JOURNAL OF COMPUTER SCIENCE AND INFORMATICS ENGINEERING
VOL. 5, NO. 4 (2026)
DOI: https://doi.org/10.55537/cosie.v5i4.1986

e-ISSN 2827-8356
p-ISSN 2827-9999

Explainable Machine Learning for Food Allergen Detection Using Tree-Based Ensembles

Bakti Amirul Jabar[1], Richard Edsel Wijaya[2], Martin Halim[3], Hansel Mitchell Tumanggor[4]
[1,2,3]*Bina Nusantara University, Jakarta, 11480, Indonesia*
[4]*National Taiwan University of Science and Technology, Taipei, 106335, Taiwan*
[1] bakti.jabar@binus.ac.id, [2]richard.wijaya007@binus.ac.id,[3]marthin.halim@binus.ac.id,
[4]hansel.tumanggor@binus.ac.id

Received: October 8, 2026 | Revised: October 11, 2026 | Accepted: October 30, 2026 | Published: November 9, 2026

COSIE is licensed *a Creative Commons Attribution-Share Alike 4.0 International License*

Abstract

Food product allergen annotations are relevant to ingredient-related classification, although dataset-label prediction must be distinguished from laboratory-confirmed allergen detection. This study compares four tree-based machine learning models XGBoost, AdaBoost, Decision Tree, and Random Forest using a secondary food ingredient dataset containing 399 records from Kaggle. An 80:20 training–testing split was used, and models were assessed using accuracy, precision, recall, and F1-score. In the reported 80-record test set, Random Forest achieved the highest accuracy (0.86), with a class-1 F1-score of 0.91, a macro-average F1-score of 0.82, and class-0 recall of 0.65. SHapley Additive exPlanations (SHAP) were used to interpret Random Forest predictions; fat/oil attributes, including olive oil and butter, were prominent model-associated features. These associations describe patterns learned from the dataset rather than causal or biologically verified effects. The findings should be viewed as preliminary, dataset-specific classification results; external validation and additional evaluation are required before any food-safety application.

Keywords: Food Allergens; Machine Learning; Classification; XAI; SHAP.

1. INTRODUCTION

Food allergy is one of the global health problems that can seriously affect, even fatal to individuals who have sensitivity to certain substances in food. According to data from the World Health Organization (WHO), the prevalence of food allergies continues to increase every year, especially in children, adolescents, and even adults. However, most of the available data comes from self-report, hospital data, or population sensitization surveys that do not always reflect the true incidence due to potential overdiagnosis or underdiagnosis. The risk of food allergy itself can be said to be low to high ranging from anxiety, skin problems, breathing, fainting, decreased blood pressure, and even death. And as a result, author concludes that early and reliable detection of allergens in food products is critical to avoiding potential health hazards.

Food allergies are categorized into two kinds depending on the sensitization pathway: class 1 and class 2 allergies. Class 1 food allergies are typically caused by a loss of oral tolerance to digestion-stable dietary proteins, and they frequently emerge during childhood when the immune system is still not fully complete. In contrast, class 2 food allergy generally results from sensitization via the respiratory route to unstable proteins such as pollen. To this date, the process of identifying food allergens generally still relies on traditional methods such as Blood Tests, Paste Tests, or Elimination Diets, which require long time, high cost, and specialized technical expertise. In the digital and big data era, machine learning-based approaches have begun to be widely used to assist classification and prediction processes in various fields, including in the food sector. However, the implementation of various machine learning methods in detecting allergens still needs to be studied more deeply and has not been widely applied in the medical world. The main problems addressed in this research are how machine learning models such as Decision Tree and Random Forest detect allergens in food products, and which model provides the most optimal results based on parameter outcomes.

The objectives of the research are as follows: to evaluate and compare the performance of several machine learning algorithms in detecting allergens, to identify the best model that can serve as the basis for an automatic allergen detection system, and to contribute to the development of machine learning-based allergen detection technology.

Due to the ability of machine learning techniques to find patterns in large amounts of data and to accurately predict based on that data, these techniques have received a great deal of attention in relation to healthcare and food safety in particular. Machine learning has been utilized in relation to food safety in regard to analyzing food quality, nutrition, diseases, and allergens. Despite many studies applying machine learning in the health and food fields, there is limited research on systematically comparing the performance of multiple machine learning models for allergen detection.

Previous research has compared machine learning methods for food product classification and evaluated validation strategies on closely related ingredient datasets [7], [9]. Accordingly, the contribution of this study is not the first application of machine learning to these records, but a focused comparison of four tree-based classifiers with global and local SHAP interpretation of Random Forest predictions. The analysis is limited to labels supplied in the dataset and does not independently verify food allergens through clinical or laboratory measurements. Wang et al. [6] proposed the use of deep learning model BERT (Bidirectional Encoder Representations from Transformers), the food allergen dataset adopted in this study comes from Allergen Nomenclature, Structural Database of Allergenic Proteins (SDAP), National Center for Biotechnology Information (NCBI) and 583 food allergens were officially reported. By proposing LightGBM and XGBoost models to predict the allergenicity of food proteins with very good results from previous studies, among the models tested the model achieved an AUC of 0.9578 and accuracy reached 0.9310. This project evaluated various machine learning algorithms using different metrics with the goal of creating a system to analyze and predict food allergies with a dataset found on Kaggle with records of 400 food products [7]. With a combination of systematic preprocessing methods and appropriate setting modifications, the Decision Tree method obtained the best accuracy of 0.984.

