Food Analytical Methods (2025) 18:2331–2343
https://doi.org/10.1007/s12161-025-02851-6

RESEARCH

Novel Artificial Intelligence Approach For nsLTP Early Detection Using NIRs Data

Alex Rodriguez-Alonso[1] · Itxasne Del Barrio[1] · Ganeko Bernardo-Seisdedos[2] · Ainhoa Osa-Sanchez[1] · Begonya Garcia-Zapirain[1]

Received: 8 April 2025 / Accepted: 1 July 2025 / Published online: 29 July 2025

Abstract
Food allergies have become a significant public health issue, particularly lipid transfer protein (LTP) allergies, which are highly stable allergens and can cause severe allergic reactions. This research aims to develop and validate an AI-driven solution for detecting LTPs in food using near-infrared spectroscopy (NIRS), exploring the feasibility of non-invasive allergen identification using AI-assisted spectroscopy. The methodology involves collecting spectral data from various food samples, building a machine learning model, and optimizing it iteratively to improve detection accuracy. The results show that the AI model achieved an accuracy of 87% and an F1-score of 89.91%, indicating its potential for enhancing food safety. In conclusion, this solution demonstrates the viability of using NIRS and AI for allergen detection, with promising future applications in healthcare.

Keywords Allergens · Near-infrared spectroscopy · Artificial intelligence · Classification

Introduction

Allergies are a critical global health issue that affects over a billion people and is projected to impact up to four billion by the 2050 s, driven by increasing exposure to environmental and dietary allergens (Global Atlas of Allergy 2025). Food allergies are a specific subset of allergic reactions derived from immune hypersensitivity to dietary proteins and can range in severity from mild local symptoms to anaphylaxis. They can be categorized into stable Class I allergens, which are resistant to heat, breakdown, and digestion and are primarily associated with gastrointestinal and systemic reactions, and Class II allergens, which are typically labile and cause milder, localized reactions. In either case, immunoglobulin E (IgE)-mediated reactions are the most common form, with rapid onset following ingestion, and involve a cascade of immune responses triggered by allergenic proteins (Global Atlas of Allergy 2025).

nsLTPs are among the most clinically significant food allergens due to their unique properties and high stability. As small, highly conserved cysteine-rich proteins found in a variety of plant species, nsLTPs are implicated in plant defense mechanisms, cell wall organization, and signal transduction (Missaoui et al. 2022). Unlike many other food allergens, nsLTPs are resistant to both heat and digestive processes, allowing them to retain their quaternary structure and thus their allergenic potential even after cooking or processing. This stability poses a particular risk for individuals with sensitivity to nsLTPs, as exposure can occur across a wide range of foods and products, from fresh and cooked fruits to processed products or cosmetics containing plant extracts.

Pru p 3, the primary nsLTP in peaches (*Prunus persica*) and the major allergen in the Rosaceae family, including peaches, nectarines, and strawberries, is one of the best-studied nsLTPs. It has been observed that this member of the

✉ Ainhoa Osa-Sanchez
ainhoa.osa.sanchez@deusto.es

Alex Rodriguez-Alonso
alexrodriguez@opendeusto.es

Itxasne Del Barrio
itxasnedel.barrio@opendeusto.es

Ganeko Bernardo-Seisdedos
ganeko.bernardo@deusto.es

Begonya Garcia-Zapirain
mbgarciazapi@deusto.es

[1] eVIDA Research Group, University of Deusto, Bilbao 48007, Spain

[2] DeustoMED Research Group, University of Deusto, Bilbao 48007, Spain

nsLTP family is often the first sensitizing allergen in patients who develop nsLTP syndrome, a condition characterized by cross-reactivity among LTPs from various plants (Maestro 2025). This cross-reactivity means that individuals sensitized to Pru p 3 may also react to other foods containing structurally similar nsLTPs, such as apples, apricots, and nuts, as well as to certain pollen. In Southern Europe and Mediterranean regions, nsLTPs allergies are notably prevalent, with Pru p 3 identified as the leading cause of severe allergic reactions. Importantly, nsLTPs like Pru p 3 are known to cause systemic allergic reactions, which are often aggravated by factors like physical exercise, alcohol, and NSAID intake, making accurate detection of these proteins in food crucial for effective allergy management (Rossi et al. 2023).

Traditional diagnostic tools for nsLTP detection, such as ELISA (enzyme-linked immunosorbent assay) and immunoblotting, are reliable but can be expensive and may require laboratory settings and extensive sample preparation. These methods, while useful in controlled environments, are not conducive to real-time monitoring or rapid decision-making in food safety contexts. Therefore, this escalating allergy trend necessitates advanced diagnostic processes, improved treatment protocols, and integrated healthcare strategies to manage and mitigate the widespread impact of allergies effectively (Mazas 2025).

The potential of near-infrared spectroscopy (NIRS) combined with machine learning has been highlighted in various studies focusing on food quality and safety. For instance, Zhenxiong Huang et al. demonstrated the effectiveness of NIRS in the non-destructive determination of diastase activity in honey, showcasing a robust methodology for assessing honey quality without altering its state (Huang et al. 2019). Similarly, Yande Liu et al. applied visible/near-infrared spectroscopy to monitor the quality of juicy peaches during storage, revealing that this technique can accurately detect changes in fruit quality over time (Liu et al. 2020). Additionally, Marina Schopf and colleagues conducted a fundamental characterization of wheat gluten, illustrating the application of NIRS in understanding the properties of gluten, which is crucial for various dietary and food safety considerations (Schopf et al. 2021).

Recent literature has further highlighted the growing synergy between spectroscopy and machine learning in food analysis. Zuo et al. (2023) investigated the combined use of spectroscopic techniques and ML algorithms for the authentication and classification of meat products, demonstrating improved accuracy and efficiency in fraud detection and quality assessment. Similarly, Zuo et al. (2025) explored the integration of near-infrared spectroscopy with advanced ML methods to predict quality parameters in processed foods, showcasing the potential of these tools for non-destructive, real-time monitoring in food industry settings. These studies underscore the current trend toward developing rapid, accurate, and sustainable analytical methods through the convergence of spectroscopy and ML, which serves as a foundation for the approach adopted in this work.

The use of NIRS in assessing the quality-related parameters in instant tea was explored by Xiaoli Bai et al., where machine learning techniques were employed to predict these parameters, demonstrating the versatility of NIRS in different food matrices (Bai et al. 2022). Furthermore, Rabie Reda et al. optimized variable selection and developed machine learning models for olive oil quality assessment using portable NIRS, providing a portable solution for on-site quality control (Reda et al. 2023). Hayato Seki et al. visualized sugar content distribution in white strawberries using NIR hyperspectral imaging, emphasizing the non-destructive evaluation of fruit quality (Seki et al. 2023). Moreover, Lars Erik Solberg et al. showcased the application of NIRS in large-scale cheese production, highlighting its efficiency in real-time quality control during the manufacturing process (Solberg et al. 2023). These studies underscore the broad applicability of NIRS across different types of aliments, enhancing both the efficiency and accuracy of food quality assessments.

