scientific reports

Check for updates

OPEN

Enhancing the detection of LTP through lyophilized protein samples and NIR spectroscopy with explainable deep learning

Ainhoa Osa-Sanchez[1]✉, Itxasne Del Barrio[1], Ganeko Bernardo-Seisdedos[2], Sara Pozo[2] & Begonya Garcia-Zapirain[1]

Lipid transfer proteins (LTPs) are clinically relevant allergens widely present in plant-based foods, and their reliable detection in complex food matrices remains a major challenge. In this study, we developed an integrated framework combining near-infrared spectroscopy (NIRS), deep learning, and explainability methods to enable accurate and interpretable identification of LTPs. A total of 11,688 spectral measurements were collect ed using a FLAME-NIR spectrometer (940–1700 nm) from homogenized food samples, purified Pru p 3 and Ara h 9 proteins, and their mixtures with LTP-free matrices such as yogurt and powdered milk. Spectral preprocessing involved first derivative transformation, Standard Normal Variate correction, and feature scaling, followed by dimensionality reduction through a 1D convolutional autoencoder, which generated 64-dimensional latent embeddings. These representations were used to train two deep learning classifiers Convolutional Neural Networks (CNNs) and TabTransformer optimized via Bayesian optimization. The inclusion of purified protein embeddings substantially improved classification performance. The CNN model achieved the highest performance with 95.8% accuracy, 97.3% precision, 96.9% F1-score, and an AUC-ROC of 0.954, outperforming the TabTransformer, which nonetheless reached 95.19% accuracy and 96.4% F1-score. Model explainability was addressed using SHAP and LIME, which identified key latent features corresponding to specific spectral regions (940–1700 nm) associated with allergenic signatures. Compared to baseline models, the protein-enhanced framework demonstrated marked improvements in specificity and overall robustness. These results highlight the value of incorporating purified protein information into AI-based spectral analysis, offering a portable, non-destructive, and interpretable strategy for allergen detection in food safety applications.

Keywords Allergens, Near-infrared spectroscopy, Artificial intelligence, Classification, Explainable artificial intelligence, Lyophilization

Nonspecific lipid transfer proteins (nsLTPs) are small, basic proteins (7–10 kDa) ubiquitously present in plant-derived foods and characterized by a conserved α-helical fold stabilized by four intramolecular disulfide bridges. This compact structure renders nsLTPs thermoresistant and protease-resistant, conferring both allergenic stability and high cross-reactivity across botanical families[1,2]. Among them, *Pru p 3* and *Ara h 9* have been identified as major allergens in peach and peanut, respectively, and are implicated in systemic allergic reactions, particularly in LTP-sensitized individuals in Southern Europe.

Due to their conformational robustness and persistence in processed food matrices, nsLTPs pose significant detection and analytical challenges. Traditional immunoassays suffer from limited sensitivity in thermally treated samples, while mass spectrometry-based techniques often require complex sample preparation. To overcome these barriers, the integration of near-infrared (NIR) spectroscopy with advanced machine learning architectures has emerged as a gold standard for non-destructive, reagent-free analysis in modern food control. Recent studies underscore the high accuracy of this combination in screening complex matrix alterations and tracking multi-target quality parameters. For instance, Lanjewar et al. developed a portable NIR setup (900–1700 nm) optimized via Savitzky-Golay, Multiplicative Scatter Correction (MSC), and Standard Normal Variate (SNV) preprocessing to successfully evaluate and predict water adulteration ranges in milk systems[3]. Beyond

[1]eVIDA Research Group, University of Deusto, Bilbao 48007, Spain. [2]DeustoMED Research Group, University of Deusto, Bilbao 48007, Spain. ✉email: ainhoa.osa.sanchez@deusto.es

liquid profiling, modern spectroscopic techniques combined with chemometrics have proven highly effective for non-destructive nutritional mapping, such as using NIR hyperspectral imaging to extract protein and fat distributions in complex meat matrices[4]. Furthermore, advanced machine learning paradigms, such as transfer learning integrated with visible near-infrared (VNIR) spectroscopy, have been successfully deployed to handle complex spectral features and evaluate chemical spoilage processes over time[5]. Extending this paradigm to chemical contaminants, Rosa et al. demonstrated that UV-Vis-NIR spectroscopy (200–1700 nm) combined with robust chemometric preprocessing can effectively isolate and quantify trace formaldehyde additions in diverse cow and buffalo milk matrices with outstanding determination coefficients ($R^2 > 0.998$)[6]. While these latest modeling frameworks validate the efficacy of targeting structural overtones to identify macro-adulterants, nutritional distributions, or chemical additives within various controlled matrices, expanding these non-targeted spectroscopic pipelines to the screening of complex trace biological components–such as specific native or allergen isoforms–remains an uncharted frontier.

In a previous study[7], we attempted to bridge this gap by introducing a novel approach integrating near-infrared spectroscopy (NIRS) with deep learning architectures to classify LTP-positive versus LTP-negative foods. Using whole-food samples and spectral data in the 940-1700 nm range, we achieved a significant classification accuracy via convolutional and transformer-based models, coupled with explainable AI tools for feature attribution. However, that approach was constrained by its reliance on natural variables, where LTP presence, concentration and isoform composition were unknown. This uncontrolled variability limited both model interpretability and the ability to relate spectral features to specific molecular signatures. In our previous work, we identified several significant limitations related to the spectral instrumentation utilized and the inherent variability present in natural food samples. Specifically, the restricted spectral range of the FLAME-NIR+ spectrometer (940–1700 nm) limited the generalizability of our results across different analytical platforms, while the relatively small dataset encompassing only 12 food categories hindered broader applicability across diverse food matrices[8]. Furthermore, while blended samples provided a degree of homogeneity, they failed to accurately reflect the complexity of intact or processed foods, where the distribution of allergens is often heterogeneous[9].

At the same time, a major outstanding problem in the application of near-infrared (NIR) spectroscopy combined with deep learning for allergen screening is the black-box nature of the models, which typically fails to provide biochemical accountability. Even when utilizing traditional explainable artificial intelligence (XAI) methods, the uncontrolled natural variability in the presence, concentration, and isoform composition of Lipid Transfer Proteins (LTPs) within raw food samples introduces confounding variables. This data complexity severely impedes precise biochemical interpretations of the models' decisions, leaving the true molecular origins of the AI's discriminant features largely unresolved.

To overcome these critical limitations, this work introduces a novel framework that bridges targeted biochemical engineering with explainable deep learning. The main contribution of this study lies in an improved methodology incorporating the use of highly purified recombinant LTPs spiked into LTP-free food matrices. Defined quantities of recombinant Pru p 3 and Ara h 9 were introduced into controlled matrices, specifically yogurt and powdered milk. These specific matrices were selected because, as animal-derived dairy products, they are inherently free of plant-derived nsLTPs (providing an absolute blank baseline), while presenting distinct physical states (a semi-solid emulsion versus a dry powder) and rich background profiles of non-target proteins and lipids to rigorously challenge the robustness of our models. This design enables the construction of a robust spectral dataset with known allergen concentrations and matrix configurations. The novelty of this strategy is twofold: first, it permits supervised deep learning training under strictly constrained biochemical conditions; second, it enables a deterministic retrospective mapping procedure that successfully attributes key latent spectral features back to true, protein-specific physical molecular vibrations (associated with C-H first overtones and N-H combination bands). The following sections describe the expression and purification of Pru p 3 and Ara h 9, the design of the spiking assay, and the computational workflow developed for spectral preprocessing, model training, and advanced explainability analysis.

Materials

The dataset employed in this investigation is an extensive compilation of spectral measurements acquired from blended food samples, particularly focusing on reflectance and absorbance.

Protein purification

Recombinant Pru p 3 and Ara h 9 proteins were produced in *Escherichia coli* BL21 (DE3) cells transformed with plasmids encoding His-tagged versions of each allergen. Cultures were grown in 1 L LB media supplemented with 50 μg/mL of kanamycin at 37 °C until reaching an OD_{600} of 0.6–0.8. Protein expression was induced with 0.5 mM IPTG and incubated overnight at 37 °C with agitation.

Cells were harvested by centrifugation and resuspended in lysis buffer (10 mM Tris-HCl pH 7.5, 0.05% NaN_3, 0.1% Triton X-100, 0.5 mM PMSF, protease inhibitor cocktail, and 250 μL DNase I per 40 mL). Lysis was performed by sonication on ice (20 s on / 40 s off, 60% amplitude, 10 min total), and the lysate was centrifuged at 4 °C for 30 min at 20000 g. The resulting supernatant was discarded, and the pellet was then resuspended in binding buffer (20 mM Tris-HCl pH 8.0, 500 mM NaCl, 5 mM imidazole, 0.05% NaN_3, 8 M urea) and incubated overnight at room temperature. A second centrifugation step yielded a supernatant (SN) containing the resolubilized protein.

Protein purification was performed using a 5 mL HisTrap TALON Crude column (Cytiva) on an *ÄKTA* system. SN was filtered (0.22 μm) and sequentially loaded. After washing with binding buffer, bound proteins were eluted using a linear gradient of imidazole (up to 500 mM) in elution buffer (same as binding buffer with 500 mM imidazole). Elution was monitored at 280 nm and fractions were pooled accordingly.

Eluted proteins were dialyzed in three steps: first against dialysis buffer (20 mM Tris-HCl pH 7.4, 150 mM NaCl, 5 mM DTT) for 5 h, then against fresh dialysis buffer overnight at $4\ ^{\circ}C$, and finally against phosphate buffer (20 mM Na_2HPO_4, pH 7.0) for another 5 h. When necessary, proteins were concentrated using Amicon Ultra centrifugal filters (MWCO 3 kDa), aliquoted, and finally lyophilized to obtain a stable powdered form suitable for analytical applications.

The final protein yield for Pru p 3 was 3.87 mg with a concentration of 47.2 μM, and for Ara h 9 was 4.24 mg at 52.3 μM.