The study had a total of 496 datasets from the Oral Food Challenge (OFC) data collection method taken based on one of Japan's national hospitals (309 males, 187 females) with an average age of 3 years [8]. Using the training dataset, and the LGBM model, the accuracy value, negative predictive value, and positive predictive value for predicting positive OFC were 0.9, 0.94, and 0.5, respectively. Narulita et al. [9], carried out research conducted on the Kaggle dataset entitled food ingredients and allergens with 399 record data conducting research experiments around validation techniques, namely 10-fold cross validation, split validation 90:10, and split validation 80:20, this study provides new insights that show that validation methods can change machine learning results with a significant scale and for their case split validation 80:20 provides optimal results. Sharma et al. [10]. This study addresses the limitations of previous work by proposing a comprehensive AI-based system that integrates various models, including NLP, GNN, supervised classifiers, and geospatial analytics, into a unified recommendation engine. The system aims to provide not only accurate allergen detection but also practical, personalized, and safe dietary guidance. Nedyalkova et al. [11]. Used a comprehensive approach to predict whether a protein from a plant could cause allergies. They linked many sorts of protein information, such as amino acid composition, structure, and chemical characteristics, with evolutionary data. Consequently, the model predicted allergenic proteins with more accuracy and reliability than other methods such as k-nearest neighbors (KNN) and naive Bayes. Ahmad et al. [12], believed that deep learning approaches such as CNN and RNN are gaining popularity due to their capacity to handle complicated data, despite their interpretability limitations and high processing requirements. To overcome this, ensemble learning models such as Bagging and Boosting have been shown to improve prediction performance. The hybrid ensemble model even showed the best performance with an accuracy of up to 96.2%. Zhang et al. [13], continued the effort by developing a predictive model based on data from 1,112 patients who underwent 1,284 Oral Food Challenges (OFC) on peanuts, eggs, and milk. They compared five algorithms, namely Naive Bayes, Logistic Regression, SVM, Random Forest, and LUCCK. The best results were achieved by the ensemble models Random Forest (for eggs) and LUCCK (Learning Using Concave and Convex Kernels) (for peanuts and milk), with AUC values reaching 0.96, 0.91, and 0.94, respectively.

Research by Bifarin [14], demonstrates that tree-based machine learning models enhanced with SHAP (SHapley Additive exPlanations) have proven effective in improving the interpretability of predictions in metabolomic data; the authors demonstrate that applying Tree SHAP to algorithms such as Random Forest and XGBoost can transparently identify the contribution of each metabolite in binary classification, thereby bridging the gap between the accuracy of black-box models and clinical interpretability. Visualization approaches using summary diagrams and dependency diagrams allow biomedical researchers to interpret prediction results in a biologically meaningful way, thereby directly creating a relevant framework for studies requiring predictive models that are also interpretable in the fields of biomedicine and health. Authors Borisov et al. [15] note that deep learning and tree-based models have emerged as the two dominant approaches to solving the problem of tabular data analysis, with varying degrees of success. Based on various surveys of the field, Borisov and his colleagues find that tree-based

DOI: https://doi.org/10.55537/cosie.v5i4.1986

models such as XGBoost, LightGBM, and Random Forest tend to outperform deep learning models - despite the fact that deep learning models are much more advanced - on tabular datasets of small and large sizes. The tree-based models can efficiently handle various types of features within the data, they can handle instances of imbalanced data, and they are inherently resistant to overfitting. Deep learning models such as TabNet, NODE, and transformer models tend to outperform tree models on large datasets that have complex feature patterns, but at the cost of requiring increased computational expenses to process the data with these models, and requiring more effort to tune the hyperparameters of those models. Borisov and his colleagues find that the choice of which models to use in analyzing data can be informed by the attributes of the data and the requirements of the scientific investigation that is to be performed with that data; Borisov himself concludes that tree-based models combined with methods of explaining the models, such as SHAP, are the primary choice of models for scientific fields that intend to make the results of those models reproduce.

Studies reviewed above address distinct tasks and units of analysis. Protein-sequence allergenicity prediction [6], [11] and patient-level oral food challenge outcomes [8], [13] should not be directly benchmarked against classification of ingredient-annotated product records. The most directly related work is product-record classification [7], [9], whereas research on SHAP [14] and tabular tree-based models [15] informs the interpretation and model choice in this study. This distinction limits cross-study performance comparisons and clarifies the incremental methodological contribution.

2. MATERIALS AND METHODS

2.1. Dataset

Information about the allergen dataset used to train the food model has been compiled in the “Food Ingredients and Allergens” dataset on Kaggle [16]. Potential allergens in different foods and items can be identified and analyzed thanks to the dataset's allergen information for various food ingredients. This dataset consists of a total of 399 records with each record representing a specific food and its associated allergen. Also, the list includes all allergens included in the foodstuff and includes foods that are not included in a specific allergy list. The data structure includes the name, food ingredient, allergen, and prediction (the food contains the product or not).

Target definition and evaluation scope. Table 1 describes Prediction as a binary label ('Contains' or 'Does not contain') determined by whether the Allergens field contains an annotation. Thus, the modeled outcome represents the supplied dataset label, not independently confirmed allergen presence or an individual's clinical allergic response. Since Allergens directly defines the target, it and any target-derived variables must not be used as classifier predictors. The manuscript does not document the exact training feature matrix; therefore, exclusion of potential target leakage cannot be verified from the present report and remains a reproducibility limitation.

