



Research article

Rapid detection and spatial visualization of wheat flour contamination in rice flour using near-infrared hyperspectral imaging coupled with PLS-DA

Sunee Jungtheerapanich^{1,*}, Kraireuk Ngowsuwan¹, Anupun Terdwongworakul², Jaruporn Rakmai¹, Pornsuda Chawengrum¹ and Jittraporn Chusrisom¹

¹ Kasetsart Agricultural and Agro-Industrial Product Improvement Institute (KAPI), Kasetsart University, Bangkok, 10900, Thailand

² Department of Agricultural Engineering, Faculty of Engineering at Kamphaeng Saen, Kasetsart University, Nakhon Pathom, 73140, Thailand

* **Correspondence:** Email: aapsnj@ku.ac.th; Tel: +66029428600.

Abstract: This study aimed to develop an effective method for detecting wheat flour contamination in rice flour using near-infrared hyperspectral imaging (NIR-HSI), with the goal of supporting raw material quality control and food safety. Rice flour samples contaminated with varying levels (0%–100%) of low-gluten and high-gluten wheat flour were analyzed using NIR-HSI, and classification models were developed using partial least squares–discriminant analysis (PLS-DA). The developed models successfully discriminated between pure rice flour and wheat flour–contaminated rice flour. The selected models based on the original spectra achieved classification accuracies of 98.18% and 100.00% for low-gluten and high-gluten wheat flour, respectively, in the independent prediction set. For the low-gluten wheat flour model, sensitivity, specificity, and F1-score values were 97.78%, 100.00%, and 98.88%, respectively, while all classification metrics for the high-gluten wheat flour model were 100.00%. The original spectra provided comparable or better classification performance than the SNV- and second derivative–pretreated spectra. In addition, color maps generated from the PLS-DA models enabled visualization of the spatial distribution of wheat flour contamination across the investigated concentration range. These findings demonstrate the potential of NIR-HSI combined with PLS-DA as a rapid and non-destructive approach for detecting and visualizing wheat flour contamination in rice flour.

Keywords: rice flour; wheat flour; contamination; near-infrared hyperspectral imaging; predictive model; chemometric

1. Introduction

Rice (*Oryza sativa* L.) is one of the most widely consumed cereal grains worldwide and a major staple food for more than half of the global population. Beyond its role as a staple food, rice is an important raw material for value-added food products because of its versatility in food processing. In particular, rice flour is widely used in gluten-free (GF) food formulations because rice is naturally free of gluten and has a mild flavor, good digestibility, and suitable functional properties. It is used in various products, including bread, cookies, cakes, muffins, noodles, and pasta, and is considered an important alternative to wheat flour in GF food production [1,2]. The increasing use of rice flour as a wheat-flour substitute consequently highlights the importance of maintaining its gluten-free status throughout the production and supply chain.

Although rice is naturally gluten-free, rice flour may become contaminated with gluten-containing cereals, particularly wheat, at various stages of the food supply chain. Unintentional cross-contamination may occur during planting, harvesting, processing, storage, transportation, and handling. For example, gluten-containing seeds or residual grains may remain in agricultural fields and be unintentionally harvested with rice, while shared storage facilities and transportation systems may provide opportunities for cross-contact. During processing, the use of shared equipment and facilities for gluten-free and gluten-containing products, together with inadequate cleaning and segregation practices, may further increase the risk of unintended gluten transfer [3,4]. Evidence from market-based studies has demonstrated that such contamination occurs in naturally gluten-free flour products. Raju et al. [5] reported gluten contamination above the Codex threshold of 20 mg/kg in flour products obtained from local markets and mills. More recently, a survey of labeled gluten-free food products marketed in China found gluten contamination in several food categories, with higher proportions observed in flour (23.1%) and biscuits and bread (18.8%) than in seasonings (4.5%) [3]. Similarly, gluten contamination exceeding 20 mg/kg was detected in 28.6% of rice flour samples in a market-based study in Turkey, particularly among naturally gluten-free and unpackaged products [4]. Collectively, these findings demonstrate that gluten contamination can occur in products that are naturally or commercially intended to be gluten-free and highlight the potential food safety risk associated with wheat-derived gluten contamination in rice flour, particularly when production, storage, and handling conditions are not adequately controlled.

The level of wheat or gluten contamination may vary depending on the source and processing conditions. Market surveys have reported gluten contamination at levels exceeding 20 mg/kg, with concentrations of 20–400 mg/kg reported in flour products obtained from local markets and mills [5]. In addition, experimental studies have investigated wheat flour contamination in rice-based matrices across a range of experimentally prepared concentrations, including 1%–10% and 5%–30% (w/w), demonstrating the feasibility of spectroscopic approaches for detecting wheat-derived contamination in rice and rice flour [6,7]. These findings highlight the need for rapid and reliable analytical methods for screening wheat contamination in rice flour, particularly for raw materials intended for GF food production.

The prevalence of gluten-related disorders has been increasing worldwide in recent years, affecting approximately 1% of the global population, being more common in women than in men [8–10]. Furthermore, children are diagnosed with it more frequently than adults [9]. These conditions can be

classified into two major groups. The first is gluten intolerance, which is not related to the immune system but results from the inability of the body to digest gluten proteins. This causes nutrient malabsorption and symptoms such as bloating and diarrhea without intestinal inflammation [11]. The second is gluten hypersensitivity, or celiac disease, which is a permanent immune-mediated enteropathy. In this case, the ingestion of gluten-containing foods triggers the production of antibodies that attack the intestinal tissue, causing inflammation and reduced absorptive capacity of the small intestine. Symptoms include pain, bloating, diarrhea, nausea, vomiting, pale and foul-smelling stools, abdominal cramps, fatigue, numbness, and skin rashes [8,9,12].

Gluten is a complex mixture of storage proteins, primarily gliadin and glutenin, found in wheat and related gluten-containing cereals such as barley and rye [8,13,14]. Because of its unique viscoelastic properties, gluten is essential to baked products such as bread, pasta, and pastries. A network structure that traps carbon dioxide during fermentation is created when wheat flour, water, and yeast combine to generate disulfide bonds between amino acid molecules. Wheat flour is widely used as the main raw material in bakery production because it produces leavened products with desirable soft and elastic textures [13]. However, for individuals with gluten-related disorders, the most effective treatment is strict avoidance of gluten-containing foods [9,13,14]. Consequently, the demand for gluten-free products, particularly bakery products, has been increasing, making them the largest segment within the gluten-free food market [10,15,16]. Alternative flours must be used in place of wheat flour for making gluten-free bakery products. There are many different types of gluten-free flours made from rice, cereals, and tuber crops. Among these, rice flour is particularly significant because of its quantity and economic importance in Thailand.