The primary goal of this paper is to develop, implement, and validate an AI-driven predictive model using NIRS to accurately detect nsLTPs in foods. By collecting data from various food types, building a comprehensive dataset, and performing in-depth statistical analysis and visualization, the paper aims to enhance the accuracy and speed of LTP detection. This model will be integrated into a user-friendly platform, allowing individuals to input measurements and predict the LTP content in their samples. The paper aims to significantly reduce false negatives and the time required for testing, making it a valuable tool for a potencial real-time allergy diagnostics. In the long term, the platform could be implemented in healthcare settings to provide professionals with quick access to crucial data. However, this prospective application would require rigorous clinical validation, regulatory approval, user training, and seamless integration with existing healthcare infrastructure. Despite current challenges such as data variability and system interoperability, our study aims to lay the groundwork for a reliable, non-invasive solution for allergy management. This paper is structured in several chapters: in the "Material and methods" section, we explain the experimental design and models used, followed by the results and discussion, where we present the most relevant findings and implications. Finally, we conclude with future work and directions for further research.

Material and Methods

Material

Experimental Setup and Data Collection

The experimental setup involved preparing food samples to measure their spectral signatures using a scientific-grade spectrometer and associated equipment. Each type of food was handled with care to avoid cross-contamination, and measurements were collected sequentially for one food item at a time, as shown in Table 1. The collected data was saved in .txt files for each measurement, ensuring the integrity and accuracy of the collected spectral data.

For the False class, there were 4050 unique data, and for the True class 3240, according to 55.5% and 44.4%, respectively, showing a good balance between classes.

Prior to handling a new food item, the equipment and workspace were thoroughly cleaned using distilled water. Tools such as knives, trays, and tweezers were sterilized or replaced between samples to avoid any residue from one sample affecting the next. Additionally, gloves were worn, further minimizing the risk of contamination. To prepare food samples for spectral analysis, a meticulous process was followed to ensure precise measurements. The samples were divided into smaller portions to facilitate the analysis.

Each piece was labeled sequentially, and every fragment cut from each sample was individually weighed and recorded. To assign ground-truth labels regarding the presence or absence of nsLTPs, each food item was first identified taxonomically and matched with entries in curated allergen databases such as AllergenOnline and the WHO/IUIS Allergen Nomenclature database. Only foods with well-documented nsLTP content based on these databases and supporting literature were labeled as "present." Conversely, food samples with no reported or confirmed nsLTP content across these authoritative sources were labeled as "absent." No assumptions were made based on food category or botanical family alone, and ambiguous cases were excluded to maintain label reliability. This ensured that each fragment could be traced back to its original source.

Table 1 Aliments purchased to produce the database

Product	Quantity of product per food item	nsLTPs present
Oranges	3	True
Apples	3	True
Crystal albumin	2±15%	False
Powdered milk	2±15%	False
Pears	2	True
Cucumbers	2	False
Kiwis	2	True
Potatoes	2	False
Strawberries	3	True
Peaches	3	True
Tomatoes	3	True
Bananas	2	True

When taking the spectral measurements, each food item was placed on a fresh surface, ensuring that no particles or moisture from previous samples were present. The samples were processed sequentially, meaning that only one food item was in the measurement area at any given time. This strict isolation ensured that no environmental factors, such as residue from other foods, could influence the spectral readings. After placing each piece in the spectrometer, a 10-s pause was observed to allow the detection system to stabilize, preventing any initial fluctuations from affecting the readings. Absorbance and reflectance measurements were taken at three distinct positions on each food sample. At each position, two types of spectral data were collected—one absorbance and one reflectance spectrum—resulting in a total of six measurements per sample. It was crucial to document every step, including the weight of each fragment and its corresponding measurement, to maintain an organized and accurate record.

All measurements were carefully saved in .txt files, one per sample, immediately after each reading to preserve the integrity and accuracy of the collected spectral data. By following these stringent protocols, the experiment ensured that the spectral data for each food sample was pure and uncontaminated, providing reliable results for further analysis.

Database Construction

After data collection, the process of constructing the database began. Handling 7490 individual .txt files manually was not feasible, so automation scripts were developed in Python. These scripts extracted the relevant information from each file, converted the raw spectral data into structured .csv files, and merged them into a single comprehensive database. This automated approach streamlined data management and ensured consistency across the entire dataset (Table 2).

Data Preprocessing and Data Visualization

Data preprocessing was essential to ensure the quality and reliability of the dataset before training the machine learning models. The process involved multiple steps: handling missing values, removing duplicates, correcting errors, and standardizing formats (e.g., replacing commas with dots in decimal notation) (1). Irrelevant columns were removed to prevent noise from degrading model performance. These steps helped eliminate inconsistencies or inaccuracies that

Table 2 Data collection distribution (total number of data) Note: Each sample was measured at 3 positions with both absorbance and reflectance modes, resulting in 6 measurements per sample

Fruit/product	Amount	Pieces	Nº of positions	Measurement method	LTP	Total data
Orange	3	10	3	2 meas. (abs/ref) at 3 pos	Yes	1080
Pear	1	10	1	2 meas. (abs/ref) at 1 pos	Yes	40
Apple	3	10	3	2 meas. (abs/ref) at 3 pos	Yes	1080
Peach	2	10	3	2 meas. (abs/ref) at 3 pos	Yes	810
Kiwi	1	10	1	2 meas. (abs/ref) at 1 pos	Yes	40
Tomato	3	10	3	2 meas. (abs/ref) at 3 pos	Yes	1080
Strawberry	6	5	3	2 meas. (abs/ref) at 3 pos	Yes	1080
Banana	1	10	1	2 meas. (abs/ref) at 1 pos	Yes	40
Cucumber	1	10	1	2 meas. (abs/ref) at 1 pos	No	40
Potato	1	10	1	2 meas. (abs/ref) at 1 pos	No	40
Albumin	30 samples	–	3	2 meas. (abs/ref) at 3 pos	No	1080
Powdered milk	30 samples	–	3	2 meas. (abs/ref) at 3 pos	No	1080
TOTAL						7490

could negatively impact the models, leading to more accurate and robust predictions.

In addition, data visualization was used to assess the class distribution and detect potential imbalances between the presence and absence of the target protein (Fig. 1). These visualizations ensured a balanced dataset and supported the validity of the training process by reducing potential bias. Although the class distribution was slightly skewed (55.5% negative vs. 44.4% positive), it was not severe enough to require oversampling or undersampling strategies, as the models were able to train effectively without exhibiting performance bias.

To standardize the input features across samples, we applied Z-score normalization to all spectral variables, using the formula shown in Eq. 1. This scaling ensured that each feature had a mean of 0 and a standard deviation of 1, improving the performance and convergence of machine learning models.

$$z = \frac{x - \mu}{\sigma} \tag{1}$$

Methods Hardware

The spectrometer used for achieving the objective is the FLAME-NIR+ Miniature Spectrometer. The Flame-NIR spectrometer from Ocean Optics is an advanced and compact device that utilizes near-infrared spectroscopy to analyze the composition of various materials effectively. Designed to be affordable, it features a small optical bench that makes it versatile and allows scientists and researchers to use it in different environments (Flame-NIR User Guide 2025).

