Application of Hyperspectral Imaging for Rapid and Nondestructive Detection of Paraffine-Contaminated Rice

Mohammad Akbar Faqeerzada[1], Tanjima Akter[2], Umuhoza Aline[1], Muhammad Fahri Reza Pahlawan[2], Byoung-Kwan Cho[1,2*]

[1]Department of Biosystems Machinery Engineering, Chungnam National University, Daejeon, Korea
[2]Department of Smart Agricultural Systems, Chungnam National University, Daejeon, Korea

Abstract. The emergence of paraffin-coated rice in China, aimed at enhancing its market appeal and achieving a translucent appearance, has given rise to a significant global food safety concern. This situation poses substantial health risks to consumers. Hyperspectral analysis, recognized as a powerful and nondestructive technique for assessing food quality and safety, offers a potential solution. This study conducted a comprehensive investigation using Visible-Near Infrared (VIS-NIR) hyperspectral imaging systems operating within the 400-1000 nm range to identify paraffin-contaminated rice. Various rice varieties from diverse regions were obtained and intentionally tainted with varying levels of paraffin. Imaged samples were further preprocessed for spectral data extraction from individual rice seeds' regions of interest (ROI). The dataset encompassed 3000 spectral records obtained from both non-contaminated and contaminated samples. The obtained spectral data were employed to develop partial least squares discriminant analysis (PLS-DA) and principal component linear discriminant analysis. The primary goal was to discriminate between contaminated and non-contaminated rice samples effectively. Notably, the results indicated that PLS-DA consistently achieved an accuracy exceeding 94% across various preprocessing techniques. Overall, this study showcased the potential of combining hyperspectral imaging with chemometrics to detect paraffin-contaminated rice seeds, providing a valuable contribution to food safety assessment in the industry.
Keywords: hyperspectral imaging, paraffine contaminated, PLS-DA, PCA-LDA, beta coefficient.

1 Introduction

Rice (*Oryza sativa L.)* has stood as a cornerstone in feeding a substantial fraction of the world's inhabitants. About half of the global populace depends on rice as a central source of sustenance. Across numerous nations, especially in Asia, rice holds a pivotal role in diets, serving as a fundamental nutritional element [1]. The adulteration of food products is a deceptive practice involving substituting high-cost ingredients with lower-grade and cheaper substitutes [2]. In recent years, rice adulteration has emerged as a significant concern, drawing attention from regulatory agencies, suppliers, and consumers alike. This adulteration trend is particularly prevalent in Chinese rice production and consumption, which was recently reported [3]. Notably, paraffin, an industrial wax, has become a central element in adulterating rice, leading to the alarming emergence of "Toxic rice." This practice involves treating rice with chemicals and applying a paraffin coating to enhance its visual appeal and marketability, resulting in a desirable translucent appearance. However, the consequences of consuming this adulterated rice are far from desirable, as it contains carcinogens [4] that pose serious health risks. Detecting adulterated rice contaminated with industrial paraffin is challenging due to the deceptive nature of its appearance. Current identification methods often rely on subjective factors, such as visual cues like color and texture, or more advanced physical and chemical tests like GC-MS [5], [6]. However, these methods offer higher precision; they are labor-intensive and require extensive separation and extraction of paraffins before qualitative and quantitative analysis can be performed.

In response to the challenges posed by traditional detection methods, emerging technologies such as hyperspectral imaging have gained attention. Hyperspectral imaging has emerged as a powerful and innovative tool for agricultural product quality inspection, which combines traditional imaging with spectral analysis, enabling the acquisition of extensive data on the surface and internal characteristics of measured samples. This technology has been widely utilized in nondestructive testing for adulterated food products, including grains, fruits, and vegetables [7]. For example, color adulterant in red chili [8], fraud detection in meat [9], and adulterated almond powder with apricots and peanuts [10]. Further, hyperspectral imaging demonstrated high potential for qualitative and quantitative analysis of agriculture crops, for instance, the determination of chemical contents of water and

*Corresponding author: Byoung-Kwan Cho (chobk@cnu.ac.kr)

tuber flour [11], prediction of anthocyanins in black rice [12], assessments of internal defects in macadamia [13] and quality analysis of stored bell peppers [14].

In this study, we capitalize on the capabilities of hyperspectral imaging technology to address the challenges posed by conventional detection methods in identifying adulterated rice. The utilization of hyperspectral imaging in detecting adulterated rice can enhance the efficiency of detection processes and ensure consumer protection from the health risks associated with consuming toxic rice. This study's Specific goal was to develop a system for nondestructive quality inspection of contaminated rice. The obtained results from this research could have broader implications for the application of hyperspectral imaging in ensuring the authenticity and quality of various food products.

