Non-destructive Detection and Spatial Mapping of Wheat Bug Infestation in Flour Using Vis-NIR Hyperspectral Imaging and Chemometric Modeling

Mahmoud Soorghali *, Sadra Karimzadeh, Bakhtiar Feizizadeh

Dept. of Remote Sensing and GIS, University of Tabriz, Iran – m.soorghali@gmail.com, sadra.karimzadeh@gmail.com, bakhtiar.feizizadeh@gmail.com

Keywords: Hyperspectral Imaging, HYSPIM, Flour Purity, Chemometric Techniques

Abstract

Wheat flour quality is critically affected by contamination from wheat bug (Eurygaster spp.), which degrades its nutritional and functional properties. Traditional detection techniques are often destructive, slow, and unsuitable for real-time quality control. This study presents a rapid and non-destructive method for detecting and spatially mapping wheat bug infestation in flour using visible–near infrared (Vis-NIR) hyperspectral imaging (HSI) integrated with advanced chemometric modeling. Hyperspectral images of three flour samples (pure, moderately infested, and heavily infested) were acquired using the HYSPIM system, covering the 400–950 nm range at 3 nm resolution. Spectral data were preprocessed using Standard Normal Variate (SNV) and Savitzky–Golay smoothing to enhance signal quality. Representative pure spectra were extracted via the SIMPLISMA algorithm and processed using mean-field independent component analysis (MF-ICA) to isolate independent spectral features. These components were used to train a Partial Least Squares Discriminant Analysis (PLS-DA) model, which was then applied to a moderately infested sample. The resulting pixel-wise classification map showed a near-equal distribution of pure (49.91%) and infested (50.09%) pixels, with spatially coherent patterns that align with expected contamination distribution. The findings underscore the effectiveness of the proposed HSI–MF-ICA–PLS-DA pipeline for semi-quantitative, spatially resolved contamination detection in flour, offering a practical tool for real-time food quality monitoring in industrial settings.

1. INTRODUCTION

Wheat flour is one of the most essential staple ingredients used in daily diets worldwide. However, its quality can be significantly compromised by contamination resulting from wheat bug (Eurygaster integriceps) infestation, which alters the flour's chemical composition and leads to undesirable changes in baking properties and nutritional value. Conventional detection methods, such as chromatography or microscopy, are often time-consuming, destructive, and require laborious sample preparation, limiting their practicality for real-time screening and industrial applications (Hashemi-Nasab & Parastar, 2022; Amirvaresi et al., 2021). In contrast, visible–near infrared (Vis-NIR) hyperspectral imaging (HSI) has emerged as a powerful and non-destructive analytical tool, capable of simultaneously capturing spatial and spectral information. When coupled with advanced chemometric techniques such as mean-field independent component analysis (MF-ICA) and partial least squares discriminant analysis (PLS-DA), HSI provides a rapid and effective means for detection and visualization of contaminants in complex food matrices (Li et al., 2019). The objective of this study is to employ HSI in combination with MF-ICA and PLS-DA to detect and spatially map wheat bug infestation in flour samples with high accuracy and resolution.
Recent studies have demonstrated the effectiveness of hyperspectral imaging for assessing food quality, particularly in grain classification and contamination detection. HSI technology has proven capable of differentiating wheat varieties and measuring contamination levels accurately (Jiang et al., 2023; Bao et al., 2019; Wang et al., 2024).

Hashemi-Nasab and Parastar (2022) developed a novel approach combining Vis-NIR hyperspectral imaging with mean-field independent component analysis (MF-ICA) and multivariate classification for saffron authentication. MF-ICA effectively extracted pure spectral and spatial profiles, followed by PLS-DA modeling that achieved 100% accuracy in distinguishing authentic and adulterated samples, even in complex mixtures. Additionally, DD-SIMCA modeling showed high sensitivity (95%) and perfect specificity (100%), confirming the method's robustness for food authenticity studies. Amirvaresi et al. (2021) compared near-infrared (NIR) and mid-infrared (MIR) spectroscopy for saffron authentication using chemometric techniques. PCA was applied to explore geographical clustering, while PLS-DA accurately classified authentic and adulterated samples. Additionally, PLSR was used to quantify adulteration levels, where NIR outperformed MIR with R^2 values up to 0.99, highlighting its superiority for both qualitative and quantitative saffron assessment.
In this study, Vis-NIR hyperspectral imaging was used to rapidly and non-destructively detect and spatially localize wheat bug infestation in wheat flour. Pure spectra of healthy and infested samples were extracted using SIMPLISMA, then analyzed with MF-ICA to obtain meaningful spectral components. These were used to train a PLS-DA model, which was applied to moderately infested flour to generate a pixel-level contamination map. This work represents one of the few studies to apply a fast and non-destructive HSI-based chemometric approach for visualizing biological contamination in flour, offering practical potential for real-time quality monitoring.