Table 1. Data Description

Feature	Description
Food Product	The name of the food product
Main Ingredient	The main ingredient of the food product
Sweetener	The type or presence of sweeteners utilized in the food product
Fat/Oil	The type or amount of fat or oil used in the food product
Seasoning	Any seasonings or spices added to enhance the flavor of the food product
Allergens	Allergens were identified using ingredient lists and categorized into standardized groups: dairy, wheat, nuts, eggs, among others
Prediction	Food product labels are classified based on their allergen and ingredient information. If the allergens feature contains a value, the predicted label is “Contains”; otherwise, if empty, the predicted label value is “Does not contain”

2.2. Data Visualization

The first step in visualizing was to convert the dataset to Data Frame format using the pandas package, and then the relevant columns such as Allergen, Sweetener, Fat/Oil, and Spice were highlighted to indicate missing values and ‘None’. Values separated by commas (,) such as in the Allergen column

DOI: https://doi.org/10.55537/cosie.v5i4.1986

were extracted and exploded into single values so that their frequencies could be calculated individually. This procedure was performed to obtain useful visualizations, such as bar charts and pie charts, which facilitated the analysis of trends such as allergies, sweeteners, fats/oils, and seasonings that frequently appear in food products. The data that has been collected and analyzed is now used to create frequency visualizations, provide more information about food composition, and identify allergens.

The comma-separated Allergen entries were expanded for frequency analysis as described above. The corresponding classification results are reported for 399 source records, not for exploded allergen entries. However, the report does not specify a complete audit of duplicate entries, the treatment of missing values versus the literal category 'None', or the exact categorical encoding used to construct model inputs. These pipeline details are necessary for full replication and should be verified against the original implementation; no unreported preprocessing procedure is assumed here.

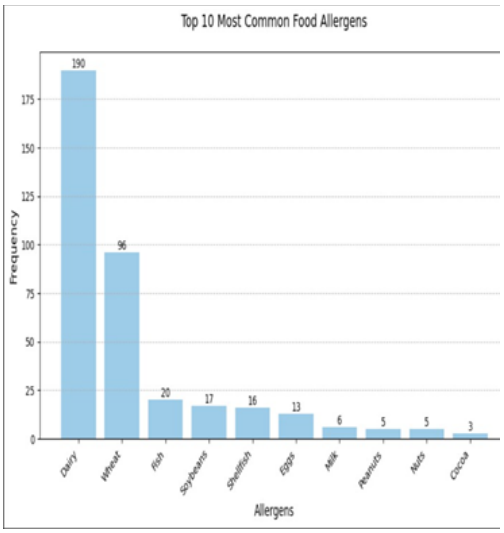


Figure 1. Most Common Food Allergens

Figure 1 shows the predominant allergens throughout the dataset. The graphics use bar charts to depict the most prevalent food allergies, with dairy ranked first, eggs sixth, and peanuts eighth.

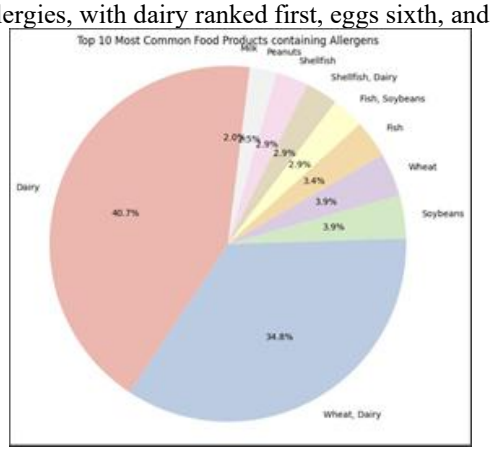


Figure 2. Most Common Food Products

Figure 2 is a pie chart that shows the distribution of the Top 10 allergens, but only for food products that are projected to include allergens based on the "Contains" value in the Prediction column. The goal is to determine which allergens appear most frequently, particularly in products that represent a risk to allergy patients.

DOI: https://doi.org/10.55537/cosie.v5i4.1986

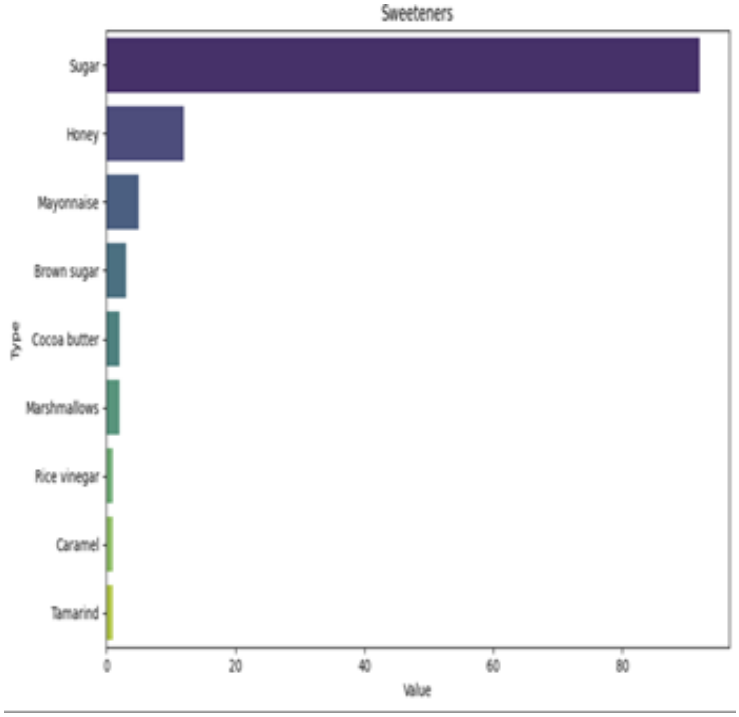


Figure 3. Most Common Sweeteners

Figure 3 focuses on the sweetening ingredients used in food products and the results are represented by graphs depicting the most widely used sweeteners, such as sugar, fructose, and artificial sweeteners as examples, and this visualization also explains and illustrates the trends in the use of sweeteners in the food business, as well as what the potential impacts on customer preferences and health impacts.

