Detection of impurities on faba bean (*Vicia faba L.*) by NIR spectroscopy

Doaa Sheahata, Ramadan ElGamal, Sherif Radwan, and Gamal Elmasry*

Department of Agricultural Engineering, Faculty of Agriculture, Suez Canal University, Ismailia 41522, Egypt.

Received: *31/5/2025*

Abstract: This Faba bean (*Vicia faba L.*) holds a significant position as one of the primary agricultural commodities globally. The presence of impurities on faba beans can lead to significant economic losses and quality deterioration that influences seed vigor and growth. Therefore, it is crucial to detect impurities seeds rapidly and non-destructively. In this study, a near-infrared (NIR) spectra acquisition device (400–1000 nm) was employed for seed quality detection. Spectral fingerprints extracted from pure faba bean seed and impurities were modeled using principal component analysis (PCA), partial least square (PLS) regression and linear discriminant analysis (LDA) to demonstrate the general overview of the spectral characteristics, predict the seed and impurities features and classify the seeds and impurities to the right categories. The results showed that impurities can be detected and classified precisely with total explained variance of 100%, with better separation of the classes. It also indicates that good statistics were obtained for prediction, cross-validation, and calibration, the PLS model achieved correlation coefficients (r) of 0.97, with minimal values of RMSE of about 2.98. LDA was utilized to classify the seeds based on their spectral fingerprints, achieving an overall classification accuracy of 84%. The model effectively distinguished between pure seeds and impurities, demonstrating its potential for rapid, non-destructive impurity detection in faba bean seeds. This study illustrates the applicability of NIR spectroscopy combined with PCA, PLS, and LDA models for accurate seed impurity detection and classification.

Keywords: Faba bean**;** seeds**;** NIR spectroscopy**;** Impurities**;** Multivariate analysis.

INTRODUCTION

Faba bean (Vicia faba L.) is an important legume crop valued for its high protein content and health-promoting bioactive compounds, making it increasingly popular in both human diets and animal feed worldwide, it's staple crops in many parts of the world, contributing significantly to food security (Johnson et al., 2020; Boccia et al., 2013; Johnson et al., 2023). Faba beans is also considered a crucial winter crop primarily cultivated in Egypt. The estimated cultivated area for faba beans in Egypt is approximately 40.31 thousand ha with total production reached 139.52 thousand tons of dry seeds (FAOSTAT, 2020). Therefore, it is very important to guarantee the purity and safety of faba bean seeds. However, the presence of various impurities such as insects, bean peels, small rocks, uneven bean pulp fractions, and seed cotyledon fractions poses significant challenges to faba bean seed quality, safety, and marketability.

The seed quality plays a pivotal role in determining crop productivity and sustainability, it is a fundamental factor that influences the overall agricultural output. Poor seed quality can lead to reduced crop yield, susceptibility to diseases, and economic losses for farmers (FAO, 2018). As commercial production expands, there is a growing demand for rapid, reliable, and cost-effective methods to assess seed quality and detect impurities, which are critical for food safety, nutritional value, and marketability (Rahman & Cho, 2016). Traditional methods for impurity detection and quality assessment of seeds are often labor-intensive, time-consuming, and require extensive sample preparation, limiting their utility for high-throughput screening in breeding programs and quality control processes (Carbas *et al.,* 2020). Due to the limitation of traditional methods, there is increasing demand for rapid, non-destructive, and reliable techniques for evaluation of seed quality. Therefore, the non-destructive testing technologies are developed and used for assessing the quality of seeds. Near-infrared spectroscopy (NIR) has emerged as a promising analytical tool for the rapid, non-invasive assessment of seed quality and impurity detection in various crops, including legumes and cereals (Plans *et al.,* 2013). This technique has been successfully applied to quantify nutritional parameters (such as protein, starch, and oil), detect adulterants, and assess bioactive compounds in seeds and flours (Lippolis *et al.,* 2024; Johnson *et al.,* 2023; Hernández-Hernández *et al.,* 2021; Carbas *et al.,* 2020). NIRS has been successfully applied to identify fumonisin B1 contamination in beans and melamine adulteration in soybean meals, highlighting its versatility for food safety applications (Haughey *et al.,* 2013). Additionally, the technique can rapidly discriminate between samples with high and low levels of phenolics or antioxidant compounds, supporting quality assurance and breeding efforts (Hernández-Hernández *et al.,* 2021). In the context of seed purity analysis, NIRS has shown remarkable potential for assessing impurity levels in crops such as cotton, maize, and rice . It has been used to identify the purity of seed cotton by converting spectral data into binary form and computing key spectral features (Reddy *et al.,* 2022). The technique also enabled the identification of pure botanical and field debris samples with an overall accuracy of 98% (Fortier *et al.,* 2010). However, challenges related to debris distribution and sample size limited its use to screening purposes (Liu *et al.,* 2024).

The integration of advanced chemometric methods such as Partial Least Squares (PLS) and Principal Component Analysis (PCA) with NIR data has significantly improved the accuracy and reliability of impurity detection and compositional analysis (Lippolis *et al.,* 2024). These statistical approaches facilitate the development of robust calibration models, enabling precise prediction of both macronutrients and micronutrients, and the discrimination of samples with varying levels of impurities or adulterants (Johnson *et al.,* 2023). The use of independent validation sets, and

*Corresponding author e-mail:**gamal.elmasry@agr.sue z.edu.eg**

clustering techniques further enhances model performance and generalizability across diverse genotypes and growing conditions (Rahman & Cho, 2016). Cross-validation is a technique utilized for evaluating predictive rule accuracy and model selection by dividing the data into multiple segments and testing the model on a validation sample (Malone *et al.,* 2022). This method allows for the verification of results derived from exploratory data analysis and the selection of the most predictive model. Despite being non-parametric, cross-validation can be applied to any automated model-building technique (Bhatti, 2018).

