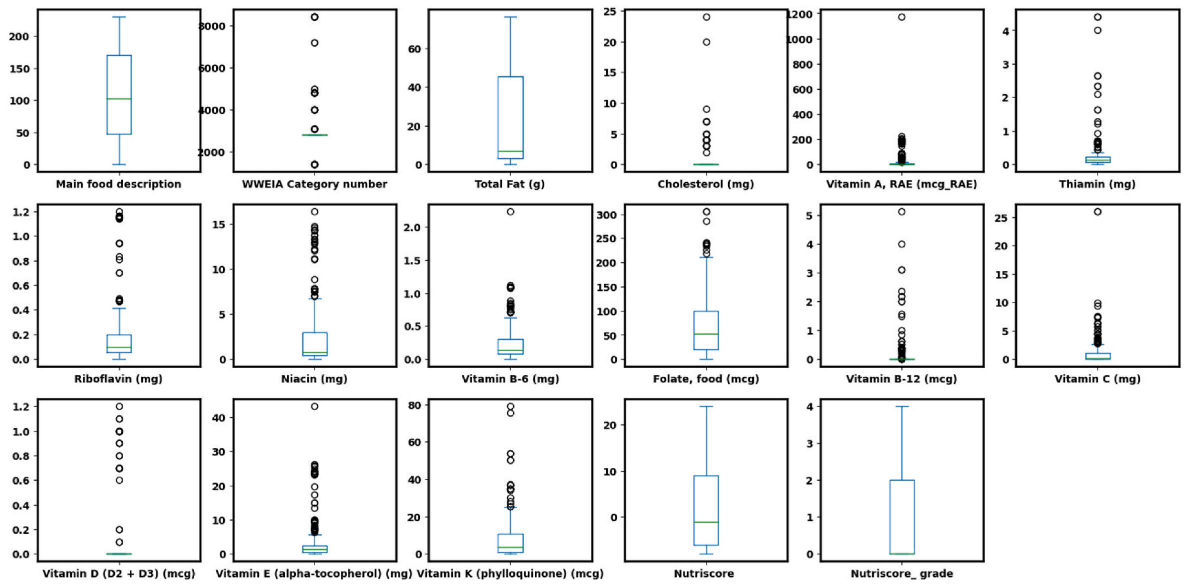

Fig. 2. Distribution of variables.

Table 1
Model comparison for predicting nutrient quality of plant-based foods

Model	Parameter	Accuracy	Precision	Recall	F1
DT	Training	0.97	0.97	0.94	0.94
	Testing	0.81	0.84	0.81	0.81
LightGBM	Training	0.99	0.99	0.96	0.96
	Testing	0.82	0.87	0.82	0.83
LR	Training	0.82	0.82	0.81	0.81
	Testing	0.83	0.85	0.84	0.84
RF	Training	0.97	0.97	0.93	0.94
	Testing	0.88	0.91	0.88	0.88

Prediction outcome for plant-based foods quality using Nutri-Score grade, micronutrients, and machine learning models; decision tree (DT), light gradient boost machine (LightGBM), logistic regression (LR) and random forest (RF).

Table 2
Comparing nutrient quality prediction (Nutri-Score) of plant-based foods

Model	Parameter	R^2	MAE	MSE
DT	Training	0.93	0.65	2.60
	Testing	0.93	1.24	5.34
LightGBM	Training	0.96	0.81	1.85
	Testing	0.96	1.00	2.13
XGBoost	Training	0.95	0.64	2.38
	Testing	0.95	0.97	2.62
RF	Training	0.96	0.72	1.40
	Testing	0.95	1.09	2.46

Coefficient of determination (R^2), mean absolute error (MAE) and mean squared error (MSE) of predicting the quality of plant-based foods using Nutri-Score and micronutrients content, and the machine learning models; decision tree (DT), light gradient-boosting machine (LightGBM), extreme gradient boost (XGR), and random forest (RF).

Overall, the models efficiently predicted the Nutri-Score of PBFs, with LightGBM showing the best performance.

The function of the correlation matrix is to assess the relationship between variables (Fig. 4). The use of variable inflation factor aided in selecting the independent variables with minimum correlation among each other. Variables with VIF>7 were excluded from the modeling leading to a reduction in multicollinearity. The correlation matrix ranges between 0 and 1 with a value above 0.7 indicating a high positive correlation. The Nutri-Score and Nutri-Score grade were positively correlated as food is graded based on Nutri-Score obtained. A strong positive correlation also existed between the total fat and Nutri-Score grade, meaning the grade increases with high fat content, and such foods are ranked as D-E, indicating a lower quality.

4. Limitations and future work

The quality of PBFs was predicted based on the Nutri-Score and micronutrients content. However, the study had some limitations. The FNDDS dietary data does not specify the brands of PBFs recorded, and other foods , such as sausages, eggs, and fish substitutes, were not included and the model cannot be generalized for other PBFs outside this data. In addition, the categories used did not have many subgroups for efficient training and prediction. More data should be included from other years to increase the number of products sampled per category. The predicted accuracy may increase as the FNDDS data is expanded to include a wider variety of PBFs. Future model optimization may result from these advancements.