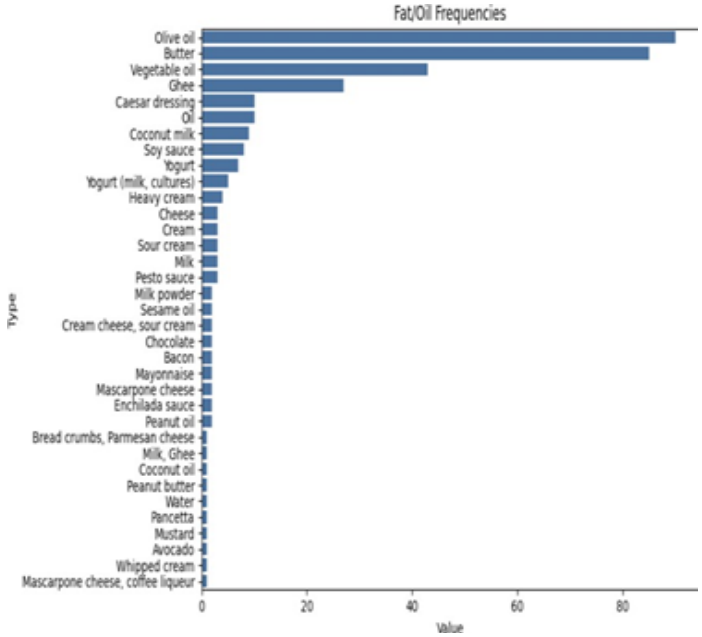


Figure 4. Most Common Fats/Oils

Figure 4 displays the use of the type of fat or oil listed in the Fat/Oil column, visualized in a bar graph displaying the commonly used ingredients. The purpose of this visualization is to provide an understanding of the patterns of fat consumption in food products that will be used for nutrition and allergy-based studies.

DOI: https://doi.org/10.55537/cosie.v5i4.1986

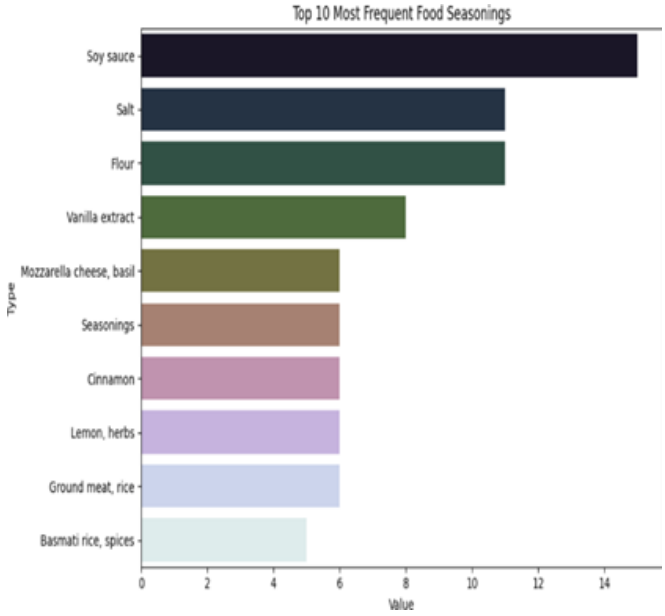


Figure 5. Most Common Food Seasonings

Figure 5 depicts the analysis of the ten most frequently used condiments displayed in a bar graph. This visualization provides information about frequently used seasoning ingredients and their potential to induce allergies or specific taste preferences.

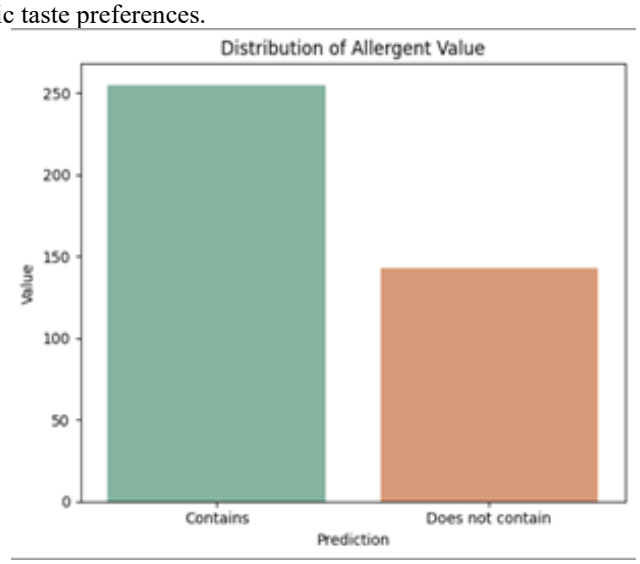


Figure 6. Distribution of Allergen Value

Figure 6 shows the distribution of food product allergen prediction categories, taken from the Prediction column. This bar graph consists of two main categories: “Contains” and “Does not contain”, which represent the number of food products predicted to contain allergens or not, respectively.

One of the main drawbacks of the Food Ingredients and Allergens dataset is the potential imbalance of the data, both in terms of prediction labels and the distribution of values in the Allergen column. The distribution of Prediction labels shows that the number of products predicted to contain allergens (Contains) is much higher than products that do not contain allergens (Does not contain). This imbalance can cause the prediction model to be biased towards the majority class, making it less able to detect products that do not contain allergens. In addition, the values in the Allergen column show an uneven distribution where some types of allergens such as milk, eggs, and peanuts dominate the frequency.