2 Materials and Methods

2.1 Sample preparation

Three varieties of uncontaminated rice samples were prepared from the local market, while the paraffin liquid was purchased from authentic companies in South Korea. Contaminated rice was made deliberately by adding paraffin to the uncontaminated rice. Six contamination concentration were used 0, 1, 2, 3, 4, and 5 %w/w. A total of 50 g contaminated rice were made for each concentration. After the paraffin liquid were added to the rice, the mixture were shaken using vortex mixer for around 5 minutes each. Every samples were kept in the room temperature to equilibrate before spectra measurement.

2.2 System development

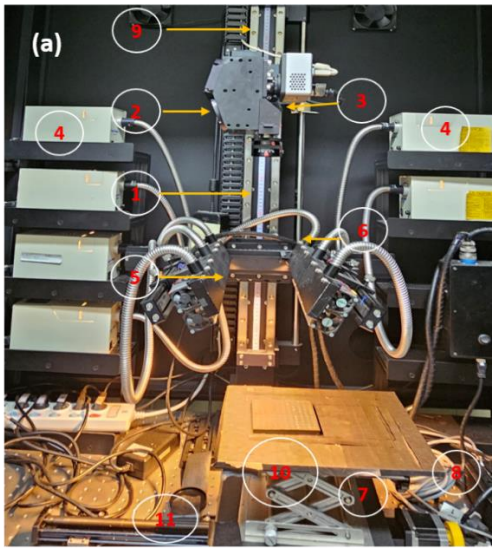

Fig. 1. Schematic representation of imaging system and components.

A line-scan hyperspectral imaging (HSI) technique was employed to assess contaminated rice non-destructively. During measurement, rice sample were placed on a 10x10 cm black plate. The system ranged the visible to near-infrared range from 400 to 1000 nm. The critical elements of the imaging setup, as depicted in Figure 1, consisted of a line-scan spectrograph provided by Headwall Photonics (Fitchburg, MA, U.S.A.), a c-mount objective lens from Schneider Optics (Hauppauge, NY, U.S.A.) with specifications of F/1.9 and 35mm, and an electron-multiplying charge-coupled device (EMCCD) camera, model MegaLuca R, manufactured by ANDOR Technology (South Windsor, CT, USA). The camera featured 1,004 spatial × 1,002 spectral pixels and operated at a maximum pixel-readout rate of 12.5 MHz. The entire system was equipped with a thermoelectric cooling mechanism, maintaining a temperature of -20 °C. The lighting setup comprised six 100 W halogen light sources (LS-F100HS) connected via optical fiber, and a stepper motor was coordinated with the camera's movements using custom software.

3 Image processing and analysis

Samples were imaged using hyperspectral imaging, and the initial hyperspectral images were converted to corresponding reflectance images to eliminate undesired noise stemming from both instruments and external sources of light; equation (1) was employed. Mainly, the correction procedure aimed to standardize the light distribution and generate images reflecting actual reflectance levels. A Teflon tile with an approximate reflectance of 99% was utilized to capture the white reference under the same lighting arrangement as the sample images. On the other hand, the dark reference was obtained by covering the lens with an opaque cap, resulting in a reflectance of approximately 0%, while the illumination unit was turned off. The refined images, denoted as Xcal, were computed by applying the formula to the measured raw hyperspectral image (Xraw), the white reference (Xwhite), and the dark reference (Xdark).

$$X_{cal} = \frac{X_{raw} - X_{dark}}{X_{ref} - X_{dark}} \qquad (1)$$

The background was removed from the rectified hyperspectral images by utilizing the mean reflectance of both the background pixels and the pixels corresponding to plant rice samples. This procedure aimed to eliminate irregular intensity and fluctuations in the background. Subsequently, the adjusted hyperspectral image was employed to extract spectral data from the ROI of the individual seed samples.

4 Data analysis

4.1 Spectral extraction and correction

After removing the background, calibrated hyperspectral images were utilized to extract spectral insights within specific areas of interest (ROIs) across the entire range of the spectra (VIS-NIR: 400-1000 nm)

for rice seed samples. The average spectrum for each seed from the ROIs of 1,500 non-contaminated and 1,500 contaminated seeds and further analysis dataset was divided into 70% for calibration and 30% for validation. The acquired raw spectral data contained random noise that resulted from variations in light direction and scattered light produced by the equipment. Consequently, preprocessing the spectral data was necessary to eliminate unwanted spectral fluctuations and unveil the valuable information within the plants. This study employed multiple preprocessing techniques, including standard normal variate (SNV), multiplicative scatter correction (MSC), smoothing, and the Savitzky–Golay first and second derivatives approach. Detailed information regarding each preprocessing method can be found in published reports [15]. Figure 2 (a) displays the initial preprocessing of raw spectral data, whereas Figure 2 (b) depicts the data after undergoing raw and SNV preprocessing and shows apparent differences between preprocessed and raw spectral data. Notably, the higher performed preprocessing was further highlighted for further analysis.