In this context, a PLS model with a correlation coefficient of 0.906 was built for estimating debris content in cotton, while Linear Discriminant Analysis (LDA) was used to classify different lint types (Li *et al.,* 2010). The application of variable selection methods like competitive adaptive reweighting algorithms (CARS) led to improved model accuracy, with the 4052–8000 cm^{-1} spectral range proving optimal (Zhang *et al.,* 2022). Another classification model using the first derivative spectra and multiplicative scatter correction (MSC) achieved 80% accuracy in predicting cotton impurities (Reddy et al., 2016). NIRS has also been applied to evaluate the purity of maize hybrids using PCA and Orthogonal Linear Discriminant Analysis (OLDA), achieving identification rates of 100% and 90% for specific varieties (Tian-xin *et al.,* 2015). In food adulteration detection, NIRS successfully identified corn mixed into Brazilian coffee at contamination levels as low as 5% using PLSR (Winkler-Moser *et al.,* 2015). Moreover, the purity of multi-grain rice seeds was determined using Vis-NIR spectroscopy. The models achieved strong predictive performance, with R^2 values of 0.920 in the short NIR region and 0.930 in the long NIR region (Zhang *et al.,* 2019).

Several studies were reported in literature for using NIR spectroscopy to detect the sed quality of different crops and legumes. However, studies using NIR spectroscopy to detect the quality of the Egyptian faba bean variety Sakha 1 seeds and the specific impurities accompanying this variety have not been reported. In this sense, developing an accessible multivariate approach using portable NIR devices to identify potential impurities in faba bean will bring great scientific-technological contribution and, consequently, will encourage Egyptian authorities entrusted with the seed quality inspection to apply such advanced technologies. Therefore, the main objective of this study was to propose a methodology for predicting and screening some impurities in faba bean seeds, based on NIR spectroscopy associated with multivariate analyses, to be implemented for the best use of this product in the industry.

MATERIALS AND METHODS

2.1. Sample collection and preparation:

The faba bean seeds of the authorized variety Sakha 1 and the accompanying impurities used in this study were obtained from the Seed Production Department, Egyptian Ministry of Agriculture and Land Reclamation. The faba bean seeds were sorted based on size and appearance to ensure uniformity within the sample set (Figure 1). The moisture content of seeds was determined using a standard oven method by drying about 10g samples at 103 °C for 72h and found to be 9% (w.b.).

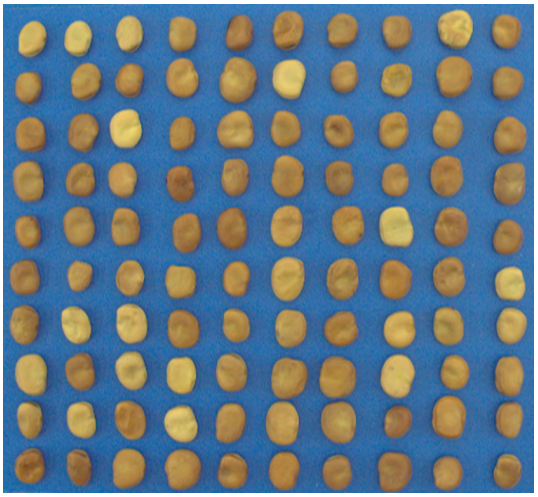


Figure (1): A representative sample of faba bean seeds of the variety Sakha 1.

Five different kinds of impurities that usually exist in faba bean seeds such as insects, bean peels, small rocks, uneven bean pulp fractions, and seed cotyledon fractions were also prepared and isolated from the bean seed lots as shown (Figure 2). Insects (*Bruchidius incarnatus Boh.*) were carefully selected to represent a species commonly found in bean seeds. Both the faba bean seeds and impurities were thoroughly cleaned to remove any external debris or contaminants. The bean seeds and impurities were packaged in airtight containers (vacuum packed bags) and stored in a cooling chamber with an average temperature of 10 °C to maintain their integrity and prevent any contamination until further analysis. Each sample was labeled with a unique identifier to maintain traceability throughout the experiment. This ensured that the measurements and observations could be accurately associated with the corresponding samples.

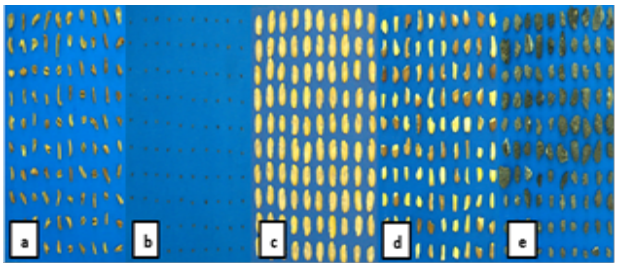


Figure (2): Representative samples of five different impurities in faba bean seeds: (a) bean peels; (b) insects; (c) equal bean seed cotyledon fractions; (d) uneven bean pulp fractions; (e) small pieces of rock.

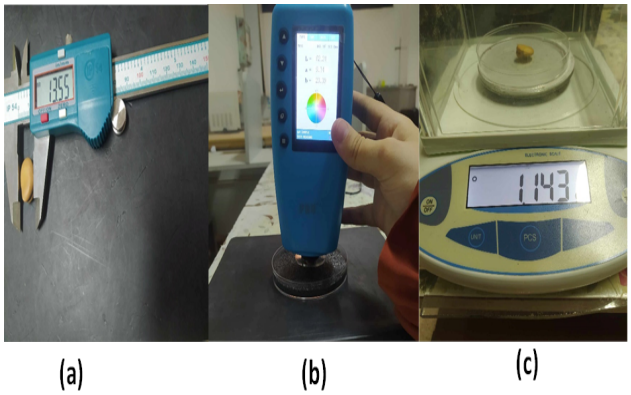
2.2. Physical properties of faba bean seeds and impurities:

Size, color, and mass, which are crucial parameters for understanding the physical properties of the samples and developing effective detection methods, were measured for faba bean seeds and the accompanying impurities. The size parameters (length, width, and thickness) of both bean seeds and impurities were measured using a digital caliper (CD-

6"C, Mitutoyo, Japan) with an accuracy of 0.01 mm (Figure 3a).