Sample acquisition and preparation

NIRS system calibration and sample preparation

Prior to data acquisition, the FLAME-NIR$^+$ spectrometer was allowed to warm up for approximately 30 minutes to stabilize its performance. Once stable, the system was calibrated using a Spectralon standard and an absolute black spectrum to ensure accurate measurements of both absorbance and reflectance. A complete calibration routine was performed daily whenever the spectrometer was used to guarantee consistent and reliable data throughout the study.

All materials and tools involved in sample preparation and measurement, including the blender (turmix), laboratory spatula, and measurement plates (white and black), were thoroughly cleaned with distilled water and, when necessary, sterilized or replaced to prevent contamination. Gloves were worn at all times to minimize external contamination, ensuring the integrity of the collected data.

For solid food samples, blending was performed to obtain homogeneous mixtures, while yogurt was homogenized manually with a sterilized spatula to preserve its natural texture. Powdered milk samples were prepared using four formulations to introduce variability:

- 60 g of powdered milk with 50 mL distilled water (30 g per 25 mL),
- 50 g with 50 mL distilled water (25 g per 25 mL),
- 50 g with 40 mL distilled water,
- 60 g with 40 mL distilled water.

Preparation of purified protein samples for AI training

To train the AI model with pure protein, the samples obtained from the purification process were first lyophilized. Two different proteins were used: Pru p 3 and Ara h 9. Before lyophilization, each protein sample contained 0.5 mg of protein in liquid form, distributed individually in Eppendorf tubes. After the lyophilization process, the recovered total solid mass (which included the remaining buffer salts from the dialysis step) was 4 mg for Pru p 3 and 5 mg for Ara h 9, each containing the 0.5 mg of absolute protein.

As a positive control, the lyophilized powder containing the protein (4 mg of Pru p 3 powder and 5 mg of Ara h 9 powder) was measured using the NIRS sensor. The measurements were performed on both white and black sample plates, and in three different sample orientations to ensure consistency and reduce positional bias, as done with the previously collected samples.

Following this, the reconstituted protein was mixed with yogurt, a food matrix that does not contain LTPs. First, the pure protein was mixed with 5 g of yogurt. Then, the mixture was sequentially diluted by adding 15 g and then 30 g of yogurt. From each resulting mixture, a 0.6 g sample was taken for NIRS analysis.

The same procedure was followed using powdered milk, another LTP-free matrix. The protein was first mixed with 5 g of powdered milk, followed by additions of 15 g and 30 g. Again, 0.6 g samples were taken from each mixture for measurement.

This protocol was applied in parallel for both Pru p 3 and Ara h 9, ensuring standardized sample preparation across all test conditions. The objective of using purified LTPs was to improve the accuracy, specificity, and interpretability of the AI model by providing it with high-quality, well-defined spectral data corresponding exclusively to the target allergens.

Data collection procedure

Once all equipment, food samples, and purified protein samples were prepared, the data collection process began with strict attention to detail to ensure accuracy and reliability. For food samples, a portion of the homogenized mixture was placed onto a piece of Parafilm positioned on the white reference plate using a sterilized laboratory spatula. The NIRS sensor was carefully positioned over the sample, and a brief stabilization period was observed to allow the system to equilibrate. Absorbance and reflectance measurements were then recorded, generating two separate text files corresponding to the white plate. The Parafilm with the sample was subsequently transferred to the black reference plate without altering the sample's orientation, and the same measurements were repeated, producing two additional text files. To account for intra-sample variability, the sample on the Parafilm was gently adjusted to expose different aspects of the mixture. Measurements were then taken first on the black plate, followed by the white plate, yielding four additional text files. A third adjustment of the sample on the white plate allowed for another set of measurements, which was then transferred to the black plate for the final set, producing a total of 12 text files per sample (six configurations with both absorbance and reflectance). This procedure was applied to 20 samples per food piece and repeated across four distinct pieces of each food type. For purified protein samples, the lyophilized proteins (Pru p 3 and Ara h 9) were measured both as positive controls and when mixed with LTP-free matrices (yogurt and powdered milk). Measurements were performed on both white and black plates and in three different sample orientations, following the same protocol as for the food samples. For each protein, 84 measurements were collected for the mixtures with yogurt and powdered milk, ensuring consistency, reproducibility, and minimization of positional bias.

Dataset composition and integration

The dataset used in this study is an integrated compilation of spectral measurements designed to detect Lipid Transfer Proteins (LTPs). It consists of two main components:

1. *Food Matrix Baseline:* A primary dataset comprising 11,520 measurements from 12 food types. These samples were collected following the protocol described in[7], providing a robust baseline of food variability.
2. *Targeted Protein Enhancement:* To improve the model's specificity, 168 new measurements of purified and lyophilized Pru p 3 and Ara h 9 proteins were added. These include both positive controls (pure protein) and spiked mixtures with LTP-free matrices (yogurt and powdered milk) at controlled concentrations.

As shown in Table 1, the final consolidated dataset includes 11,688 total samples. This hybrid approach allows the AI model to learn from both the complex, natural environment of food matrices and the high-purity spectral signatures of the target allergens.

For each sample analyzed, a total of 139 variables were collected, encompassing spectrometer settings, measurement conditions, spectral data, and identification metadata. However, only the spectral data–comprising absorbance and reflectance values at specific pixel positions corresponding to individual wavelengths–were used as input features for AI model training. The remaining variables were recorded for documentation and reproducibility purposes but were excluded from the modeling process. This structure ensured that the AI model received only relevant spectral information while maintaining complete records of all measurement parameters for reproducibility and quality control.

Flame-NIR device

The Flame-NIR[10], spectrometer has been used in our study to take the measurements required for allergen detection in various food matrices. This device integrates NIR technology into a compact and modular system, and is equipped with a Hamamatsu G8160-03 InGaAs detector operating in the 940–1700 nm wavelength range. This spectral region is highly informative for identifying molecular overtones and combination bands associated with functional groups such as O–H, N–H, and C–H, which are relevant for the characterization of allergenic compounds.

One of the major advantages of the Flame-NIR is its portability, which facilitates its use outside of controlled laboratory environments. With a weight of only 265 grams and dimensions of 89.1 mm $\times$ 63.3 mm $\times$ 31.9 mm, it is well-suited for on-site, non-destructive analysis. In our study, this allowed for rapid spectral acquisition in diverse settings, including industrial, domestic, and potentially remote locations, enhancing the practical applicability of NIR spectroscopy for real-world allergen screening.

To fully document each measurement, a total of 139 variables were collected for every analyzed sample, encompassing spectrometer settings, measurement conditions, spectral data, and identification metadata. These variables were organized into four categories:

- *Spectrometer Settings* (7 features):
  - Integration Time (sec)
  - Scans to average
  - Trigger mode
  - Nonlinearity correction enabled
  - Boxcar width
  - Storing dark spectrum
  - Spectrometer

Item	LTP (Y/N)	Number of samples
Pru p 3 protein	Y	84
Ara h 9 protein	Y	84
Potato	N	960
Cucumber	N	960
Powdered milk	N	960
Natural yogurt	N	960
Banana	Y	960
Orange	Y	960
Apple	Y	960
Pear	Y	960
Kiwi	Y	960
Strawberry	Y	960
Peaches	Y	960
Tomato	Y	960
Total		11,688

Table 1. Collected samples.

- *Measurement Conditions* (2 features):
  - X-Axis mode
  - Number of pixels in the spectrum
- *Spectral Data* (128 features):
  - Absorbance and reflectance values at different pixel positions, each corresponding to a specific wavelength
- *Identification and Metadata* (2 features):
  - User ID
  - Tag (sample identifier)

Methodology

Tabular data is used to train the models in the proposed framework, which is seen in Fig. 1. Hyperparameters are adjusted via Bayesian optimization. CNN and TabTransformer are the two classifier types that are examined in this study.

Lastly, strategies of interpretation and explanation are applied. By helping to validate automatic detection and offering insights into model decision-making processes, SHAP and LIME techniques enhance the interpretability of the framework.

Data preprocessing

To enhance signal quality and get the data ready for training, data preparation was done in many steps. The following describes the procedure's primary steps, as seen in Fig. 2:

1. *Data Labeling*: The various measurements of each food were first combined into a `.csv` file because they are acquired in text files. Columns containing user and device information were removed in the next phase. The following coding was then used to give binary labels based on the existence of LTP:
   - *Positive LTP (1)*: Samples containing proteins such as *Ara h 9* or *Pru p 3*. Examples of foods with positive LTP include kiwi, apple, peach, orange, banana, tomato, strawberry, and pear.
   - *Negative LTP (0)*: Samples without detectable LTPs. Examples of foods with negative LTP include yogurt, potatoes, cucumbers, and powdered milk.
2. *Data type conversion*: In order to ensure appropriate interpretation of the decimal separators, a conversion to numerical values was carried out upon detecting numerical values saved as text.