The Codex Alimentarius Commission (2015) [17] states that the maximum amount of gluten allergens in gluten-free foods is 20 mg/kg, which is considered a relatively low threshold. To ensure the authenticity and safety of gluten-free products, reliable analytical methods for detecting adulteration and contamination are essential. Quantifying gluten in food products typically requires specific analytical kits, such as the R5 Mendez enzyme-linked immunoassay (ELISA), which involves chemical reagents and is labor-intensive, expensive, time-consuming, and destructive [18,19]. In contrast, near-infrared hyperspectral imaging (NIR-HSI) is a highly efficient analytical technique that is rapid, accurate, non-destructive, and requires minimal sample preparation [20–22]. It is environmentally friendly and capable of analyzing heterogeneous samples, large numbers of samples, and qualitative and quantitative data. NIR-HSI has become a potent instrument that integrates spectroscopy with imaging, providing simultaneous spectral and spatial information [11,22,23]. This technique produces a hypercube, a three-dimensional dataset that combines two spatial dimensions (X and Y pixel coordinates) with one spectral dimension (wavelengths). Each pixel contains spectral information related to the chemical composition of the sample, thereby linking spectroscopic data with quality attributes and spatial information [20,23,24]. The food industry requires reliable methods to monitor food safety and quality, particularly for the detection of contamination and adulteration. The agricultural and food industries have effectively used NIR-HSI for quality assessment, adulteration detection, and contamination monitoring. Previous studies have demonstrated the successful use of NIR-HSI to identify melamine contamination in milk powder [25], detect the adulteration of wheat flour with sorghum, oat, and corn flours [26], and discriminate among oats, barley, wheat, and rye [11]. NIR-HSI has also been used to identify peanut contamination in wheat flour [18], detect peanut flour adulteration in chocolate powder [19], and reveal cumin powder adulteration with peanut, pecan, and walnut shells [27]. Furthermore, NIR-HSI has been applied to detect talcum powder in wheat flour [28] and identify adulteration in wheat grains [29].

However, to the best of our knowledge, the application of NIR-HSI for detecting wheat flour

contamination in rice flour has not been sufficiently investigated. Therefore, this study aimed to develop a rapid and non-destructive NIR-HSI-based method for detecting wheat flour contamination in rice flour. The developed approach has the potential to support efficient screening and quality control of gluten-free raw materials, thereby contributing to food safety.

2. Materials and methods

2.1. Material and sample preparation

Milled rice was ground using dry milling equipment (FOSS, Cyclotec 1093, Denmark), sieved through a 0.1 mm sieve, packed in a sealed aluminum foil bag, and stored at 25 °C. Two types of wheat flour with different gluten contents were purchased from a supermarket in Bangkok, Thailand. According to the manufacturer's certificate of analysis (COA), the wheat flour samples contained 10.8% protein and 29.5% wet gluten (low-gluten wheat flour) and 14.1% protein and 40.0% wet gluten (high-gluten wheat flour). Wheat flour was mixed with rice flour at various proportions (0%–100%). All samples were subsequently analyzed using NIR-HSI and were used to develop calibration and prediction models for the classification of wheat flour contamination in rice flour.

2.2. Spectral acquisition

The samples were analyzed using NIR-HSI (VLNIR-CL-100-N17E, SPECIM SisUHEMA, Spectral Imaging Ltd., Oulu, Finland) in the reflectance mode. All spectra were collected in the wavelength range of 900–1700 nm with a resolution of 3.2 nm. The setup included a 20-watt tungsten halogen lamp kit (DECOSTAR, Carrollton, GA, United States), a CCD camera (Xeva 992, Xenics Infrared Solution Ltd., Oulu, Finland), and processing software (LUMO Software Suite, Spectral Imaging Ltd., Oulu, Finland). To acquire the images, the samples were placed on a motorized translation stage and scanned at a speed of 9 mm/s. Spectral and spatial information were acquired and recorded using the LUMO Software Suite (Spectral Imaging Ltd., Oulu, Finland) in the form of three-dimensional data, referred to as a hypercube. Each hypercube consisted of an x-axis (320 pixels) and a y-axis corresponding to the length of the sample as it moved through the imaging field. Each pixel contained a single spectral profile, with 256 reflectance values recorded per spectrum.

2.3. NIR spectroscopic model development and validation

Spectral image data of rice flour samples mixed with wheat flour at varying levels (0%–100%) obtained from NIR-HSI were analyzed in terms of absorbance. The hyperspectral data were analyzed using Evince software (Prediktera, Umeå Sweden). After that, the masked area of the sample image to identify the regions of interest (ROIs) was removed from the background regions, and averaged spectral data from all pixels within each ROI were used to obtain a mean spectrum representing each sample. The dataset was systematically partitioned into a calibration set and an independent prediction set by ordering the samples according to their wheat-flour contamination levels from the lowest to the highest level and applying a repeating 1,1,1,2 allocation scheme, in which “1” represented the calibration set, and “2” represented the prediction set. This systematic partitioning was designed to maintain representative coverage of the wheat-flour contamination range in the calibration set while

retaining an independent set of samples for subsequent model evaluation. The prediction set was not used during model development. The resulting calibration and prediction sets consisted of 165 and 55 samples, respectively. The Unscrambler version 9.8 software (CAMO AS, Trondheim, Norway) was used to create chemometric models for sample classification (rice flour and rice flour contamination with wheat flour) using partial least squares–discriminant analysis (PLS-DA) with both the original and pretreated spectra. Cross-validation was performed during model development to evaluate model performance and determine the appropriate model complexity. The number of latent variables (LVs) was determined based on the residual validation variance evaluated as a function of the number of components during cross-validation. The cross-validation results were used to determine the number of LVs used for the final PLS-DA model.

A Y value of 0 indicates that the sample is made of 100% rice flour. Rice flour is mixed with wheat flour in varying proportions; the Y value is set to 1. A decision boundary of 0.5 was predefined as the midpoint between the two class labels (0 and 1). The prediction set was then used to test the model, and the prediction results were interpreted as follows:

1. If the predicted Y value is less than 0.5, it indicates that the test sample consists of 100% rice flour (class 0).
2. If the predicted Y value is greater than or equal to 0.5, it means that the test sample is a mixture of rice flour and wheat flour at different levels (class 1).