The spectrometer is equipped with a high-quality, uncooled InGaAs (Indium Gallium Arsenide) array detector, providing accurate measurements across a spectral range of approximately 900 to 1700 nm. This versatile instrument is ideal for applications such as food quality analysis or agricultural

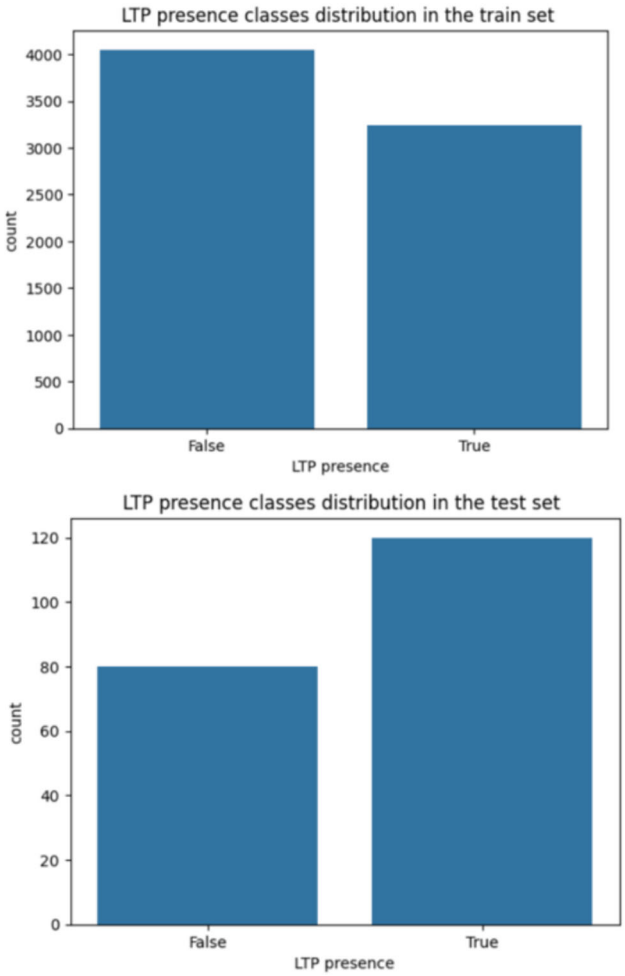

Fig. 1 Distribution of the data classes (positive and negative cases)

research among others (FLAME-NIR+ 2024; Flame-NIR 2025).

In addition to the spectrometer, several other components are essential for a complete NIRS system (Fig. 2).

On the one hand, a TC-DR-PROBE has been used. The TC-DR-PROBE is a specialized probe designed for measuring diffuse reflectance, integrating a halogen light source and focusing optics into a single unit. This combination allows for efficient illumination of the sample with bright, continuous light, essential for accurate spectroscopic analysis. By measuring the light reflected at a 45° angle, the probe captures how much light is scattered in various directions from the material, providing a realistic representation of its optical properties. This capability is particularly valuable in NIRS, where understanding the light interaction with samples is essential for obtaining reliable analytical results (Ocean Photonics 2025a).

On the other hand, the optical fiber serves as a critical link in the NIRS system, enabling the efficient transmission of light between the spectrometer and the sample. Ocean Insight's laboratory optical fiber (previously known as Ocean Optics) offers a reliable and affordable solution, coming in multiple lengths and designs. Each end of the fiber is fitted with standard SMA905 connectors. The design minimizes light loss and maintains the integrity of the spectral data, ensuring that accurate and reliable results are obtained during the analysis (Ocean Photonics 2025b).

Moreover, the system needs some calibration features and standards to ensure accurate and reliable measurements. One essential component is Spectralon, a calibration standard known for its high reflectance across a wide range of wavelengths. By using Spectralon, effective calibration of the instruments has been done, providing a consistent reference point for analyzing the samples. Additionally, the system requires the capability to capture the dark spectrum as well, which involves measuring the detector's response when no light is present. This measurement is crucial for correcting any background noise or signal interference, allowing for a clearer representation of the sample's true spectral data. Together, these calibration features are vital for enhancing the overall performance and reliability of the system in NIRS applications (Flame NIR Manual 2025).

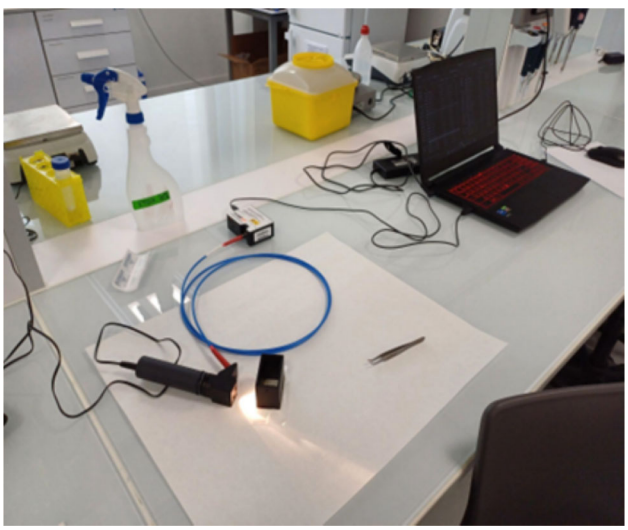

Fig. 2 Setup for data collection

Finally, it is essential to have a computer device with the software installed. The software used has been the OceanView. OceanView software is an advanced tool designed to work with spectrometers. It allows easy management and analysis of reflectance and absorbance measurements, key measurements needed for the detection of the LTP protein in food. For reflectance, the software helps compare the light reflected from a sample against a reference standard, the previously mentioned Spectralon. In terms of absorbance, OceanView processes the data to calculate how much light is absorbed by the sample.

Once all these components are connected and the sensor calibration has been completed, the device will be ready to start collecting data. The flowchart in Fig. 3 provides a step-by-step summary of the sample collection process, structured into distinct stages for clarity and efficiency.

The process begins with initial preparation, which involves assembling the necessary hardware components to ensure all required equipment is available. Next, the equipment undergoes calibration to guarantee accurate and consistent measurements. Additionally, all essential tools must be gathered and prepared before proceeding to the next steps.

In the food evaluation stage, it is determined whether a new type of food is being analyzed. If a new food item is introduced, the tools must be sterilized to prevent cross-contamination. Otherwise, if the food type remains the same, the current tools can continue to be used without sterilization.

The next step, sample preparation, involves dividing the food sample into smaller portions to facilitate handling and measurement. Each portion is then labeled to ensure proper identification and to avoid any confusion. Finally, the portions are placed on a clean and suitable surface for measurement.

During the measurements stage, the absorbance and reflectance of each portion are recorded using calibrated equipment. To ensure accuracy, measurements must be taken at three different positions within the sample. If all three positions have not yet been measured, the sample's position is adjusted, and additional measurements are taken. Once all required positions have been measured, the data is renamed and saved appropriately.