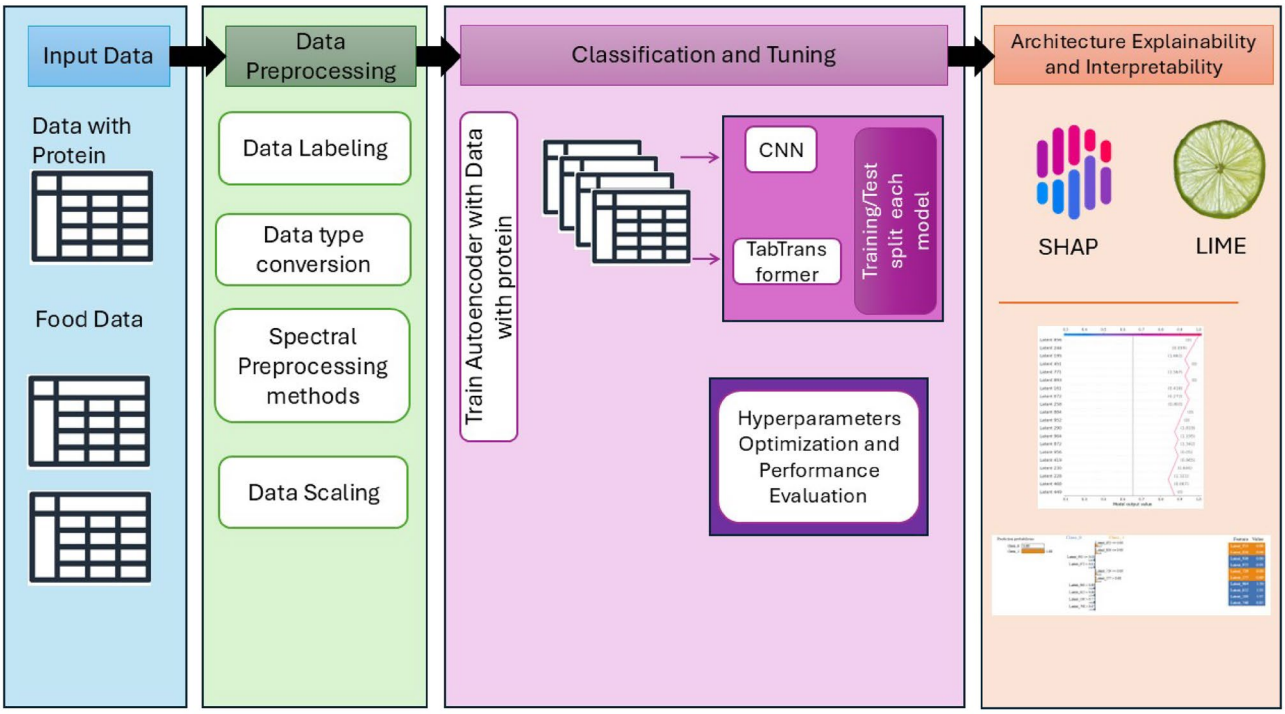

Fig. 1. Overview of the proposed framework for LTP detection, tuning, and explainability.

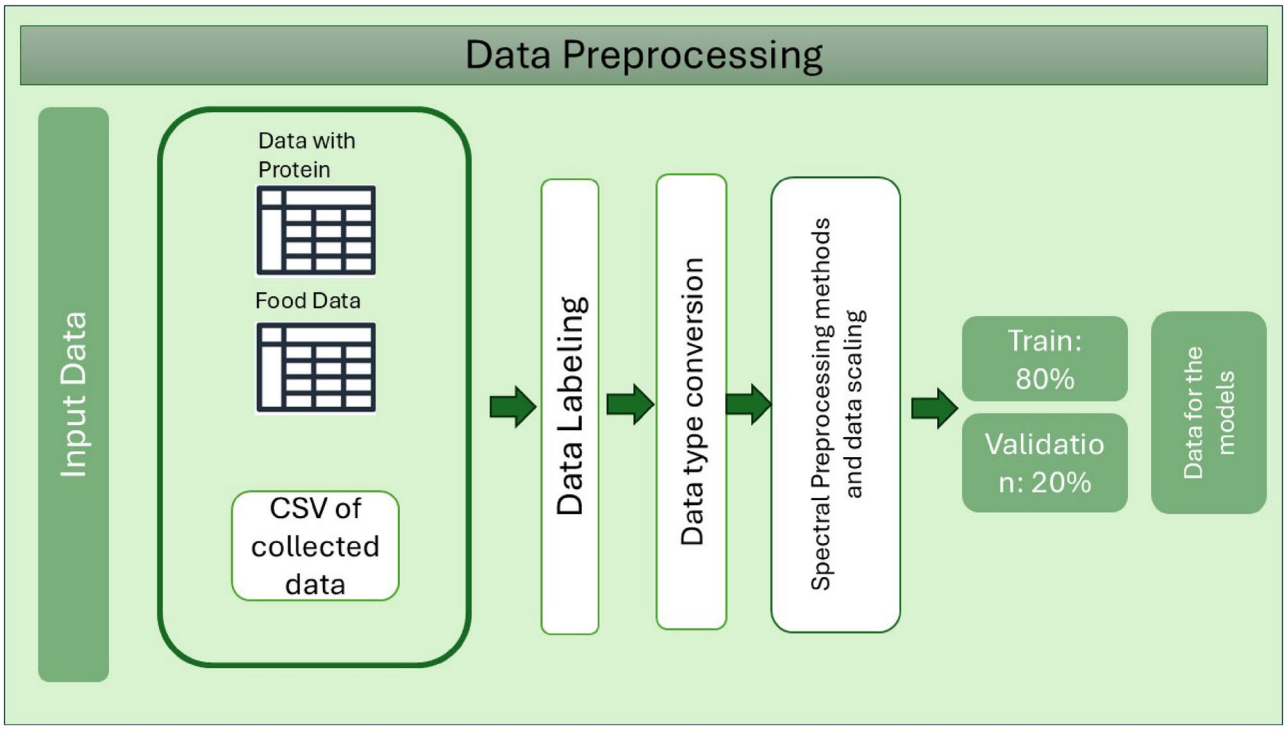

Fig. 2. The utilized preprocessing approach.

3. *Application of spectral preprocessing methods and data scaling*: In the current study, we apply spectral preprocessing methods for LTP classification tasks: first derivative to emphasize spectral features, standard normal variate for dispersion normalization, and StandardScaler to achieve zero mean and unit variance on the entire dataset.

    - *First Derivative (FD):* Widely recognized as a robust preprocessing strategy to resolve overlapping peaks, eliminate baseline shifts, and enhance subtle spectral features. Taking the derivative of spectral data accentuates key biochemical variations that might otherwise be masked by broader background signals, thereby facilitating better discrimination among different classes of materials based on their specific molecular vibrations[11]. Furthermore, combining FD with subsequent normalization has been shown to significantly outperform isolated methods in predicting components within complex biological matrices[12].
    - *Standard Normal Variate (SNV):* Particularly effective in normalizing spectra to reduce physical scatter effects caused by variations in sample density, particle size, and optical path length. Crucially, SNV operates row-wise (horizontally) on each individual spectrum independently. Wang et al. provided extensive evidence of SNV's effectiveness in removing multiplicative interference and achieving baseline stability across diverse spectral measurements, making it an essential tool for enhancing prediction model performance[11]. By minimizing interferences caused by these extraneous physical factors, SNV streamlines the data, reduces inherent intra-sample variability, and substantially improves the generalization capabilities of classification algorithms[11].
    - *StandardScaler:* In contrast to SNV, this technique operates column-wise (vertically) across the entire dataset for each independent wavelength channel. It standardizes the features by removing the mean and scaling them to unit variance. This step is a strict requirement for deep learning architectures (such as CNNs and Transformers) to ensure that all input dimensions share the same statistical scale, preventing high-amplitude spectral regions from disproportionately dominating the gradient descent during optimization[13]. Proper column scaling ensures that the neural network treats all wavelength variables equitably, notably augmenting the predictive and generalization power of machine learning models applied to spectroscopy[14].

Consequently, the configuration of preprocessing methods implemented in this research, comprising the First Derivative, SNV, and StandardScaler, collectively yielded optimal results for LTP classification tasks. The synergistic effects of these preprocessing techniques help in transforming raw spectral data into a format that is not only more interpretable but also significantly boosts the performance of classification models, embodying a critical step in spectral data analysis (Fig. 3).

Classification and tuning

The proposed framework employed two distinct models utilizing tabular data. The hyperparameters were also adjusted using Bayesian optimization.

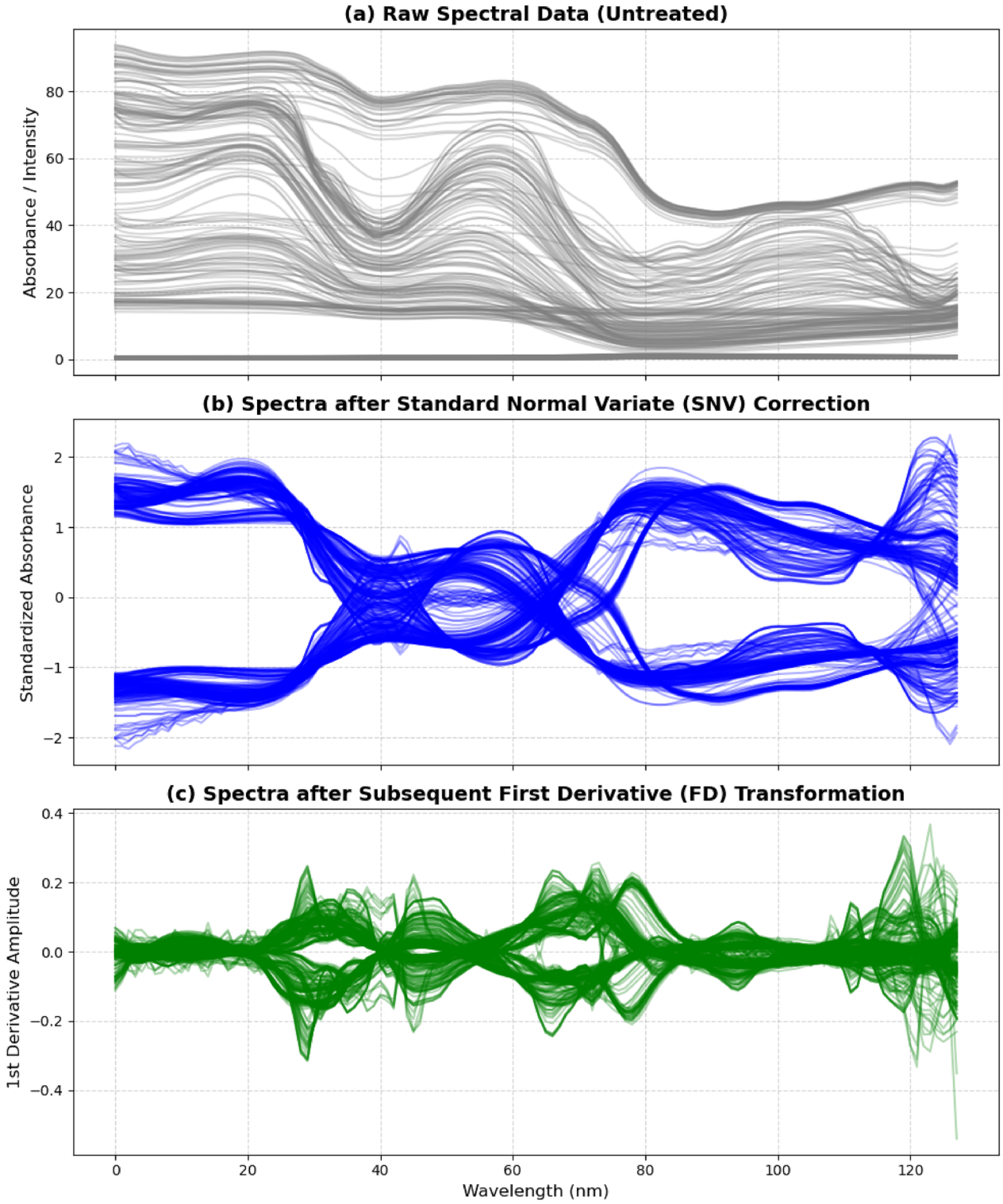

Fig. 3. Sequential stages of the spectral preprocessing pipeline for the 11,688 near-infrared (NIR) measurements: (**a**) Raw, untreated spectra showing significant baseline shifts and multiplicative scatter effects; (**b**) Spectral profiles following row-wise Standard Normal Variate (SNV) correction, which effectively standardizes the physical path length variations; and (**c**) Final enhanced spectra after applying the subsequent First Derivative (FD) transformation, resolving overlapping bands and accentuating localized molecular vibrations before feature extraction.