* Corresponding author

2. MATERIALS AND METHODS

2.1 Dataset

2.1.1 Hyperspectral Images

Hyperspectral images were acquired using the HYSPIM system, covering 580 spectral bands across the 400–950 nm range with a spectral resolution of 3 nm. Imaging was performed under standardized conditions with uniform illumination and a fixed camera height of 25 cm (Figure 1). Three types of wheat flour samples including Pure, moderately infested, and heavily infested with wheat bug (Eurygaster spp.) were captured using identical settings to ensure consistency and comparability (Figure 2).

Figure 1. The HYSPIM hyperspectral imaging

Figure 2. Wheat flour samples

Figure 3 shows the spectral response curves of the three wheat flour samples: pure, moderately infested, and heavily infested.

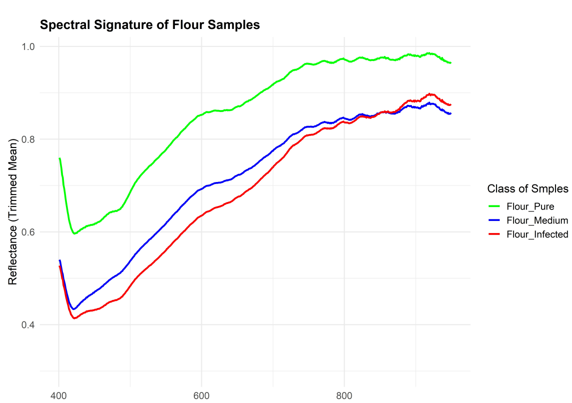

Figure 3. Spectral signatures of the three wheat flour samples

To facilitate the use of the hyperspectral images captured by the HYSPIM system (originally in MATLAB format) in other platforms such as ENVI, the images were converted to the standard ENVI format using the HYSPIM-CONVERTER V.1.0.1 developed at the Remote Sensing Laboratory, University of Tabriz (https://github.com/M-SOURGHALI/HYSPIM-CONVERTER).

2.2. Preprocessing

To enhance the quality of the spectral data and minimize external variability, two common preprocessing techniques were applied. Standard Normal Variate (SNV) was used to correct for multiplicative scatter effects and baseline shifts by centering and scaling each individual spectrum. This method improves comparability between samples by reducing the influence of particle size and surface scattering. In addition, Savitzky-Golay smoothing was employed to reduce high-frequency noise while preserving the original shape and features of the spectral curves. This filter applies a local polynomial regression across neighboring wavelengths, providing a smooth and noise-reduced signal suitable for further analysis.

2.3. Data Extraction, Processing, and Modeling

2.3.1. SIMPLISMA algorithm

To extract the most informative spectral data from the samples, the SIMPLISMA (SIMPle-to-use Interactive Self-modeling Mixture Analysis) algorithm was applied to the pure (healthy) and contaminated flour images. SIMPLISMA is an unsupervised method that identifies the most representative "pure" pixels in hyperspectral data by maximizing spectral contrast and minimizing interference from noise and mixed signals. From each class, 20 pure spectra were selected as inputs for the subsequent MF-ICA modeling step. These spectra served as reliable references for distinguishing between healthy and infested flour in the classification process.

2.3.2. MF-ICA algorithm

The pure spectra extracted from healthy and infested flour samples were used as input for the MF-ICA algorithm. MF-ICA is a powerful unsupervised method for blind source separation that decomposes hyperspectral data into spatially and spectrally independent components. This method enables the isolation of chemically meaningful signals even in complex and noisy datasets. In this study, two main components were extracted, corresponding to healthy and infested flour, which were subsequently used in the classification step.

