Check for updates

RESEARCH ARTICLE

REVISED OpenHSR: An open, FAIR-compliant dataset for modeling the health star rating system

[version 2; peer review: 1 approved with reservations, 1 not approved]

Flore N'kam Suguem [1-3], Valentin Lafargue [1,2,4-6]

[1]Institut de Mathematiques de Toulouse, Toulouse, Occitanie, 31062, France
[2]Mathematics and Informatics, Toulouse University, Toulouse, 31062, France
[3]Physics and Astronomy, University of Bologna, Bologna, 40127, Italy
[4]Institut de Recherche en Informatique de Toulouse, Toulouse, Occitanie, 31062, France
[5]ANITI2 Toulouse, Toulouse, 31400, France
[6]INRIA Bordeaux, Bordeaux, 33405, France

v2 **First published:** 05 Dec 2025, **5**:374
https://doi.org/10.12688/openreseurope.21837.1
Latest published: 01 Oct 2026, **5**:374
https://doi.org/10.12688/openreseurope.21837.2

Abstract

Background

Nutritional awareness is increasing worldwide due to rising obesity rates and the widespread use of nutrition-tracking technologies. However, nutritional labels on food packaging remain difficult for many consumers to interpret. To improve clarity, public authorities have introduced simplified front-of-pack labeling systems, such as the Health Star Rating (HSR) used in Australia and New Zealand, which summarizes a product's overall nutritional quality. Despite its adoption, the lack of open data on Health Star Rating labeled products limits independent research, reproducibility, and data-driven policy evaluation.

Methods

We introduce OpenHSR, the first open, FAIR-compliant dataset (Findable, Accessible, Interoperable, and Reusable) dedicated to the Health Star Rating system. The dataset includes 246 unique food products collected from Australian supermarket websites and brand databases. Using these data, we trained and compared several regression models linear, tree-based, and neural approaches to predict Health Star Rating values from nutritional attributes, thereby developing a transparent, interpretable ("white-box") Health Star Rating calculator.

Open Peer Review

Approval Status ? ✗

	1	2
version 2 (revision) 01 Oct 2026		✗ view
version 1 05 Dec 2025	? view	✗ view

1. **Mohammad Nazrul Islam Bhuiyan**, BCSIR Chattogram Laboratories, BCSIR, Chattogram, Bangladesh

2. **Giovanni Cangelosi**, Polo Medicina Sperimentale e Sanità Pubblica, Camerino, Italy

Any reports and responses or comments on the article can be found at the end of the article.

Results

Machine learning models achieved strong predictive performance, with the Support Vector Regression model yielding the best Mean Squared Error = 0.28. The trained models were then applied to predict Health Star Rating values for 6,554 Open Food Facts products, demonstrating scalability and generalizability of the approach.

Conclusions

OpenHSR provides the first openly available, standardized dataset linking nutritional composition and Health Star Rating values. Its transparent design supports reproducible research, automated Health Star Rating estimation, and cross-dataset integration. This resource enables future studies in nutrition science, food labeling policy, and machine learning applications for public health.

Keywords
Dataset, Health star rating (HSR), Nutri-score, Regression, Machine Learning, Benchmark, Interpretability



This article is included in the Marie-Sklodowska-Curie Actions (MSCA) gateway.



This article is included in the Horizon Europe gateway.



This article is included in the Health Sciences gateway.

Corresponding author: Flore N'kam Suguem (flore.n_kam_suguem@math.univ-toulouse.fr)

Author roles: N'kam Suguem F: Conceptualization, Data Curation, Formal Analysis, Investigation, Methodology, Writing – Original Draft Preparation; **Lafargue V**: Conceptualization, Methodology, Software, Writing – Original Draft Preparation

Competing interests: No competing interests were disclosed.

Grant information: This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No 101120449 (FAIROmics) and the France 2030 programme under Grant Agreement No. ANR-23-IACL-0002.
The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Copyright: © 2026 N'kam Suguem F and Lafargue V. This is an open access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

How to cite this article: N'kam Suguem F and Lafargue V. **OpenHSR: An open, FAIR-compliant dataset for modeling the health star rating system [version 2; peer review: 1 approved with reservations, 1 not approved]** Open Research Europe 2026, **5**:374 https://doi.org/10.12688/openreseurope.21837.2

First published: 05 Dec 2025, **5**:374 https://doi.org/10.12688/openreseurope.21837.1

REVISED **Amendments from Version 1**

The revised version includes introduction part in revised and Health star rating system introduction was updated Table 1 was revised and discussion part was revised as In our benchmarking evaluation, SVM and XGBoost yielded the best performance metrics. This aligns with established machine learning behavior, as kernel methods and ensemble approaches consistently perform well on small tabular datasets. These algorithms can efficiently capture complex relationships among nutritional variables without immediately overfitting, making them highly robust when training data is scarce. However, it must be noted that their low Mean Squared Error (MSE) reflects optimized loss on the training data rather than guaranteed generalized accuracy, highlighting the need for cautious interpretation.

Any further responses from the reviewers can be found at the end of the article

1. Introduction

The global prevalence of diet-related chronic non-communicable diseases, particularly obesity, cardiovascular disease, and type 2 diabetes, represents a critical clinical and public health challenge.[1] To mitigate this medical burden, healthcare professionals and public authorities increasingly advocate for preventative, population-level interventions. Among these, front-of-pack labeling (FOPL) systems have emerged as a vital tool for facilitating healthier dietary choices by translating complex nutritional information into easily interpretable visual cues at the point of purchase.[2]

The Health Star Rating (HSR) system, implemented in Australia and New Zealand, is one such FOPL scheme. Similar to other systems, such as the Nutri-Score widely used in European countries, HSR aims to support consumer decision-making and reduce the incidence of metabolic disorders by providing a clear, standardized indication of a product's overall nutritional profile. These labels typically appear on packaged foods and range from 0.5 to 5 stars. Despite the broad clinical and policy relevance of HSR, access to large-scale, structured datasets of HSR-labeled products remains limited, restricting the potential for comprehensive research and data-driven healthcare policy evaluation.

Existing datasets such as Open Food Facts, FoodSwitch, and AUSNUT 2023 primarily focus on providing nutrient information, either as open data or through consumer-oriented applications that promote healthier food choices. However, a gap remains, as none of these datasets specifically emphasize or comprehensively capture HSR information.

To address this gap, we introduce OpenHSR, the first open and publicly available dataset dedicated to the Health Star Rating system. The dataset aggregates information from multiple Australian supermarkets and includes each product's name, nutritional composition per 100 g or 100 mL, HSR category, and assigned Health Star Rating. Our dataset has been developed in accordance with the FAIR principles Findable, Accessible, Interoperable, and Reusable to ensure high-quality data management and research reproducibility. By making this dataset openly accessible, our goal is to enable reproducible analyses of nutritional trends, facilitate evaluation of the HSR system's scoring consistency, and support further research in nutrition science, food policy, and machine learning applications for health-related data.