2.3. Machine Learning Classification Models

Decision Tree is a tree-based classification algorithm that works by dividing the data into branches based on the most informative features. Decision trees are easy to interpret, but prone to overfitting as a drawback. Random Forest is an extension of Decision Tree that works by creating multiple decision trees and combining their predictions using bagging (bootstrap aggregating). XGBoost (Extreme Gradient Boosting) is a boosting algorithm that works by incrementally modeling where each new model tries to

DOI: https://doi.org/10.55537/cosie.v5i4.1986

correct the errors of the previous model. AdaBoost (Adaptive Boosting) forms the final prediction model from a combination of several weak models.

2.4. Training and Testing

The dataset is divided into a ratio of 8:2 where 80% is used for training and 20% is used for testing. Each model uses a random state of 42. The class_weight="balanced" parameter from Scikit-learn is utilized in Decision Tree and Random Forest to neutralize the effect of class imbalance. Evaluation metrics include Accuracy, Precision, Recall, and F1-Score.

The reported hold-out test set includes 80 observations (23 in Class 0 and 57 in Class 1), with 319 records in the training partition. The manuscript does not establish whether splitting was stratified or describe additional model random seeds, exact tree and boosting hyperparameters, package versions, or any tuning procedure. Accordingly, the comparison is interpreted as the reported single-split evaluation rather than a repeated or fully tuned benchmark. A repeated stratified cross-validation study, with all model selection confined to training data, would be needed to quantify split-dependent variability without reusing the final test set for optimization.

2.5. Shapley Additive Explanations (SHAP)

SHAP is an explainable artificial intelligence (XAI) method based on Shapley game theory that calculates the Shapley value for each feature consistently to produce local and global explanations. Tree SHAP is applied to tree-based models to measure the magnitude and direction of each input variable's contribution. For interpreting signed SHAP values, the positive or negative direction must be referenced to a specified model output and class; a negative value means a reduction in the explained output relative to its baseline, not an intrinsic negative effect on allergen biology. The original experimental report does not specify the SHAP package version, the precise explainer configuration, or its reference output/background data. Accordingly, the directional interpretation of individual SHAP values should be treated as provisional until those settings are documented.

3. DISCUSSION

Four tree-based models AdaBoost, XGBoost, Random Forest, and Decision Tree—were evaluated for classification of the dataset-defined allergen labels. Tables 2–5 present the reported classification metrics on the common-sized test set of 80 observations. Similarity in test-set performance across algorithms does not establish an absence of overfitting, because corresponding training and repeated validation metrics are not reported. The results below therefore describe performance in this particular held-out partition.

Table 2. XGBoost Classification Report

XGBoost Model	Precision	Recall	F1-Score	Support
Class 0	0.70	0.70	0.70	23
Class 1	0.88	0.88	0.88	57
Accuracy			0.82	80
Macro Avg	0.79	0.79	0.79	80
Weighted Avg	0.82	0.82	0.82	80

Table 2 displays the performance of the XGBoost model for binary classification. For the first class (Class 0), the precision, recall, and F1-score for the model were each 0.70 with 23 samples classified by the model. For the second class (Class 1), each of these metrics reached 0.88 with 57 samples classified by the model. For both classes, accuracy is 0.82. For both classes, the macro average for precision, recall, and F1-score is 0.79. The weighted average for each of these metrics is 0.82, indicating that the model performed better with the larger group of samples, or Class 1.

Table 3. AdaBoost Classification Report

AdaBoost Model	Precision	Recall	F1-Score	Support
Class 0	0.62	0.65	0.64	23
Class 1	0.86	0.84	0.85	57
Accuracy			0.79	80
Macro Avg	0.74	0.75	0.74	80
Weighted Avg	0.79	0.79	0.79	80

Table 3 displays the classification performance for the AdaBoost model. For Class 0, the precision, recall, and F1-score for the model reached 0.62, 0.65, and 0.64, respectively, with 23 samples classified by the model. For Class 1, the precision, recall, and F1-score reached 0.86, 0.84, and 0.85, respectively, with

DOI: https://doi.org/10.55537/cosie.v5i4.1986

57 samples classified by the model. For both classes, the accuracy for the AdaBoost model is 0.79. For both classes, the macro average for precision, recall, and F1-score is approximately 0.74 and 0.75. The weighted average for each of these metrics is 0.79, indicating that the AdaBoost model performed better with the majority class of samples (Class 1).

Table 4. Decision Tree Classification Report

AdaBoost Model	Precision	Recall	F1-Score	Support
Class 0	0.79	0.65	0.71	23
Class 1	0.88	0.93	0.90	57
Accuracy			0.85	80
Macro Avg	0.83	0.79	0.81	80
Weighted Avg	0.85	0.85	0.85	80

Table 4 presents the classification performance of the Decision Tree model. For class 0, the model found precision of 0.79, recall of 0.65, and an F1-score of 0.71 with 23 samples classified. For class 1, the model found precision of 0.87, recall of 0.93, and an F1-score of 0.90 with 57 samples classified. For both classes, the model found an overall accuracy of 0.85. Furthermore, the macro average precision, recall, and F1-score values are 0.83, 0.79, and 0.81, respectively. Finally, the model found the weighted average precision, recall, and F1-score to each class to be 0.85, indicating its strong performance with the majority class.