5. Conclusion

PBFs and their corresponding nutrient were compiled from the FNDDS dietary intake data, which contains foods consumed in the US using a 24-h recall. Four supervised ML classifiers and regressor models were evaluated using ten-fold cross-validation and their precision, accuracy, recall, MSE, and R^2 score were reported and based on the selected metrics. RF (0.88) and LightGBM (0.95) performed best in determining the Nutri-Score grade and predicting the exact Nutri-Score for the PBFs. These two models were the most suitable to determine PBF quality based on Nutri-Score and micronutrients. The Nutri-Score grade highly correlated with the total fat content in PBFs.

Author contributions

Conceptualization, Project administration, Funding acquisition, Validation, Writing - Review and Editing: ANAA; Investigation, Formal analysis, Methodology, Writing - Original draft: CT; Writing - Review & Editing: NAT. All authors have approved the final version to be published and have agreed to be accountable for all aspects of the work.

Disclosure statement

Authors declare no conflict of interests.

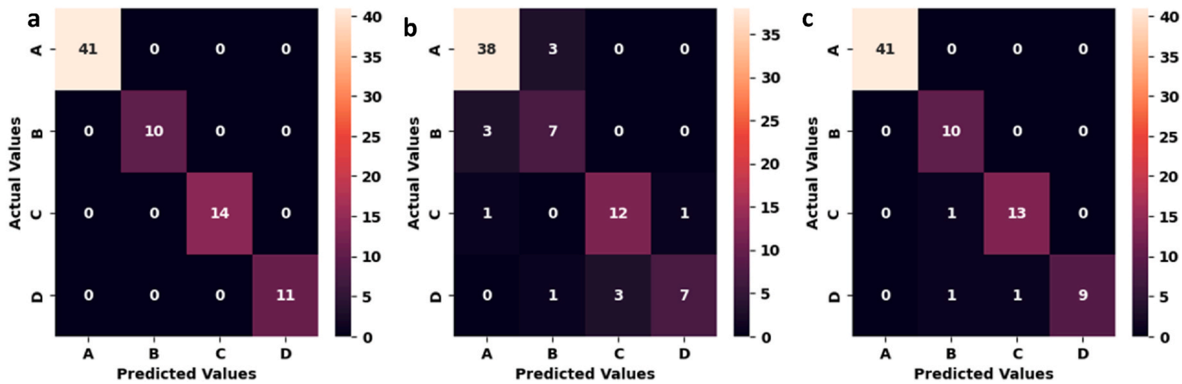

Fig. 3. Confusion matrix for classifying plant-based foods based on the quality for all the models (a) light gradient boost machine (LightGBM) and decision tree (DT), (b) logistic regression (LR) and (c) random forest (RF).

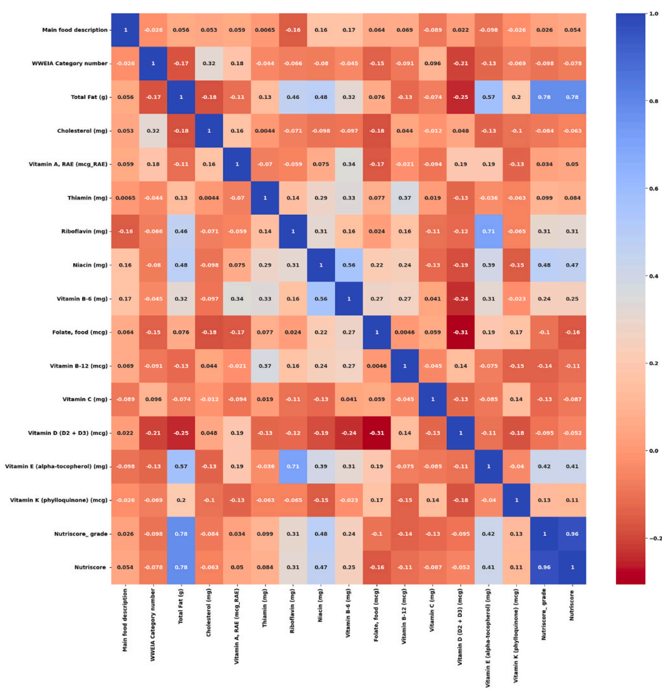

Fig. 4. Correlation between variables.

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.

Data availability

Data will be made available on request.

Acknowledgement

The project was supported by USDA-NIFA-AFRI; Award number: 2021-67022-34148.

References

Agarwal, S., 2014. Data mining: data mining concepts and techniques. In: Proceedings - 2013 International Conference on Machine Intelligence Research and Advancement, vol. 2013. ICMIRA. https://doi.org/10.1109/ICMIRA.2013.45.

Ananey-Obiri, D., Sarku, E., 2020. Predicting the presence of heart diseases using comparative data mining and machine learning algorithms. Int. J. Comput. Appl. 176 (11), 17–21. https://doi.org/10.5120/ijca2020920034.