2.3.3. Chemometric Analysis

Following the extraction of independent components via MF-ICA, a supervised classification model was constructed using Partial Least Squares Discriminant Analysis (PLS-DA). PLS-DA is a widely used multivariate statistical method that projects high-dimensional spectral data into a lower-dimensional latent space while maximizing the separation between predefined classes. In this study, the two independent spectral components derived from healthy and infested flour samples were used to train the PLS-DA model. The goal was to develop a robust classifier capable of accurately discriminating between healthy and contaminated regions in the hyperspectral images. PLS-DA was chosen due to its effectiveness in handling collinearity, high dimensionality, and noise inherent in hyperspectral datasets. Model training was performed using the pure component data, to evaluate the performance and class separation ability of the PLS-DA model, the plotIndiv function was used to visualize the distribution of samples in the latent variable space. The result was a pixel-level classification map, in which each pixel was assigned to one of the predefined classes (pure or infested), allowing for the spatial visualization of contamination patterns across the flour sample. The classification performance of the model was evaluated based on its ability to generalize from pure training spectra to spatially heterogeneous and partially infested samples.

3. Result

Figure 4 displays the pure spectral signatures extracted from healthy and infested wheat flour samples using the SIMPLISMA algorithm. The green and red curves represent the spectral responses of pure and infested samples, respectively, across the 400–950 nm wavelength range. Notably, negative reflectance values are observed in some regions of the spectra, particularly in the 400–500 nm range. This phenomenon is due to the application of the Standard Normal Variate (SNV) preprocessing technique, which standardizes each spectrum by centering (subtracting the mean) and scaling (dividing by the standard deviation). As a result, the absolute reflectance values lose their physical meaning, but the relative differences and spectral shapes become more comparable, enhancing class separation.
The overall pattern reveals a clear divergence between the two classes, particularly in the 500–900 nm range, where infested samples show higher reflectance values. These spectral differences indicate changes in the optical properties of flour due to wheat bug contamination, which are effectively captured and emphasized by the SNV transformation and the SIMPLISMA extraction process.

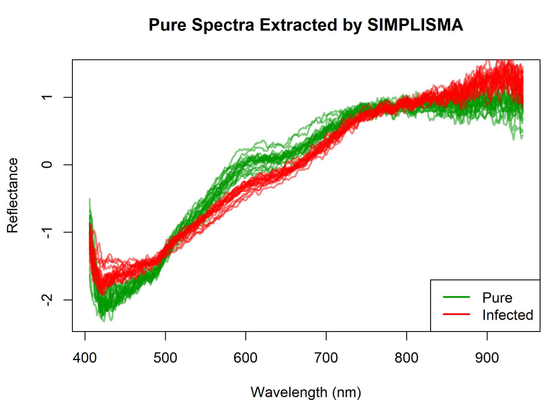

Figure 4. Pure spectral signatures extracted from healthy and infested wheat flour samples using the SIMPLISMA algorithm (20 spectra per class).

Figure 5 illustrates the scatter plot of the two independent components (IC1 and IC2) extracted using the MF-ICA algorithm, based on 40 pure spectra (20 from pure flour and 20 from infested samples). Each point represents a sample projected in the two-dimensional latent space formed by IC1 and IC2. The blue points correspond to pure flour spectra, while the red points represent spectra from infested flour. The plot demonstrates clear and distinct clustering between the two classes, with minimal overlap. The pure samples are mostly distributed in the lower half of the plot (negative IC2 values), whereas the infested samples are primarily located in the upper half (positive IC2 values). This well-defined separation indicates that the independent components extracted by MF-ICA successfully capture the underlying spectral differences between pure and infested flour, which is essential for robust classification. The result validates the effectiveness of MF-ICA in distinguishing subtle chemical or physical changes caused by wheat bug infestation and provides a strong foundation for downstream modeling using supervised classifiers such as PLS-DA.

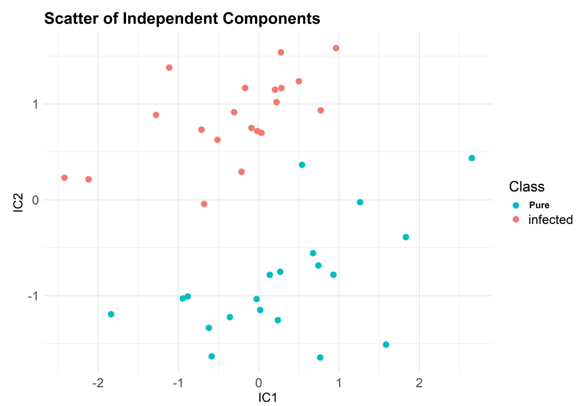

Figure 5. Scatter plot of the two independent components

To build the classification model, the two independent components extracted by MF-ICA were used as input variables in the Partial Least Squares Discriminant Analysis (PLS-DA). The plotIndiv function was used to visualize the model performance and assess class separation in the latent variable space.
Figure 6 displays the score plot of the first two PLS components, which together explain 100% of the variance in the data (50% by each component). The blue points represent pure (healthy) flour spectra, and the orange points represent infested flour spectra. Each point corresponds to one of the 40 pure spectra used for training the model.
The plot reveals a clear and distinct separation between the two classes along both latent variables, indicating that the PLS-DA model was successful in capturing the discriminatory information between healthy and infested samples. This strong separation confirms the effectiveness of the spectral features extracted via MF-ICA and supports the model's ability to perform accurate classification. The absence of misclassified samples in this training set highlights the robustness of the model and its potential for reliable detection of contamination in new samples.