The final stage, completion and repetition, determines whether additional samples need to be analyzed. If more samples are available, the process returns to the step of dividing

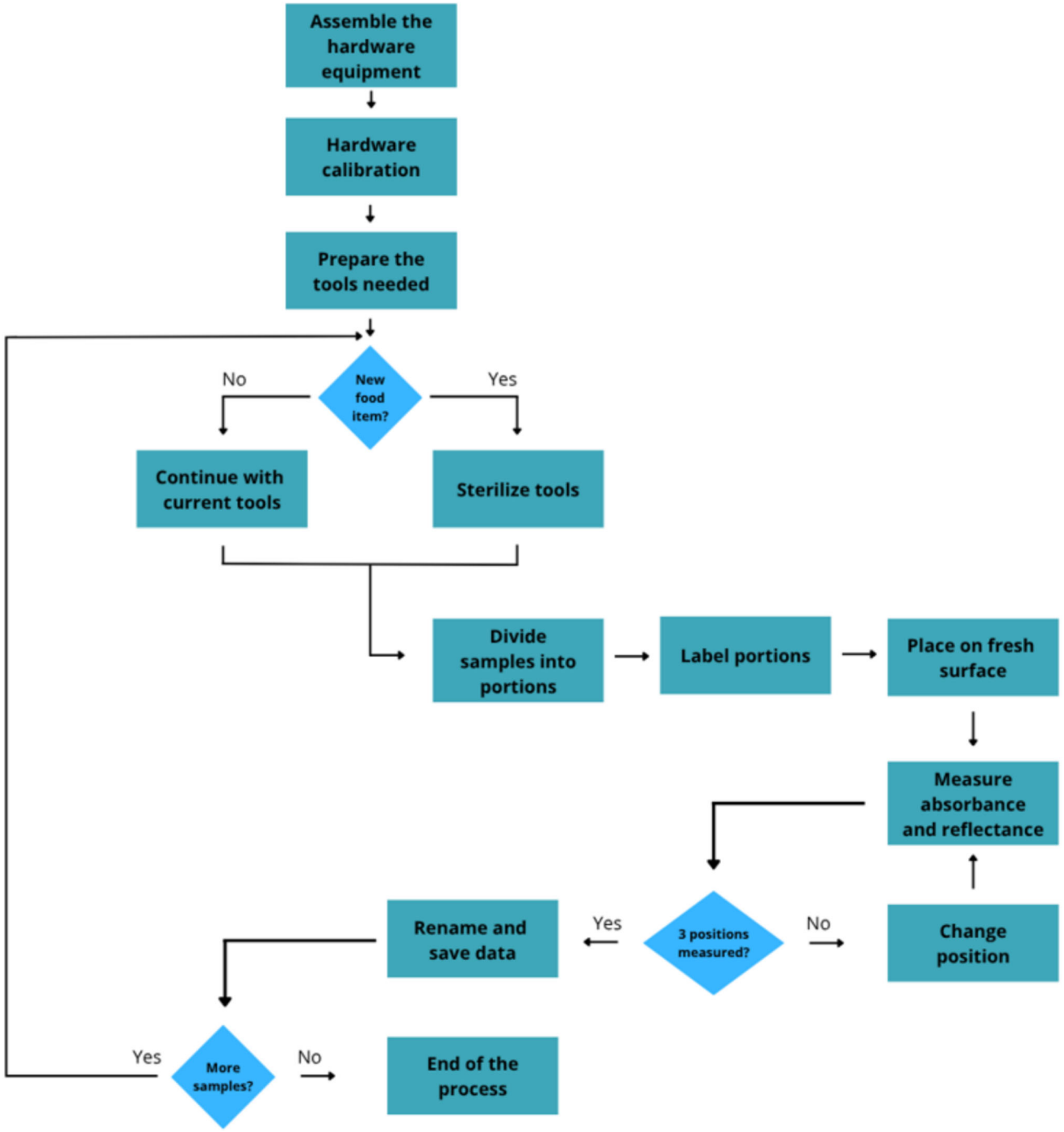

Fig. 3 Flowchart of sample collection

them into portions. If there are no further samples, the process concludes.

Methods About Artificial Intelligence

Logistic Regression

Logistic regression is a fundamental statistical method used in machine learning and data science for binary classification tasks. Unlike linear regression which predicts a continuous output, logistic regression predicts the probability of a binary outcome (0 or 1) (Shewhart and Wilks 2013; Peduzzi et al. 1996).

The logistic function, or sigmoid function, defined as

$$\sigma(z) = \frac{1}{1 + e^{-z}} \tag{2}$$

maps any real-valued number into the range [0, 1], making it suitable for probability estimation. The model estimates the probability that a given input, x, belongs to a particular class by using the logistic function on a linear combination of input features (Shewhart and Wilks 2013; Peduzzi et al. 1996):

$$\text{logit}(P) = \ln\left(\frac{P}{1-P}\right) = \mathbf{w}^T\mathbf{x} + b \tag{3}$$

where P is the probability of the positive class, w is the vector of weights, and b is the bias term. The model parameters are typically estimated using maximum likelihood estimation (MLE). Logistic regression is widely appreciated for its simplicity, interpretability, and efficiency, particularly in high-dimensional feature spaces (Shewhart and Wilks 2013; Peduzzi et al. 1996).

Bernoulli Naive Bayes

Bernoulli Naive Bayes is a probabilistic classifier particularly well-suited for binary/boolean features. It is based on Bayes' Theorem, which describes the probability of an event based on prior knowledge of conditions that might be related to the event. For Bernoulli Naive Bayes, the presence or absence of a feature is treated as a binary condition. The algorithm assumes that each feature is conditionally independent of the others given the class label, which simplifies the computation. The formula for the conditional probability is Bishop (2006)

$$P(C_k \mid \mathbf{x}) = \frac{P(C_k)\prod_{i=1}^{n} P(x_i \mid C_k)^{x_i}(1 - P(x_i \mid C_k))^{1-x_i}}{P(\mathbf{x})} \tag{4}$$

where $P(Ck---x)$ is the posterior probability of class Ck given feature vector x, $P(Ck)$ is the prior probability of class Ck, $P(xi---Ck)$ is the likelihood of feature xi given class Ck, and $P(x)$ is the probability of the feature vector (Bishop 2006; Brownlee 2016).

Decision Tree

A decision tree is a powerful machine learning model used for classification and regression tasks. It works by recursively splitting the data into subsets based on the value of input features, forming a tree-like structure. Each internal node represents a decision rule on a feature, each branch corresponds to an outcome of the rule, and each leaf node represents a class label or a continuous value. The goal is to create a model that predicts the target variable by learning simple decision rules inferred from the data features. Key metrics for making these splits include Gini impurity and information gain for classification and mean squared error (MSE) for regression (Breiman et al. 2017; Quinlan 1986).

The Gini impurity for a node, t, is calculated as

$$G(t) = 1 - \sum_{i=1}^{C} p_i^2 \tag{5}$$

where pi is the proportion of instances of class i in node t. Information is defined as the reduction in entropy from a split and is calculated by

$$IG(S, A) = H(S) - \sum_{v \in \text{Values}(A)} |S_v||S|H(S_v) \tag{6}$$

where $H(S)$ is the entropy of set S, and Sv is the subset of S for which attribute A has value v. For regression trees, the mean squared error (MSE) at a node t is given by

$$MSE(t) = \frac{1}{N}\sum_{i=1}^{N}(y_i - \bar{y})^2 \tag{7}$$

where yi are the actual values, and y is the mean of the values in node t (Breiman et al. 2017; Quinlan 1986).