Davies, T., Louie, J.C.Y., Scapin, T., Pettigrew, S., Wu, J.H.Y., Marklund, M., Coyle, D.H., 2021. An innovative machine learning approach to predict the dietary fiber content of packaged foods. Nutrients 13 (9). https://doi.org/10.3390/nu13093195.

Du, L., Lu, W., Gao, B., Wang, J., Yu, L.L., 2019. Authenticating raw from reconstituted milk using fourier transform infrared spectroscopy and chemometrics. J. Food Qual. 2019 https://doi.org/10.1155/2019/5487890.

Ducrot, P., Julia, C., Serry, A.J., 2022. Nutri-score: awareness, perception and self-reported impact on food choices among French adolescents. Nutrients 14 (15). https://doi.org/10.3390/nu14153119.

Engelhardt, A., Kanawade, R., Knipfer, C., Schmid, M., Stelzle, F., Adler, W., 2014. Comparing classification methods for diffuse reflectance spectra to improve tissue specific laser surgery. BMC Med. Res. Methodol. 14 (1), 1–15. https://doi.org/10.1186/1471-2288-14-91.

Hawley, K.L., Roberto, C.A., Bragg, M.A., Liu, P.J., Schwartz, M.B., Brownell, K.D., 2013. The science on front-of-package food labels. Publ. Health Nutr. 16 (3), 430–439. https://doi.org/10.1017/S1368980012000754.

Huybrechts, I., Murphy, N., Julia, C., Hercberg, S., Srour, B., Kesse-guyot, E., Latino-martel, P., Biessy, C., Casagrande, C., Jenab, M., Ward, H., Weiderpass, E., Dahm, C. C., Overvad, K., Kyr, C., Kaaks, R., Ku, T., Boeing, H., Schwingshackl, L., et al., 2018. Nutritional Quality of Food as Represented by the FSAm-NPS Nutrient Profiling System Underlying the Nutri-Score Label and Cancer Risk in Europe : Results from the EPIC Prospective Cohort Study, pp. 1–21.

Kahleova, H., Levin, S., Barnard, N., 2017. Cardio-metabolic benefits of plant-based diets. Nutrients 9 (8). https://doi.org/10.3390/nu9080848.

Khan, M.I.H., Sablani, S.S., Nayak, R., Gu, Y., 2022. Machine learning-based modeling in food processing applications: state of the art. Compr. Rev. Food Sci. Food Saf. https://doi.org/10.1111/1541-4337.12912.

Ma, P., Li, A., Yu, N., Li, Y., Bahadur, R., Wang, Q., Ahuja, J.K., 2021. Application of machine learning for estimating label nutrients using USDA Global Branded Food Products Database, (BFPD). J. Food Compos. Anal. 100, 103857 https://doi.org/10.1016/j.jfca.2021.103857.

Mondal, A., 2022. LightGBM in Python | complete guide on how to use LightGBM in Python. Analytica Vidya. https://www.analyticsvidhya.com/blog/2021/08/complete-guide-on-how-to-use-lightgbm-in-python/.

Pratt, N., 2020. Formulating Plant-Based Foods - Nutrition Challenges and Opportunities. Kerry Health and Nutrition Institute.

Publique, Sante, 2022. Nutri-Score. https://www.santepubliquefrance.fr/en/nutri-score.

Qian, J., Tian, F., Luo, Y., Lu, M., Zhang, A., 2022. A novel multisensor detection system design for low concentrations of volatile organic compounds. IEEE Trans. Ind. Electron. 69 (5), 5314–5324. https://doi.org/10.1109/TIE.2021.3080218.

Szabo De Edelenyi, F., Egnell, M., Galan, P., Druesne-Pecollo, N., Hercberg, S., Julia, C., 2019. Ability of the Nutri-Score front-of-pack nutrition label to discriminate the nutritional quality of foods in the German food market and consistency with nutritional recommendations. Arch. Publ. Health 77 (1), 1–9. https://doi.org/10.1186/s13690-019-0357-x.

Tachie, C., Nwachukwu, I.D., Aryee, A.N.A., 2023. Trends and innovations in the formulation of plant-based foods. Food Production, Processing and Nutrition 5 (1), 16. https://doi.org/10.1186/s43014-023-00129-0.

Tachie, C.Y., Obiri-Ananey, D., Tawiah, N.A., Aryee, A.N., 2023. Predicting fatty acid classes in popular US snacks using NHANES data and machine learning approaches, pp. 1–25. In press.

Vanga, S.K., Raghavan, V., 2018. How well do plant based alternatives fare nutritionally compared to cow's milk? J. Food Sci. Technol. 55 (1), 10–20. https://doi.org/10.1007/s13197-017-2915-y. Springer India.

Yao, X., Steven Xu, X., Yang, Y., Zhu, Z., Zhu, Z., Tao, F., Yuan, M., 2021. Stratification of population in NHANES 2009–2014 based on exposure pattern of lead, cadmium, mercury, and arsenic and their association with cardiovascular, renal and respiratory outcomes. Environ. Int. 149, 106410 https://doi.org/10.1016/j.envint.2021.106410.