4.2 Model Development

Processed spectral data were utilized to develop distinct models of principle component analysis combined with linear discrimination analysis (PCA-LDA) and partial least square discrimination analysis (PLS-DA). The spectral data were divided into 2000 (1000 contaminated and 1000 non-contaminated) spectral data for calibration and 1000 (500 contaminated and 500 non-contaminated) for validation, all derived from raw preprocessed spectral data.

PCA-LDA analysis, a statistical technique, synergistically merges principal component analysis (PCA) and linear discriminant analysis (LDA) to analyze intricate datasets. PCA effectively diminishes the dimensionality of high-dimensional data by identifying inherent patterns and correlations. Data is repositioned Through this transformation into a new coordinate system where the first principal component captures the most variance, and the subsequent components decrease variance. This dimensionality reduction serves to simplify data analysis. Conversely, LDA pinpoints the linear amalgamation of features that distinctly classify predefined groups or classes within the dataset. It achieves this by optimizing between-class variance while minimizing within-class variance. In the PCA-LDA analysis process, PCA takes precedence by initially reducing data dimensionality by identifying critical patterns and correlations. The altered data then undergoes LDA to determine the optimal linear amalgamation of features for effectively segregating the data into predefined groups or classes. Combining these two techniques in PCA-LDA enhances the precision of data classification by reducing dimensionality and enhancing data distinctiveness.

Furthermore, we have developed a PLS-DA model, a modified version of partial least square regression (PLS-R). PLS-DA is specifically designed for situations where the goal is to classify data points into different classes or categories based on their features. In PLS-DA, the method adapts the principles of PLS to classification problems. It constructs a linear relationship between the predictor variables (independent variables) and a set of dummy variables representing the class labels (dependent variables). The goal is to find latent variables that maximize the covariance between predictor variables and class labels and optimize class separation. Further information and implementation on the NIR spectral data can be found in previously published reports [16].

4.3 Models performance evaluation

The analysis was adapted to encompass two distinct groups of rice seeds: non-contaminated and contaminated samples. The evaluation process focused on the performance assessment of the PCA-LDA and PLS-DA models. Essential metrics, including sensitivity, specificity, precision, accuracy, and error rate, were computed using equations (2 to 6) derived from the confusion matrix.

$$Sensitivity = \left(\frac{TP}{TP+FN}\right) * 100 \quad (2)$$

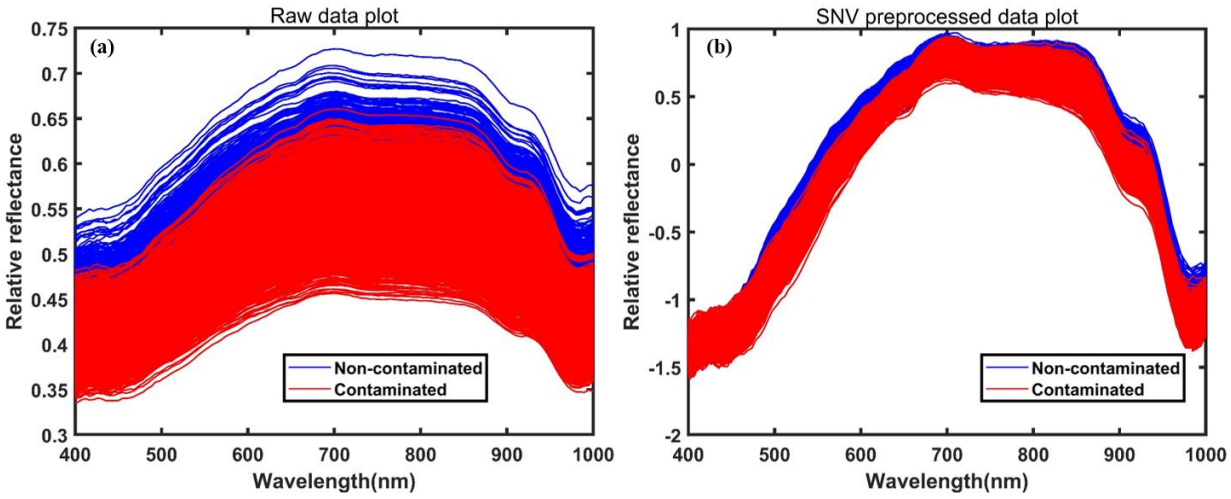

Fig. 2. Plots show (a) raw spectral data and (b) preprocessed spectral data.