Table 5. Random Forest Classification Report

AdaBoost Model	Precision	Recall	F1-Score	Support
Class 0	0.83	0.65	0.73	23
Class 1	0.87	0.95	0.91	57
Accuracy			0.86	80
Macro Avg	0.85	0.80	0.82	80
Weighted Avg	0.86	0.86	0.86	0.80

Table 5 summarizes all of the performance of the Random Forest model in binary classification. For Class 0, the model achieved a precision of 0.83, recall of 0.65, and F1-score of 0.73 based on 23 samples. For Class 1, the model showed higher performance with a precision of 0.87, recall of 0.95, and F1-score of 0.91 from 57 samples. The overall accuracy is 0.86. The macro average scores are 0.85 for precision, 0.80 for recall, and 0.82 for F1-score, while the weighted averages are consistently 0.86, indicating strong overall performance, particularly in detecting the majority class.

Among the four models, Random Forest produced the highest reported test accuracy (0.86), followed by Decision Tree (0.85), XGBoost (0.82), and AdaBoost (0.79). Random Forest achieved a class-1 F1-score of 0.91 and macro-average F1-score of 0.82, compared with a class-0 F1-score of 0.73 and class-0 recall of 0.65. Decision Tree achieved a class-1 F1-score of 0.90. These results indicate stronger performance on the majority class than on the minority class; the differences should not be generalized beyond this small single-split evaluation without repeated validation.

The test set is imbalanced, with 57 records in Class 1 and 23 in Class 0. Decision Tree and Random Forest were configured with class_weight='balanced', whereas this weighting is not reported for AdaBoost or XGBoost. This was a modeling choice intended to account for unequal class frequencies. As no controlled weighted-versus-unweighted comparison is provided, the present results do not establish that class weighting improved recall or stabilized accuracy. The reported Class 0 recall of 0.65 for Random Forest highlights the remaining minority-class classification limitation.

DOI: https://doi.org/10.55537/cosie.v5i4.1986

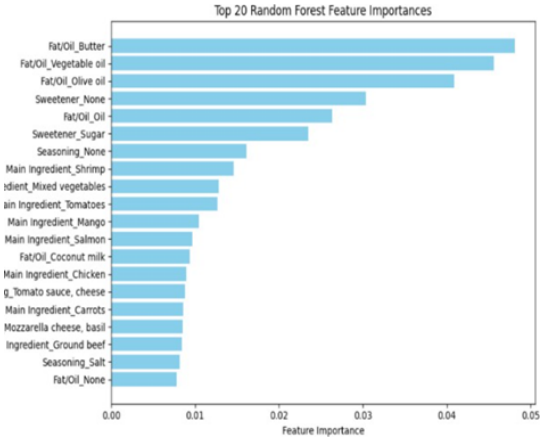


Figure 7. Random Forest Feature Importances

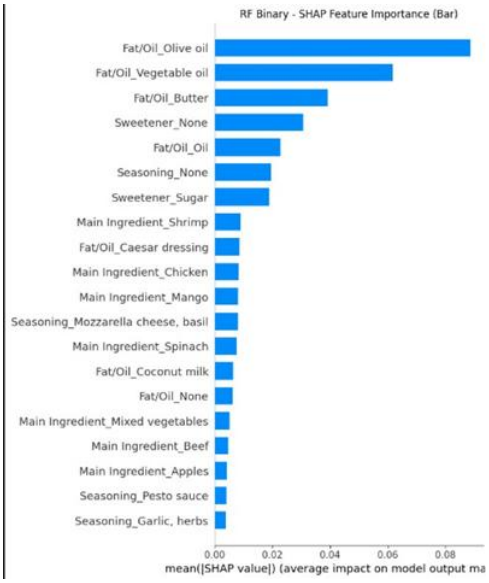


Figure 8. Random Forest SHAP Feature Importances

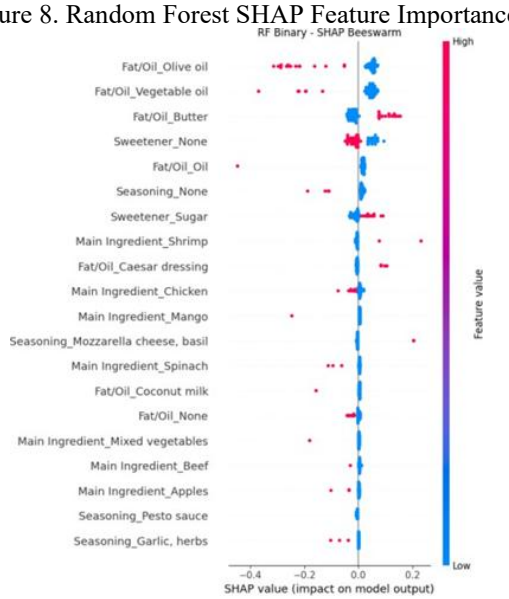


Figure 9. Random Forest SHAP Beeswarm Plot

DOI: https://doi.org/10.55537/cosie.v5i4.1986

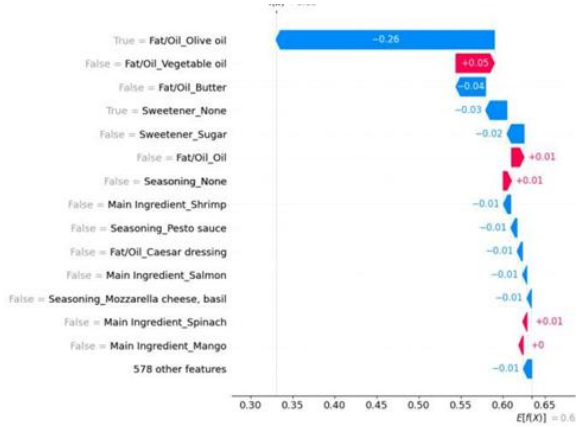


Figure 10. Random Forest SHAP Waterfall Plot

As shown in Figure 7, the most important features in food allergen prediction are Fat/Oil_Butter, Fat/Oil_Vegetable oil, and Fat/Oil_Olive oil features have the most contribution, ingredient categories such as Sweetener_None, Sweetener_Sugar, and Seasoning_None also have a strong influence. Some key ingredients such as Shrimp, Tomatoes, Mango, and Salmon appear as important indicators, which is biologically appropriate as some of them are common allergens. The importance ranking of features is not trivial and has a correlation to this topic as the condition of the ingredients used to process food is directly related to the presence or absence of allergens. So directly focusing on this element not only improves the interpretability of the model but also directs future projects on feature engineering and data collection.