Support Vector Machine

Support vector machines (SVM) are a set of supervised learning methods used for classification, regression, and outlier detection. SVM works by finding the hyperplane that best separates the data into different classes. The key idea is to maximize the margin, which is the distance between the hyperplane and the nearest data point of any class. The optimal hyperplane is found by solving the following optimization problem (Cortes and Vapnik 1995; Burges 1998):

$$\min_{w,b} \frac{1}{2}\|w\|^2 \quad \text{subject to} \quad y_i(w \cdot x_i + b) \geq 1 \tag{8}$$

For non-linearly separable data, SVM can use kernel functions (such as polynomial, radial basis function) to map the input features into higher-dimensional space where a hyperplane can be used to separate the data (Cortes and Vapnik 1995; Burges 1998).

XGBoost

XGBoost (eXtreme Gradient Boosting) is a powerful and efficient implementation of the gradient boosting algorithm, designed to enhance speed and performance for machine learning tasks. It combines predictions from multiple decision trees, each correcting errors from the previous ones, to form a strong predictive model (Chen and Guestrin 2016).

The objective function in XGBoost includes a regularization term to prevent overfitting, mathematically expressed as

$$L(\theta) = \sum_{i=1}^{n} l(y_i, \hat{y}_i) + \sum_{k=1}^{K} \Omega(f_k) \tag{9}$$

where

$$\Omega(f) = \gamma T + \frac{1}{2}\lambda \sum_{j=1}^{T} w_j^2 \tag{10}$$

This regularization, along with second-order Taylor approximation for optimization, allows XGBoost to achieve high accuracy and fast convergence, making it a favorite in data science competitions and real-world applications (Chen and Guestrin 2016).

Deep Neural Network

Deep neural networks (DNNs) are powerful machine learning models that mimic the human brain's structure, consisting of layers of interconnected neurons. Each neuron performs a weighted sum of its inputs, applies an activation function, and passes the output to the next layer (Goodfellow et al. 2016a; LeCun et al. 2015).

Mathematically, this process can be described as

$$z_j^{(l)} = \sum_i w_{ij}^{(l)} a_i^{(l-1)} + b_j^{(l)} \tag{11}$$

and

$$a_j^{(l)} = \sigma\left(z_j^{(l)}\right). \tag{12}$$

DNNs learn complex patterns by adjusting weights and biases through backpropagation, minimizing error via gradient descent (Burges (1998); Chen and Guestrin (2016)):

$$w_{ij}^{(l)} \leftarrow w_{ij}^{(l)} - \eta \frac{\partial w_{ij}^{(l)}}{\partial L} \tag{13}$$

Confusion Matrix

The confusion matrix is a table that describes the performance of a classification model, displaying the number of true positives (TP), false positives (FP), true negatives (TN), and false negatives (FN) (Bishop 2006).

Accuracy

Accuracy measures the percentage of correct predictions, considering both positive and negative classes (Peduzzi et al. 1996).

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \tag{14}$$

Precision

Precision measures the proportion of true positive predictions out of the total positive predictions (Cortes and Vapnik 1995).

$$\text{Precision} = \frac{TP + FN}{TP} \tag{15}$$

Recall (Sensitivity)

Recall, or sensitivity, measures the proportion of actual positives correctly identified by the model (Shewhart and Wilks 2013).

$$\text{Recall} = \frac{TP}{TP + FN} \tag{16}$$

F1-Score

The F1-score is the harmonic mean of precision and recall. It is useful when there is a need for balance between the two (Chen and Guestrin 2016).

$$F1 = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \tag{17}$$

AUC-ROC (Area Under the Curve-Receiver Operating Characteristic)

AUC-ROC is a measure that describes the model's ability to distinguish between classes. The AUC value ranges from 0.5 (random) to 1 (perfect) (Goodfellow et al. 2016b).

$$\text{AUC} = \int_0^1 \text{TPR}(\text{FPR})\, d(\text{FPR}) \tag{18}$$

Hyperparameter Fine-Tuning

The fine-tuning process of a deep neural network (DNN) was a critical step in optimizing its performance and ensuring the model's robustness in various applications. Fine-tuning involves iterative modifications to the network's architecture, hyperparameters, and training dynamics to improve accuracy, reduce overfitting, and enhance generalization capabilities on unseen data.

One of the first aspects of fine-tuning was adjusting the architecture of the neural network. This process included experimenting with the number and size of hidden layers, as well as the neurons within each layer (Jossa-Bastidas et al. 2023a). Selecting the right number of layers and neurons was

crucial, as too few could limit the model's learning capacity, while too many could lead to overfitting. In this process, different activation functions, such as ReLU, Sigmoid, or Tanh, were also tested to determine which yielded the best non-linear transformations for the given dataset.

Another key element of fine-tuning was the optimization algorithm. Commonly used optimizers such as Adam, Stochastic Gradient Descent (SGD), and RMSprop were employed to update the model's weights efficiently. Each optimizer was tested with varying learning rates to find the most suitable combination that balanced fast convergence with stable learning. The learning rate was a particularly sensitive parameter; if it is set too high, the model may miss optimal solutions, whereas a low learning rate may result in unnecessarily long training times.

In addition to the standard optimization process, the use of custom loss functions is sometimes necessary, especially in cases where specific tasks require tailored performance metrics. For example, in imbalanced classification problems, custom loss functions can be implemented to focus on improving precision or recall in minority classes. This ensures that the model is fine-tuned not only for overall accuracy but also for specific application needs.

Threshold optimization was another important part of the fine-tuning process. In classification problems, a default threshold of 0.5 is often used to make decisions. However, this threshold can be adjusted to optimize the balance between false positives and false negatives, depending on the use case. Adjusting this threshold can lead to improvements in precision, recall, or the F1-score, particularly when the model's predictions need to be carefully balanced between sensitivity and specificity.

Finally, the number of training epochs was fine-tuned. While increasing the number of epochs can allow the model to learn more complex patterns, it also risks overfitting, especially if the model continues to train after reaching its optimal performance. To address this, early stopping techniques are frequently employed, where training is automatically halted once the performance on the validation set ceases to improve.

The final network architecture consisted of 7 fully connected hidden layers, each composed of 128 neurons and using the ReLU activation function. To prevent overfitting, a dropout layer (rate = 0.1) was applied after each hidden layer, and L2 regularization ($\lambda = 0.6$) was incorporated. The output layer used a sigmoid activation function appropriate for binary classification. The network was trained using the Adadelta optimizer with a learning rate of 0.01, over 400 epochs.

In conclusion, fine-tuning a DNN was an iterative process that required careful adjustments to the network architecture, optimization algorithms, learning rate, and loss functions, all while monitoring the model's performance on both the training and validation sets. By systematically refining these components, the model became more accurate, generalized better to unseen data, and became suitable for real-world applications.

Results

The experiment was conducted by selecting samples from the database, focusing on food items such as fruits, vegetables, and proteins that are known to have a higher likelihood of containing nsLTPs. The selected dataset included mixed values to ensure a comprehensive analysis of the detection capability of the AI model. The validation process covered both the overall platform and the individual machine learning models. Evaluation metrics included the confusion matrix, accuracy, recall, F1-score, and AUC-ROC curve. These metrics were calculated using separate validation and independent test sets, which enabled an assessment of the models' learning capabilities and their generalization to unseen data. While no cross-validation was applied to the final DNN model, all classical machine learning models (e.g., XGBoost, SVM) were evaluated using 5-fold cross-validation to ensure baseline consistency. For the DNN, we opted for an independent test set evaluation after extensive architecture tuning, aiming to simulate real-world deployment and avoid overfitting from internal re-use of validation data. Consequently, no standard deviations or confidence intervals are provided. An independent test set was used to assess the final performance of the models. This dataset was not involved in training or validation and thus provided an unbiased measure of generalization. Performance metrics calculated during testing included accuracy, F1-score, recall, AUC, and confusion matrices.