Models

Two potential models are employed in the categorization framework that this study presents. To evaluate NIR spectral data, we used CNNs and TabTransformer, two deep learning architectures.

- *CNN*: This kind of neural network is used to evaluate grid-structured data. CNNs use a method called convolution to extract pertinent characteristics from incoming data[15]. CNNs can extract both local and hierarchical patterns, which makes them suitable for spectral data analysis. Equation 1 defines the convolution operation as follows:

$$y_i = f\left(\sum_{j=0}^{k-1} w_j x_{i+j} + b\right) \tag{1}$$

When the input spectrum is denoted by x, the convolutional filter of size k is represented by w, the bias term is denoted by b, and the activation function is denoted by $f(\cdot)$. Important spectral fluctuations were preserved while dimensionality was decreased through the use of pooling layers. The collected characteristics were further processed for classification by fully linked layers. CNNs can improve classification across classes by efficiently modeling local correlations in spectral data.

- *Latent Feature Extraction with Autoencoder1D*: In the context of improving feature representation and reducing the dimensionality of NIR data, the implementation of a one-dimensional convolutional autoencoder (Autoencoder1D) is a notable approach trained with the data with the proteins. Autoencoders are fundamentally unsupervised neural networks designed to reconstruct input data while learning compact latent representations that encapsulate the most salient information contained within the data[16,17]. By utilizing this mechanism, the autoencoder achieves an efficient reduction in dimensionality, which is essential given the high-dimensional nature of hyperspectral data[16]. The architecture of Autoencoder1D consists of an encoder and a decoder. The encoder is composed of convolutional layers with ReLU activation functions, followed by pooling operations that systematically reduce the dimensionality of the input while capturing high-level features[17]. This progressive extraction of features through convolutional layers allows the autoencoder to learn relevant data representations that enhance the quality of subsequent classification tasks[9]. The resulting compressed feature maps are flattened and passed through a fully connected layer, establishing a latent space of 64 dimensions, which serves as the foundation for downstream models. Conversely, the decoder is structured to mirror the encoder, employing transposed convolutions and a suitable activation function to effectively reconstruct the original input signal[18]. This dual structure ensures that the latent representations retain the essential features for reconstruction, thus underscoring the importance of dimensionality reduction in feature learning. The model is trained using a mean squared error (MSE) loss function in conjunction with the Adam optimizer, which is known for its efficiency in training deep neural networks through adaptive learning rates[19]. Incorporating early stopping methods further mitigates the risk of overfitting, ensuring that the model generalizes well on unseen data[20]. Upon completion of the training phase, the encoder's capability to extract the 64-dimensional latent embeddings from the NIR spectra is leveraged. These embeddings, characterized as compressed and denoised representations, are then utilized by classification models such as CNN and other deep learning architectures[16]. This method of employing compressed representations not only enhances the performance of classifiers but also improves robustness due to the reduced noise interference in the data[17]. The interplay between latent feature extraction via Autoencoder1D and subsequent classifier application underscores the significance of dimensionality reduction methods in machine learning pipelines focused on hyperspectral imaging and classification tasks. The term 'protein embeddings' refers to the 64-dimensional latent vectors generated by the Autoencoder1D when specifically trained/fed with purified Pru p 3 and Ara h 9 spectral data.
- *TabTransformer*: TabTransformer is intended for organized tabular data, making it appropriate for cases that include NIR spectral characteristics. TabTransformer is based on self-attention Transformers [21]. The model processes numerical features through normalization and categorical features via parametric embeddings, which are then contextualized using Transformer encoder layers. The multi-head attention mechanism calculates relationships between categorical features through Eq. 2.

$$\text{Attention}(Q, K, V) = \text{softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V \tag{2}$$

where Q, K, and V are the query, key, and value matrices, respectively, and d_k is the dimension of the key vectors. The self-attention mechanism allows the model to learn global spectral dependencies, capturing complex interactions between wavelengths. This approach is particularly useful for NIR spectroscopy, where subtle spectral variations across the entire range contribute to classification performance. For final predictions, these contextual embeddings are concatenated with normalized numerical data and then run through an MLP. According to empirical findings, TabTransformer performs better than alternative deep tabular models.

Hyperparameter tuning

Bayesian optimization with Gaussian processes was used to train and tune the classifier models[22]. By incorporating previous assessments, Bayesian optimization improves the accuracy and efficiency of hyperparameter tuning

in contrast to random or traditional grid search. Equation 3 provides a mathematical description of Bayesian optimization.

$$x = \arg\max_{x \in X} f(x) \tag{3}$$

The main architectural components of each classifier model were adjusted. These gadgets were all made with unique features to enhance work performance.

Performance evaluation assessment

To evaluate and systematically quantify the classification performance of the deep learning models, several standard statistical metrics were utilized. These measurements offer crucial insights into the efficacy, reliability, and safety constraints of the proposed allergen detection framework[23].

Accuracy represents the proportion of correctly identified instances (both true positives and true negatives) out of the total number of evaluated samples. A higher accuracy implies a superior overall capacity of the model to correctly identify the presence or absence of nsLTPs across the tested food matrices.

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \tag{4}$$

Sensitivity (Recall) measures the model's capacity to correctly identify true positive instances out of all actual positive samples. In the context of food allergen screening, maximizing sensitivity is of paramount clinical importance, as it directly minimizes dangerous False Negatives that could compromise consumer safety.

$$\text{Sensitivity} = \frac{TP}{TP + FN} \tag{5}$$

Specificity evaluates a classifier's capacity to correctly recognize true negative instances (e.g., blank or LTP-free matrices). It is computed as the percentage of true negative samples relative to the total number of actual negative cases. High specificity values are essential to avoid unnecessary False Positives and over-warning.

$$\text{Specificity} = \frac{TN}{TN + FP} \tag{6}$$

Precision indicates the percentage of correctly classified positive cases among all instances predicted as positive by the model. It is calculated as the ratio of true positives to the total sum of true positives and false positives. A higher precision score denotes a reduced rate of false positive errors.

$$\text{Precision} = \frac{TP}{TP + FP} \tag{7}$$

F1-Score is defined as the harmonic mean of precision and sensitivity. It provides a balanced assessment of the model's performance, making it particularly beneficial when dealing with skewed or unbalanced data distributions where a compromise between false alarms and missed detections must be evaluated.

$$\text{F1-Score} = 2 \times \frac{\text{Precision} \times \text{Sensitivity}}{\text{Precision} + \text{Sensitivity}} \tag{8}$$

AUC-ROC (Area Under the Receiver Operating Characteristic curve) evaluates the model's ability to distinguish between classes across all possible classification thresholds. Superior discriminative performance is indicated by a higher AUC score, where a value of 1.0 denotes a perfect classifier and 0.5 represents a performance equivalent to a random guess[24].

The Matthews Correlation Coefficient (MCC) is a robust and balanced metric that takes into account all four quadrants of the confusion matrix (true positives, true negatives, false positives, and false negatives). It is widely regarded as one of the most reliable metrics for evaluating models on unbalanced datasets. MCC values range from -1 (total disagreement) to +1 (perfect prediction), where 0 represents a random guess[25].

$$\text{MCC} = \frac{(TP \times TN) - (FP \times FN)}{\sqrt{(TP + FP)(TP + FN)(TN + FP)(TN + FN)}} \tag{9}$$

Architecture explainability and interpretability

AI models need to be comprehensible and interpretable in order to produce precise and comprehensible predictions when applied to food and NIR data. This may be accomplished using a variety of techniques, such as LIME (Local Interpretable Model-agnostic Explanations) and SHAP (Shapley Additive Explanations).

- *SHAP:* SHAP values use cooperative game theory to allocate a model's prediction to its features, providing a consistent measure of feature importance[26]. Equation 10 determines the SHAP value i for a feature i. N is the set of all features, S is a subset of N that does not contain feature i, and $f(S)$ is the model's prediction with features in subset S.

$$i = \frac{S!(N-S-1)!}{N!}\left(f(S \cup \{i\}) - f(S)\right) \tag{10}$$

- *LIME:* LIME uses a local approximation of the model to an interpretable one in order to explain individual predictions[26]. Adjusting the input data and then analyzing the changes in the predictions is the basic notion. Equation 11 shows that the explanatory model g is trained to minimize the loss function (e.g., mean squared error), f is the original model, π_x is the locality measure around instance x, and $\Omega(g)$ is the model's complexity.

$$g(x) = \arg\min_{g \in G} \mathcal{L}(f, g, x) + \Omega(g) \tag{11}$$

Experiments

Experiments Configurations: A dedicated local workstation equipped with an NVIDIA GeForce RTX 5090 graphics card was used to train the images produced from the spectral data and classification models. There is enough processing power on this card to support deep learning. This hardware setup was created to fully utilize the GPU's parallel processing power, which is essential for speeding up model training and enabling effective file system manipulation. Python 3.12 and PyTorch were used for model building, training, validation, testing, and picture production. The Python artificial intelligence package Scikit-Learn was used to construct evaluation measures. The Python packages Matplotlib and SeaBorn were used to analyze the data and produce graphics suitable for publishing. By making it easier to create visual representations including performance metric plots, confusion matrices with classification results, feature extraction findings, and activation maps, these tools enhanced the study's data's interpretability and clarity. To stop the model from preferring the majority class, we first computed class weights because our dataset was imbalanced. During training, these weights were added to the `CrossEntropyLoss` function and computed using Scikit-Learn's `compute_class_weight` function.