The classification performance of the PLS-DA models was evaluated using accuracy, sensitivity (recall), specificity, precision, F1-score, and the confusion matrix based on the independent prediction set. In the confusion matrix, true positive (TP) represented wheat flour–contaminated samples correctly classified as contaminated, true negative (TN) represented pure rice flour samples correctly classified as pure, false positive (FP) represented pure rice flour samples incorrectly classified as contaminated, and false negative (FN) represented wheat flour–contaminated samples incorrectly classified as pure rice flour. These metrics were calculated from the confusion metrics as follows [30–32]:

$$\text{Accuracy (\%)} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{FP} + \text{TN} + \text{FN}} \times 100$$

$$\text{Sensitivity (recall)(\%)} = \frac{\text{TP}}{\text{TP} + \text{FN}} \times 100$$

$$\text{Specificity (\%)} = \frac{\text{TN}}{\text{TN} + \text{FP}} \times 100$$

$$\text{Precision (\%)} = \frac{\text{TP}}{\text{TP} + \text{FP}} \times 100$$

$$\text{F1 - score (\%)} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \times 100$$

Regression-based statistics, including the coefficient of determination (R^2), standard error of calibration (SEC), standard error of prediction (SEP), and bias, were also calculated to characterize the agreement between the reference and predicted Y values. The model was developed and optimized using the calibration set. The final model was selected from the models developed using the calibration set and was subsequently evaluated using the independent prediction set. Finally, distribution maps of wheat flour contamination at different levels (0%–100%) were generated using MATLAB R2021b software (MathWorks Co., US), enabling the representation of contaminant distribution within rice flour samples.

3. Results and discussion

3.1. NIR spectral characteristics of rice flour and wheat flour-contaminated samples

The NIR spectra of rice flour samples contaminated with two types of wheat flour (low-gluten wheat flour and high-gluten wheat flour) at various contamination levels were determined using NIR-HSI in the wavelength range of 900–1700 nm with reflectance mode; the original spectral profiles of the samples are plotted in Figure 1(a). According to the manufacturer's certificate of analysis (COA), the low-gluten wheat flour contained 29.5% wet gluten, whereas the high-gluten wheat flour contained 40.0% wet gluten. As shown in Figure 1(a), the original spectra exhibited similar overall spectral patterns among the rice flour, wheat flour-contaminated rice flour, and wheat flour samples; however, differences in spectral intensity were observed across the wavelength range. In addition, Figure 1(b) demonstrated the NIR spectra of samples (100% rice flour, rice flour contaminated with 50% wheat flour, and 100% wheat flour) with standard normal variate pretreatment to reduce the effect of scattering. After SNV pretreatment, the spectral profiles became more comparable, indicating that variations associated with light scattering and baseline effects were substantially reduced, while the characteristic spectral features were retained.

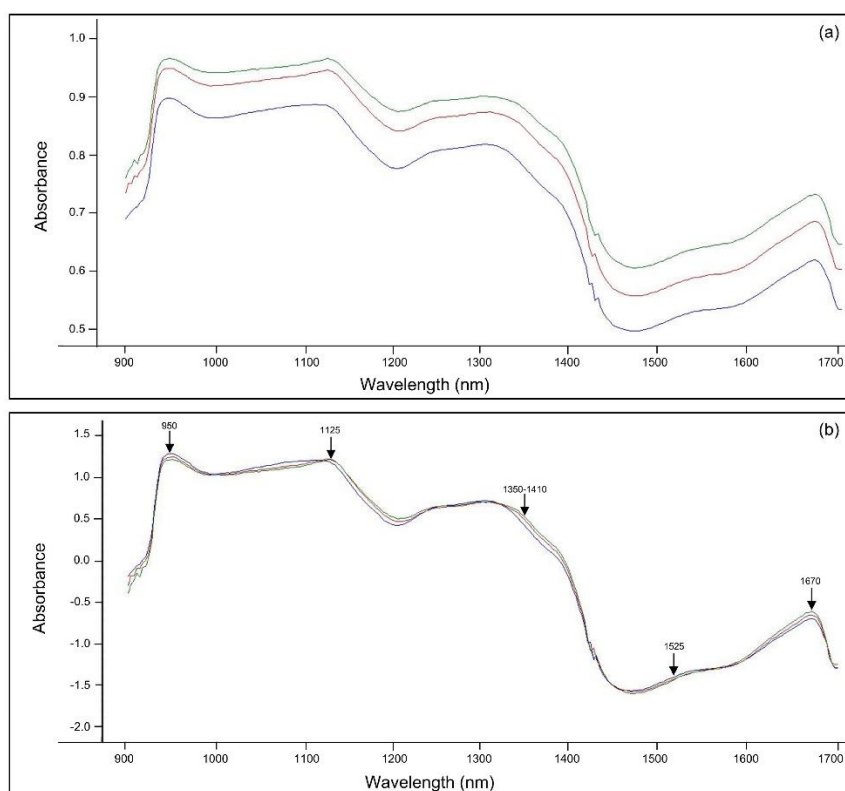


Figure 1. (a) Original NIR spectra of rice flour, rice flour mixed with wheat flour, and wheat flour samples; (b) spectra after pretreatment.

Band assignment data are displayed in Table 1. The prominent spectral features were observed at approximately 950, 1125, 1350–1410, 1525, and 1670 nm, which were related to the O–H first overtone, C–H stretch second overtone, C–H stretch third overtone, C–H combination, N–H stretch

first overtone, N–H stretch second overtone, O–H stretch first overtone, and O–H stretch second overtone, corresponding to the characteristic absorptions associated with water, starch-based carbohydrates, proteins, and lipids [33–35]. These spectral responses provide the basis for detecting wheat flour contamination because the incorporation of wheat flour changes the relative proportions of the major chemical constituents of rice flour, particularly starch and protein. In addition, differences in the protein content associated with the different wet gluten contents of the two wheat flour types may contribute to differences in protein-related spectral responses. Therefore, although NIR-HSI does not directly measure gluten, the resulting changes in the overall spectral pattern can be used as chemical fingerprints to discriminate pure rice flour from rice flour containing wheat flour. The different wet gluten contents of the two wheat flour types may also contribute to differences in the spectral responses observed among the contaminated rice flour samples.

Table 1. Band assignments of significant NIR regions of rice flour and wheat flour.