Table 3 presents the key results of the experiment, including accuracy, precision, recall, and F1-score of the model applied to the dataset. The most relevant results highlight the model's ability to achieve an accuracy of 87% and an F1-score of 89.91%, demonstrating strong predictive power in detecting LTPs in various food samples.

In addition to the quantitative results, a confusion matrix is provided in Fig. 4 to visually represent the performance of the model, showing the true positive, true negative, false positive, and false negative classifications. These results indicate the model's ability to minimize false negatives, a crucial factor in ensuring food safety for individuals with severe allergies. The performance of simpler models like logistic regression, decision trees, and Naive Bayes showed limitations, as they struggled to generalize to unseen data, with logistic regression and decision trees overfitting the validation set and Naive Bayes providing more consistent but still suboptimal results. Advanced models like SVM and XGBoost improved classification, but were hindered by the large dataset and time constraints, limiting hyperparameter

Table 3 Model performance results

Model	Best hyperparameters	Validation results	Test results
Logistic Regression	C=10000.0 solver='liblinear'	Accuracy 76% F1-score: negative 80% positive 69% AUC: 0.86	Accuracy 41% F1-score: negative 53% positive 19% AUC: 0.42
Decision Tree	criterion: entropy max depth: None min samples leaf: 1 min samples split: 2	Accuracy 96% F1-score: negative 96% positive 95% AUC: 0.96	Accuracy 41% F1-score: negative 53% positive 19% AUC: 0.47
Naive Bayes	alpha=0.1 binarize=1.5	Accuracy 60% F1-score: negative 45% positive 68% AUC: 0.64	Accuracy 57% F1-score: negative 20% positive 71% AUC: 0.50
SVM	kernel: rbf gamma: auto C: 100.0	Accuracy 78% F1-score: negative 78% positive 77% AUC: 0.86	Accuracy 46% F1-score: negative 49% positive 42% AUC: 0.47
XGBoost	subsample: 1.0 n estimators: 200 max depth: 5 learning rate: 0.2 gamma: 0 colsample bytree: 1.0	Accuracy 97% F1-score: negative 97% positive 96% AUC: 1.0	Accuracy 41% F1-score: negative 53% positive 20% AUC: 0.48
DNN	Hidden layers: 7 optimizer: Adadelta learning rate: 0.01 epochs: 400 custom loss: (0.125, 0.875) dropout: 0.1 L2 rate: 0.6	Gradually incrementing according to the epochs executed reaching 91%, indicating the model was learning patterns	Accuracy 87% F1-score: negative 82% positive 90% AUC: 0.94

tuning. Ultimately, a DNN emerged as the most effective model, handling complex, non-linear relationships in the data while controlling overfitting through techniques like dropout and L2 regularization. Its robustness across validation and test sets demonstrated the suitability of deep learning to capture intricate patterns in the data.

In order to quantify the model's performance with respect to undetected allergenic cases, the *false negative rate (FNR)* was computed based on the confusion matrix (Fig. 4). This metric reflects the proportion of true positive cases that were incorrectly classified as negative. From the test set results, the model yielded a total of 116 true positives and 4 false negatives. Therefore, the FNR is calculated as

$$\text{FNR} = \frac{\text{False Negatives (FN)}}{\text{True Positives (TP)} + \text{False Negatives (FN)}} = \frac{4}{116+4} = 0.033$$

This corresponds to a false negative rate of approximately *3.3%*, indicating that the model successfully identified nearly all allergenic samples, which is essential for applications in food safety and allergy diagnostics.

Discussion

The significance of early detection of nsLTP extends beyond food safety and plays an important role in the management of a challenging and persistent allergic condition. nsLTPs, and particularly Pru p 3, are highly stable cysteine-rich proteins with resistance to degradation (Durvasulu 2021). Their robustness is compounded by cross-reactivity, which affects individuals with nsLTP sensitivity across multiple food groups. Current diagnostic methods, while accurate, are limited by their complexity and time requirements, rendering them impractical for rapid response needs in public or industrial settings (Álvarez et al. 2024). A significant innovation of this research lies in its application of NIRS specifically for detecting nsLTPs, which are major allergens in various foods.

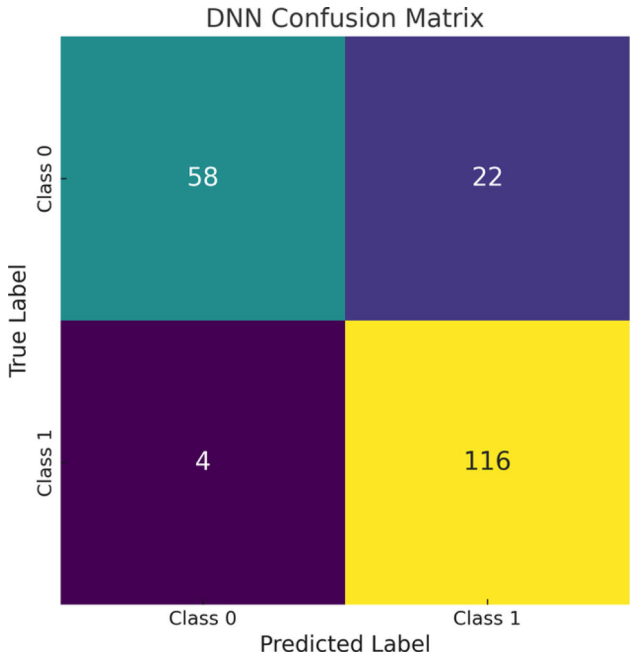

Fig. 4 Confusion matrix of the DNN model with the test set

Previous studies, such as those by Huang et al. (2019) and Liu et al. (2020), primarily focused on general food quality parameters like diastase activity in honey or changes in fruit quality during storage. In contrast, this research uniquely targets the detection of specific allergens, addressing a crucial yet underexplored area in food safety. The focus on allergen detection has the potential to significantly impact public health, particularly for individuals with severe food allergies. The research further distinguishes itself through the sophisticated use of machine learning algorithms to enhance the predictive power of NIRS. While many studies, including those by Bai et al. (2022) and Reda et al. (2023), have employed machine learning in conjunction with NIRS for quality assessment, this research takes a step further by optimizing these models specifically for allergen detection. The integration of diverse models, such as DNNs and XGBoost, coupled with fine-tuning techniques, reflects a higher level of sophistication in data analysis compared to prior work. This methodological advancement enables more accurate and reliable predictions, which is critical when dealing with allergenic proteins that can have serious health consequences. Comparable applications of NIRS for allergen detection have been explored in recent literature. For instance, Jossa-Bastidas et al. (2023) achieved an accuracy of approximately 80% when detecting gluten using portable NIRS devices and deep learning techniques (Jossa-Bastidas et al. 2023). Similarly, peanut allergen detection using spectral analysis combined with chemometrics has shown accuracies ranging from 76 to 85% depending on matrix complexity (Zhou et al. 2021). In light of these benchmarks, our model's performance—with 87% accuracy and 89.9% F1-score—demonstrates competitive results, especially considering the complexity of nsLTP detection. Another innovative aspect of this research is the creation of a large-scale, automated database for handling spectral data. The automation process, which involves converting raw spectral data into structured CSV files using Python scripts, greatly enhances the efficiency of data management. In contrast, earlier studies often dealt with smaller datasets or relied on manual processing, which could lead to errors and inefficiencies (Murzyn et al. 2022). The ability to efficiently manage and analyze large volumes of data allows this research to explore patterns and correlations at a scale not previously feasible, leading to more robust models and insights (Durvasulu 2021). Finally, the research's emphasis on developing a prototype for a potencial real-time allergen detection further sets it apart from existing literature. Although studies like those by Solberg et al. (2023) focused on online quality control in industrial settings, this research aims to create a portable and user-friendly device that can be used in everyday environments, such as homes or restaurants. This approach not only broadens the applicability of NIRS technology but also directly addresses a market gap for consumer-accessible food safety tools.