The SHAP bar plot (Figure 8) displays the mean absolute SHAP value as a more accurate and unbiased measure of feature importance compared to the Gini index. The results show that Fat/Oil_Olive oil with the value of 0.08 is the most dominant feature followed by Fat/Oil_ Vegetable with the value of 0.06 and Fat/Oil_Butter oil with value of 0.04 and there is a no significant difference between Figure 7 and Figure 8. In the Gini index, Fat/Oil_Butter ranks first whereas SHAP places Fat/Oil_Olive oil as the most influential, thus indicates that SHAP provides a more representative picture of feature contributions because it comprehensively considers interactions between features in each prediction.

The beeswarm plot in Figure 9 displays the SHAP values for the top 20 features of the Random Forest model, with each point representing one sample, and with the color indicating that a red feature value means ‘high/True’ and a blue value means ‘low/False’. The feature with the highest importance is Fat/Oil_Olive oil, indicated by the red points that are scattered far to the left along the plot. Sweetener_None appears to have varying effects on the model, indicated by the way that the points for that feature are scattered in both the positive and negative direction on the plot. Fat/Oil_Vegetable oil and Fat/Oil_Butter have similar patterns to olive oil indicating that each of these features having a True value will lead to a negative prediction for the model. Finally, the features that have the lowest importance include features like Food Product_Mango Spinach, Beef, and several of the Main Ingredients, as the points for each of these features are all concentrated around the zero line on the plot. Thus, each of these features has a weak effect upon the predictions of the Random Forest model.

Figure 10’s waterfall plot displays the model’s prediction for the single sample of the individual. The Fat/Oil_Olive oil feature makes the largest negative contribution to the model’s prediction of the sample’s class, with a value of -0.26. Following behind is Fat/Oil_Vegetable oil with +0.05, and then Fat/Oil_Butter with -0.04. Some features exhibit small positive contributions, such as Fat/Oil_Oil oil with a value of False and +0.01, and Seasoning_None makes a small positive contribution of False and +0.01. However, neither of these small positive contributions are significant enough to impact the model’s prediction of the sample’s class. Thus, the model’s classification of this single sample was essentially dominated by the presence of the olive oil-based fat.

The waterfall contribution of −0.26 should be interpreted relative to the explained model output and baseline rather than as a biological effect. Without an explicit specification of the output class, scale, and reference value, the sign cannot establish an unambiguous change in the probability of allergen presence. Figures 7–10 thus provide model-specific explanatory evidence, not causal validation or independent allergen confirmation.

4. CONCLUSION

This study compared XGBoost, AdaBoost, Decision Tree, and Random Forest for binary

classification of the supplied food product allergen-related labels. In the reported single 80-record test set, Random Forest had the highest accuracy (0.86), with a class-1 F1-score of 0.91 and macro-average F1-score of 0.82. However, its Class 0 recall was 0.65, demonstrating weaker sensitivity for the minority class. Class weighting was applied to Decision Tree and Random Forest, but its independent effect on performance cannot be established without an unweighted comparison.

SHAP explanations for Random Forest identified fat/oil variables, including olive oil, vegetable oil, and butter, among the influential features for its dataset-label predictions. Their contributions describe model associations and should not be interpreted as evidence of ingredient-level causality or clinically verified allergenicity. The analysis is limited by the small sample size (399 records), reliance on labels derived from existing allergen annotations, the single train–test split, incomplete reporting of feature preprocessing and hyperparameters, and the absence of external validation. Moreover, the reported method does not confirm whether target-derived predictors were excluded, which must be checked before treating these results as leakage-free estimates. Future work should prioritize a documented leakage-free pipeline, repeated stratified validation, explicit SHAP output settings, and independent validation on larger datasets. The current findings support exploratory classification and interpretation of dataset labels; they do not establish a validated consumer-facing allergen detection system.

AUTHOR CONTRIBUTIONS

Bakti Amirul Jabar: Methodology, Validation, Writing – Review & Editing. Richard Edsel Wijaya: Conceptualization, Methodology, Data Curation, Writing – Original Draft. Martin Halim: Conceptualization, Methodology, Data Curation, Visualization, Writing – Original Draft, Software. Hansel Mitchell Tumanggor: Validation, Writing – Review & Editing.

DATA AVAILABILITY STATEMENT

The dataset used in this study, namely the Food Ingredients and Allergens dataset, was obtained from Kaggle and published by Laksika Tharmalingam. The dataset is publicly accessible at: https://www.kaggle.com/datasets/uom190346a/food-ingredients-and-allergens.

REFERENCES

[1] S. H. Sicherer and H. A. Sampson, “Food allergy,” J. Allergy Clin. Immunol., vol. 125, no. 2, pp. S116–S125, Feb. 2010, doi: 10.1016/j.jaci.2009.08.028.

[2] V. Sampath et al., “Food allergy across the globe,” J. Allergy Clin. Immunol., vol. 148, no. 6, pp. 1347–1364, Dec. 2021, doi: 10.1016/j.jaci.2021.10.018.