In Section 2, we describe the HSR system in detail. Section 3 introduces the datasets available for nutritional analysis, including publicly accessible ones for the Nutri-Score, but none for HSR. In Section 4, we present our dataset and its diversity in HSR coverage. Finally, Section 6 introduces machine learning–based regression algorithms to predict HSR from nutritional values, enabling the creation of our own HSR calculator and checking of the coherence between the actual calculation and the theoretical guidelines.

2. Health star rating system introduction

The Health Star Rating (HSR) system, jointly implemented by the Australian and New Zealand governments, assigns packaged foods a rating from 0.5 to 5 stars.[3] Higher ratings indicate a healthier overall nutritional profile. The HSR is computed using the official HSR Calculator, which applies the Nutrient Profiling Scoring Criterion (NPSC) to determine a product's category-specific score.

To account for fundamental differences in how nutrient profiles should be evaluated across different types of products, the algorithm first assigns a food item to one of six primary categories:

- **Category 1:** Non-dairy beverages, jellies, and water-based ice confections
- **Category 1D:** Milk and dairy beverages
- **Category 2:** General foods

- **Category 2D:** Dairy foods
- **Category 3:** Oils and spreads
- **Category 3D:** Cheese

Once categorized, the system routes the product into one of two distinct calculation pathways to determine its baseline score:

Pathway A (Categories 1D, 2, 2D, 3, and 3D): For general foods, dairy, oils, and cheeses, the baseline score is calculated by evaluating four "negative" components associated with increased risk of chronic diseases: energy, saturated fat, sodium, and total sugars.

Pathway B (Category 1): For non-dairy beverages, the nutritional expectations are structurally different. Therefore, the baseline score is determined by evaluating only two negative components: energy and total sugars.

For both pathways, the calculated baseline score is subsequently adjusted by subtracting points for the presence of "positive" components, specifically protein content, dietary fiber, and the proportion of fruits, vegetables, nuts, and legumes (FVNL). This final adjusted score is then mapped to the standardized 0.5–5 star scale, yielding a single value that reflects the product's overall nutritional quality.

3. Related datasets

Open Food Facts is an open-source, publicly available dataset accessible at https://fr.openfoodfacts.org/,[4] containing metadata, Nutri-Score, NOVA classification, and Green Score for a wide variety of food products. The primary method of data collection relies on user submissions from around the world. While Health Star Rating (HSR) information is available for some products, there are often missing values in the metadata. Moreover, HSR data is not included in the downloadable version of the dataset, which limits its reusability for research purposes.

FoodSwitch is primarily a mobile phone application designed to help consumers make healthier food choices while enabling crowdsourcing of national food composition data.[5] Although the app includes Front-of-Pack nutrition labels such as the Health Star Rating (HSR), the underlying dataset is not directly accessible. Access requires submitting a request and demonstrating a research purpose, which limits its usability for broader projects. Consequently, the dataset does not fully comply with FAIR principles, as it is not readily findable, accessible, or reusable for a wide range of users.

The AUSNUT dataset is a comprehensive collection of detailed food supply data that supports the 2023 National Nutrition and Physical Activity Survey[6] and is accessible at https://www.foodstandards.gov.au/science-data/food-nutrient-databases/ausnut. Classification or coding systems are employed to group similar foods for reporting intakes and tracking trends over time. The study utilized three classification schemes to report:

- Foods, dietary supplements, and nutrient intakes
- Intake of discretionary versus non-discretionary foods
- Nutrient intakes relative to the Australian Dietary Guidelines

However, this dataset does not include Health Star Rating (HSR) information, limiting its applicability for analyses that rely on front-of-pack nutritional labeling metrics.

4. Dataset description

4.1 Overview

The OpenHSR dataset comprises 246 unique food products available in Australian retail chains in 2025.[7] Each record contains approximately 20 fields, which can be grouped into three main categories:

Nutrient composition includes energy (kJ and kcal), protein, total fat, saturated fat, carbohydrate, sugars, fiber, and sodium. Metadata includes product name, brand, product type, category, serving size, package weight, country of origin, retailer, date collected, allergens, ingredients, and data source. Health Star Rating (HSR) is provided as a score from 0.5 to 5, reflecting overall nutritional quality.

4.2 Data collection process

The data was manually obtained from the package images of products from retailers websites and Brand-specific online pages such as Coles, Woolworths, Kellogg's, Aldi and Fourntwenty. A part of the data was also similarly obtained from Open data facts' website. Using the packaging information is known to be more reliable especially in the case of open data facts where the written information on their website differed from the information on the packaging in some instances, decreasing the risk of human-induced errors.

Manual collection ensured accuracy and consistency in variable representation, while metadata fields were harmonized to allow cross-product comparisons.

4.3 Data validation

The dataset is provided in a tabular format (CSV), where each row represents a product and each column corresponds to a distinct variable. The dataset has no missing values in numerical entries.

Both manual and computational double-checks of nutritional panels and Health Star Ratings (HSR) were conducted to ensure a high-quality dataset.

The dataset also integrates the FAIR principles to support reproducibility and facilitate future research use:

Findable: Each product record is uniquely identified, and the dataset is assigned a DOI via Open Research Europe.

Accessible: The dataset is openly available under a CC BY 4.0 license, with direct download access.

Interoperable: Standardized units, consistent variable names, and harmonized categories enable integration with other datasets and computational pipelines.

Reusable: Comprehensive metadata, data dictionaries, and clear documentation support reproducibility, machine learning applications, and policy evaluations.

We described every variable gathered in the OpenHSR dataset in the software code on our github repository of the Appendix.

5. Exploratory analysis

The dataset exhibits sufficient diversity in HSR, which has enabled the development of a transparent, interpretable white-box HSR calculator.

The distribution of product categories across HSR reveals how nutritional quality varies within and between categories, as observed in Figure 1a. Some categories tend to achieve higher HSRs like the 1D category, while other like the 2D category tend to have lower ratings, reflecting differences in their nutrient profiles. This analysis highlights the relationship between product type and nutritional quality.

Figure 1b shows the categories' distribution in the dataset, it provides insight into its overall composition. The distribution of the different categories are heterogeneous; however, this is not a characteristic of our dataset but the general distribution in supermarket. Indeed, we do not expect the same number of product in the '2' category being Food as in the '3D' category, which is Cheese.

We refer to Figure 2a for the nutritional values' distributions across the dataset, they confirm the variability among the recorded food items. Moreover, a clear relationship is observed between energy content and HSR, products with higher calories per 100 g tend to have lower HSRs, and conversely, products with lower calorie content tend to have higher HSRs. In Figure 2b, lighter-colored points represent lower HSRs, which occur less frequently but correspond to higher energy content per 100 g.