Sensitivity, as expressed in Equation (2), delineated the models' efficacy in correctly identifying samples belonging to the contaminated group. This metric gauged the models' ability to discern actual positive cases while mitigating false negatives. Sensitivity served as a vital indicator of the models' capacity to accurately spot actual instances of contamination within the rice seed samples.

$$Specificity = \left(\frac{TN}{FP+TN}\right) * 100 \quad (3)$$

Equation (3) specificity measures parameter highlighted the models' adeptness in correctly recognizing true negatives while minimizing instances of false positives. Specificity provided insights into the models' capability to label negative cases accurately.

$$Precision = \left(\frac{TP}{TP+FP}\right) * 100 \quad (4)$$

Precision, outlined in Equation (4), assessed the models' precision in accurately predicting the number of accurate optimistic predictions within the contaminated group. It illuminated the models' accuracy in identifying contaminated instances among the predicted positives.

$$Accuracy = \left(\frac{FP+FN}{TP+FP+FN+TN}\right) * 100 \quad (5)$$

$$Error\ rate = \left(\frac{TP+TN}{TP+TN+FP+FN}\right) * 100 \quad (6)$$

Furthermore, the overall effectiveness of the models' classification was encapsulated by accuracy, denoted by Equation (5). Accuracy presented a comprehensive view of how well the models correctly categorized non-contaminated and contaminated rice seed samples. Conversely, the error rate, calculated using Equation (6), examined the models' classification performance by quantifying the proportion of inaccurately assigned samples concerning the total number of samples.

The terms are defined as follows: True Positive (TP) represented the instances where contaminated rice seeds were correctly identified as such. It corresponded to a positive response for an actual positive instance. False Positive (FP) indicated cases where non-contaminated rice seeds were wrongly classified as contaminated. It corresponded to a positive response for a negative instance. True Negative (TN) denoted the accurate recognition of non-contaminated rice seeds. It corresponded to a negative response for an actual negative instance. False Negative (FN) encompassed instances where contaminated rice seeds were erroneously identified as non-contaminated. It corresponded to a negative response for a favorable instance. The adapted analysis delved into the performance evaluation of the PCA-LDA and PLS-DA models, employing these metrics to comprehensively assess their accuracy in classifying non-contaminated and contaminated rice seed samples.

5 Results and Discussions

5.1 Spectral information of rice samples

Figure 3 (a) presents the raw mean spectral data and (b) demonstrates the SNV preprocessed of contaminated rice samples. The mean spectral shows an apparent difference between contaminated and non-contaminated spectra, in which the contaminated samples show lower intensity due to absorbing more light, and the non-contaminated spectra show higher intensity. Contaminated rice absorbs more light might due to presence of hydrocarbon bonds in parrafin [17]. However, both contaminated and non-contaminated spectral patterns demonstrated a similar pattern, and slight changes appear between 448-507 nm and 903-932 nm.

5.2 PCA-LDA and PLS-DA results

The PCA-LDA and PLS-DA models' effectiveness in classifying contaminated and non-contaminated rice samples into two classes was appraised using different preprocessing methods on spectral data. Nevertheless, the accuracy of the PCA-LDA and PLS-DA models varied based on the chosen preprocessing approach. The PCA-LDA and PLS-DA methods consistently demonstrated the lowest root mean square error of cross-validation (RMECV) for each preprocessing technique. To identify the optimal preprocessing methods, the ones resulting in the minimum RMSECV and the most appropriate number of latent variables (LVs) were chosen. The number of LVs was determined through a 10-fold cross-validation process. The results from the models are summarized in Table 1, depicting the calibration and prediction outcomes. The raw and range normalization (RN) preprocessing yielded enhanced accuracy across all models, including PCA-LDA and PLS-DA.

Comparing the performance of PCA-LDA and PLS-DA models, it becomes evident that they both achieved commendable levels of accuracy. However, the outcomes presented in Table 1, where these analysis methods were compared against highly effective preprocessing techniques, reveal distinct patterns. PCA-LDA displayed varying prediction set accuracies, ranging from 90% to 77%, and an elevated error rate surpassing 22%. In contrast, leveraging range normalization, PLS-DA showcased the highest accuracy, reaching 94.10% in the prediction set and a meager error rate of 5.9%. The classification outcomes of PLSDA for contaminated and non-contaminated rice samples, based on the application of range normalization, are visually depicted in Figure (4).