[3] W. A. Żukiewicz-Sobczak, P. Wróblewska, P. Adamczuk, and P. Kopczyński, “Causes, symptoms and prevention of food allergy,” Adv. Dermatol. Allergol., vol. 2, pp. 113–116, 2013, doi: 10.5114/pdia.2013.34162.

[4] S. R. Nimmagadda and R. Evans, “Allergy: Etiology and Epidemiology,” Pediatr. Rev., vol. 20, no. 4, pp. 110–116, Apr. 1999, doi: 10.1542/pir.20-4-110.

[5] B. K. Ballmer-Weber, “Value of Allergy Tests for the Diagnosis of Food Allergy,” Dig. Dis., vol. 32, no. 1–2, pp. 84–88, 2014, doi: 10.1159/000357077.

[6] L. Wang, D. Niu, X. Zhao, X. Wang, M. Hao, and H. Che, “A Comparative Analysis of Novel Deep Learning and Ensemble Learning Models to Predict the Allergenicity of Food Proteins,” Foods, vol. 10, no. 4, p. 809, Apr. 2021, doi: 10.3390/foods10040809.

[7] H. Shaukat, A. Sultan, and H. Salahuddin, “Enhancing Food Safety: A Machine Learning Approach for Accurate Detection and Classification of Food Allergens.”

[8] T. Arima et al., “Prediction of oral food challenge outcome using machine learning methods,” J. Allergy Clin. Immunol., vol. 155, no. 2, p. AB258, Feb. 2025, doi: 10.1016/j.jaci.2024.12.798.

[9] Siska Narulita, Sekarlangit Sekarlangit, and Milka Putri Novianingrum, “Deteksi Alergen pada Produk Pangan Menggunakan Algoritma Support Vector Machines (SVM),” Bridge J. Publ. Sist. Inf. Dan Telekomun., vol. 3, no. 1, pp. 64–76, Feb. 2025, doi: 10.62951/bridge.v3i1.393.

[10] P. Sharma, M. T. Siddiqi, S. Agarwal, K. Kaur, and K. G, “Food Allergen Detection and Recommendation,” SSRN Electron. J., 2025, doi: 10.2139/ssrn.5198646.

[11] M. Nedyalkova, M. Vasighi, A. Azmoon, L. Naneva, and V. Simeonov, “Sequence-Based Prediction of Plant Allergenic Proteins: Machine Learning Classification Approach,” ACS Omega, vol. 8, no. 4, pp. 3698–3704, Jan. 2023, doi: 10.1021/acsomega.2c02842.

[12] Junaid Ahmad et al., “UTILIZING DEEP LEARNING TECHNIQUES FOR DETECTING AND ANALYZING FOOD ALLERGIES,” Kashf J. Multidiscip. Res., vol. 2, no. 03, pp. 46–60, Mar. 2025, doi: 10.71146/kjmr332.

[13] J. Zhang et al., “Prediction of oral food challenge outcomes via ensemble learning,” Inform. Med. Unlocked, vol. 36, p. 101142, 2023, doi: 10.1016/j.imu.2022.101142.

[14] O. O. Bifarin, “Interpretable machine learning with tree-based shapley additive explanations: Application to metabolomics datasets for binary classification,” PloS One, vol. 18, no. 5, p. e0284315, 2023, doi: 10.1371/journal.pone.0284315.

[15] V. Borisov, T. Leemann, K. Seßler, J. Haug, M. Pawelczyk, and G. Kasneci, “Deep Neural Networks and Tabular Data: A Survey,” 2021, doi: 10.48550/ARXIV.2110.01889.

[16] L. Tharmalingam, “Food Ingredients And Allergens.” Accessed: May 05, 2025. [Online]. Available: https://www.kaggle.com/datasets/uom190346a/food-ingredients-and-allergens/data

[17] Y.-Y. Song and Y. Lu, “Decision tree methods: applications for classification and prediction,” Shanghai Arch. Psychiatry, vol. 27, no. 2, pp. 130–135, Apr. 2015, doi: 10.11919/j.issn.1002-0829.215044.

[18] M. Pal, “Random forest classifier for remote sensing classification,” Int. J. Remote Sens., vol. 26, no. 1, pp. 217–222, Jan. 2005, doi: 10.1080/01431160412331269698.

[19] P. Zhang, Y. Jia, and Y. Shang, “Research and application of XGBoost in imbalanced data,” Int. J. Distrib. Sens. Netw., vol. 18, no. 6, p. 155013292211069, Jun. 2022, doi: 10.1177/15501329221106935.

[20] Y. Cao, Q.-G. Miao, J.-C. Liu, and L. Gao, “Advance and Prospects of AdaBoost Algorithm,” Acta Autom. Sin., vol. 39, no. 6, pp. 745–758, Jun. 2013, doi: 10.1016/S1874-1029(13)60052-X.

[21] F. Pedregosa et al., “Scikit-learn: Machine Learning in Python,” Mach. Learn. PYTHON.

[22] G. Naidu, T. Zuva, and E. M. Sibanda, “A Review of Evaluation Metrics in Machine Learning Algorithms,” in Artificial Intelligence Application in Networks and Systems, vol. 724, R. Silhavy and P. Silhavy, Eds., Cham: Springer International Publishing, 2023, pp. 15–25. doi: 10.1007/978-3-031-35314-7_2.

[23] L. Arrighi, I. A. de Moraes, M. Zullich, M. Simonato, D. F. Barbin, and S. B. Junior, “Explainable Artificial Intelligence Techniques for Interpretation of Food Models: a Review,” 2025, arXiv. doi: 10.48550/ARXIV.2504.10527.