6. Health star rating calculator: benchmark

We used a diverse set of machine learning (ML) regression models to predict the Health Star Rating (HSR). As summarized in Table 1, the models include three fully white-box and interpretable algorithms (LR,[8,9] DT,[9–11] kNN[9,12]), two ensemble methods (RF,[9,13] XGBoost[9,14]) whose individual base learners are interpretable although the overall ensemble is not, and three black-box models (SVM,[9,15,16] MLP,[17,18] FTN[17,18]). A detailed description of our model implementation and deployment strategy is available on our GitHub repository.

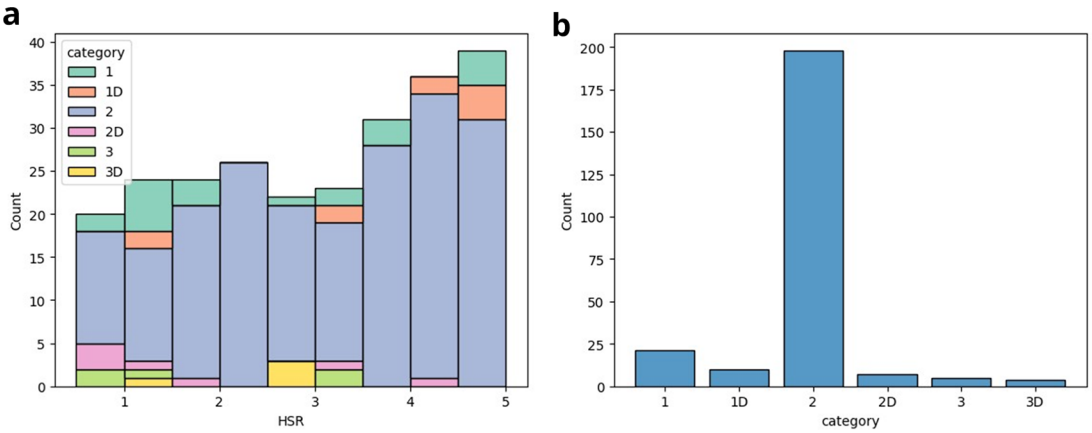

Figure 1. Descriptive statistics of the OpenHSR. (a) Distribution of product categories across HSR. (b) Distribution of categories across the dataset.

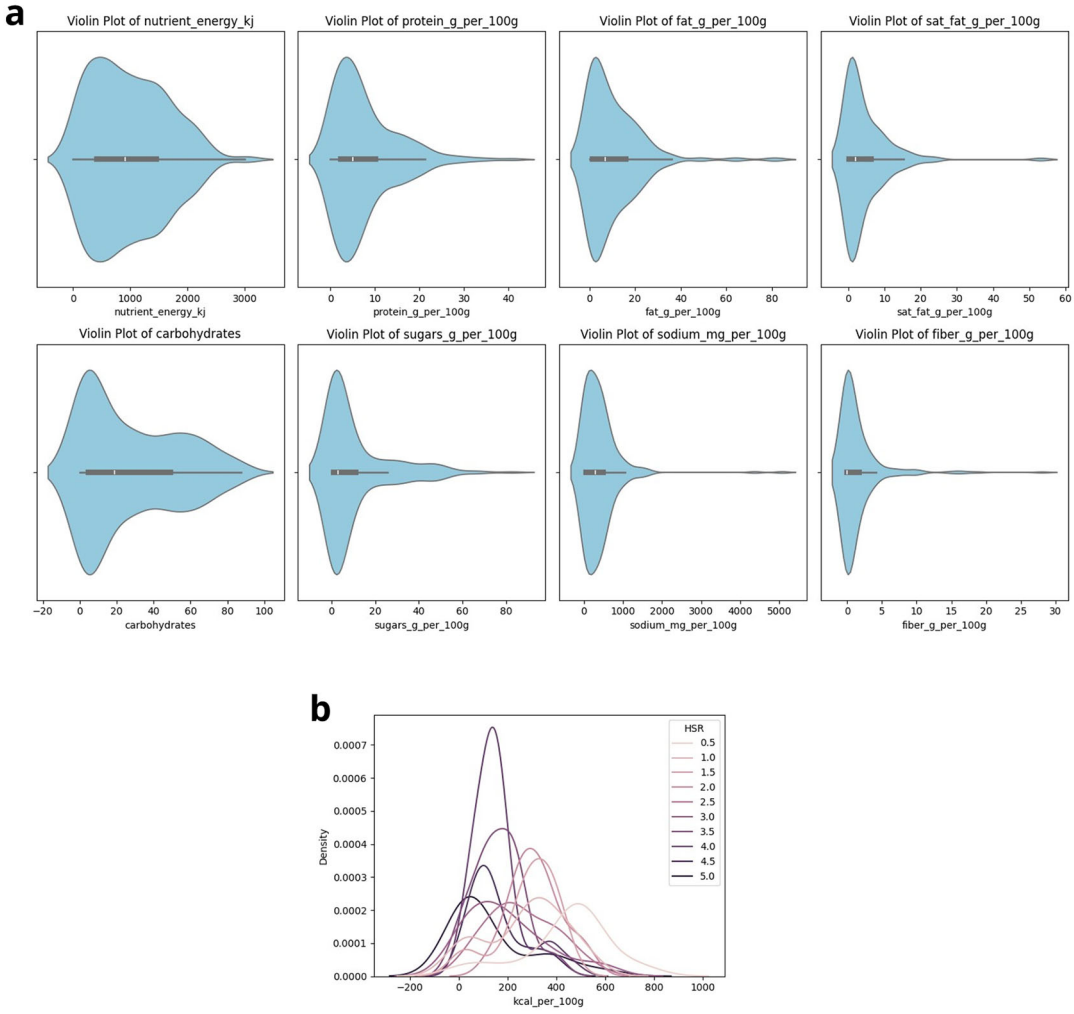

Figure 2. Descriptive statistics of the nutritional values within the OpenHSR. (a) Distribution of nutritional values across the dataset. (b) Density plot of kcal per 100 g of products per HSR.

Table 1. Performance of various machine learning regression algorithms for HSR prediction.

ML regression algorithm	MSE	acc	by-half-acc	Interpretable
Linear (LR)	0.82	0.42	0.76	✓
Decision Tree (DT)	0.55	0.34	0.76	✓
K-Nearest Neighbor (kNN)	0.40	0.42	0.86	✓
Random Forest (RF)	0.33	0.36	**0.92**	–
XGBoost	0.35	0.48	**0.92**	–
Support Vector Machine (SVM)	**0.28**	**0.62**	0.86	×
Multilayer Perceptron (MLP)	0.67	0.40	0.78	×
Fine-tuned network (FTN)	0.86	0.26	0.50	×

Each algorithm was evaluated based on the Mean Squared Error (MSE), exact match rate (acc) after rounding regression outputs to the nearest 0.5-star increment, and lenient match rate (by-half-acc), which tolerates a 0.5-star deviation. In bold are the best results, the minimum for the MSE, and the maximum for the exact match rate (acc) and lenient match rate (by-half-acc).