The color values of the faba bean seeds and their impurities were determined using a colorimeter (WR10QC, X-Rite, USA) as shown in Figure (3b). Before recording the measurements, the colorimeter was calibrated to ensure accuracy. This involved using a standard white tile, which served as a reference point for all subsequent measurements. The colorimeter was zeroed on this tile to ensure that any subsequent measurements were not affected by any potential bias in the instrument. Carful was taken to ensure that the sample was clean and free from any dust or other contaminants that could potentially affect color measurements. The sample was placed in a consistent position on the measuring stage of the device to ensure consistency. The colorimeter measured three parameters for each sample in the LAB color space: L* (lightness), a* (the red/green axis), and b* (the yellow/blue axis) as shown in Figure (3b). The measurements were recorded and used to calculate the average color properties for each type of sample. This provided a comprehensive understanding of the color characteristics of faba bean seeds and impurities, which are crucial for developing effective detection methods.

Figure (3): Measuring physical features of bean seeds using



(a) a digital caliper; (b) a colorimeter; and (c) a digital balance.

The mass of each seed sample was measured using a precise weighing scale. Each sample was carefully placed on the center of the weighing scale (Techplast, SAB – 220, Egypt), ensuring that it was clean and dry as shown in Figure (3c). The mass of each sample was recorded to the nearest milligram. The mass measurement provided valuable information about the mass of the samples, which could be used to assess the density and uniformity of the seeds and the impurities.

The data collected on the size, color, and mass of the bean seeds and their impurities served as important parameters for subsequent analysis and comparison. These measurements provided valuable insights into the physical characteristics of the samples, aiding in the development of effective detection of impurities in faba bean seeds. Besides determining size, color and mass, the density of the bean seeds and the five impurities were also calculated by dividing the mass by the volume of the sample. Each density measurement was repeated three times for each sample to ensure consistency and the average density for each sample was calculated.

2.3. NIRS data acquisition

A hand- held NIR spectrometer (LinkSquare1, Link Square Inc., Korea) shown in Figure (4a) was used in the spectral range of $400-1000$ nm to scan the samples spectrally with optical resolution of 1 nm to 44 nm of 552 wavelengths. It was used to extract the spectral fingerprints of the bean seeds and the five impurities. The reflectance of the samples at each wavelength was recorded, creating a spectral profile for each sample. These spectral profiles were then analyzed to identify and distinguish between the seeds and impurities. The solid form of the samples (seeds and impurities) allowed direct and unaltered analysis of their spectral characteristics, providing valuable insights into their composition and structure, and enabling the effective detection of impurities. During scanning, the faba bean seeds were placed in a Petri dish and then directly scanned in a completely dark room by the Link Square spectrometer held in a vertical position. In total, 600 spectral fingerprints of the solid seeds. After taking the spectral fingerprints of all samples, they were exported in excel sheets via the LinkSquare Collector software (LinkSquare Co., Korea) installed in the computer as shown in Figure (4b).

Figure (4): The hand-held LinkSquare NIR spectrometer (a) LinkSquare 1 spectrometer, and (b) User interface of the LinkSquare collector software.

2.4. Spectral data analysis:

2.4.1. Preparing and preprocessing of spectral data:

Spectral fingerprints for the seeds and impurities samples were exported to spreadsheet Excel files using the LinkSquare Collector software. Spectral preprocessing techniques encompass a collection of discretionary mathematical procedures that are conducted on the spectra prior to the construction of a calibration model. The mathematical pretreatment of spectra serves to diminish noise or extraneous data, as accomplished through smoothing techniques, while concurrently augmenting the signal originating from the chemical information via differentiation as accomplished by standard normal variate (SNV) preprocessing. After the preprocessing step, each type of the spectral data was then imported into the Unscrambler software (Unscrambler x, Camo Analytics, Denmark) to be analyzed separately using different multivariate modeling methods (PCA, PLS and LDA) to predict the seed features or to classify the seeds and impurities.

2.4.2. PCA model:

Principal component analysis (PCA) was performed to reduce the dimensionality of the spectral data while preserving most of the variance. It identifies the principal components (PCs) that explain the most variance in the data. These PCs are linear combinations of the original spectral variables. The variance explained by each PC and the cumulative variance explained by multiple PCs were examined. A scree plot or variance explained plot helps in determining the optimal number of PCs to retain. PCA helped in identifying clusters or patterns that could indicate the presence of impurities based only on the spectra fingerprints of the bean seeds and the impurities. This analysis technique provided insights into the overall structure of the data and aided in the interpretation of the results.

2.4.3. PLS model:

Partial Least Squares (PLS) regression was employed to investigate the underlying relationships between the spectral data (X), acquired across various wavelengths, and the corresponding physical attributes (Y) of the samples, including color parameters, mass, and dimensional measurements (length, width, and thickness). This multivariate analysis technique facilitated the modeling of complex correlations between spectral signatures and tangible physical characteristics, thereby enabling the effective discrimination between bean seeds and foreign impurities. By establishing a predictive model, PLS allowed for the classification of samples based on their spectral profiles, contributing to a more accurate and objective identification of impurities within the dataset.

In the sample set (both faba bean seeds and five unique impurity categories, a total of 600 spectral fingerprints were employed for calibration, validation, and predictive modeling using NIR spectroscopy. Spectral preprocessing methods, specifically smoothing techniques, were implemented to mitigate baseline drift prior to conducting regression modeling. Spectral fingerprints were systematically extracted from faba bean seeds as well as from the five impurities at 552 wavelengths (variables) within the spectral range extending from 400 to 1100 nm. The complete spectral range was utilized to develop PLS models for each color feature based on the average spectra derived from the 600 spectral fingerprints during the calibration phase. The dataset was also randomly partitioned into 70% for calibration and 30% for prediction, to facilitate the construction of the PLS model.