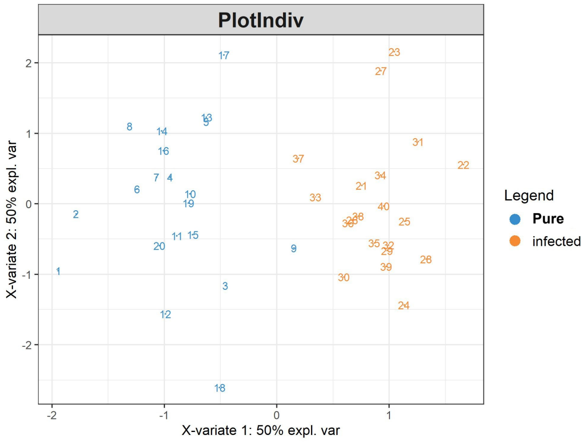

Figure 6. Score plot of the PLS-DA model based on two components

To further evaluate the generalization performance of the developed PLS-DA classification model, the hyperspectral image of the moderately infested wheat flour sample which was not included in the model training was subjected to pixel-wise classification. This image was preprocessed using the same SNV and smoothing techniques as the training samples and then projected onto the PLS-DA model previously built using pure spectra (pure and infested). Each pixel spectrum in the test image was independently classified into one of the two predefined classes. The resulting classification map is presented in Figure 7, where each pixel has been labeled as either “pure” or “infested” based on the model’s prediction.

The pixel-wise classification map obtained by applying the PLS-DA model to the moderately infested wheat flour sample reveals an almost equal distribution between pure (49.91%) and infected (50.09%) pixels. This balanced classification indicates the model’s high sensitivity in detecting subtle spectral variations associated with intermediate levels of contamination particularly noteworthy given that this sample was not included in the model training phase.
The presence of clustered and spatially coherent infected regions, rather than random or scattered classifications, suggests that the model has not merely captured numerical differences in spectral values but has also successfully recognized realistic contamination patterns across the flour surface. In other words, the model has accurately identified genuinely infected areas as distinct from healthy ones, without producing arbitrary or unrealistic segmentation. This spatial consistency reflects the high robustness of the model and the discriminative power of the spectral components extracted by the MF-ICA algorithm.
Moreover, the model’s ability to classify a sample with an unknown level of contamination at this degree of accuracy and spatial resolution highlights its potential for semi-quantitative contamination estimation a feature that is highly valuable in food quality monitoring systems, especially for early detection of infestation.