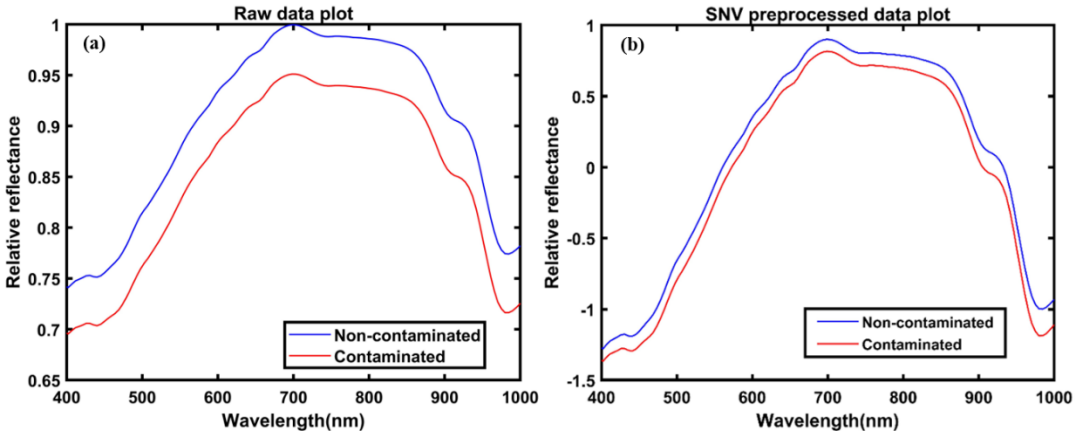

Fig. 3. Raw and preprocessed spectral data of contaminated and non-contaminated samples.

Models		Calibration set (1500) Sen	Spec	Pre	Acc	E-rate	Prediction set (1000) Sen	Spec	Pre	Acc	E-rate
PCA-LDA	Raw	92.40	92.40	92.42	92.40	7.60	90.10	90.10	90.24	90.10	9.90
	SNV	79.45	79.45	79.62	79.45	20.55	77.80	77.80	77.80	77.80	22.20
	RN	86.60	86.60	86.62	86.60	13.40	83.60	83.60	83.78	83.60	16.40
PLS-DA	Raw	94.90	94.90	94.90	94.90	5.10	93.50	93.50	93.54	93.50	6.50
	SNV	88.80	88.80	88.82	88.80	11.20	88.00	88.00	88.00	88.00	12.00
	RN	94.95	94.95	94.95	94.95	5.05	94.10	94.10	94.11	94.10	5.90

Table 1 summarizes the analysis results of the PCA-LDA and PLS-DA models.

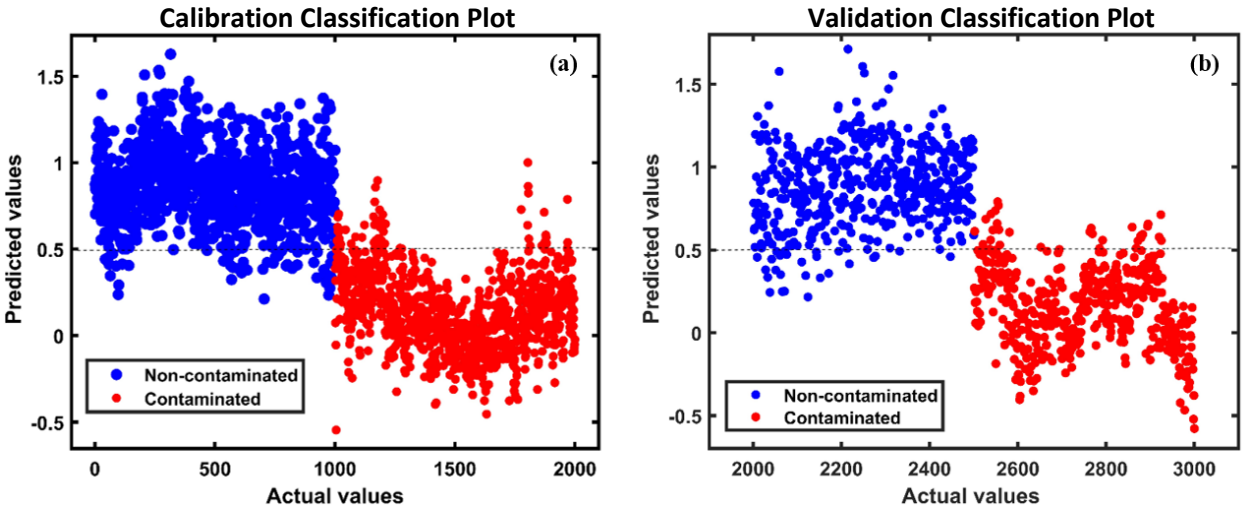

Fig. 4. PLS-DA classification analysis results (a) calibration and (b) validation, respectively.

Furthermore, to present a comprehensive overview of the outcomes for each class, confusion matrices were generated for the most effective models, as illustrated in Figure 5. The two axes of the matrices encompassed two categories: contaminated and non-contaminated samples. The labels assigned by the model were represented on the horizontal axis, while the vertical axis depicted the actual data labels. In particular, Figure 5 encompassed three subfigures: (a) showcasing the calibration results and (b) displaying the validation outcomes. These matrices pertained to the performance of the PLS-DA model, which employed range-normalization preprocessing, highlighting its remarkable achievements.