Wavelength (nm)	Band assignment	Substance
950	O–H first overtone	Water [33]
1110–1200	C–H stretch second overtone, C–H stretch third overtone	Carbohydrates, lipids [33–35]
1350–1410	O–H first overtone, C–H stretch second overtone, C–H combination	Carbohydrates, lipids [18,33]
1490–1540	N–H stretch first overtone, N–H stretch second overtone, O–H stretch first overtone, O–H stretch second overtone	Carbohydrates, proteins [18,33]
1670	O–H stretch second overtone, C–H stretch second overtone	Carbohydrates, lipids [33]

3.2. PLS-DA model development and classification performance

These spectra were employed as input data for developing classification models using PLS-DA in the Unscrambler software. The model statistics, including the coefficient of determination (R^2), standard error of calibration (SEC), standard error of prediction (SEP), and bias, obtained from the calibration and independent prediction sets, are summarized in Table 2. In addition, the classification performance of the models was evaluated using the independent prediction set based on accuracy, sensitivity, specificity, precision, F1-score, and the confusion matrix (Table 3). As shown in Table 2, the original and pretreated spectra produced generally comparable PLS-DA performance; however, the original spectra provided the most consistent overall performance for the selected models, particularly when considering both the model statistics and classification results. For low-gluten wheat flour, the model developed using the original spectra showed an R^2 of 0.794, SEC of 0.175, SEP of 0.215, and a correct detection rate of 98.18%. The second-derivative model showed a higher R^2 (0.880) and lower SEC (0.134), but its correct detection rate was slightly lower (96.36%). The SNV-treated model showed an R^2 of 0.758, SEC of 0.190, SEP of 0.228, and a correct detection rate of 98.18%. The relatively good performance of the original spectra may be attributed to the fact that the spectral information relevant to discrimination was already adequately preserved in the measured spectra. The

flour samples were finely ground and passed through a 0.1 mm sieve before spectral measurement, which may have reduced variations in particle size and, consequently, some particle size-related scattering effects among samples. Consequently, the correction of multiplicative scatter effects by SNV may have provided limited additional benefit and could have altered some intensity-related spectral information useful for discrimination [36]. Similarly, second-derivative preprocessing enhances spectral features and can improve the resolution of overlapping bands [37]; however, it may also amplify high-frequency noise and reduce the robustness of the model when the original spectra already provide sufficient spectral separation between the classes. Therefore, the application of additional spectral pretreatment did not necessarily improve the classification performance in the present dataset. This interpretation is consistent with the present results, where the original spectra achieved a higher correct detection rate than the second-derivative spectra for the low-gluten wheat flour samples (98.18% vs. 96.36%), while the three spectral treatments produced the same classification accuracy for the high-gluten wheat flour samples (100.00%). The correct detection rate represents the percentage of prediction samples that were correctly classified by the PLS-DA model. In the present study, the prediction set consisted of 55 samples; therefore, the correct detection rates of 98.18% and 100.00% obtained for the selected original-spectrum models corresponded to 54/55 and 55/55 correctly classified samples, respectively. For high-gluten wheat flour, the PLS-DA model developed using the original spectra showed an R^2 of 0.900, SEC of 0.122, SEP of 0.122, and a correct detection rate of 100.00%. As shown in Table 3, this model correctly classified all 55 independent prediction samples, yielding an accuracy, sensitivity, specificity, precision, and F1-score of 100.00%, with 45 true positives, 10 true negatives, and no false positives or false negatives.

The detailed classification performance of the PLS-DA models in the independent prediction set is presented in Table 3. For low-gluten wheat flour, the original-spectrum and SNV models achieved an accuracy of 98.18%, a sensitivity of 97.78%, a specificity of 100.00%, a precision of 100.00%, and an F1-score of 98.88%. Both models correctly classified 44 of the 45 contaminated samples and all 10 pure rice flour samples, resulting in one false negative and no false positives. The second-derivative model showed slightly lower performance, with an accuracy and sensitivity of 96.36% and an F1-score of 97.73%, resulting from two false-negative classifications. In contrast, all three spectral treatments for high-gluten wheat flour achieved perfect classification, with accuracy, sensitivity, specificity, precision, and F1-score of 100.00%. For the selected original-spectrum model for low-gluten wheat flour, the only misclassified sample was a wheat flour-contaminated sample with a predicted Y value close to the predefined decision boundary of 0.5. This result indicates that the model provided excellent discrimination between pure rice flour and wheat flour-contaminated rice flour, although a small degree of overlap between the two classes may occur for individual samples. The decision boundary of 0.5 was predefined as the midpoint between the two class labels (0 and 1). Therefore, the reported classification metrics reflect the performance of the model under the predefined classification criterion.

The consistently higher classification performance observed for high-gluten wheat flour may be related to its higher wet gluten content (40.0%) compared with the low-gluten wheat flour (29.5%). The greater contribution of protein-related spectral responses may have increased the spectral distinction between pure rice flour and wheat flour-contaminated rice flour, thereby facilitating discrimination by the PLS-DA model. However, this interpretation should be considered as a possible explanation because other compositional differences between the two wheat flour types may also contribute to the observed differences in model performance. Despite the slightly lower classification performance observed for the low-gluten wheat flour, the selected model still achieved 98.18%

accuracy, with 97.78% sensitivity and 100.00% specificity in the independent prediction set. These results demonstrate that the NIR-HSI-based PLS-DA approach was capable of discriminating wheat flour-contaminated rice flour from pure rice flour, even when the contaminating wheat flour had a relatively lower wet gluten content.

Table 2. Statistical values of the PLS-DA equation used to distinguish between 100% rice flour and rice flour mixed with varying levels of wheat flour.

Samples	Pretreatment	No. of LVs	PLS-DA calibration (n = 165)			Prediction (n = 55)		Correct detection (%)
			R ²	SEC	Bias	SEP	Bias	
Rice flour mixed with varying levels of wheat flour (low gluten)	None*	6	0.794	0.175	-2.666×10^{-7}	0.215	0.034	98.18
	2D	12	0.880	0.134	-9.753×10^{-9}	0.220	0.028	96.36
	SNV	6	0.758	0.190	2.153×10^{-7}	0.228	0.027	98.18
Rice flour mixed with varying levels of wheat flour (high gluten)	None*	5	0.900	0.122	-3.641×10^{-7}	0.122	0.012	100.00
	2D	10	0.898	0.124	4.552×10^{-8}	0.150	0.033	100.00
	SNV	7	0.898	0.124	-2.420×10^{-7}	0.150	0.009	100.00

*Note: Pretreatments: 2D, second derivatives, SNV, standard normal variate. LVs, latent variables. PLS-DA model statistics: R², coefficient of determination, SEC, standard error of calibration, SEP, standard error of prediction. *The selected model. Correct detection (%) = (number of correctly classified prediction samples / total number of prediction samples) × 100.

Table 3. Classification performance of the PLS-DA models for discrimination between 100% rice flour and rice flour mixed with varying levels of wheat flour in the independent prediction set.