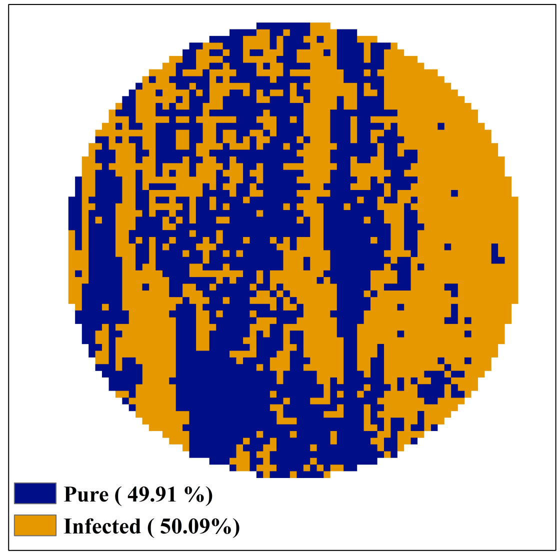

Figure 7. Pixel-wise classification map

4. Discussion and Conclusion

In this study, a non-destructive and rapid method based on visible–near infrared hyperspectral imaging (Vis-NIR HSI) combined with chemometric modeling was successfully developed for the detection and spatial localization of wheat bug (Eurygaster spp.) contamination in wheat flour. By applying preprocessing techniques such as Standard Normal Variate (SNV) and Savitzky–Golay smoothing, spectral noise was minimized and baseline variability corrected, improving the quality of input data.
The use of the SIMPLISMA algorithm allowed for the extraction of pure spectral signatures from pure and infested samples, which served as reliable inputs for the MF-ICA algorithm. MF-ICA effectively decomposed the hyperspectral data into two independent components, clearly distinguishing between the two classes. These components were then used to construct a PLS-DA model, which demonstrated excellent discrimination between pure and infested flour spectra, as visualized in the plotIndiv score plot.
Critically, the model was evaluated on an unseen sample with moderate infestation. The resulting pixel-wise classification map revealed a nearly balanced distribution between pure and infected pixels, reflecting the model’s sensitivity to intermediate levels of contamination. Moreover, the spatial coherence observed in the classified map indicated that the model captured realistic contamination patterns, not just numerical variation, thereby confirming the robustness and spatial reliability of the approach.
The findings of this study on the detection of wheat bug contamination in wheat flour using hyperspectral imaging and chemometric modeling are in line with results from previous research focused on saffron authentication. Specifically, Hashemi-Nasab and Parastar (2022) developed a hybrid method combining Vis-NIR hyperspectral imaging with MF-ICA and PLS-DA, which achieved 100% classification accuracy in distinguishing authentic and adulterated saffron samples—even in complex mixtures. Additionally, their use of the DD-SIMCA model yielded 95% sensitivity and 100% specificity, highlighting the robustness and reliability of their approach in food authenticity assessments.

Similarly, Amirvaresi et al. (2021) compared near-infrared (NIR) and mid-infrared (MIR) spectroscopy coupled with chemometrics for saffron authentication. They applied PCA for geographical clustering, PLS-DA for binary classification, and PLSR for quantifying adulteration levels. Their results demonstrated the superiority of NIR over MIR, with R^2 values up to 0.99 for predicting adulteration levels, thus confirming its utility for both qualitative and quantitative analysis.
In comparison, the present study introduces several key innovations. First, it focuses on a more challenging and sensitive matrix wheat flour where contamination is typically non-uniform and less visually detectable. Second, unlike the saffron studies, this work not only performs spectral classification but also generates a pixel-wise contamination distribution map for a moderately infested sample. Importantly, the test image was not included in the training phase, yet the model successfully identified contamination patterns, demonstrating strong generalization capability and the discriminative power of MF-ICA-derived features. These findings extend the applicability of HSI-based chemometric models beyond saffron authentication to semi-quantitative spatial detection of contaminants in powdered food matrices.

The results highlight the potential of HSI combined with MF-ICA and PLS-DA as a powerful, fast, and non-destructive method for semi-quantitative detection of contamination in flour-based products. This approach offers promising applications for real-time quality control in the food industry, particularly in early detection of partial infestations that are often missed by traditional methods.

5. References

Amirvaresi, A., Nikounezhad, N., Amirahmadi, M., Daraei, B., Parastar, H., 2021. Comparison of near-infrared (NIR) and mid-infrared (MIR) spectroscopy based on chemometrics for saffron authentication and adulteration detection. *Food Chemistry*, 344, 128647.

Bao, Y., Mi, C., Wu, N., Liu, F., He, Y., 2019. Rapid classification of wheat grain varieties using hyperspectral imaging and chemometrics. *Applied Sciences*, 9(19), 4119.

Hashemi-Nasab, F. S., Parastar, H., 2022. Vis-NIR hyperspectral imaging coupled with independent component analysis for saffron authentication. *Food Chemistry*, 393, 133450.

Jiang, X., Bu, Y., Han, L., Tian, J., Hu, X., Zhang, X., Luo, H., 2023. Rapid nondestructive detecting of wheat varieties and mixing ratio by combining hyperspectral imaging and ensemble learning. *Food Control*, 150, 109740.

Li, J., Zhang, Y., Liu, M., Chen, J., Xue, L., 2019. Rapid detection and visualization of mechanical bruises on “Nanfeng” mandarin using the hyperspectral imaging combined with ICA_LSQ method. *Food Analytical Methods*, 12(9), 2025-2034.

Wang, Y., Ou, X., He, H. J., Kamruzzaman, M., 2024. Advancements, limitations and challenges in hyperspectral imaging for comprehensive assessment of wheat quality: An up-to-date review. *Food Chemistry*: X, 21, 101235.

https://doi.org/10.5194/isprs-annals-X-4-W8-2025-737-2026