For all machine learning methods, hyperparameters were selected via grid search using scikit-learn's default 5-fold cross-validation. The dataset was partitioned into a training set (80%) and an independent test set (20%). For classical machine learning algorithms, hyperparameter tuning was performed on the entire training set using 5-fold cross-validation. For neural network models, the training set was further split into training (64%) and validation (16%) subsets.

It is also important to note that data standardization was applied selectively, depending on the algorithm's sensitivity to feature scaling. While standardization had a negligible impact on scale-invariant models such as the Decision Tree Regression algorithm, it was essential for distance-based and margin-based methods, including K-Nearest Neighbors and Support Vector Machines. Our neural network results were obtained using the methods of Defazio et al. Schedule-Free optimizer.[19]

The hyperparameters different from the scikit-learn default ones are as follows:

1. Decision Tree:'max_depth': 4,'max_features': 8,'min_samples_split': 10
2. K-Nearest Neighbor:'n_neighbors': 7
3. Random Forest:'max_depth': 6,'max_features': 8,'min_samples_split': 2
4. XGBoost'n_estimators': 500,'max_depth': 4,'min_samples_split': 10,'learning_rate': 0.01,'loss": "squared_error'
5. Support Vector Machine:'C': 100,'epsilon': 0.01,'kernel':'rbf'
6. Multilayer Perceptron: 5 loops of'BatchNorm''Dense' of 32 neurons and'ReLU' layers, with starting learning rate equal to 0.01.
7. Fine-tuned network: Multilayer Perceptron architecture above, with the last dense layer replaced, with a starting learning rate equal to 0.01.

Overall, the models achieved strong predictive performance for HSR. For example, XGBoost achieved a half-accuracy of 92% with a mean squared error below 0.4. In contrast, fully interpretable models such as the Decision Tree (see Figure 3) produced lower predictive accuracy compared to less interpretable approaches, notably the Support Vector Machine (SVM) regression, which was among the best performing models.

Performance could likely be improved, particularly for models such as SVM and Multilayer Perceptron (MLP), with access to additional labeled data. To address data limitations, we attempted to pre-train a foundation model to predict Nutri-Scores, subsequently using its latent representations to predict HSR. Note that our neural network results were obtained using 3. However, this approach did not yield satisfactory results for HSR prediction (Table 1), despite performing well for Nutri-Score prediction (MSE = 0.24, after transforming Nutri-Scores to a 1–5 numerical scale).

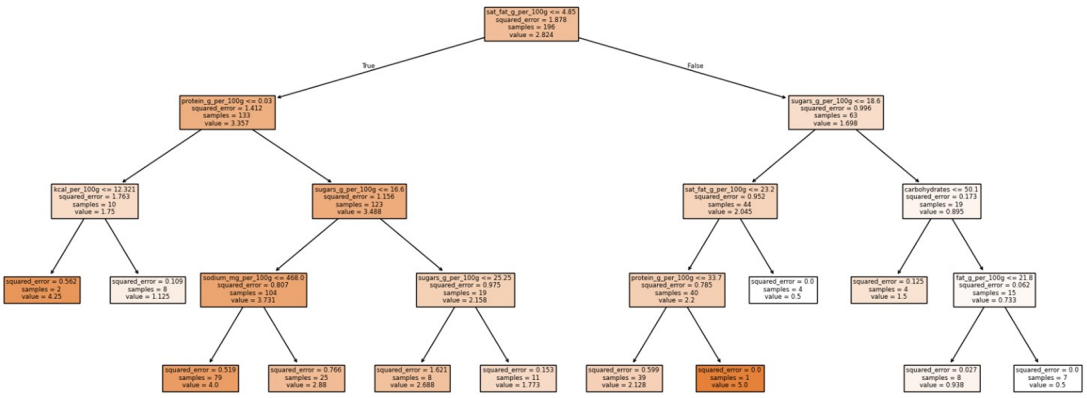

Figure 3. Decision-making process of the decision tree Regression algorithm. It shows which variable it uses, and the HSR predicted in consequences. The lighter-colored the cells are, the lower the HSR predicted will be.

From an applied perspective, although a half-star difference in HSR may not substantially alter consumer judgments, prioritizing interpretability over marginal performance gains can be justified. However, for research applications or large-scale predictive analyses, the SVM regression model is recommended due to its superior performance (lowest MSE). To demonstrate the generalizability of our approach, Section 7 presents HSR predictions on an independent nutritional dataset.

As illustrated in Figure 3, the decision tree identifies energy and sugar content as primary tree splitting criteria, indicating their predominant role in HSR estimation. This structural hierarchy is consistent with the official HSR algorithm, in which energy and total sugars serve as the primary negative components that determine the baseline score.

7. Health star rating predictions

We applied our trained models to the OpenFoodFacts dataset by selecting all products containing the nutritional attributes required by the algorithms described in the previous section. The only missing feature was the category variable introduced in Section 2. To address this, we constructed a mapping dictionary linking the most common categories in our training data to their corresponding categories in the HSR classification system.

Using the Support Vector Regression (SVR) model, we then predicted HSR values for the selected products. Since no ground-truth HSR labels were available for these Open Food Facts products, the predicted values cannot be quantitatively validated and should be interpreted as exploratory estimates.

The HSR calculator surrogate model demonstrates strong computational efficiency, producing HSR predictions for 6,554 products in only 0.1 seconds, suggesting its suitability for large-scale applications.

8. Discussion

OpenHSR addresses an important gap in front of label packaging research by providing the first open, FAIR-compliant dataset dedicated to the Health Star Rating (HSR) system. Existing resources have different strengths but do not necessarily meet the standards for data openness and FAIRness. Open Food Facts offers a large open food product database, yet HSR information is not consistently available for all products, making it challenging to construct dedicated HSR datasets. FoodSwitch captures extensive front-of-pack labeling information but provides more controlled access to its underlying database than fully open repositories. Meanwhile, AUSNUT is primarily intended for research on nutrient composition and dietary intake rather than for HSR analysis. By linking nutritional composition directly to HSR values in a standardized and openly accessible format, OpenHSR provides a transparent baseline for reproducible research on front-of-pack nutrition labeling.