The analysis involved decomposing the predictor matrix (spectral data) and the response matrix (seed and impurities features) into latent structures. These latent variables (factors) are extracted to capture the most relevant information that correlates with the response variables. Full cross-validation was employed to determine the optimal number of latent factors. Cross-validation helps in evaluating the model's performance on unseen data and prevents overfitting by ensuring that the model generalizes well to new data. The coefficient of determination (R^2), and the root mean squared error (RMSE) were used to assess the accuracy of PLS regression models. High R^2 values and low RMSE values indicate a model that accurately captures the relationship between spectral data and seed features (Williams & Norris, 2001).

The value of R^2 measures the proportion of variance in the response variables that is explained by the predictor variables and can be calculated as:

$$R^2 = 1 - \frac{\sum_{i=1}^{n} (y_i - \hat{y}_i)^2}{\sum_{i=1}^{n} (y_i - \bar{y})^2} \quad (2.1)$$

where: y_i = observed values, $\hat{y}_i$ = predicted values, $\bar{y}$ = mean of observed values, n = number of observations.

RMSE measures the average magnitude of the prediction error, providing a measure of how well the model's predictions match the actual observations and can be calculated for calibration as:

$$RMSE_C = \sqrt{\frac{\sum_{i=1}^{n} (y_i - \hat{y}_i)^2}{n}} \quad (2.2)$$

where: y_i = observed values in the calibration set, $\hat{y}_i$ = predicted values in the calibration set, n = number of observations in the calibration set.

For cross-validation:

$$RMSE_{CV} = \sqrt{\frac{\sum_{i=1}^{n} (y_i - \hat{y}_{CV,i})^2}{n}} \quad (2.3)$$

where: y_i = observed values in the validation set, $\hat{y}_{CV,i}$ = predicted values in the cross- validation set, n = number of observations in the validation set.

And, for prediction:

$$RMSE_P = \sqrt{\frac{\sum_{i=1}^{n} (y_i - \hat{y}_{P,i})^2}{n}} \quad (2.4)$$

where: y_i = observed values in the prediction set, $\hat{y}_{P,i}$ = predicted values in the prediction set, n = number of observations in the prediction set.

2.4.4. LDA model

Linear discriminant analysis (LDA) was utilized to classify and distinguish between different impurity types. The LDA is a statistical technique that aims to find a linear combination of features that maximally separates the classes and to create a discriminant function that could accurately classify the bean seeds and impurities into their respective categories. This analysis technique allowed for the identification of impurities based on their unique characteristics and spectral profiles, improving the accuracy of impurity detection. A confusion matrix was built to evaluate the classification performance of the LDA model. It displays the number of correct and

incorrect classifications for each class, allowing for the calculation of accuracy, sensitivity, and specificity. The performance of the LDA model was assessed using several metrics, including accuracy. Accuracy measures the proportion of correctly classified samples among all samples.

RESULTS

3.1. PCA analysis:

The PCA was performed on the raw spectra of faba bean seeds and the selected five different kinds of impurities (Figure 1 and 2). The PCA analysis effectively reduced the dimensionality of the spectral data, summarizing the variance across the wavelengths into a smaller number of principal components (PCs). The first two principal components (PC1 and PC2) accounted for 100% of the total variance, indicating that these components captured most of the meaningful information in the spectral data as shown in Figure (5). These score plots of the samples show that there is a clear clustering of the spectral data observed in the PCA score plot, with distinct two groups representing the faba bean seeds and the different impurity types (Figure 5).

The PC1 accounted for 67% of the total variance, primarily capturing differences between bean seeds and other impurity types based on their spectral features. The PC2 explained 33% of the variance, distinguishing subtle variations within impurity types. The high level of data explanation highlights the effectiveness of the analysis in converting the original dataset and the absorbances linked to the vibrational modes of the NIR spectra into principal components. The dispersion of the five different impurities in faba bean seeds as shown in PC1-PC2 coordinates shows that PC1 is the component that most effectively explains the data distribution. In other words, PC1 is the key factor responsible for differentiating the clusters.

A tightly grouped clusters for faba bean seeds in the PCA score plot reflect high uniformity in the spectral characteristics of the seeds. Distinct grouping between faba bean seeds and impurities validates the effectiveness of PCA in capturing spectral differences between faba. minimal overlap observed between some impurity clusters, such as seed cotyledon fractions and uneven pulp, indicates areas where higher-resolution data or additional preprocessing may enhance separation. The distinct clusters in the score plot highlight PCA's capability in detecting impurities non-destructively, making it a valuable preprocessing step for further analyses like PLS or LDA.

The PCA score plot Figure (6) reveals two well-separated clusters. The cluster on the right side of the plot represents the spectral signatures of faba bean seeds, while the distinct cluster on the left corresponds to the insect samples. The clear distinction between the two clusters validates the effectiveness of near-infrared (NIR) spectroscopy in detecting insects. The first principal component (PC1) accounted for 99% of the total variance, indicating that it captures the primary spectral differences between the faba bean seeds and the insect-infested samples. The second principal component (PC2) explained only 1% of the variance, suggesting that it captures minor variations within each group. As shown in Figure (7) PCA was applied to the spectral data of faba bean seeds and small rocks to reduce dimensionality and highlight significant variance patterns. The first principal component (PC1) accounted for 98% of the total variance, while the second principal component (PC2) explained only 2%. The high variance captured by PC1 suggests that it effectively differentiates between faba bean seeds and small rocks based on their spectral characteristics. The cluster on the right represents the spectral signatures of small rocks, while the cluster on the left corresponds to faba bean seeds. The strong separation along PC1 indicates that the primary spectral differences arise from variations in material composition. Small rocks, being inorganic, exhibit distinct spectral absorbance patterns compared to the organic faba bean seeds, which likely contribute to the high variance explained by PC1. PC2, which explains only 2% of the variance, captures minor variations within each group, suggesting that the spectral responses of the materials are relatively consistent within their respective clusters. The PCA score plot Figure (8) demonstrates two distinct clusters. The cluster on the right corresponds to faba bean seeds, while the cluster on the left represents bean peels. The first principal component (PC1) accounted for 99% of the total variance, while the second principal component (PC2) accounted for only 1%. The dominant variance captured by PC1 indicates that the primary spectral differences between these two materials are well represented along this axis.