The confusion matrices illustrated a situation where a slightly higher number of contaminated samples were inaccurately classified as non-contaminated, and a small number of non-contaminated samples were misclassified as contaminated and showed the model error rate. Additionally, in multivariate analysis, the resulated beta coefficient from the models reveals the connection between each composition's absorption and specific spectral bands. Its importance stems from assisting in selecting informative wavelengths and simplifying the interpretation of outcomes. Figure (6) displays the derived PLS-DA beta coefficient using the raw spectral data, showcasing the contaminated and non-contaminated spectral insights linked to rice samples.

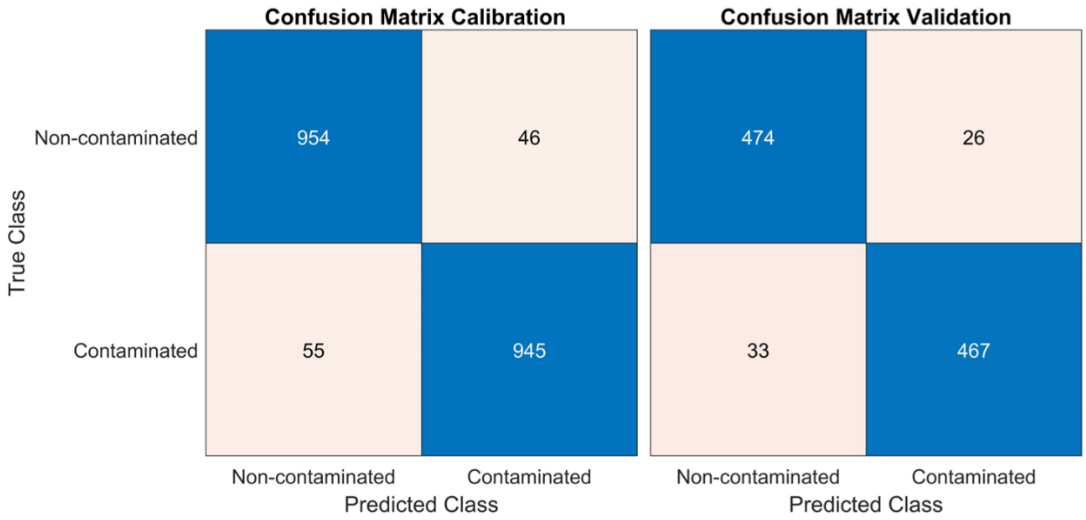

Fig. 5. Confusion matrices analysis of calibration and validation set.

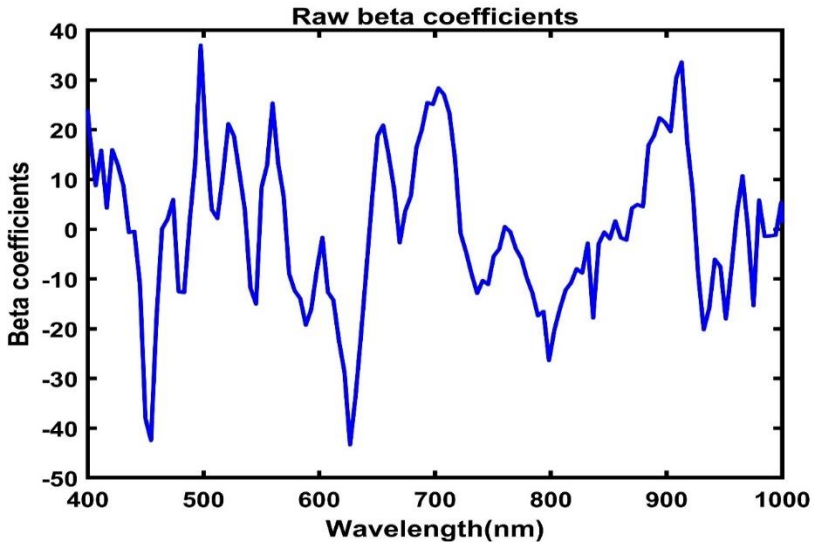

Fig. 6. Beta coefficient of the calibration model

The absorbed energy causes bonds to vibrate and stretch, leading to characteristic absorption bands in the VNIR spectrum. The energy levels associated with C-H, O-H, and N-H bonds align well with the energy range of VNIR light, making these regions particularly informative for analysis. The C-H bonds are prevalent in organic compounds, including fats, oils, carbohydrates, and proteins [18]. The O-H bonds are commonly found in water and various functional groups like alcohols and phenols [19]. Similarly, N-H bonds are present in molecules such as amino acids and proteins. As a result, VIS-NIR spectroscopy can be used to detect, quantify, and characterize a wide range of organic compounds based on their unique spectral signatures in the VNIR region. Within the spectral range encompassing VNIR wavelengths (400 to 1000 nm), we encountered spectra containing vital chemical details of rice constituents. These constituents include carbohydrates, proteins, amylose, and moisture, all of which correspond to the third overtone absorption of C–H bonds at 880 nm, as well as the second overtone absorption of O–H bonds in the range of 750 to 900 nm and N–H bonds spanning from 962 to 1000 nm [20], [21].