In our benchmarking evaluation, SVM and XGBoost yielded the best performance metrics. This aligns with established machine learning behavior, as kernel methods and ensemble approaches consistently perform well on small tabular datasets. These algorithms can efficiently capture complex relationships among nutritional variables without immediately overfitting, making them highly robust when training data is scarce. However, it must be noted that their low Mean Squared Error (MSE) reflects optimized loss on the training data rather than guaranteed generalized accuracy, highlighting the need for cautious interpretation.

Conversely, deep learning architectures, including the Multilayer Perceptron and the fine-tuned foundation network, underperformed significantly. This limitation is entirely driven by the severe constraint imposed by our sample size, which comprises only 246 unique food products. Neural networks and foundation models inherently depend on vast quantities of data to accurately map complex parameter spaces. Applying such data-hungry models to an extremely small dataset fundamentally limits their reliability and generalizability, making them far less effective than traditional statistical algorithms on this benchmark. Their inclusion in the benchmark nonetheless serves a dual purpose: to ensure methodological completeness, as both architectures would represent state-of-the-art solutions given sufficient data, and to provide a foundation for future research by offering a readily available open-source codebase from which subsequent studies can build.

The creation of this open dataset carries immediate implications for independent reproducibility and data-driven nutrition policy. By directly linking nutrient profiles to their standardized front-of-pack scores, OpenHSR empowers researchers and policymakers to transparently evaluate scoring consistency and dietary adherence. Providing unobstructed access to this data ensures that public health guidelines can be independently audited and modeled, fostering greater accountability and transparency in national food labeling systems.

9. Conclusion

In this work, we presented the creation of a new dataset linking Health Star Ratings (HSR) with detailed product metadata for Australian retail foods. We provided a thorough analysis of the dataset, including its structure, data quality, and potential applications. Additionally, we demonstrated how machine learning models trained on this dataset can be used to predict HSR values for products where this information is not available, highlighting its utility for research and consumer applications.

The dataset is publicly accessible online, and its design adheres to the FAIR principles, ensuring it is findable, accessible, interoperable, and reusable. We anticipate that this resource will facilitate further studies on nutrition labeling, dietary assessment, and predictive modeling in food and health research.

9.1 Study design and reporting guidelines

This research employs a retrospective observational study design to aggregate and analyze the nutritional profiles of packaged foods associated with HSR. To ensure methodological rigor and transparency, the development and evaluation of the machine learning prediction models were guided by the principles of the TRIPOD[20] (Transparent Reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis) statement, an established EQUATOR Network guideline used here for nutritional dataset benchmarking.

9.2 Limitations

One of the main limitations of our dataset is its relatively small size (246 entries), which may affect the generalizability and performance of machine learning models. In addition, the dataset focuses exclusively on products available in the Australian market, limiting its applicability to other regions. Due to the fact that the data were collected manually, it may be subject to packaging updates and labeling changes over time. Furthermore, serving-size variability is not included, which may influence nutrient comparisons across products.

9.3 Future directions

To address these limitations, we plan to expand the dataset by increasing the number of product entries and incorporating data from multiple regions. Future versions will aim to automate data collection processes, include serving-size information, and ensure periodic updates to maintain accuracy and relevance.

Ethics and consent

Ethical approval and consent were not required.

Declaration of generative AI and AI-assisted technologies in the writing process

During the preparation of this work the authors used ChatGPT in order to enhance the readability and language. After using the tool, the authors reviewed and edited the content as needed and take full responsibility for the content of the publication.

Data availability

Underlying data

Zenodo: OpenHSR. [https://doi.org/10.5281/zenodo.17469191].[7]

The project contains the following underlying data:

- OpenHSR.csv (OpenHSR.csv contains raw, unaveraged product-level nutrition data for Australian breakfast cereals, including nutrient composition, Health Star Ratings, ingredients, allergens, and retailer information collected from manufacturer and supermarket sources.).

Data are available under the terms of the Creative Commons Attribution 4.0 International license (CC-BY 4.0).

Software availability statement

Source code available from: https://github.com/ValentinLafargue/HealthStarDataset

Archived software available from: https://doi.org/10.5281/zenodo.17556223

License: MIT License.

References

1. World Health Organization: *Noncommunicable Diseases Progress Monitor 2022.* World Health Organization; 2022.
Reference Source
2. Song J, Brown MK, Tan M, *et al.*: **Impact of Color-Coded and Warning Nutrition Labelling Schemes: A Systematic Review and Network Meta-Analysis.** *PLoS Medicine.* 2021; **18**(10): e1003765.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**
3. Health Star Rating Unit, Nutrition Policy Section: **HSR System Calculator and Style Guide.** 2025.
4. Open Food Facts Team: **Open Food Facts – a Collaborative, Open Database of Food Products from Around the World.** 2018.
Reference Source
5. Dunford E, Trevena H, Goodsell C, *et al.*: **FoodSwitch: A Mobile Phone App to Enable Consumers to Make Healthier Food Choices and Crowdsourcing of National Food Composition Data.** *JMIR mHealth and uHealth.* 2014; **2**(3): e37.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**
6. (FSANZ), Food Standards Australia New Zealand: *AUSNUT 2023 – the 2023 National Nutrition and Physical Activity Study Food and Nutrient Database.* 2025.
Reference Source
7. N'kam Suguem F, Lafargue V: Open Health Star Rating (OpenHSR) (Version V1). Data set. *Zenodo.* 2025.
Publisher Full Text
8. Legendre A-M: **Nouvelles Méthodes Pour La Détermination Des Orbites Des Comètes.** *Courcier.* 1805.
Reference Source
9. Pedregosa F, Varoquaux G, Gramfort A, *et al.*: **Scikit-Learn: Machine Learning in Python.** *Journal of Machine Learning Research.* 2011; **12**: 2825–2830.
10. Breiman L, Friedman JH, Olshen RA, *et al.*: *Classification and Regression Trees.* Wadsworth International Group; 1984.
Reference Source
11. Hastie T, Tibshirani R, Friedman J: *The Elements of Statistical Learning: Data Mining, Inference, and Prediction.* 2nd ed. Springer Series in Statistics; 2009.
Publisher Full Text
12. Fix E, Hodges JL Jr: **Discriminatory Analysis: Nonparametric Discrimination – Consistency Properties. No. 4.** *USAF School of Aviation Medicine.* 1951.
Reference Source
13. Breiman L: **Random Forests.** *Machine Learning.* 2001; **45**: 5–32.
Publisher Full Text
14. Chen T, Guestrin C: **XGBoost: A Scalable Tree Boosting System.** *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD'16).* 2016; pp. 785–794.
Publisher Full Text
15. Drucker H, Burges CJC, Kaufman L, *et al.*: **Support Vector Regression Machines.** *Advances in Neural Information Processing Systems 9 (NeurIPS 1996).* 1997; pp. 155–161.
Reference Source
16. Crammer K, Singer Y: **On the Algorithmic Implementation of Multiclass Kernel-Based Vector Machines.** *JMLR.* 2002; **2**: 265–292.
17. Ansel J, Yang E, He H, *et al.*: **PyTorch 2: Faster Machine Learning Through Dynamic Python Bytecode Transformation and Graph Compilation.** *Proceedings of the 29th ACM International Conference on Architectural Support for Programming Languages and Operating Systems.* Vol. 2 .New York, NY, USA: ASPLOS'24; 2024; pp. 929–947.
Publisher Full Text
18. Schmidhuber J: **Deep Learning in Neural Networks: An Overview.** *Neural Networks.* 2015; **61**: 85–117.
PubMed Abstract | **Publisher Full Text**
19. Defazio A, Yang XA, Khaled A, *et al.*: **The Road Less Scheduled.** *The Thirty-Eighth Annual Conference on Neural Information Processing Systems.* 2024.
Reference Source
20. Collins G, Moons K, Dhiman P, *et al.*: **TRIPOD+AI Statement: Updated Guidance for Reporting Clinical Prediction Models That Use Regression or Machine Learning Methods.** *BMJ.* 2024; **385**: e078378.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