Autoencoder

Baseline in previous work

In the previous study, all classification models were trained directly on the raw protein spectral data, without any dimensionality reduction. This setup did not include a dedicated unsupervised feature extraction step and allowed the evaluation of each classifier's inherent capacity to learn from high-dimensional input. However, this approach was prone to overfitting and suffered from high computational cost due to the large number of input features relative to the dataset size.

Latent feature extraction with Autoencoder1D

In this work, we improved upon the previous methodology by introducing a 1D Convolutional Autoencoder (CAE) as a feature extraction and dimensionality reduction module. The Autoencoder was trained in an unsupervised manner on the consolidated spectral dataset to compress the 128-feature input into a 64-dimensional latent representation. These latent vectors were subsequently used as input for both the CNN and TabTransformer classifiers, providing a more compact and informative feature space that enhances generalization and robustness.

Network Architecture: The Autoencoder1D architecture consists of a symmetric encoder-decoder structure:

- *Encoder:* Comprises two convolutional stages. The first layer uses 16 filters with a kernel size of 5 and padding of 2, followed by ReLU activation and MaxPool1d (size 2). The second layer increases to 32 filters (kernel 5, padding 2) with a second MaxPool1d layer.
- *Bottleneck:* The output of the convolutional layers is flattened and passed through a Linear layer ($1024 \rightarrow 64$) to produce the 64-dimensional latent vector. This 64-dimensional configuration was empirically selected after evaluating tighter compression levels, including 16 and 32 dimensions. Although smaller latent spaces forced more compact representations, they resulted in a significant increase in reconstruction error (MSE) and a subsequent drop in downstream classification performance, suggesting that compressing the spectral features beyond 64 dimensions discards subtle but vital molecular variations required for accurate allergen discrimination.
- *Decoder:* The bottleneck is expanded via a Linear layer ($64 \rightarrow 1024$) and reconstructed using two ConvTranspose1d layers (16 filters and 1 filter respectively, kernel size 2, stride 2). The final layer employs a Sigmoid activation to map the reconstructed signal back to the original normalized spectral range.

Training Parameters: The model was optimized using the Mean Squared Error (MSE) loss function and the Adam optimizer with a learning rate of 10^{-3}. Training was conducted for 100 epochs with a batch size of 32, ensuring the stabilization of the reconstruction error before freezing the encoder for the classification stage.

Hyperparameter tuning of classifiers

To ensure optimal training stability and computational efficiency across the various models, hyperparameter tuning was performed using the Optuna framework. The optimization objective was evaluated using a 5-fold stratified cross-validation scheme. Crucially, to prevent data leakage and account for the high correlation among multiple spectra acquired from the same sample (e.g., across different orientations, reference plates, and measurement configurations), the cross-validation partitions were strictly constructed at the independent physical sample level (sample-level split). Specifically, all spectral data rows corresponding to a single physical specimen were constrained to the same fold, ensuring that no highly correlated repetitions from the same sample

Hyperparameter	Search space	Best value found
Hidden dimension	Integer $\in [32, 256]$	253
Dropout	Float $\in [0.0, 0.5]$	0.0076
Epochs	Integer $\in [20, 300]$	295
Learning rate (Lr)	Float $\in [10^{-5}, 10^{-3}]$ (log scale)	8.34×10^{-4}

Table 2. Hyperparameter search space and best configuration found using Optuna for the CNN classifier on latent representations.

Hyperparameter	Search space	Best value found
Dim	Integer $\in [8, 128]$	57
Depth	Integer $\in [1, 12]$	12
Heads	Integer $\in [1, 8]$	2
attn_dropout	Float $\in [0.0, 0.3]$	0.272
ff_dropout	Float $\in [0.0, 0.3]$	0.084
mlp_hidden_mult1	Integer $\in [2, 4]$	3
mlp_hidden_mult2	Integer $\in [1, 2]$	1
Epochs	Integer $\in [50, 300]$	289
Learning Rate (Lr)	Float $\in [10^{-5}, 10^{-3}]$ (log scale)	4.55×10^{-5}
Batch size	Integer $\in [16, 64]$	28

Table 3. Hyperparameter search space and best configuration found using Optuna for the TabTransformer on latent representations.

were split between the training and validation sets, thereby guaranteeing an unbiased assessment of the models' true generalization capability.

Based on the search algorithms, hyperparameters such as learning rates were dynamically optimized, and early halting was utilized to prevent overfitting. The most accurate hyperparameter configurations were discovered after 100 trials.

CNN

To optimize the performance of the classifier trained on latent representations extracted by the 1D Autoencoder, we used Optuna, a state-of-the-art hyperparameter optimization framework based on Bayesian optimization via Tree-structured Parzen Estimators (TPE). The objective function was defined to maximize the weighted F1-score averaged across a 5-fold stratified cross-validation. During training within each fold, early stopping with a patience of 10 epochs was employed to prevent overfitting. We used a class-weighted CrossEntropyLoss to account for class imbalance, and the Adam optimizer for gradient updates.

Table 2 summarizes the search space explored for each hyperparameter, as well as the optimal values identified by Optuna.

This tuning process enabled the discovery of an architecture and training configuration that maximized classifier performance on the latent features, thereby contributing significantly to the robustness and generalization capability of the final model.

TabTransformer

To optimize the performance of the TabTransformer trained on latent representations obtained from a 1D Autoencoder, we employed Optuna, a powerful hyperparameter optimization framework that utilizes Bayesian optimization with TPE. The objective function was defined to maximize the weighted F1-score over a 5-fold stratified cross-validation. The latent representations were extracted from the encoder of a pretrained 1D Autoencoder, and were used as the input to the TabTransformer. As the dataset contained no categorical features, the categories parameter was left empty and only continuous inputs were processed. A class-weighted CrossEntropyLoss was used to mitigate class imbalance, and the Adam optimizer was employed to train the model. Each fold was trained independently, and no early stopping was used to allow full exploitation of each configuration.

Table 3 summarizes the search space explored by Optuna and the best hyperparameter configuration discovered.

The Optuna study was configured to perform 40 trials, and the optimization process successfully identified a robust architecture that leverages the expressiveness of attention mechanisms in low-dimensional latent space. This contributed to improved generalization performance on the binary classification task.

Model architecture	Accuracy (%)	Sensitivity (Recall) (%)	Precision (%)	F1-score (Weighted) (%)
Decision Tree (DT)	85.60	89.63	88.99	85.57
Support Vector Machine (SVM)	83.33	81.28	92.99	83.71
Random Forest (RF)	91.12	95.56	91.56	91.00
k-Nearest Neighbors (k-NN)	91.20	95.43	91.78	91.10
TabTransformer + Protein	95.19	96.67	96.15	96.41
CNN + Protein	95.80	96.41	97.29	96.85

Table 4. Performance metrics for traditional and proposed deep learning models with protein embeddings.

Model	Accuracy (%)	Recall (%)	F1 score (%)	Specificity (%)
Decision Tree (DT)	85.60	89.63	85.57	80.12
Support Vector Machine (SVM)	83.33	81.28	83.71	86.40
Random Forest (RF)	91.12	95.56	91.00	85.20
k-Nearest Neighbors (k-NN)	91.20	95.43	91.10	85.61
CNN (baseline)	90.98	90.00	89.50	90.00
CNN + Protein Embeddings	95.80	96.41	96.85	94.55
TabTransformer (baseline)	83.02	91.74	89.55	65.13
TabTransformer + Protein Embeddings	95.19	96.67	96.42	92.15

Table 5. Comparison of results with previous studies and traditional machine learning methods.

Reported results

To evaluate and quantify each model's performance on the binary classification task, we assessed a range of well-established metrics including accuracy, specificity, precision, sensitivity (recall), F1 score, MCC, and AUC-ROC. These indicators provide comprehensive insight into the behavior and robustness of the models, especially under imbalanced class conditions. The results are shown in Table 4.

The TabTransformer with protein embeddings achieved an accuracy of 95.19%, a specificity of 92.15%, a precision of 96.15%, and a recall of 96.67%, resulting in a F1 score of 96.41%. The MCC for this model was 0.885, indicating strong overall performance, even in the presence of potential class imbalance. The AUC-ROC was approximately 0.944, reflecting a high discriminative capability across thresholds.

In comparison, the CNN with protein embeddings demonstrated slightly superior performance, with an accuracy of 95.80%, specificity of 94.55%, precision of 97.29%, and recall of 96.41%, yielding a F1 score of 96.85%. Its MCC was 0.902, and the AUC-ROC reached 0.954, suggesting a marginal but consistent improvement over the TabTransformer variant.

These results clearly indicate that the integration of protein-based embeddings significantly enhances the predictive power of both architectures. Notably, the CNN model benefitted the most from this integration, outperforming the TabTransformer across nearly all evaluation metrics. The observed improvements reinforce the relevance of biologically meaningful information in aiding deep learning models for complex classification tasks.

Performance comparison with previous work

To contextualize the effectiveness of integrating protein embeddings and benchmark against traditional approaches, we compared the current models with those from our previous publication as well as conventional machine learning classifiers. Table 5 summarizes the performance of both CNN and TabTransformer architectures, with and without protein embeddings, alongside the baseline performance of conventional models.