Samples	Pretreatment	Accuracy	Sensitivity	Specificity	Precision	F1-score	TP	TN	FP	FN
		(%)	(%)	(%)	(%)	(%)				
Rice flour mixed with varying levels of wheat flour (low gluten)	None*	98.18	97.78	100.00	100.00	98.88	44	10	0	1
	2D	96.36	95.56	100.00	100.00	97.73	43	10	0	2
	SNV	98.18	97.78	100.00	100.00	98.88	44	10	0	1
Rice flour mixed with varying levels of wheat flour (high gluten)	None*	100.00	100.00	100.00	100.00	100.00	45	10	0	0
	2D	100.00	100.00	100.00	100.00	100.00	45	10	0	0
	SNV	100.00	100.00	100.00	100.00	100.00	45	10	0	0

*Note: Pretreatments: 2D, second derivatives, SNV, standard normal variate. *The selected model.

3.3. Classification threshold and spatial visualization of wheat flour contamination

Figure 2 shows the scatterplots between the reference Y-values and the predicted Y-values obtained from the PLS-DA models. Using the predefined decision boundary of 0.5, samples with predicted Y-values less than 0.5 were classified as pure rice flour (100%), whereas samples with predicted Y-values greater than or equal to 0.5 were classified as rice flour contaminated with wheat flour at varying levels. The predicted values were generally distributed on the expected sides of the decision threshold, resulting in high correct detection rates for both low-gluten and high-gluten wheat flour contamination. A clear separation between the two groups was observed, supporting the high discriminative ability of the developed models. These results demonstrate that the PLS-DA models were able to distinguish pure rice flour from wheat flour-contaminated rice flour based on differences in their NIR spectral patterns.

Furthermore, Figure 3 illustrates the color maps generated using MATLAB software, which visualize the spatial distribution of wheat flour contamination across different contamination levels (0%–100%). These images clearly demonstrate the capability of NIR-HSI, combined with PLS-DA, to spatially discriminate wheat flour-contaminated rice flour from pure rice flour.

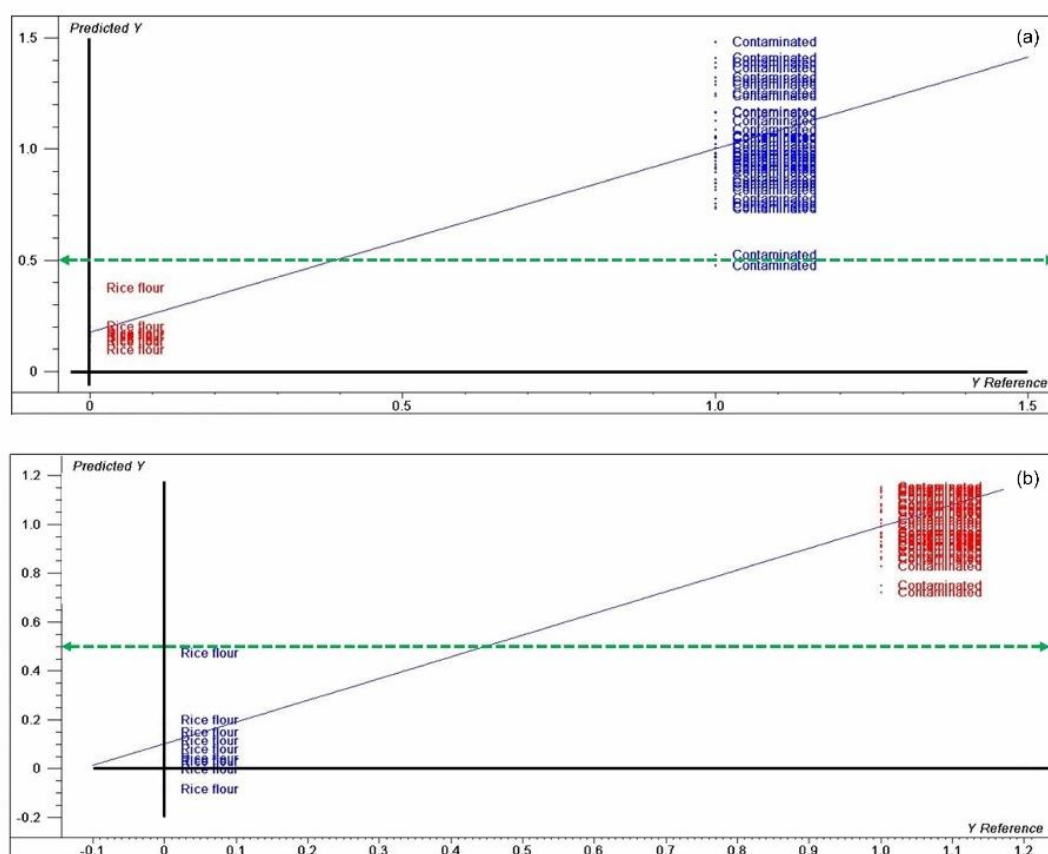


Figure 2. Scatterplot between reference Y-values and predicted Y-values obtained from PLS-DA models for the classification of pure rice flour ($Y < 0.5$) and rice flour contaminated with wheat flour at different levels ($Y \geq 0.5$): (a) Low-gluten wheat flour and (b) high-gluten wheat flour.