Open Peer Review

Current Peer Review Status: ? ✗

Version 2

Reviewer Report 07 October 2026

https://doi.org/10.21956/openreseurope.27379.r82318

© **2026 Cangelosi G.** This is an open access peer review report distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

✗

Giovanni Cangelosi
Polo Medicina Sperimentale e Sanità Pubblica, Camerino, Italy

Dear Authors,
The relevance of the topic and the support received through international funding warrant an in-depth scientific discussion of the concerns raised. However, the response letter does not fully address several issues identified in the previous review and offers interpretations that appear subjective rather than sufficiently precise and responsive to the specific comments. In several instances, the responses appear to narrow the scope of the concerns without adequately discussing their methodological and interpretative implications, thereby limiting the effectiveness of the peer-review process. In particular, significant concerns remain that hinder a comprehensive assessment and appreciation of the manuscript's scientific contribution. The title still does not clearly identify the study design. This ambiguity extends to the main text, making the objectives and methodological framework less readily understandable. Reference to the TRIPOD recommendations suggests a study involving a prediction model but does not, in itself, unequivocally establish the design adopted. The abstract remains excessively long and insufficiently balanced in its allocation of content. The background occupies a disproportionate amount of space relative to the methodological information and interpretation of the findings. Conversely, the conclusions require a more informative formulation that clarifies the potential implications for clinical practice without extending beyond what the data support. The methods require a clearly identifiable section and a structure that allows readers to assess the alignment between the research question, study design, procedures, and analyses. In its current form, the fragmented presentation of methodological information limits the transparency and verifiability of the research process. The placement of Sections 9.1–9.3 after the conclusions also requires revision. Given the nature of their content, Section 9.1 should be incorporated into the Methods section, which is currently not explicitly delineated in the manuscript, whereas Sections 9.2 and 9.3 should be placed after the Discussion and before the Conclusions. This reorganisation would improve the logical progression of the manuscript and allow the conclusions to serve as its final synthesis. The relevance of the content and the underlying project is not in question. These concerns relate exclusively to the scholarly presentation of the work, which does not yet meet the applicable international standards.

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: Public Health

I confirm that I have read this submission and believe that I have an appropriate level of expertise to state that I do not consider it to be of an acceptable scientific standard, for reasons outlined above.

Version 1

Reviewer Report 29 January 2026

https://doi.org/10.21956/openreseurope.23620.r67738

© **2026 Cangelosi G.** This is an open access peer review report distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Giovanni Cangelosi
Polo Medicina Sperimentale e Sanità Pubblica, Camerino, Italy

Dear Authors,

The topic addressed in the manuscript is unquestionably timely and of high scientific relevance, fully aligned with the thematic priorities promoted by the Europe programmes, and therefore deserves appropriate attention and a high level of methodological and conceptual rigor. However, from the perspective of academic scholarship and international scientific communication, the manuscript presents numerous and substantial shortcomings that warrant an almost complete re-evaluation of the work.

First, key methodological elements of strong scientific impact are lacking. A structured and robust description of the research methodology is missing, beginning with a clear identification of the study design, which is not evident either from the title or from the subsequent sections of the manuscript. The title itself requires substantial revision, as it lacks clarity and makes use of non-standardized or unexplained acronyms, thereby hindering immediate comprehension by an international readership.

The abstract is excessively lengthy and redundant; alignment with international editorial standards is strongly recommended, with an indicative length of 200–250 words. Similarly, the number of keywords should be limited to a maximum of 4–5, preferably including the study design, in order to improve indexing and discoverability.

The introduction is deficient in several respects: it lacks a clear epidemiological framework of the phenomenon under investigation, and the research objectives are not stated explicitly and unequivocally. Furthermore, relevant clinical and care-related references are absent, which is

particularly critical if the study aims to have implications for clinical practice or healthcare systems. Finally, the concluding sentence of the introduction does not conform to conventional scientific writing standards and should be reformulated in a more formal manner, clearly oriented toward the presentation of the study objectives.

The methods section is not adequately detailed. In the context of international scientific dissemination, the adoption of an established reporting guideline (such as those promoted by the EQUATOR Network) is strongly recommended to ensure transparency, reproducibility, and methodological robustness in the presentation of the research.

The results appear to constitute one of the strengths of the manuscript; however, a truly accurate and reliable assessment is only possible once the previously highlighted methodological and structural issues have been adequately addressed.

Unless constrained by specific editorial requirements, the limitations section should be presented after the discussion and not anticipated. In the current version of the manuscript, the discussion is effectively absent and is inappropriately incorporated into the conclusions, which is not consistent with standard practices in scientific reporting. It is therefore recommended to develop a separate, critical, and well-argued discussion section, more closely oriented toward comparison with the existing literature and toward the scientific and clinical implications of the findings.

Finally, the reference list appears largely outdated, with numerous citations older than five years, representing a significant limitation in light of widely accepted academic recommendations. Moreover, the bibliography provides limited insight from a clinical practice perspective and offers insufficient support for the applicability of the study findings.

In conclusion, while acknowledging the relevance of the topic addressed, the manuscript requires a thorough and systematic revision in order to meet the scientific and methodological standards expected of a European and international-level publication.

Is the work clearly and accurately presented and does it cite the current literature?
No

Is the study design appropriate and does the work have academic merit?
No

Are sufficient details of methods and analysis provided to allow replication by others?
No

If applicable, is the statistical analysis and its interpretation appropriate?
Yes

Are all the source data underlying the results available to ensure full reproducibility?
Partly

Are the conclusions drawn adequately supported by the results?
No

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: Public Health

I confirm that I have read this submission and believe that I have an appropriate level of expertise to state that I do not consider it to be of an acceptable scientific standard, for reasons outlined above.