Based on the results obtained, it may conclude that the near infrared reflectance spectroscopy (NIRS) can be used to detect faba bean impurities. The NIRS combined with PCA models can be utilized to discriminate and cluster between faba bean seeds and their impurities. These results are in agreement with those reported by Oliveira et al.(2023), who used NIR spectroscopy to detect impurities in cocoa shell powder through PCA analysis. In their study, the data obtained with a benchtop device showed that the first two principal components (PCs) explained 99.75% of the total variance, with 92.86% attributed to PC1 and 6.69% to PC2. For the portable device data, PC1 accounted for 87.57% and PC2 for 11.01%, together explaining 98.58% of the variance in the data.

3.2. Prediction of color using PLS model:

The results, as presented in Table 1, show clear distinctions in the color properties between pure faba bean seeds and the various impurities. For instance, the **L** value (lightness) of pure seeds is significantly higher than that of insect-contaminated seeds, which have lower lightness values due to discoloration.

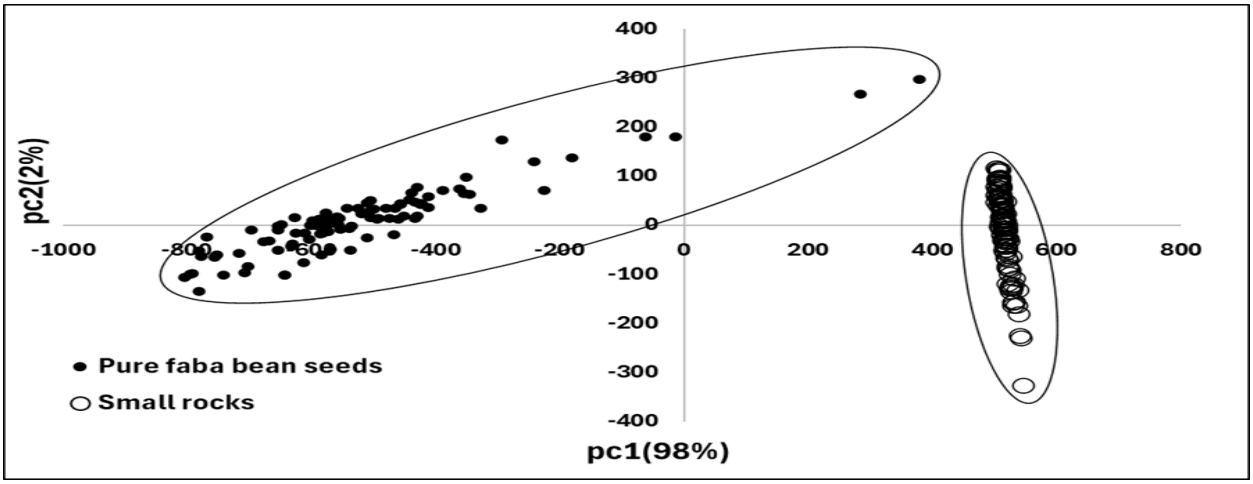


Figure (5): Scatter plot of the PCA scores of raw spectral data for pure faba bean seeds and five different kinds of impurities.

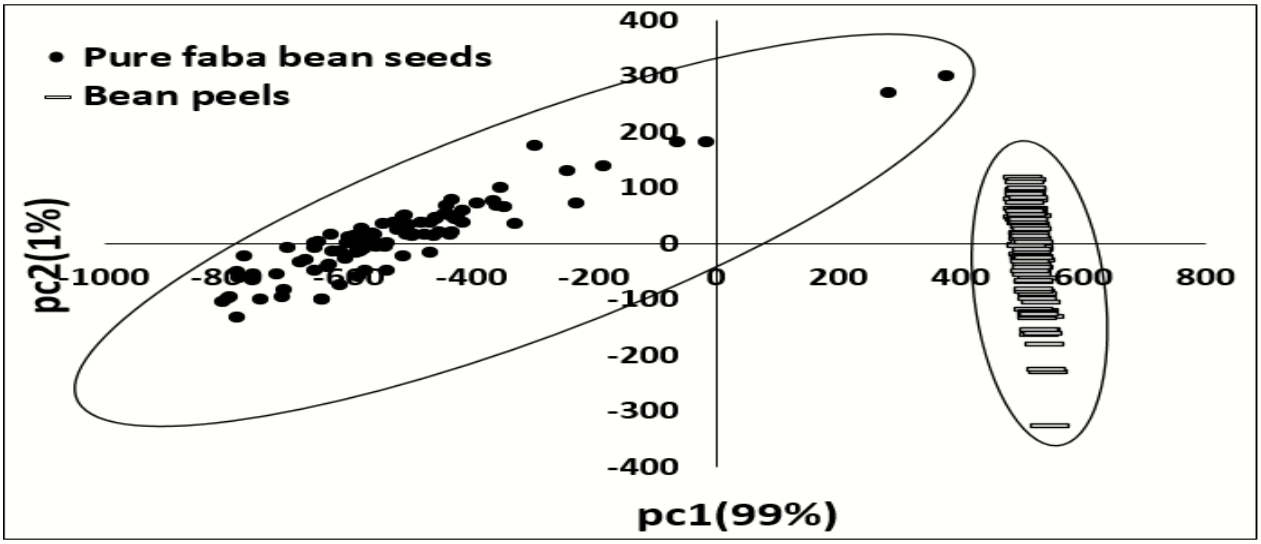


Figure (6): Scatter plot of the PCA scores of raw spectral data for pure faba bean seeds and insects.