Furthermore, the amide bonds in proteins can cause absorption bands in the 600-900 nm range (Nirmal Thirunavookarasu, 2022). Amino acids have specific absorption features around 950 nm [22]. Starch and carbohydrates may contribute to absorption bands in the 800-1000 nm region due to their molecular vibrations [23]. Water absorption is significant in the 900-1000 nm range [24].

6 Conclusion

In conclusion, this study focused on applying hyperspectral imaging combined with chemometric analysis to detect paraffin-contaminated rice seeds, addressing a pressing food safety concern. The emergence of paraffin-coated rice in the market has

raised alarms due to potential health risks, making the need for reliable detection methods crucial. Moreover, employed Visible-Near Infrared (VIS-NIR) hyperspectral imaging and advanced analytical techniques showcased promising results. Through a meticulous investigation involving different rice varieties from diverse regions contaminated with varying levels of paraffin, the study employed Partial Least Squares Discriminant Analysis (PLS-DA) and Principal Component Analysis Linear Discriminant Analysis (PCA-LDA) models for classification. The effectiveness of these models was evaluated using different preprocessing methods on the spectral data, including raw data, standard normal variate (SNV) transformation, and range normalization (RN). The results highlighted the potential of both PCA-LDA and PLS-DA models in accurately classifying contaminated and non-contaminated rice samples. However, a comparative analysis demonstrated that PLS-DA exhibited superior performance, especially with range normalization preprocessing. This preprocessing approach consistently yielded high accuracy rates, reaching 94.10% in the prediction set, with a minimal % error rate of 5.9%. At the same time, while generally accurate, PCA-LDA showed more variability in accuracy, ranging from 90% to 77%, and a higher error rate exceeding 22% based on different preprocessing techniques. This research underscores the potential of hyperspectral imaging and chemometric approaches, particularly PLS-DA with range normalization preprocessing, as a viable solution for detecting paraffin-contaminated rice. Furthermore, the study's findings contribute significantly to the field of food safety assessment, providing a valuable tool for the industry to ensure the quality and safety of rice products in the market. With the potential for further advancement, refinements in spectral data extraction and classification methodologies could enhance detection accuracy, making such techniques even more robust and reliable for real-world applications.

References

[1] FAO, “Rice Market Monitor, Food Agric. Organ. United Nations,” *Food Agric. Organ. United Nations*, vol. XXI, no. 1, pp. 1–38, 2018.

[2] R. Johnson, “Food fraud and ‘Economically motivated adulteration’ of food and food ingredients,” *Food Fraud Adulterated Ingredients Background, Issues, Fed. Action*, pp. 1–56, 2014.

[3] X. Feng, Q. Zhang, P. Cong, and Z. Zhu, “Preliminary study on classification of rice and detection of paraffin in the adulterated samples by Raman spectroscopy combined with multivariate analysis,” *Talanta*, vol. 115, pp. 548–555, 2013, doi: 10.1016/j.talanta.2013.05.072.

[4] K. Renu *et al.*, “Molecular mechanism of heavy metals (Lead, Chromium, Arsenic, Mercury, Nickel and Cadmium) - induced hepatotoxicity – A review,” *Chemosphere*, vol. 271, p. 129735, 2021, doi: 10.1016/j.chemosphere.2021.129735.

[5] J. F. X. Fen, L. Shiyao, C. Hui, and A. Di, “Determination of illegally added paraffin in rice by GC-MS/MS,” *Hubei Prov. Eng. Technol. Res. Cent. Food Qual. Saf. Test, Hubei Prov. Inst. Food Superv. Test, Wuhan 430075, China)*, 2019, doi: 10.11882/j.issn.0254-5071.2019.03.038.

[6] D. Huang *et al.*, “Concentrations of and risks posed by short-chain and medium-chain chlorinated paraffins in soil at a chemical industrial park on the southeast coast of China,” *Environ. Pollut.*, vol. 258, p. 113704, 2020, doi: 10.1016/j.envpol.2019.113704.

[7] U. Aline, T. Bhattacharya, M. A. Faqeerzada, M. S. Kim, I. Baek, and B. Cho, “Advancement of non-destructive spectral measurements for the quality of major tropical fruits and vegetables : a review,” *Front. Plant Sci.*, no. August, pp. 1–18, 2023, doi: 10.3389/fpls.2023.1240361.

[8] M. Hussain Khan, Z. Saleem, M. Ahmad, A. Sohaib, H. Ayaz, and M. Mazzara, “Hyperspectral imaging for color adulteration detection in red chili,” *Appl. Sci.*, vol. 10, no. 17, 2020, doi: 10.3390/app10175955.