The results demonstrate a clear performance boost when protein embeddings are incorporated into both architectures. Notably:

- CNN + Protein Embeddings achieved the best overall performance across most metrics, especially in precision and F1-score, outperforming both traditional machine learning models and the raw baseline networks.
- Traditional classifiers like k-NN (91.20%) and Random Forest (91.12%) achieved solid baseline performance on raw spectra, proving that the datasets carry strong discriminative information, though they remain inferior to the proposed latent deep learning architectures.
- TabTransformer + Protein Embeddings showed a significant improvement over its baseline, particularly in specificity (from 65.13% to 92.15%) and accuracy (from 83.02% to 95.19%).
- While the TabTransformer baseline had strong recall, it lacked in specificity and overall balance, which was substantially corrected with the integration of protein features.

Explainability and interpretability

To better understand the internal decision-making of the CNN and TabTransformer models, we used two complementary explainability techniques: SHAP and LIME. Figs. 4 and 5 present a side-by-side comparison of

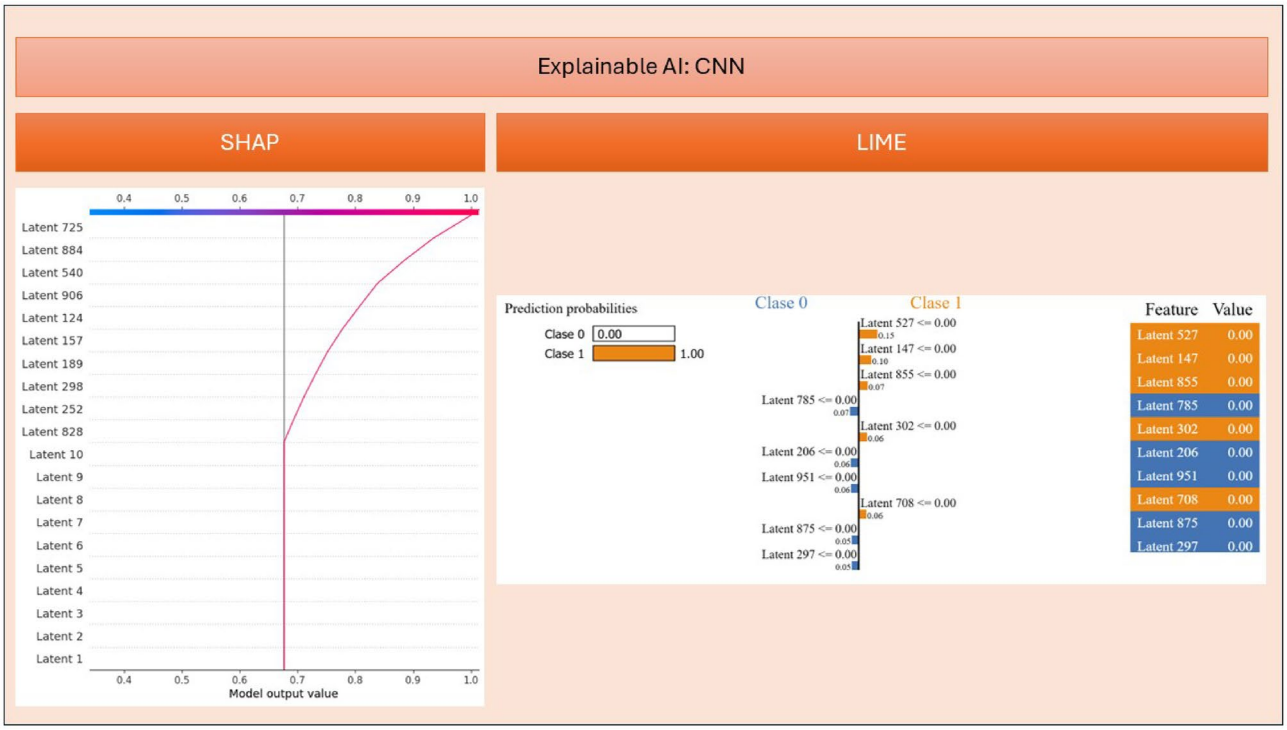

Fig. 4. Explainability and interpretability of the CNN model. The numbers in parentheses represent the SHAP value for each point. LIME shows the list of features with their corresponding values.

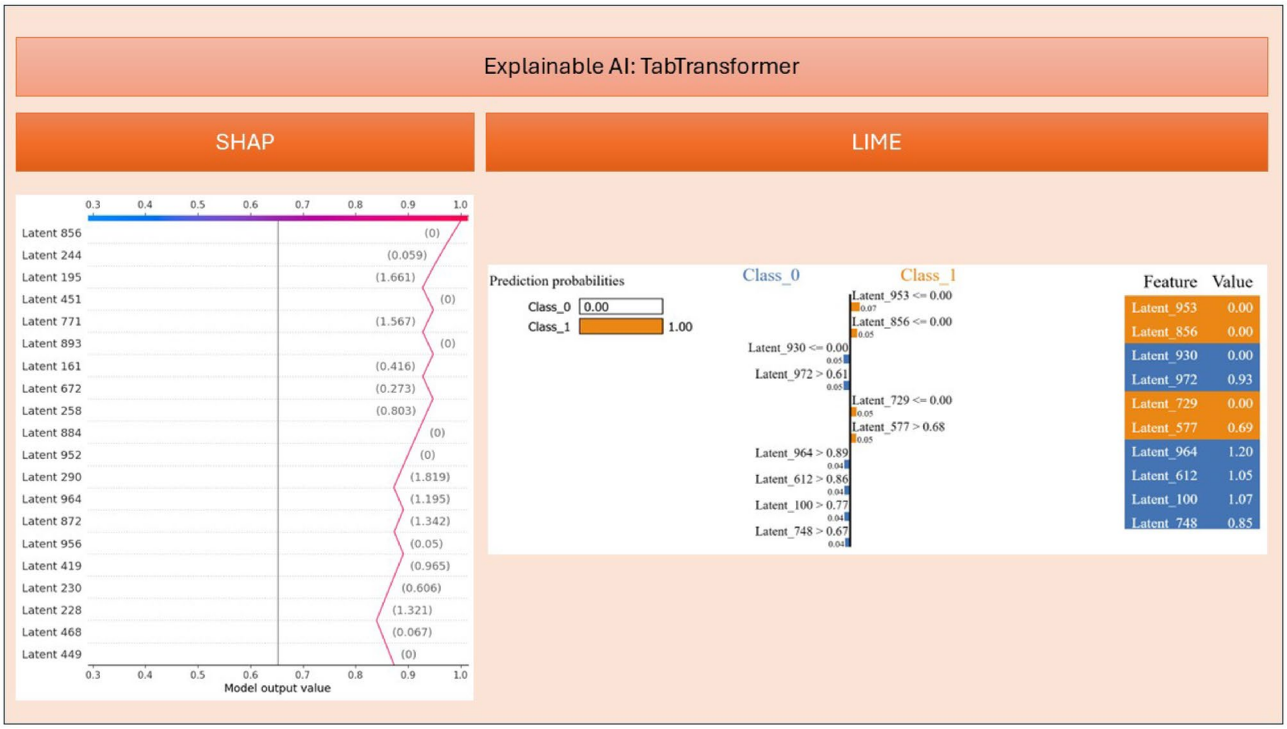

Fig. 5. Explainability and interpretability of the TabTransformer model. The numbers in parentheses represent the SHAP value for each point. LIME shows the list of features with their corresponding values.

the two methods for one highly confident prediction (probability = 1.00) from each model: the CNN (Fig. 4) and the TabTransformer (Fig. 5).

The SHAP plots (left side of each figure) reveal how each latent feature contributes additively from a baseline (typically 0.5) to the final model output of 1.0. Positive SHAP values push the prediction toward class 1 (LTP Positive), while negative values indicate resistance toward class 0. To map these latent coordinates back to physical wavelength bands, the effective receptive field of the 1D Convolutional Encoder was calculated deterministically based on its spatial dimensions and sliding window constraints.

In the CNN, the key latent features driving the class 1 prediction converged dynamically on the protein-specific spectral regions. These include Latent 725, Latent 884, and Latent 540, which map back to the 1064–1153 nm, 1088–1177 nm, and 1068–1158 nm sub-bands, respectively. This tightly clustered convergence aligns with the first overtone and combination bands of C-H and N-H molecular vibrations characteristic of the nsLTP protein backbone. Notably, no latent features exhibited meaningful negative contributions; variables located at the edges of the spectral range, such as Latent 2 and Latent 3 (940–1035 nm), displayed near-zero SHAP values, indicating negligible opposition.

In the TabTransformer, Latent 195 (1058–1147 nm) and Latent 771 (1035–1124 nm) contributed positively to the class 1 decision, anchoring the transformer's multi-head attention mechanism to the same biochemically relevant domain. Conversely, several latent features showed temporary resistance to this classification path. These include Latent 244 (1076–1165 nm) with an impact of –0.059, Latent 161 (1029–1118 nm) at –0.416, Latent 258 (1082–1171 nm) at –0.273, and Latent 228 (1070–1159 nm) at –1.321. This indicates that while the self-attention layers internally evaluated features competing with the positive assignment, these negative weights were ultimately outweighed by the overwhelming presence of the targeted allergenic molecular fingerprints.

The LIME plots show the local influence of the top latent features on the prediction. Orange bars indicate support for class 1, while blue bars signal pushback toward class 0.

For the CNN model, latent features 527 [1038–1128 nm], 147 [1062–1152 nm], and 855 [1086–1176 nm] strongly promoted class 1. Conversely, latents 206 [1032–1122 nm], 708 [972–1062 nm], 297 [1002–1092 nm], 875 [1014–1104 nm], 951 [1086–1176 nm], and 785 [1050–1140 nm] were aligned with class 0, with 206 and 708 appearing repeatedly–likely due to feature redundancy or correlated signals.

In the TabTransformer, LIME highlighted latents 953 [1098–1188 nm], 856 [1092–1182 nm], 972 [1020–1110 nm], 729 [1098–1188 nm], and 577 [954–1044 nm] as strong drivers toward class 1. Features 964 [972–1055 nm] (1.20), 612 [972–1056 nm] (1.05), 100 [972–1055 nm] (1.07), 748 [1020–1098 nm] (0.85), and 930 [960–1050 nm] (0.09) attempted to steer predictions toward class 0 but exerted insufficient influence.