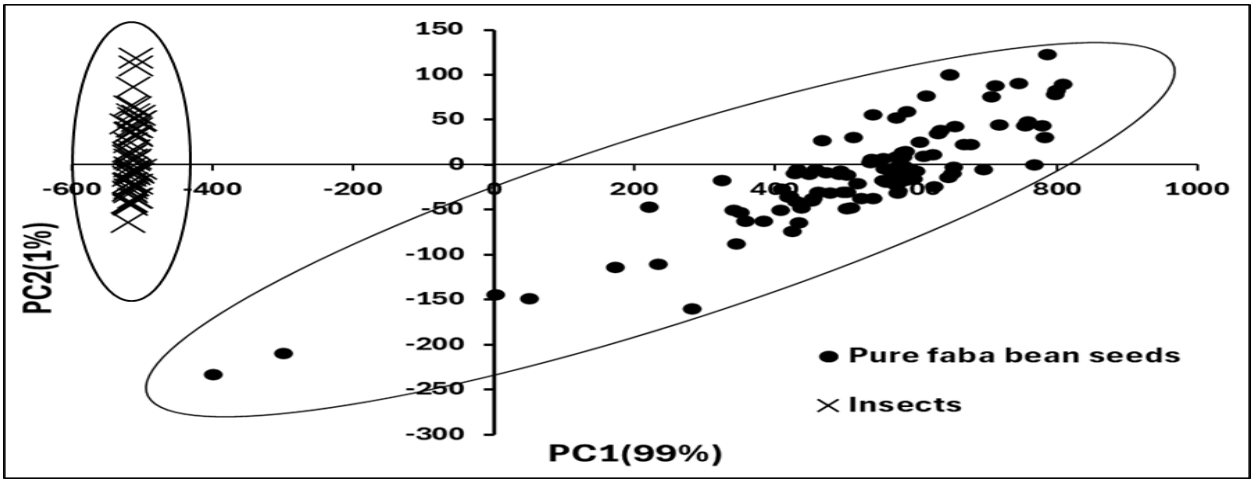


Figure (7): Scatter plot of the PCA scores of raw spectral data for pure faba bean seeds and small rocks.

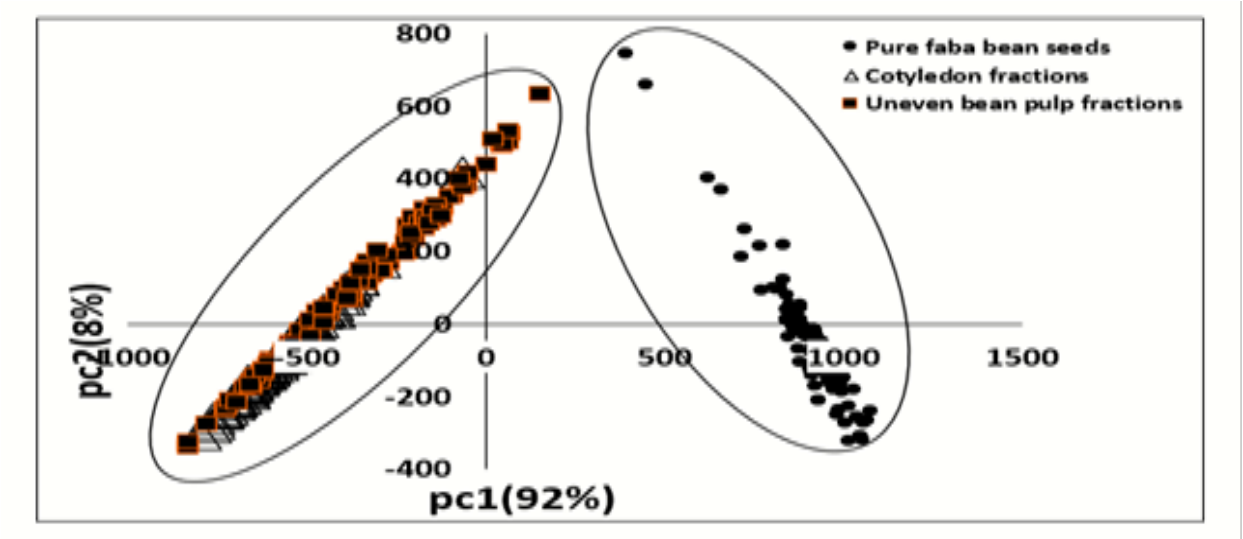


Figure (8): Scatter plot of the PCA scores of raw spectral data for pure faba bean seeds and bean peels.