Author Response 17 Sep 2026

Flore N'kam Suguem

We thank the reviewer for the detailed and rigorous assessment, and for recognising the timeliness and scientific relevance of the work and its alignment with European research priorities. In response, we have substantially restructured the manuscript to bring it into line with international reporting and editorial standards. We address each point below; section references correspond to the revised version.

1. Study design and methodological structure.
We agree that the study design should be stated explicitly. We have added a dedicated "Study Design and Reporting Guidelines" section that identifies the work as a retrospective observational study aggregating and analysing the nutritional profiles of packaged foods associated with the Health Star Rating (HSR), together with a machine-learning prediction component. The design is now stated unambiguously rather than left implicit in the title and body.

2. Title.
We have carefully considered the reviewer's comment. We have retained the title because it incorporates the name of the dataset we created and release, OpenHSR, and expands its only acronym within the title itself, as "the health star rating system." No undefined or non-standardised acronym therefore remains in the title, which we believe accurately identifies both the resource and its subject. We would of course be glad to revise it further at the editor's discretion.

3. Abstract length.
We appreciate the reviewer's attention to concision and indexing. The abstract follows the Open Research Europe structured format (Background, Methods, Results, Conclusions) and, at approximately 280 words, falls within the platform's stated limit of up to 300 words for a structured abstract. We have therefore retained the current length so as not to lose the structured content the platform requires, while reviewing the text to remove avoidable redundancy. We remain happy to condense it further should the editorial office prefer.

4. Keywords.
The keywords conform to the Open Research Europe guidance, which permits up to eight relevant terms. We have reviewed the set to ensure the terms accurately describe the subject and support discoverability. We would gladly adjust the number or wording at the editor's discretion.

5. Introduction.
We have substantially rewritten the Introduction to address the reviewer's concerns. It now opens with an epidemiological framing of diet-related chronic non-communicable diseases, specifically obesity, cardiovascular disease, and type 2 diabetes, as a clinical and public-health burden, supported by relevant references (WHO, 2022; Song et al., 2021). We have positioned front-of-pack labelling and the HSR system as preventative, population-level interventions, making the clinical and policy relevance explicit, and we have stated the study objectives explicitly in the paragraph introducing OpenHSR.

6. Methods and reporting guideline (EQUATOR Network).
We agree that adopting an established reporting guideline strengthens transparency and reproducibility. In the new "Study Design and Reporting Guidelines" section we now state that the development and evaluation of the prediction models were guided by the TRIPOD statement, an EQUATOR Network guideline for the transparent reporting of multivariable prediction models, applied here to nutritional dataset benchmarking. In addition, the Methods (Section 6) now report the full modelling procedure to allow replication: an 80/20 train/test split, hyperparameter selection by grid search with scikit-learn's 5-fold cross-validation, a further 64/16 training/validation split for the neural models, selective feature standardisation according to each algorithm's sensitivity to scaling, and a complete list of the non-default hyperparameters used for every model.

7. Results.
We thank the reviewer for recognising the results as a strength of the manuscript. We hope that, with the methodological and structural revisions now in place, the results can be assessed on the firmer footing the reviewer indicated was needed.

8. Discussion.
We agree that the discussion should stand on its own. We have added a separate, critical Discussion section that compares OpenHSR with existing resources (Open Food Facts, FoodSwitch, and AUSNUT) with respect to openness and FAIR compliance, interprets the benchmark results (why kernel- and ensemble-based methods perform well on a small tabular dataset while deep and fine-tuned models underperform at n = 246), and discusses the implications for independent reproducibility and data-driven nutrition and food-labelling policy. The Discussion is now clearly separated from the Conclusion.

9. Placement of the Limitations.
As recommended, the Limitations now follow the Discussion rather than preceding it, consistent with standard scientific reporting structure.

10. Reference list.
We have updated and broadened the references to include recent and clinically relevant sources, among them WHO (2022), Song et al. (2021), the TRIPOD statement (Collins et al., 2024), and current methodological and implementation sources (Pedregosa et al., 2011; Defazio et al., 2024; Ansel et al., 2024). We have retained the foundational references for the classical algorithms (e.g., least squares, k-nearest neighbours, and classification and regression trees), as citing the originating works is standard convention in the machine-learning literature; these are now complemented by the modern implementation

references so that the actual methods used are properly documented. We are grateful to the reviewer for guidance that has led to a substantially more rigorous, better structured, and more clearly reported manuscript, and we hope the revised version now meets the standard expected for publication.

Competing Interests: No competing interests were disclosed.

Reviewer Report 19 January 2026

https://doi.org/10.21956/openreseurope.23620.r67739

© 2026 Bhuiyan M. This is an open access peer review report distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

?

Mohammad Nazrul Islam Bhuiyan
BCSIR Chattogram Laboratories, BCSIR, Chattogram, Bangladesh

Overall Assessment: The article presents a valuable and timely contribution by creating the first open, FAIR-compliant dataset for the Health Star Rating (HSR) system. The work addresses a clear gap in nutritional research. However, the manuscript requires major revisions due to significant methodological inconsistencies, presentation issues, and a lack of critical validation before it can be considered for indexing. It is not yet in an acceptable state.

Specific Mistakes, Inconsistencies, and Needed Improvements
1. Abstract & Results Inconsistency:

Lines 26-28 (Abstract): "Machine learning models achieved strong predictive performance, with the Support Vector Regression model yielding the best overall accuracy (Mean Squared Error = 0.28)."
Issue: The term "accuracy" is incorrectly used for a regression task. Mean Squared Error (MSE) is a loss metric, not accuracy. The authors later define "accuracy" as rounding to the nearest 0.5-star, but this is not standard and is conflated in the abstract. This should be rephrased to state "best performance" or "lowest error."

2. Sample Size and Methodology:

Lines 115-116: "The OpenHSR dataset comprises 246 unique food products..."
Issue: A sample size of n=246 is extremely small for training and comparing eight machine learning models, especially neural networks and foundation model fine-tuning. This severely limits the generalizability and reliability of the benchmark results (Table 1). The paper must explicitly highlight this as a critical limitation and temper claims about model performance. The statement "Performance could likely be improved... with access to additional labeled data" is an understatement.

3**. Incomplete and Confusing HSR System Description:**

Section 2: The description of the HSR algorithm is fragmented and confusing. It lists six categories (1, 1D, 2, 2D, 3, 3D) and then states they are "reclassified into two different categories." The logic and necessity of this two-step process are not clearly explained for a reader unfamiliar with the system. A clearer, more linear explanation of the official algorithm is needed.

4. Lack of Critical Validation in Key Experiment:

Section 7 (Health star rating predictions): This is the most significant flaw.

The authors apply their model to 6,554 Open Food Facts products without any ground-truth HSR labels.