[9] M. Kamruzzaman, “Fraud Detection in Meat Using Hyperspectral Imaging,” *Meat Muscle Biol.*, vol. 5, no. 3, 2021, doi: 10.22175/mmb.12946.

[10] M. A. Faqeerzada *et al.*, “Hyperspectral shortwave infrared image analysis for detection of adulterants in almond powder with one-class classification method,” *Sensors (Switzerland)*, vol. 20, no. 20, p. 5855, 2020, doi: 10.3390/s20205855.

[11] R. E. Masithoh, L. M. Kandpal, S. Lohumi, W. S. Yoon, H. Z. Amanah, and B. K. Cho, “Shortwave Infrared Hyperspectral Imaging for the Determination and Visualization of Chemical Contents of Wheat and Tuber Flour,” *Int. J. Adv. Sci. Eng. Inf. Technol.*, vol. 12, no. 4, pp. 1574–1579, 2022, doi: 10.18517/ijaseit.12.4.14266.

[12] H. Z. Amanah *et al.*, “Near-infrared hyperspectral imaging (Nir-hsi) for nondestructive prediction of anthocyanins content in black rice seeds,” *Appl. Sci.*, vol. 11, no. 11, 2021, doi: 10.3390/app11114841.

[13] A. Rahman, Q. Wu, H. Chang, S. Wang, J. Yan, and H. Xu, “Assessment of Intact Macadamia Nut Internal Defects Using Near-Infrared Spectroscopy,” *ASABE Annu. Int. Meet.*, pp. 3–8, 2020.

[14] A. Rahman *et al.*, “Quality analysis of stored bell peppers using near-infrared hyperspectral imaging,” *Trans. ASABE*, vol. 61, no. 4, pp. 1199–1207, 2018, doi: 10.13031/trans.12482.

[15] Å. Rinnan, F. van den Berg, and S. B. Engelsen, "Review of the most common pre-processing techniques for near-infrared spectra," *TrAC - Trends in Analytical Chemistry*. 2009, doi: 10.1016/j.trac.2009.07.007.

[16] J. Yasmin, M. R. Ahmed, S. Lohumi, C. Wakholi, M. S. Kim, and B. K. Cho, "Classification method for viability screening of naturally aged watermelon seeds using FT-NIR spectroscopy," *Sensors (Switzerland)*, vol. 19, no. 5, 2019, doi: 10.3390/s19051190.

[17] H. Chung, "Applications of near-infrared spectroscopy in refineries and important issues to address," *Applied Spectroscopy Reviews*, vol. 42, no. 3. pp. 251–285, 2007, doi: 10.1080/05704920701293778.

[18] S. Serranti, D. Cesare, and G. Bonifazi, "The development of a hyperspectral imaging method for the detection of Fusarium-damaged, yellow berry and vitreous Italian durum wheat kernels," *Biosyst. Eng.*, vol. 115, no. 1, pp. 20–30, 2013, doi: 10.1016/j.biosystemseng.2013.01.011.

[19] Z. Wu, G. Ouyang, X. Shi, Q. Ma, G. Wan, and Y. Qiao, "Absorption and quantitative characteristics of C-H bond and O-H bond of NIR," *Opt. Spectrosc. (English Transl. Opt. i Spektrosk.*, vol. 117, no. 5, pp. 703–709, 2014, doi: 10.1134/S0030400X1411023X.

[20] S. Weng *et al.*, "Hyperspectral imaging for accurate determination of rice variety using a deep learning network with multi-feature fusion," *Spectrochim. Acta - Part A Mol. Biomol. Spectrosc.*, vol. 234, p. 118237, 2020, doi: 10.1016/j.saa.2020.118237.

[21] L. Ma *et al.*, "Systematic discovery about NIR spectral assignment from chemical structural property to natural chemical compounds," *Sci. Rep.*, vol. 9, no. 1, pp. 1–17, 2019, doi: 10.1038/s41598-019-45945-y.

[22] J. Hang, D. Shi, J. Neufeld, K. E. Bett, and J. D. House, "Prediction of protein and amino acid contents in whole and ground lentils using near-infrared reflectance spectroscopy," *Lwt*, vol. 165, no. June, p. 113669, 2022, doi: 10.1016/j.lwt.2022.113669.

[23] R. Pourdarbani, S. Sabzi, S. Jarolmasjed, and T. Panagopoulos, "Determination of the most effective wavelengths for prediction of fuji apple starch and total soluble solids properties," *Appl. Sci.*, vol. 10, no. 22, pp. 1–16, 2020, doi: 10.3390/app10228145.

[24] J. H. Lam, K. J. Tu, and S. Kim, "Narrowband diffuse reflectance spectroscopy in the 900–1000 nm wavelength region to quantify water and lipid content of turbid media," *Biomed. Opt. Express*, vol. 12, no. 6, p. 3091, 2021, doi: 10.1364/boe.425451.