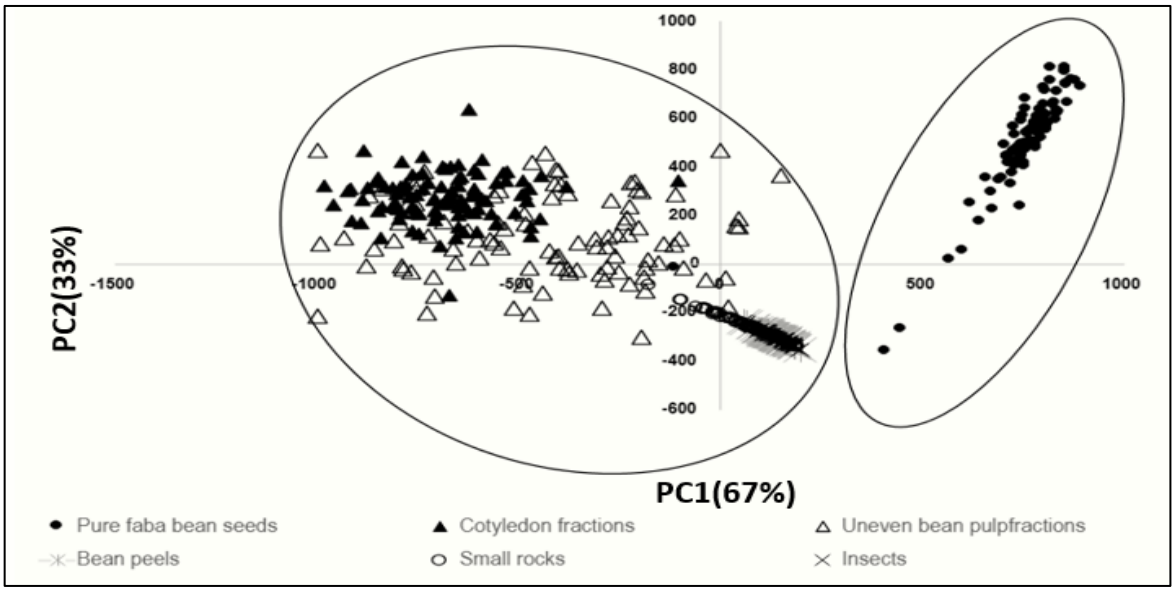


Figure (9): Scatter plot of the PCA scores of raw spectral data for pure faba bean seeds and small rocks.

The red-green component **a** shows variability across different impurities, with pure seeds showing the highest value, indicating a more balanced red-green intensity, whereas insect-contaminated and rock samples exhibit significantly lower values. For the yellow-blue component **b**, pure seeds exhibit the highest values, clearly differentiating them from insect-damaged seeds, which show much lower values. The one-way ANOVA results reveal statistically significant differences ($P \leq 0.05$) between pure seeds and all impurity classes across color features.

Color features such as L, a, and b show significant differences between pure seeds and impurities, especially in the case of insects and rocks, which display significantly different color profiles. This analysis demonstrates that color features are effective for detecting impurities in faba bean seeds. The significant differences observed between the pure seeds and various impurities across all evaluated features highlight the reliability of these features for impurity detection. The incorporation of ANOVA for feature selection enhances the robustness of this method, allowing for accurate classification of seed impurities. Table (1): Color attributes of faba bean seeds and impurities "Values followed by different letters are significantly different ($P<0.05$)".

Table (1): Average values of measured color attributes of faba bean seeds and impurities

Attributes	L	a	b
Pure faba bean seeds	69.4[a]	11.8[a]	74.08[a]
Cotyledon fraction	51.1[b]	11.0[a]	74.77[a]
Un even faba bean fraction	50.8[b]	7.26[b]	74.31[a]
Faba bean peels	41.4[c]	6.41[b]	68.18[a]
Small rocks	33.9[d]	1.93[c]	46.16[c]
Insects	24.1[d]	4.85[d]	34.82[d]

-Values in same column followed by different letters are significantly different (P<0.05)

PLS regression models were built to predict color intensity values for L* (Lightness), a* (Green Red), and b* (Blue Yellow). The relationship between the measured values of color parameters (L*, a* and b*) and those predicted by the PLS models are shown in Figures 10. The results indicate that the PLS models were highly accurate in predicting L* (Lightness), achieving correlation coefficients (R^2) higher than 0.96 and RMSE lower than 2.98. Similarly, predictions of a* (Green Red) and b* (Blue-Yellow) yielded an equal R^2 value of about 0.93 and RMSE values ranged from 0.83 to 2.72. The high R^2 values across calibration and validation phases underscore the model's accuracy and reliability, making it suitable for impurity detection and enabling timely intervention during seed sorting and processing. This capability is particularly valuable for quality control systems in the agricultural industry, where rapid, non-destructive monitoring can enhance efficiency and reduce waste. The integration of NIR spectral data and physical measurements offers scalability for automated impurity detection systems in seed processing plants. While the PLS model effectively reduced dimensionality and identified key features particularly for complex or overlapping impurity profiles.

The results also showed that the analysis needed 2 latent factors for the color variables (L*, a*, b*) which reflects the straightforward relationship between these attributes and impurity presence. The low number of latent factors suggests that color-based variations were relatively simple and well-defined, making them easy to model with minimal complexity. This optimization ensured that the model effectively captured relevant patterns while minimizing the risk of overfitting. The optimized number of latent factors for each Partial Least Squares (PLS) model plays a crucial role in detecting impurities in bean seeds. With an optimized number of latent factors, the model's ability to detect impurities without overfitting is improved, making it a valuable tool for non-invasive, rapid quality control in different agricultural applications.

3.3. Classification of faba bean seeds impurities using the linear discriminant analysis (LDA) model

Spectral fingerprints extracted from faba bean seeds and impurities at 552 wavelengths (variables) in the spectral range from 400–1100 nm was used to develop linear discriminant analysis (LDA) models to classify the seeds into different categories based on the presence of impurities. The LDA model developed for classifying faba bean seeds and their impurities achieved an overall accuracy of 84%. The model utilized 20 components and employed a quadratic discriminant function to classify the samples into six categories.