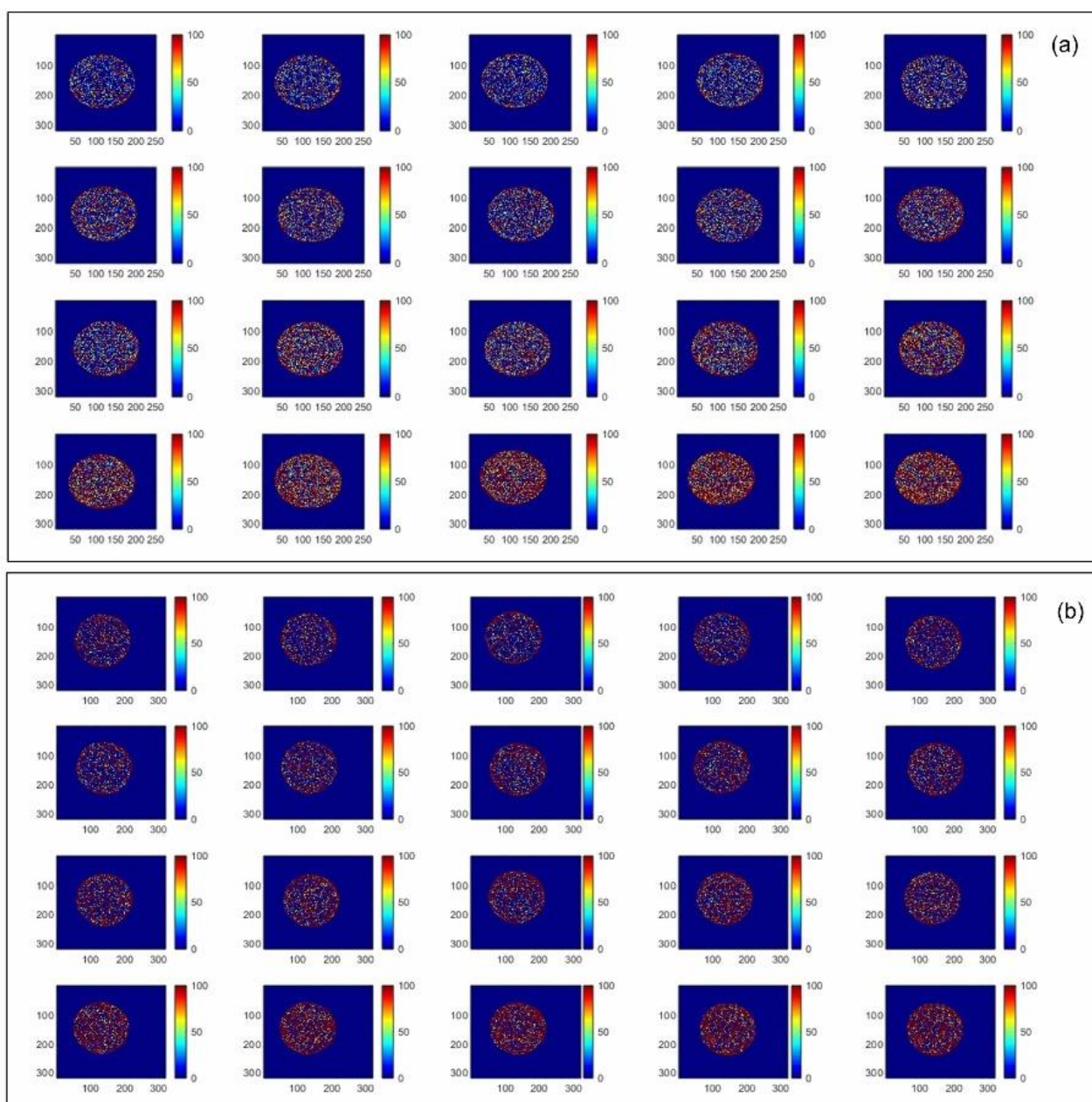


Figure 3. Color map illustrating the spatial distribution of wheat flour contamination in rice flour samples: (a) Low-gluten wheat flour and (b) high-gluten wheat flour.

3.4. Comparison with previous studies and implications

Previous studies have consistently demonstrated the potential of near-infrared hyperspectral imaging (NIR-HSI) combined with chemometric analysis for detecting adulteration in powdered food matrices. For instance, Verdú et al. (2016) [26] applied a short-wave NIR-HSI (SW-NIR-HSI, 400–1000 nm) to detect adulterations in wheat flour and bread with low-cost grains (sorghum, oats, and corn). Wheat flour was adulterated at four different levels (2.5%, 5%, 7.5%, and 10%), and flour was prepared and used to make bread. Their results indicated that the important peak was mainly associated with organic constituents such as carbohydrates; they found that the SW-NIR-HSI technique could successfully detect adulterations in samples at relatively low adulteration levels, highlighting the

feasibility of NIR-HSI techniques for flour authentication. Extending the spectral range to longer wavelengths, Zhao et al. (2019) [18] identified peanut powder contamination in two types of wheat flour (spring wheat flour and winter wheat flour) at low levels of contamination (0.01%–10% w/w) using NIR-HSI in the spectral range of 935.61–1720.23 nm with a spectral resolution of 3.45 nm. The partial least squares regression (PLS) model with standard normal variate (SNV) and Savitzky–Golay first derivative had the best performance, with coefficients of determination of prediction (R^2) of 0.993 and 0.991 and root mean square error of prediction (RMSEP) of 0.251% and 0.285% for the contaminated spring wheat flour and winter wheat flour samples, respectively. The significant wavelengths identified in their study were associated with the C–H second overtone, C–H combination, O–H first overtone, N–H stretch first overtone, and O–H stretch first overtone, which are directly linked to the protein and starch components of flour. Prediction maps based on PLS models demonstrated the advantage of NIR-HSI in visualizing the spatial distribution of contaminants. Similar observations were reported by Laborde et al. (2021) [19]; their study detected peanut flour adulteration in cocoa powder (0%, 0.1%, 1%, 10%, and 100%) using NIR-HSI with a wider spectral range (1000–2500 nm) combined with multivariate curve resolution–alternating least squares (MCR-ALS). Their results showed that very low levels of adulteration (0.1% of peanut) could be detected, with important wavelengths (1200, 1208, 1435, 1474, and 1491 nm) corresponding to the CH, CH₂, and CH₃ second overtone, N–H second overtone, and C–H stretching related to water and protein. In starch-based systems, Khamsopha et al. (2021) [35] used the NIR-HSI to detect adulterants in tapioca starch at wavelengths in the range of 935–1720 nm. Pure tapioca starch was adulterated with limestone powder across a wide concentration range. The informative wavelengths reported in their results are mainly attributed to starch-related absorption bands (1200, 1360, and 1432 nm). This study showed that the model gave excellent results, with the high correlation coefficient obtained underlining the robustness of NIR-HSI for quantitative adulterant analysis in tapioca starch with rapid detection. More recent studies have further reinforced the potential of NIR-HSI for flour authentication. He et al. (2023) [28] investigated adulteration of talcum powder in wheat flour using an NIR-HSI in the 900–1700 nm range. The results showed that the SNV-PLS model had the best performance and was suitable for integration into portable detection systems. Likewise, Saha et al. (2023) [38] demonstrated the capability of reflectance-mode NIR-HSI in the 900–2500 nm wavelength range for quantifying metanil yellow adulteration in chickpea flour. Pure chickpea flour was adulterated with metanil yellow in a range of 0%–2% (w/w). The results highlight the potential of NIR-HSI for industrial implementation in food safety monitoring.

The results of the present study are in agreement with previous studies. In particular, the absorption bands identified at 950, 1125, 1350–1410, 1525, and 1670 nm are consistent with the wavelengths reported by Zhao et al. (2019) [18], Laborde et al. (2021) [19], and Khamsopha et al. (2021) [35], which correspond to bands of O–H, C–H, and N–H functional groups related to water, starch, proteins, and lipids. This suggests that the compositional difference between rice flour and wheat flour is governed by similar molecular features to those observed in other flour adulteration systems, regardless of the specific adulterant or matrix. Moreover, the high classification performance achieved in this study, with accuracies exceeding 95%, is comparable to the predictive accuracies reported in earlier works [18,27,28]. This agreement indicates that the spectral range of 900–1700 nm with reflectance mode is sufficient for reliable discrimination between pure and contaminated flour samples. Unlike some previous studies that relied on spectral pretreatments to optimize model performance, the present results demonstrate that the original spectra provide robust classification. In addition, the visualization results obtained in

this study support and extend previous research. Similar to the reports of Zhao et al. (2019) [18] and Laborde et al. (2021) [19], the color maps generated successfully illustrated the spatial distribution of wheat flour contamination across a wide concentration range. However, the present study expands on earlier work by demonstrating that such spatial discrimination remains effective for both low- and high-gluten wheat flour. Finally, the spectral features, model performance, and visualization outcomes observed in this study and those reported previously confirm the robustness and power of NIR-HSI for detecting a wide range of contamination in flour. These findings further reinforce the applicability of NIR-HSI as a rapid, non-destructive, and reliable tool for flour contamination detection.