Conclusions

This research has achieved significant advances in the detection of nsLTPs using artificial intelligence and NIRS. The main achievement of this study is the successful development of a non-invasive AI system capable of detecting LTPs in food with an accuracy of 87% and an F1-score of 89.91%. This confirms that it is indeed possible to integrate AI and NIRS to improve allergen detection in practical settings, providing a robust solution to a growing public health issue.

The system not only contributes to food safety by offering a reliable way to prevent allergic reactions but also demonstrates the potential for broader applications in healthcare and food diagnostics. These results highlight the project's success in developing a scalable and user-friendly platform that can be adopted in various contexts, from homes to public eating establishments, ensuring the safety of individuals with food allergies.

Limitations of this study include the variability in NIRS data, which was influenced by the shape, type, and condition of the tested foods, potentially reducing the consistency of the model's performance. To mitigate this, we included different food types and conditions to help the model generalize better to unseen samples. Rigorous quality controls were implemented throughout the experimental process, including frequent sensor calibration and standardization of measurement procedures. This was complemented by increased

replicates and improved sample preparation to ensure reliable data acquisition.

While classical machine learning models such as XGBoost and SVM were evaluated using 5-fold cross-validation, the final deep neural network (DNN) was assessed using a hold-out approach with an independent test set. This decision was made to avoid optimism bias after extensive hyperparameter tuning and to better simulate real-world deployment. Future iterations will include cross-validation of the final DNN model to further validate its robustness across folds.

In addition, although the dataset includes a variety of food types, further work is needed to systematically assess the model's robustness across different matrix complexities. High-moisture or heterogeneous foods may introduce variability in spectral signatures, which could affect prediction accuracy. Future studies should stratify samples by physicochemical properties (e.g., water content, texture) to evaluate potential performance differences and improve model generalization across diverse food environments.

From a biochemical perspective, it is important to note that the analyzed samples were fruit extracts potentially containing a complex mixture of proteins and other organic compounds. These may alter the NIRS spectral signals of nsLTPs, and the current AI model does not yet quantify protein concentrations, which could be relevant for assessing allergenic potency.

Future work will focus on refining the model to reduce false negatives further and ensure more consistent results across different food types and conditions. To improve specificity, we aim to purify Pru p 3 through recombinant expression to create a standard reference for NIRS calibration and spectral fingerprinting. This will also support future in situ quantification strategies.

Finally, while this prototype outlines the conceptual framework for a potential real-time allergen detection system in environments such as homes, schools, and restaurants, the current version remains at a proof-of-concept stage. Achieving practical deployment will require substantial progress in hardware miniaturization, real-time data processing, and regulatory validation. Applications in public or clinical environments, although promising, would necessitate further studies on user acceptance, training protocols, and compatibility with existing healthcare infrastructure.

Author Contributions A.R.A and I.D.B wrote the draft of the manuscript; B.G.Z, A.O.S, and G.B.S improved the scientifc content of the manuscript. All the authors contributed the proper organization; summary and grammatical error corrections and edited the manuscript critically for important intellectual content and made a meaningful contribution to the work. All authors have approved the manuscript and agree with its submission to the Journal of Health informatics.

Funding Open Access funding provided thanks to the CRUE-CSIC agreement with Springer Nature. This study was funded by smartJAN 2023 of the project "smartJAN: Solución tecnológica para la detección y seguimiento de la calidad de los alimentos para personas con intolerancias, alergias o riesgo de intoxicación alimentaria".

Data Availability No datasets were generated or analyzed during the current study.

Declarations

Ethics Approval and Consent to Participate Not applicable

Consent to Participate Not applicable

Conflict of Interest The authors declare no competing interests.

References

Álvarez P, Aguado R, Molina J, Trujillo-Aguilera A, Villalba M, Díaz-Perales A, Oeo-Santos C, Chicano E, Blanco N, Navas A, Ruiz-León B, Jurado A (2024) Pollen-food allergy syndrome: from food avoidance to deciphering the potential cross-reactivity between Pru p 3 and Ole e 7. Nutrients 16(17):2869. https://doi.org/10.3390/nu16172869. Number: 17 Publisher: Multidisciplinary Digital Publishing Institute. Accessed 24 Mar 2025

Bai X, Zhang L, Kang C, Quan B, Zheng Y, Zhang X, Song J, Xia T, Wang M (2022) Near-infrared spectroscopy and machine learning-based technique to predict quality-related parameters in instant tea. Sci Rep 12(1):3833. https://doi.org/10.1038/s41598-022-07652-z. Nature Publishing Group. Accessed 24 Mar 2025

Bishop CM (2006) Pattern recognition and machine learning. Springer, New York. Accessed 24 Mar 2025. https://link.springer.com/book/9780387310732

Breiman L, Friedman J, Olshen RA, Stone CJ (2017) Classification and regression trees. Chapman and Hall/CRC, New York. https://doi.org/10.1201/9781315139470

Brownlee J (2016) Machine learning in Python: essential techniques for predictive analysis. Wiley.com. Accessed: 24 Mar 2025. https://www.wiley.com/en-us/Machine+Learning+in+Python%3A+Essential+Techniques+for+Predictive+Analysis-p-9781118961742

Burges CJC (1998) A tutorial on support vector machines for pattern recognition. Data Min Knowl Discov 2(2):121–167. https://doi.org/10.1023/A:1009715923555. Accessed 24 Mar 2025

Chen T, Guestrin C (2016) XGBoost: a scalable tree boosting system. In: Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining. KDD '16, pp 785–794. Association for Computing Machinery, New York,