Interpreting internal conflicts

Both SHAP and LIME confirm that a small subset of latent variables attempted to favor class 0, but their influence was considerably weaker than that of class 1-promoting features. This asymmetry explains why the models confidently assigned a final score of 1.00 for class 1 in the analyzed instances.

In the CNN, LIME detected latents (e.g., 206 [1032–1122 nm], 708 [972–1062 nm]) attempting to counteract the class 1 prediction, though SHAP did not register them as impactful. In the TabTransformer, both SHAP (via negative values) and LIME (via features such as 964 [972–1055 nm], 100 [972–1055 nm], 748 [1020–1098 nm]) revealed minor resistance toward class 0, which was easily outweighed by stronger class 1 activators.

Discussion

The characterization and detection of nsLTPs in food matrices are critical for ensuring food safety and minimizing allergenic risks. A deep understanding of molecular vibrations is pivotal in spectroscopic analyses, specifically in the near-infrared (NIR) region. The newly developed models leverage NIRS to capture molecular signatures associated with nsLTPs, which are small proteins abundant in specific food sources. The model's latent variables, especially those associated with indices 725, 884, 540, 906, and 124, repeatedly converged on a dense region of the NIRS spectrum between 1068 and 1164 nm. This aligns with known first overtone and combination bands of C-H and N-H molecular vibrations, which are critical for the structural attributes of proteins like Pru p 3 and Ara h 9[27]. As previously reported, these wavelengths are biochemically meaningful, as they correspond to vibrational modes of functional groups that are particularly abundant and structurally relevant in LTPs. Indeed, LTPs such as Pru p 3 and Ara h 9 are small proteins characterized by a high content of hydrophobic amino acids, eight conserved cysteine residues, and a tightly folded four alpha-helix bundle stabilized by four disulfide bonds. This unique architecture gives rise to a stable hydrophobic core rich in C-H and N-H bonds, particularly in aliphatic side chains and the protein backbone.

In this context, the C-H first overtone region (1070–1150 nm) is likely capturing variations in methyl, methylene, and methine stretching modes that arise from aliphatic residues such as leucine, isoleucine, valine, and alanine, which are overrepresented in LTPs[28,29]. Similarly, the N-H combination band region (1140–1200 nm) reflects the presence of amide linkages within the peptide backbone and potential hydrogen bonding interactions that define secondary structure elements, especially alpha-helices[30]. The combination of these signals provides a molecular fingerprint that is both rich in structural specificity and robust to environmental noise introduced by the food matrix or the presence of non-allergenic proteins such as Bovine Serum Albumin. Crucially, by targeting this 1068–1164 nm window, the deep learning framework effectively bypasses the dominant water absorption band located around 1400–1450 nm (O-H first overtone). This confirms that the model is leveraging true, protein-specific spectral features rather than merely learning moisture differences between high-water matrices (e.g., yogurt) and low-water systems (e.g., powdered milk).

The ability of the model to extract these features despite using a non-targeted, label-free approach highlights the strength of combining NIR spectroscopy with deep learning. The latent features do not rely on individual peak intensities but instead leverage combinational patterns across wavelength windows, allowing the algorithm

to detect subtle shifts caused by the presence of LTPs, even at trace concentrations. This is particularly important in the context of allergen detection, where cross-reactivity, protein isoforms, and food processing-induced matrix changes can complicate conventional antibody-based assays.

Clinical implications and risk-weighting of error metrics

From a consumer protection and clinical diagnostics perspective, the operational consequences of classification errors are asymmetric. A False Positive (FP) result leads to unnecessary food waste or over-warning, which constitutes a manageable economic and logistical inconvenience. Conversely, a False Negative (FN) result–where the system erroneously flags an LTP-contaminated sample as safe—poses an immediate, life-threatening risk of anaphylaxis to sensitized individuals. Consequently, minimizing FNs by maximizing Sensitivity (Recall) must be heavily weighted in real-world deployments. The optimized CNN model achieved a remarkable Sensitivity of 96.41% (with traditional baselines like k-NN and RF reaching up to 95.43% and 95.56%, respectively), successfully restricting dangerous FNs. To achieve a zero-tolerance threshold for industrial screening, the decision boundary of the deep learning classifier can be dynamically shifted (e.g., reducing the classification threshold from 0.50 to 0.20). This adjustment forces the model into a conservative, safety-first operational paradigm that maximizes public health protection by eradicating missed detections at the expense of a slight, clinically safe increase in false alarms.

Hyperparameter sensitivity and algorithmic transferability

A key challenge in implementing deep learning pipelines for spectroscopic food analysis is hyperparameter sensitivity. Optimal configurations discovered for a specific dataset are inherently bound to its unique resolution, instrument parameters, and matrix characteristics; thus, they seldom transfer directly to alternative spectral environments without a drop in accuracy. To address this reproducibility barrier, this framework intentionally avoided static, hardcoded architectural selections, employing automated Bayesian optimization via Optuna instead. This optimization scheme evaluated hyperparameters continuously across a strict sample-level 5-fold stratified cross-validation boundary, ensuring that training stability was driven by true generalization rather than data leakage. While researchers deploying this methodology on distinct commodities or using different NIR sensors will likely select divergent learning rates, filter counts, or regularization depths, the underlying optimization framework remains highly transferable, automating the discovery of optimal model parameters for cross-dataset adaptation.

Beyond comparisons with our previous approach, the present findings also align with and extend the existing literature on NIR-based food classification. Reference[31] proposed a spectrum correction method (CMSSA) to reduce spectral variance caused by surface irregularities in hyperspectral images. While Yuan et al. focused on improving accuracy in the classification of healthy versus moldy peanuts, our application addresses a more complex task: distinguishing LTP-containing foods, where molecular diversity and allergen-related variability introduce additional challenges that necessitate tailored preprocessing strategies.

Similarly[32], employed deep learning for NIR spectroscopy in grain analysis, using Stacked Sparse Autoencoders (SSAE) and Attention-based Extreme Learning Machines (AT-ELM). Our results parallel this trend by showing that deep learning architectures, particularly CNN and TabTransformer, excel at spectral feature extraction and classification. However, unlike Le et al.'s approach, which prioritized dimensionality reduction, our strategy emphasizes spectral preprocessing and feature integration through protein embeddings. This distinction highlights how methodological choices can be tailored to specific biochemical targets and allergen detection tasks.

Reference[33] demonstrated successful food adulteration detection using NIR spectroscopy combined with chemometric models, reporting very high classification accuracy (100% for binary classification). Although their study addressed adulteration rather than allergen detection, the comparable performance achieved by our CNN and TabTransformer models underscores the broader applicability of NIR spectroscopy in food safety and quality control. Together, these studies highlight the versatility of NIR spectroscopy and machine learning for tackling diverse challenges in food analysis, while our work extends this applicability to the critical domain of allergen detection.

Modular component contribution

To justify the layout of the proposed diagnostic architecture, the individual contribution of each modular block was analyzed. The pipeline's stability relies on three cooperative layers: first, the spectral preprocessing block (incorporating first-derivative filtering and Standard Normal Variate transformation) removes matrix-induced light scattering noise and baseline drift, yielding a steady baseline accuracy improvement. Second, the 1D-CAE dimensionality reduction block handles the feature space by compressing 128 raw wavelengths into 64 dense latent variables, eliminating uninformative high-frequency background noise and preventing network overfitting. Third, the integration of targeted purified protein embeddings anchors these compressed vectors to real biochemical profiles rather than environmental moisture fluctuations. This final block resolves the performance imbalances observed in standalone tabular architectures, providing the structural regularizer required for accurate allergen screening.

Conclusions

The integration of purified protein embeddings (Pru p 3 and Ara h 9) significantly enhanced the predictive performance of both CNN and TabTransformer models for LTP detection in food matrices. The CNN achieved the highest overall metrics, including 95.80% accuracy, 96.85% F1 score, and an AUC-ROC of 0.954, demonstrating a consistent improvement over traditional baselines and raw spectral networks. Notably, the incorporation of protein information improved the balance between specificity and recall, particularly for the TabTransformer,

which increased specificity from 65.13% to 92.15% and accuracy from 83.02% to 95.19%, thereby reducing false positives without compromising sensitivity. Traditional machine learning architectures like k-NN and Random Forest verified the data richness by yielding robust baseline accuracies of 91.20% and 91.12%, though they remain structurally surpassed by deep learning latent representations.

Interpretability analyses using SHAP and LIME revealed that key latent features consistently contributed to class 1 predictions, particularly within the 1068–1164 nm and 1140–1200 nm spectral regions. These regions correspond to C-H and N-H vibrational modes, reflecting structural elements characteristic of LTPs, such as hydrophobic residues, amide linkages, and disulfide-stabilized alpha-helices. This correspondence confirms that the models capture biochemically meaningful patterns rather than relying on spurious correlations.

Furthermore, the models exhibited robustness to matrix effects, accurately detecting LTPs in complex food matrices including yogurt and powdered milk, demonstrating their potential for practical, non-destructive, and label-free allergen screening. The explainability analyses confirmed that minor resistance from class 0 features was outweighed by strong class 1 activators, ensuring confident predictions and providing transparency in model decision-making. Overall, this study highlights the complementary value of protein-based embeddings combined with NIR spectroscopy and deep learning, offering a reliable, interpretable, and high-performing framework for rapid detection of lipid transfer proteins in diverse food products, matching the validation trends established in current chemometric literature[3].

While the proposed NIR spectroscopy and deep learning framework demonstrated high accuracy and interpretability for detecting LTPs, several limitations warrant consideration. The study primarily evaluated controlled food matrices and purified proteins, leaving the generalizability to highly processed foods, multi-ingredient products, or trace-level allergens untested. Additionally, only two major LTP allergens, Pru p 3 and Ara h 9, were included, limiting the immediate applicability to other allergenic LTPs or structurally similar proteins. Environmental factors such as temperature, moisture, or sample heterogeneity could also influence spectral measurements, potentially affecting prediction performance in field conditions. Furthermore, while SHAP and LIME offered valuable insights into feature importance, these methods approximate latent feature contributions and may not fully capture complex interactions within the models.