4. Conclusions

The results demonstrated that NIR-HSI combined with PLS-DA was effective for discriminating pure rice flour from rice flour contaminated with wheat flour over a wide contamination range of 0%–100%, including both low-gluten and high-gluten wheat flour. The characteristic NIR spectral responses associated with water, carbohydrates, proteins, and lipids provided chemical information that enabled discrimination between pure and contaminated rice flour. Among the evaluated spectral treatments, the original spectra provided the most consistent classification performance. For low-gluten wheat flour, the selected model based on the original spectra achieved 98.18% accuracy, with 97.78% sensitivity, 100.00% specificity, and an F1-score of 98.88% in the independent prediction set. For high-gluten wheat flour, the selected model achieved 100.00% accuracy, sensitivity, specificity, precision, and F1-score. The comparable performance of the original spectra with or better than the pretreated spectra suggests that additional spectral preprocessing was not necessary for effective discrimination in the present dataset, possibly because the finely milled and sieved flour samples had relatively consistent particle-size characteristics, and the relevant spectral information was sufficiently preserved in the original spectra. In addition, the NIR-HSI-based PLS-DA models enabled spatial visualization of wheat flour contamination through color maps, demonstrating the capability of the technique to provide both classification and spatial information. Overall, these findings support the potential of NIR-HSI coupled with PLS-DA as a rapid, non-destructive, and practical approach for screening and visualizing wheat flour contamination in rice flour.

Author contributions

S. Jungtheerapanich contributed to the main conception, acquired funding, conducted formal data analysis, performed methodology and investigation, prepared the manuscript, and managed the project. K. Ngowsuwan assisted in NIR-HSI methodology. A. Terdwongworakul assisted in NIR-HSI methodology and guidance. J. Rakmai assisted in sample preparation. P. Chawengrum assisted in sample preparation. J. Chusrisom assisted in the investigation. All authors have read and agreed to the published version of the manuscript.

Use of AI tools declaration

The authors declare they have not used Artificial Intelligence (AI) tools in the creation of this article.

Acknowledgments

The authors deeply thankful the Kasetsart University Research and Development Institute (KURDI) for the financial support to project FF(KU)26.68.

Conflict of interest

The authors declare no conflict of interest.

References

1. Park J, Kim HS (2023) Rice-based gluten-free foods and technologies: A review. *Foods* 12: 4110. <https://doi.org/10.3390/foods12224110>
2. Tian Y, Sun J, Li J, et al. (2024) Effects of milling methods on rice flour properties and rice product quality: A review. *Rice Sci* 31: 33–46. <https://doi.org/10.1016/j.rsci.2023.11.002>
3. Li Y, Liu Q, Godefroy S, et al. (2025) Gluten contamination of labelled gluten-free food products marketed in China. *Foods* 14. <https://doi.org/10.3390/foods14122025>
4. Öcal HY, Özel HG (2025) Naturally gluten-free flours are commonly contaminated, while commercially produced gluten-free flours are relatively safe: A market-based study in Turkey. *Front Nutr* 12: 1707584. <https://doi.org/10.3389/fnut.2025.1707584>
5. Raju N, Joshi AKR, Vahini R, et al. (2020) Gluten contamination in labelled and naturally gluten-free grain products in southern India. *Food Addit Contam A* 37: 531–538. <https://doi.org/10.1080/19440049.2020.1711970>
6. Xu L, Yan SM, Cai CB, et al. (2013) Untargeted detection of illegal adulterations in Chinese glutinous rice flour (GRF) by NIR spectroscopy and chemometrics: Specificity of detection improved by reducing unnecessary variations. *Foods Anal Methods* 6. <https://doi.org/10.1007/s12161-013-9575-y>
7. Radman M, Jurina T, Benković M, et al. (2018) Application of NIR spectroscopy in gluten detection as a cross-contaminant in food. *Croat J Food Technol Biotechnol Nutr* 13: 120–127. <https://doi.org/10.31895/hcptbn.13.3-4.4>
8. Singla D, Malik T, Singh A, et al. (2024) Advances in understanding wheat-related disorder: A comprehensive review on gluten-free products with emphasis on wheat allergy, celiac and non-celiac gluten sensitivity. *Food Chem Adv* 4: 100627. <https://doi.org/10.1016/j.focha.2024.100627>
9. Vázquez-Polo M, Navarro V, Perez-Junkera G, et al. (2024) Assessment of a training course for cookery students regarding celiac disease and gluten-free Diet. *Heliyon* 10: e39060. <https://doi.org/10.1016/j.heliyon.2024.e39060>
10. Zerbini C, De Canio F, Martinelli E, et al. (2024) Are gluten-free products healthy for non-celiac consumers? How the perception of well-being moderates gluten-free addiction. *Food Qual Prefer* 118: 105183. <https://doi.org/10.1016/j.foodqual.2024.105183>
11. Erkinbaev C, Henderson K, Paliwal J (2017) Discrimination of gluten-free oats from contaminants using near infrared hyperspectral imaging technique. *Food Control* 80: 197–203. <https://doi.org/10.1016/j.foodcont.2017.04.036>