NY, USA. https://doi.org/10.1145/2939672.2939785. https://dl.acm.org/doi/10.1145/2939672.2939785. Accessed 24 Mar 2025
Cortes C, Vapnik V (1995) Support-vector networks. Mach Learn 20(3):273–297. https://doi.org/10.1007/BF00994018. Accessed 24 Mar 2025
Durvasulu MBT (2021) Building efficient storage architectures with Python. Int J Adv Res Educ Technol 08(03). Tech & Apps Mgmt Spec. III, Automatic Data Processing, Inc. NJ, USA. https://doi.org/10.15680/IJARETY.2021.0803018. Accessed 24 Mar 2025
Flame NIR Manual (2025) manual_flamenir.pdf. Accessed 24 Mar 2025. https://www.optosirius.co.jp/OceanOptics/technical/manual_flamenir.pdf
Flame-NIR. https://spectrecology.com/product/flame-nir/. Accessed 24 Mar 2025
Flame-NIR User Guide (2025) https://spectrecology.com/wp-content/uploads/2020/11/Flame-NIR-User-Guide.pdf. Accessed 24 Mar 2025
FLAME-NIR+ (2024). https://ara.ae/spect/spectroscopy-solutions/general-purpose-spectrometers/flame-nir-spectrometer/flame-nir-intsma25/. Accessed 24 Mar 2025
Global Atlas of Allergy (2025) English Version. https://eaaci-cdn-vod02-prod.azureedge.net/KnowledgeHub/education/books/Global%20Atlas%20of%20Allergy%20-%20English%20Version.pdf. Accessed 24 Mar 2025
Goodfellow I, Bengio Y, Courville A (2016a) Deep learning. MIT Press. Accessed 24 Mar 2025. https://mitpress.mit.edu/9780262035613/deep-learning/
Goodfellow I, Bengio Y, Courville A (2016b) Deep learning. Online. https://www.deeplearningbook.org/. Accessed 24 Mar 2025
Huang Z, Liu L, Li G, Li H, Ye D, Li X (2019) Nondestructive determination of diastase activity of honey based on visible and near-infrared spectroscopy. Molecules 24(7). https://doi.org/10.3390/molecules24071244
Jossa-Bastidas O, Sanchez AO, Bravo-Lamas L, Garcia-Zapirain B (2023a) IoT system for gluten prediction in flour samples using NIRS technology, deep and machine learning techniques. Electronics 12(8). https://doi.org/10.3390/electronics12081916
Jossa-Bastidas O, Sanchez AO, Bravo-Lamas L, Garcia-Zapirain B (2023b) IoT system for gluten prediction in flour samples using NIRS technology, deep and machine learning techniques. Electronics 12(8):1916. https://doi.org/10.3390/electronics12081916
LeCun Y, Bengio Y, Hinton G (2015) Deep learning. Nature 521(7553):436–444. https://doi.org/10.1038/nature14539. Accessed 24 Mar 2025
Liu Y, Zhang Y, Jiang X, Liu H (2020) Detection of the quality of juicy peach during storage by visible/near infrared spectroscopy. Vib Spectrosc 111:103152. https://doi.org/10.1016/j.vibspec.2020.103152. Accessed 24 Mar 2025
Maestro ES (2025) ¿Qué es la alergia a la proteína LTP? https://www.quironsalud.com/blogs/es/alergologia-infantil/alergia-proteina-ltp. Accessed 2025-03-24
Mazas YA Implicación de las LTP en la alergia alimentaria. Accessed: Mar. 24, 2025 (2025). https://www.arapap.es/wp-content/uploads/2020/04/Implicacion-de-las-LTP-en-la-alergia-alimentaria.pdf
Missaoui K, Gonzalez-Klein Z, Pazos-Castro D, Hernandez-Ramirez G, Garrido-Arandia M, Brini F, Diaz-Perales A, Tome-Amat J (2022) Plant non-specific lipid transfer proteins: an overview. Plant Physiol Biochem 171:115–127. https://doi.org/10.1016/j.plaphy.2021.12.026. Accessed 2025-03-24
Murzyn CM, Jans ER, Clemenson MD (2022) SPEARS: a database-invariant spectral modeling API. J Quant Spectrosc Radiat Transf 277:107958. https://doi.org/10.1016/j.jqsrt.2021.107958. Accessed 24 Mar 2025
Ocean Photonics (2025a) Integrated halogen light source diffuse reflection probe TC-DR-PROBE | Product Information | Ocean Photonics Co., Ltd. https://www.oceanphotonics.com/product/oceanoptics_000739.html. Accessed 24 Mar 2025
Ocean Photonics (2025b) Laboratory-grade patch cord optical fiber | Product Information | Ocean Photonics Co., Ltd. https://www.oceanphotonics.com/product/oceanoptics_000065.html. Accessed 24 Mar 2025
Peduzzi P, Concato J, Kemper E, Holford TR, Feinstein AR (1996) A simulation study of the number of events per variable in logistic regression analysis. J Clin Epidemiol 49(12):1373–1379. https://doi.org/10.1016/s0895-4356(96)00236-3
Quinlan JR (1986) Induction of decision trees. Mach Learn 1(1):81–106. https://doi.org/10.1007/BF00116251. Accessed 2025-03-24
Reda R, Saffaj T, Bouzida I, Saidi O, Belgrir M, Lakssir B, El Hadrami EM (2023) Optimized variable selection and machine learning models for olive oil quality assessment using portable near infrared spectroscopy. Spectrochim Acta Part A Mol Biomol Spectrosc 303:123213. https://doi.org/10.1016/j.saa.2023.123213. Accessed 24 Mar 2025
Rossi CM, Lenti MV, Merli S, Licari A, Marseglia GL, Di Sabatino A (2023) Immunotherapy with Pru p 3 for food allergy to peach and non-specific lipid transfer protein: a systematic review. Clin Mol Allergy 21(1):3. https://doi.org/10.1186/s12948-023-00184-5. Accessed 24 Mar 2025
Schopf M, Wehrli MC, Becker T, Jekle M, Scherf KA (2021) Fundamental characterization of wheat gluten. Eur Food Res Technol 247(4):985–997. https://doi.org/10.1007/s00217-020-03680-z. Accessed 24 Mar 2025
Seki H, Ma T, Murakami H, Tsuchikawa S, Inagaki T (2023) Visualization of sugar content distribution of white strawberry by near-infrared hyperspectral imaging. Foods 12(5). https://doi.org/10.3390/foods12050931
Shewhart WA, Wilks SS (2013) Wiley series in probability and statistics. In: Applied logistic regression. John Wiley & Sons, Ltd, pp 501–510. https://doi.org/10.1002/9781118548387.scard. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/9781118548387.scard. https://onlinelibrary.wiley.com/doi/abs/10.1002/9781118548387.scard. Accessed 24 Mar 2025
Solberg LE, Wold JP, Dankel K, Øyaas J, Måge I (2023) In-line near-infrared spectroscopy gives rapid and precise assessment of product quality and reveals unknown sources of variation–a case study from commercial cheese production. Foods 12(5):1026. https://doi.org/10.3390/foods12051026. Number: 5, Multidisciplinary Digital Publishing Institute. Accessed 24 Mar 2025
Zhou Y, Wu X, Liu Y, Zhang Y (2021) Detection of peanut allergens using infrared spectroscopy and chemometric models. Food Control 130:108367. https://doi.org/10.1016/j.foodcont.2021.108367
Zuo J, Peng Y, Li Y, Zou W, Chen Y, Huo D, Chao K (2023) Nondestructive detection of nutritional parameters of pork based on NIR hyperspectral imaging technique. Meat Sci 202:109204
Zuo J, Peng Y, Li Y, Chen Y, Yin T, Chao K (2025) Integrating transfer learning and spectroscopy for enhanced pork spoilage assessment using correlation analysis. Food Chem 465:142117