Building on the promising results of this study, future work could explore the quantitative potential of NIR spectroscopy for LTP detection. Specifically, since the NIR signal intensity is expected to be proportional to protein concentration, it may be possible to move beyond binary classification and develop regression-based models that estimate the actual amount of LTP present in a sample.

Data availability

The data that support the findings of this study are available on request from the corresponding author.

Received: 12 February 2026; Accepted: 3 June 2026
Published online: 10 June 2026

References

1. Wolters, P., Ostermann, T. & Hofmann, S. Lipid transfer protein allergy: characterization and comparison to birch-related food allergy. *JDDG J. Der Deutschen Dermatologischen Gesellschaft* **20**, 1430–1438, https://doi.org/10.1111/ddg.14881 (2022).
2. Lopes, J. et al. Allergy to lipid transfer proteins (ltp) in a pediatric population. *Eur. Ann. Allergy Clin. Immunol.* **55**, 86. https://doi.org/10.23822/eurannaci.1764-1489.229 (2023).
3. Lanjewar, M. G., Parab, J. S. & Kamat, R. K. Machine learning based technique to predict the water adulterant in milk using portable near infrared spectroscopy. *J. Food Compos. Anal.* **131**, 106270. https://doi.org/10.1016/j.jfca.2024.106270 (2024).
4. Zuo, J. et al. Nondestructive detection of nutritional parameters of pork based on nir hyperspectral imaging technique. *Meat Sci.* **202**, 109204. https://doi.org/10.1016/j.meatsci.2023.109204 (2023).
5. Zuo, J. et al. Integrating transfer learning and spectroscopy for enhanced pork spoilage assessment using correlation analysis. *Food Chem.* **465**, 142117. https://doi.org/10.1016/j.foodchem.2024.142117 (2025).
6. Rosa, D. G., Malik, V. V., Patle, L. B., Parab, J. S. & Lanjewar, M. G. Detection and quantification of formaldehyde adulteration in cow and buffalo milk using uv-vis-nir spectroscopy with machine learning. *Food Chem.* **492**, 145485. https://doi.org/10.1016/j.foodchem.2025.145485 (2025).
7. Osa-Sanchez, A., Del Barrio, I., Bernardo-Seisdedos, G., Mendez-Zorrilla, A. & Garcia-Zapirain, B. Detection of ltp in food using near-infrared spectroscopy and explainable artificial intelligence. *Food Anal. Methods* 1–22 (2025).
8. Huang, Y. et al. Rapid and non-destructive geographical origin identification of chuanxiong slices using near-infrared spectroscopy and convolutional neural networks. *Agriculture* **14**, 1281. https://doi.org/10.3390/agriculture14081281 (2024).
9. Lun, Z., Wu, X., Dong, J. & Wu, B. Deep learning-enhanced spectroscopic technologies for food quality assessment: convergence and emerging frontiers. *Foods* **14**, 2350. https://doi.org/10.3390/foods14132350 (2025).
10. Ocean Insight. Nirquest512-2.5 spectrometer | ocean insight. (Accessed 03 July 2023). https://www.oceaninsight.com/products/spectrometers/near-infrared/nirquest2.5/ (2023).
11. Wang, J., Ding, J., Abulimiti, A. & Cai, L. Quantitative estimation of soil salinity by means of different modeling methods and visible-near infrared (vis-nir) spectroscopy, ebinur lake wetland, Northwest China. *PeerJ* **6**, e4703. https://doi.org/10.7717/peerj.4703 (2018).
12. Xiao, Q. Spectral preprocessing combined with deep transfer learning to evaluate chlorophyll content in cotton leaves. *Plant Phenomics* **2022** 9813841. https://doi.org/10.34133/2022/9813841 (2022).
13. Xu, Y., Yang, W., Wu, X., Wang, Y. & Zhang, J. Resnet model automatically extracts and identifies ft-nir features for geographical traceability of polygonatum kingianum. *Foods* **11**, 3568. https://doi.org/10.3390/foods11223568 (2022).
14. Helin, R., Indahl, U. G., Tomić, O. & Liland, K. H. On the possible benefits of deep learning for spectral preprocessing. *J. Chemom.* **36**, https://doi.org/10.1002/cem.3374 (2021).
15. DePaoli, D. T., Tossou, P., Parent, M., Sauvageau, D. & Côté, D. C. Convolutional neural networks for spectroscopic analysis in retinal oximetry. *Sci. Rep.* **9**, 11387. https://doi.org/10.1038/s41598-019-47621-7 (2019).
16. Tung, P., Sheikh, H. A., Ball, M., Nabiei, F. & Harrison, R. J. Sigma: Spectral interpretation using gaussian mixtures and autoencoder. *Geochem. Geophys. Geosystems* **24**, https://doi.org/10.1029/2022gc010530 (2023).
17. Chen, X., Ma, L. & Yang, X. Stacked denoise autoencoder based feature extraction and classification for hyperspectral images. *J. Sens.* **2016**, 1–10. https://doi.org/10.1155/2016/3632943 (2016).

18. Torti, E., Fontanella, A., Plaza, A., Plaza, J. & Leporati, F. Hyperspectral image classification using parallel autoencoding diabolo networks on multi-core and many-core architectures. *Electronics* **7**, 411. https://doi.org/10.3390/electronics7120411 (2018).
19. Chen, Y. et al. Variety identification of orchids using fourier transform infrared spectroscopy combined with stacked sparse auto-encoder. *Molecules* **24**, 2506. https://doi.org/10.3390/molecules24132506 (2019).
20. Shevchenko, D., Ugryumov, M. & Artiukh, S. V. Monitoring data aggregation of dynamic systems using information technologies. *Innov. Technol. Sci. Solut. Ind.* 123–131. https://doi.org/10.30837/itssi.2023.23.123 (2023).
21. Huang, X., Khetan, A., Cvitkovic, M. & Karnin, Z. Tabtransformer: Tabular data modeling using contextual embeddings. (Accessed 13 March 2025). https://arxiv.org/abs/2012.06678 (2020).
22. Mahboubi, N., Xie, J. & Huang, B. Point-by-point transfer learning for bayesian optimization: An accelerated search strategy. *Comput. Chem. Eng.* **194**, 108952. https://doi.org/10.1016/j.compchemeng.2024.108952 (2025).
23. Osa-Sanchez, A. et al. Explainable ai-based approach for age-related macular degeneration (amd) detection via fundus imaging. *IEEE Access* **13**, 341–360. https://doi.org/10.1109/ACCESS.2024.3522862 (2025).
24. Receiver operating characteristic curve analysis in diagnostic accuracy studies: A guide to interpreting the area under the curve value. (Accessed 07 July 2025). https://journals.lww.com/tjem/_layouts/15/oaks.journals/downloadpdf.aspx?an=01949575-202323040-00001.
25. Chicco, D. & Jurman, G. The advantages of the matthews correlation coefficient (mcc) over f1 score and accuracy in binary classification evaluation. *BMC Genomics* **21**, 6. https://doi.org/10.1186/s12864-019-6413-7 (2020).
26. Osa-Sanchez, A. et al. A cascading approach with vision transformers for age-related macular degeneration diagnosis and explainability. In Antonacopoulos, A. et al. (eds.) *Pattern Recogn.*, 250–265. https://doi.org/10.1007/978-3-031-78398-2_17 (Springer Nature Switzerland, 2025).
27. Firth, R. A., Dimino, T. L. & Gichuhi, W. K. Negative ion photoelectron spectra of deprotonated benzonitrile isomers via computation of franck–condon factors. *J. Phys. Chem. A* **126**, 4781–4790. https://doi.org/10.1021/acs.jpca.2c03220 (2022).
28. Matsumura, F. et al. Does the sum-frequency generation signal of aromatic c–h vibrations reflect molecular orientation?. *J. Phys. Chem. B* **127**, 5288–5294. https://doi.org/10.1021/acs.jpcb.3c01225 (2023).
29. Brathwaite, A. D. et al. Coordination and spin states in fe+(c2h2)n complexes studied with selected-ion infrared spectroscopy. *J. Phys. Chem. A* **126**, 9680–9690. https://doi.org/10.1021/acs.jpca.2c07556 (2022).
30. Zhang, Y., Xu, F. & Li, Z. The influence of piezoelectric on the nonlinear stochastic vibration of bn nanoresonator. *Phys. Scripta* **99**, 095904. https://doi.org/10.1088/1402-4896/ad65c6 (2024).
31. Yuan, D., Jiang, J., Qiao, X., Qi, X. & Wang, W. An application to analyzing and correcting for the effects of irregular topographies on NIR hyperspectral images to improve identification of moldy peanuts. *J. Food Eng.* **280**, 109915. https://doi.org/10.1016/j.jfoodeng.2020.109915 (2020).
32. Le, B. T. Application of deep learning and near infrared spectroscopy in cereal analysis. *Vib. Spectrosc.* **106**, 103009. https://doi.org/10.1016/j.vibspec.2019.103009 (2020).
33. Leng, T. et al. Quantitative detection of binary and ternary adulteration of minced beef meat with pork and duck meat by NIR combined with chemometrics. *Food Control* **113**, 107203. https://doi.org/10.1016/j.foodcont.2020.107203 (2020).

Author contributions

All authors contributed to the conception and design of the study. Material preparation, data collection, and analysis were carried out by A.O.-S., I.D.B., G.B.-S., and S.P. The first draft of the manuscript was written by A.O.-S., I.D.B., and G.B.-S., and all authors provided feedback on previous versions. Supervision was provided by B.G.-Z. All authors read and approved the final version of the manuscript.

Funding

This study has been partially funded by the Department of Industry, Energy Transition and Sustainability of the Basque Country, under the grant ZL-2025/00598.

Declarations

Competing interests

The authors declare no competing interests.

Additional information

Correspondence and requests for materials should be addressed to A.O.-S.