The only "validation" presented is that the predicted distribution "appears well diversified" (Fig 4). This is not scientific validation. It is a circular argument; the distribution's shape does not prove the predictions are correct.

This section should either be removed or reframed as a hypothesis-generating or methodological demonstration with explicit, strong caveats about the complete lack of validation. Statements about "reliability and generalizability" (Line ~210) are unsupported by the presented evidence.

5**. Presentation and Referencing Errors:**

Table 1 & References: The table lists algorithms like "Linear (Legendre, 1805)" and "K-Nearest Neighbor (Fix, 1951)." Citing century-old foundational papers for standard implementations is inappropriate and obscures the actual libraries/methods used (e.g., scikit-learn). References should point to contemporary, relevant methodological sources or the implementation documentation.

Figure 3 Reference: The text references "Figure 3" (the decision tree) but provides no interpretation or analysis of what the visualized tree shows about feature importance in HSR prediction, missing an opportunity for the "interpretability" discussed.

Grammar/Clarity: Several sentences are awkward or unclear.

Line ~58: "The HSR are computed..." -> "The HSR is computed..."

Line ~209: "supporting the reliability and generalizability of our HSR estimation approach" -> This claim is too strong, as noted above.

Line 72-73: "all of which contributes" -> "all of which contribute"

6. **Limitations Section - Adequacy:**

Section 8 (Limitations): The authors correctly identify the main limitations: small dataset size, regional focus (Australia), manual collection, and lack of serving-size variability. This is good. However, they must strengthen the first point regarding the profound impact of n=246 on the machine learning benchmark. The limitation regarding the lack of validation for the Open Food Facts predictions should also be added here.

1. Clarify Methodology: Rewrite the HSR system description for clarity. Correct the misuse of "accuracy" in the abstract and results.
2. Address Sample Size Critique: Prominently discuss the severe constraint of the small dataset (n=246) on the machine learning benchmark, reframing conclusions to be more tentative.
3. Reframe or Remove Unvalidated Predictions: Entirely rework Section 7. Remove claims of "reliability and generalizability." Present it strictly as a proof-of-concept application with explicit warnings that the predictions are unvalidated. Ideally, the authors should attempt to find a small

subset of Open Food Facts products with true HSR labels for a minimal validation.
4. Correct Presentation: Update Table 1 references to modern sources. Provide brief analysis of Figure 3. Fix grammatical errors.
5. Strengthen Limitations: Add the lack of validation for the large-scale predictions to the limitations section and emphasize the sample size issue.
The core idea—creating an open HSR dataset—is excellent and FAIR-aligned. However, in its present state, the manuscript contains a major methodological overreach (Section 7) and underestimates the impact of its primary limitation (small n). These issues fundamentally weaken the study's conclusions. Once these are robustly addressed, the article will make a solid contribution.

Is the work clearly and accurately presented and does it cite the current literature?
Partly

Is the study design appropriate and does the work have academic merit?
Partly

Are sufficient details of methods and analysis provided to allow replication by others?
Partly

If applicable, is the statistical analysis and its interpretation appropriate?
Partly

Are all the source data underlying the results available to ensure full reproducibility?
Partly

Are the conclusions drawn adequately supported by the results?
Partly

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: Food Science and Technology

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard, however I have significant reservations, as outlined above.

Author Response 17 Sep 2026

Flore N'kam Suguem

We thank you for the thorough and constructive assessment, and for recognising the value of the first open, FAIR-compliant HSR dataset. We have carried out major revisions addressing each point below; section references correspond to the revised manuscript.

1. Misuse of "accuracy" in the Abstract and Results.
We agree. In the Abstract we no longer describe MSE as "accuracy"; the sentence now reads

that the Support Vector Regression model yielded the best Mean Squared Error (= 0.28). In Table 1 and its caption we have replaced "accuracy" and "lenient accuracy" with "exact match rate (acc)" and "lenient match rate (by-half-acc)", and we now state explicitly that these are match rates after rounding to the nearest 0.5-star increment, keeping MSE as the loss/error metric throughout.

2. Small sample size (n = 246) and its impact on the ML benchmark.
We agree this is a critical constraint and have stopped understating it. The new Discussion section now states that the low MSE of the strongest models reflects optimised loss on a small training set rather than guaranteed generalisation, and that the underperformance of the neural and fine-tuned models is "entirely driven by the severe constraint imposed by our sample size" of 246 products. We have tempered the performance claims accordingly and retained the sample size as the first point in the Limitations.

3. Fragmented HSR system description.
Section 2 has been rewritten as a clear, linear description. The algorithm first assigns each product to one of six primary categories (1, 1D, 2, 2D, 3, 3D), then routes it into one of two calculation pathways: Pathway A (categories 1D, 2, 2D, 3, 3D) using four negative components (energy, saturated fat, sodium, total sugars), and Pathway B (category 1, non-dairy beverages) using two (energy, total sugars), before adjusting for the positive components (protein, fibre, FVNL). We have added the official calculator reference.

4. Lack of validation in Section 7 (predictions on Open Food Facts).
We agree this was an overreach and have reframed the section as an explicitly exploratory application. We removed the claim that the diversified distribution supported the "reliability and generalizability" of the approach. The text now states that, because no ground-truth HSR labels were available for these Open Food Facts products, the predicted values cannot be quantitatively validated and should be interpreted as exploratory estimates. Figure 4 is now presented only as an illustration of the predicted distribution, and the remaining claim is limited to computational efficiency (6,554 predictions in ~0.1 s).

5. Presentation and referencing.

- *Table 1 references.* The table now identifies algorithms by name and abbreviation only (LR, DT, kNN, RF, XGBoost, SVM, MLP, FTN). In the accompanying text we cite the actual implementations used, including scikit-learn (Pedregosa et al. 2011), alongside the original method references, so the century-old citations no longer stand in for the implementation.
- *Figure 3 interpretation.* We added an interpretation of the decision tree: it identifies energy and sugar content as the primary splitting criteria, consistent with the official HSR algorithm in which energy and total sugars are the primary negative components. (Please note we will correct the internal cross-reference so this paragraph points to Figure 3.)
- *Grammar/clarity.* Corrected, including "The HSR is computed"; the positive-components sentence has been rewritten, resolving the "all of which contributes" agreement issue, and we removed the overstated "reliability and generalizability" wording.

6. Adequacy of the Limitations.

We retained the four original limitations and strengthened the first: the new Discussion makes the impact of n = 246 on the benchmark explicit, and the lack of validation for the large-scale Open Food Facts predictions is now stated directly in Section 7. We believe these revisions address the methodological overreach and the underweighting of the sample-size limitation the reviewer identified, and we thank them again for feedback that has materially improved the manuscript.

Competing Interests: No competing interests were disclosed.