12. Martín-Masot R, Correa-López N, Herrador-López M, et al. (2025) Immunomodulatory effects of a short-term gluten-free diet on pediatric celiac disease: Findings from a single-cell transcriptomics study. *J Nutr Biochem* 146: 110063. <https://doi.org/10.1016/j.jnutbio.2025.110063>
13. Moreno A, Hernandez S, Sanchis A, et al. (2025) Peptidomic insights into gluten peptide degradation: Exploring bacterial strains as a palliative strategy for celiac disease. *Food Biosci* 68: 106759. <https://doi.org/10.1016/j.fbio.2025.106759>
14. Tovo PA, Armandi A, Bruno M, et al. (2025) Adults with celiac disease exhibit overexpression of endogenous retroviruses, TRIM28, and SETDB1 despite gluten-free diet. *Virus Res* 359: 199613. <https://doi.org/10.1016/j.virusres.2025.199613>
15. Biesiekierski JR (2025) Clinical utility of gluten restriction in gastrointestinal disorders: Evidence, practice, and reappraisal of non-coeliac gluten sensitivity. *Best Pract Res Cl Ga* 79: 102086. <https://doi.org/10.1016/j.bpg.2025.102086>
16. Huong NTM, Tien NNT, Phi NTL, et al. (2025) Quality and digestibility of gluten-free cookies made from rice flour substituted with highly enzyme-resistant mung bean starch. *Int J Food Sci* 1: 7689819. <https://doi.org/10.1155/ijfo/7689819>
17. Codex Alimentarius (2015) Standard for foods for special dietary use for persons intolerant to gluten (CXS 118-1979), Rome.
18. Zhao X, Wang W, Ni X, et al. (2019) Utilising near-infrared hyperspectral imaging to detect low-level peanut powder contamination of whole wheat flour. *Biosyst Eng* 184: 55–68. <https://doi.org/10.1016/j.biosystemseng.2019.06.010>
19. Laborde A, Puig-Castellví F, Jouan-Rimbaud Bouveresse D, et al. (2021) Detection of chocolate powder adulteration with peanut using near-infrared hyperspectral imaging and multivariate curve resolution. *Food Control* 119: 107454. <https://doi.org/10.1016/j.foodcont.2020.107454>
20. Zhang J, Guo Z, Ren Z, et al. (2023) Rapid determination of protein, starch and moisture content in wheat flour by near-infrared hyperspectral imaging. *J Food Compos Anal* 117: 105134. <https://doi.org/10.1016/j.jfca.2023.105134>
21. Jimoh KA, Hashim N, Shamsudin R, et al. (2025) Hyperspectral imaging for detection of macronutrients retained in glutinous rice under different drying conditions. *Cur Res Food Sci* 10: 100963. <https://doi.org/10.1016/j.crf.2024.100963>
22. Cui J, Cao S, Hao J, et al. (2025) A novel rapid technique for measuring wheat protein content using near-infrared hyperspectral imaging. *J Food Compos Anal* 148: 108415. <https://doi.org/10.1016/j.jfca.2025.108415>
23. Miao YH, Zong LY, Su WH (2025) Machine learning in VIS/NIR spectroscopy and hyperspectral imaging for rapid detection of staple foods quality and safety: A review. *J Food Compos Anal* 148: 108422. <https://doi.org/10.1016/j.jfca.2025.108422>
24. Lin N, Ren Y, Sun DW (2025) Advancements in novel polysaccharide analysis: Emerging spectral technology approaches. *Trends Food Sci Tech* 162: 105037. <https://doi.org/10.1016/j.tifs.2025.105037>
25. Huang M, Kim MS, Delwiche SR, et al. (2016) Quantitative analysis of melamine in milk powders using near-infrared hyperspectral imaging and band ratio. *J Food Eng* 181: 10–19. <https://doi.org/10.1016/j.jfoodeng.2016.02.017>
26. Verdú S, Vásquez F, Grau R, et al. (2016) Detection of adulterations with different grains in wheat products based on the hyperspectral image technique: The specific cases of flour and bread. *Food Control* 62: 373–380. <http://dx.doi.org/10.1016/j.foodcont.2015.11.002>

27. Florián-Huamán J, Cruz-Tirado JP, Barbin DF, et al. (2022) Detection of nutshells in cumin powder using NIR hyperspectral imaging and chemometrics tools. *J Food Compos Anal* 108: 104407. <https://doi.org/10.1016/j.jfca.2022.104407>
28. He HJ, Chen Y, Li G, et al. (2023) Hyperspectral imaging combined with chemometrics for rapid detection of talcum powder adulterated in wheat flour. *Food Control* 144: 109378. <https://doi.org/10.1016/j.foodcont.2022.109378>
29. Zhu J, Rao Z, Ji H (2024) Fast and simultaneous detection of wheat kernel adulteration using hyperspectral imaging technology and deep convolutional neural network. *J Food Saf* 44: e13133. <https://doi.org/10.1111/jfs.13133>
30. Kasemsumran S, Boondaeng A, Ngowsuwan K, et al. (2023) Mid-infrared and near-infrared spectroscopies to classify improper fermentation of pineapple wine. *Chem Pap* 77: 335–349. <https://doi.org/10.1007/s11696-022-02472-x>
31. Lee NY, Na IS, Lee KW, et al. (2024) Detection of physical hazards from fruit processed products using hyperspectral imaging and prediction based on PLS-DA and logistic regression machine learning models. *Appl Food Res* 4: 100506. <https://doi.org/10.1016/j.afres.2024.100506>
32. Yang Z, Lu X, Chen L (2025) Discriminating the adulteration of varieties and misrepresentation of vintages of Pu'er tea based on Fourier transform near infrared diffuse reflectance spectroscopy. *Front Chem* 13: 1546702. <https://doi.org/10.3389/fchem.2025.1546702>
33. Beć KB, Grabska J, Huck CW (2020) Near-infrared spectroscopy in bio-applications. *Molecules* 25: 2948. <https://doi.org/10.3390/molecules.25122948>
34. Wang Y, Lv Y, Wei Y, et al. (2018) Identification of maize haploid kernels based on hyperspectral imaging technology. *Comput Electron Agr* 153: 188–195. <https://doi.org/10.1016/j.compag.2018.08.012>
35. Khamsoha D, Woranitta S, Teerachaichayut S (2021) Utilizing near infrared hyperspectral imaging for quantitatively predicting adulteration in tapioca starch. *Food Control* 123: 107781. <https://doi.org/10.1016/j.foodcont.2020.107781>
36. Jahani R, Yazdanpanah H, Van Ruth SM, et al. (2020) Novel application of near-infrared spectroscopy and chemometrics approach for detection of lime juice adulteration. *Iran J Pharm Res* 19: 34–44. <https://doi.org/10.22037/ijpr.2019.112328.13686>
37. Fodor M, Matkovits A, Benes EL, et al. (2024) The role of near-infrared spectroscopy in food quality assurance: A review of the past two decades. *Foods* 13: 3501. <https://doi.org/10.3390/foods13213501>
38. Saha D, Senthilkumar T, Singh CB, et al. (2023) Quantitative detection of metanil yellow adulteration in chickpea flour using line-scan near-infrared hyperspectral imaging with partial least square regression and one-dimensional convolutional neural network. *J Food Compos Anal* 120: 105290. <https://doi.org/10.1016/j.jfca.2023.105290>



AIMS Press

© 2026 the Author(s), licensee AIMS Press. This is an open access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0>)