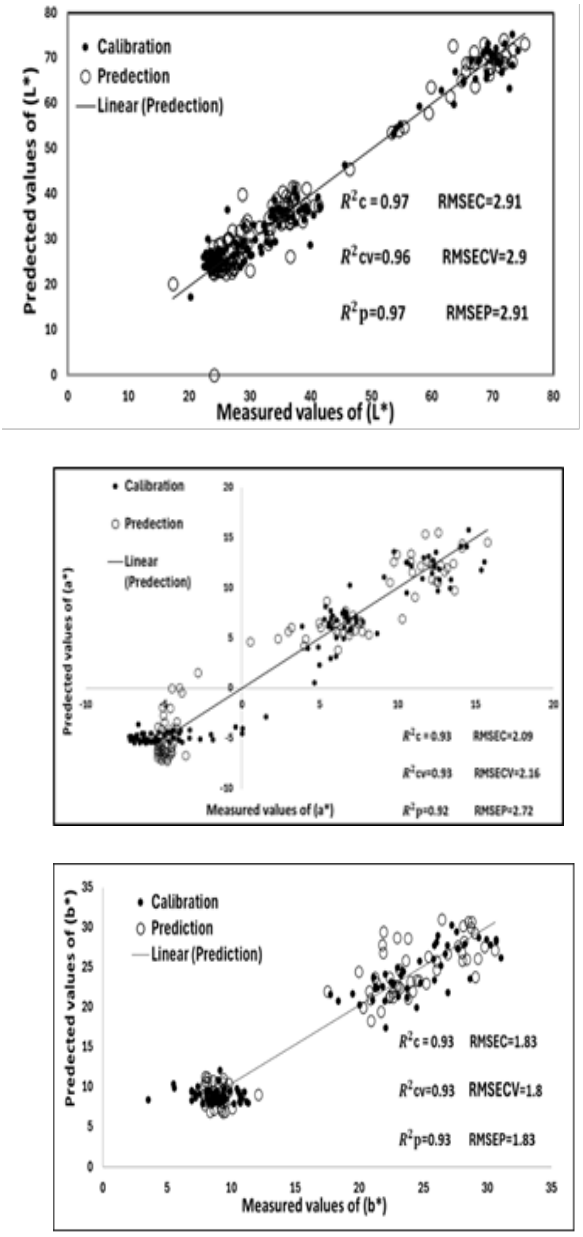


Figure (10): Measured and predicted (L*, a*, b*) values for calibration and prediction sets using PLS models.

These categories included pure seeds, insects, bean peels, small rocks, uneven bean pulp fractions, and cotyledon fractions. The 84% accuracy indicates a high level of classification performance, suggesting the model effectively distinguished most of the categories based on their spectral fingerprints in the range of 400–1100 nm. As shown in table (2), the confusion matrix provides a detailed breakdown of how well the model performed in categorizing the seeds into six classes. The confusion matrix presented above summarizes the classification performance of the Linear Discriminant Analysis (LDA) model used for detecting impurities in faba bean seeds. The matrix evaluates the actual categories (rows) against the predicted categories

(columns), showing the number of correctly and incorrectly classified samples. The classification performance of the model highlights important insights into the detection of faba bean seed impurities. The key observations based on the confusion matrix and model evaluation can be summarized as follows: The model achieved 100% accuracy in identifying pure seeds, successfully distinguishing them from all other impurity categories. This indicates that the spectral fingerprints of pure seeds are highly distinct, allowing for clear differentiation from contaminants. The model performed flawlessly in identifying pure seeds (again), confirming the strong spectral contrast between untainted seeds and other types of impurities, ensuring no misclassifications within this category, the model identified cotyledon fractions with 99% accuracy. However, there was a slight misclassification, with 16 cotyledon fraction samples being incorrectly classified as uneven pulp fractions. This indicates a minor overlap between the spectral characteristics of these two impurity classes, The model correctly identified 84% of uneven pulp fractions. However, 1 sample was misclassified as a cotyledon fraction, which suggests that both categories share some spectral characteristics that caused this confusion, The accuracy for bean peels was lower, at 58%. Several samples were misclassified as small rocks (7%) and insects (26%). This suggests that bean peels have spectral similarities with small rocks and insect-damaged seeds, which made them harder to distinguish in this model, small rocks were identified with an accuracy of 91%, although 5% of the samples were misclassified as uneven pulp fractions and 2% as insect-damaged seeds. The model performed well in distinguishing small rocks, but some overlap with other impurities still exists and the model identified insect-damaged seeds with 72% accuracy. A significant portion (37%) of insect-damaged samples was misclassified as bean peels, indicating a strong overlap between insect-damaged seeds and bean peels in terms of spectral characteristics. A small percentage (2%) of insect samples were misclassified as small rocks.

Table (2): Confusion matrix of LDA analysis to classify faba bean seeds and impurities into six categories with 84% accuracy

		Predicted class					
		Pure Seeds	Cotyledon	Uneven Pulp	Bean Peels	Small Rocks	Insects
Actual class	Pure Seeds	100	0	0	0	0	0
	Cotyledon	0	99	16	0	0	0
	Uneven Pulp	0	1	84	0	0	0
	Bean Peels	0	0	0	58	7	26
	Small Rocks	0	0	0	5	91	2
	insects	0	0	0	37	2	72

DISSCUSSION

The results obtained from the PLS model in this study recorded R^2 higher than that obtained by *Chen et al.*(2024) who utilized near-infrared spectroscopy (NIRS) combined with partial least squares regression (PLSR) to detect low-price rice adulteration in high-price rice blends, achieving correlation coefficients exceeding 87% for accurate estimation of contamination levels in rice samples. Moreover, the results obtained from this study are in agreement with those reported by Han et al.(2023) who applied NIRS combined with PLS to analyze spectral data from cotton seeds and achieved high correlation coefficients (R^2 = 0.98) for impurity detection in cotton seeds. The results also in agreement with Xue et al.(2023) who utilized Partial Least Squares Discriminant Analysis (PLS-DA) to classify hybrid maize seeds achieving a recognition accuracy of 84.4% to 100%. Furthermore, the result are also close to those obtained by Zhang *et al.*(2022) who applied PLS-DA model using transmittance hyperspectral imaging, achieving high accuracy in detecting impurities in hybrid wheat seeds, with accuracy rates of 95.69%, 98.25%, and 97.25% for different hybrid varieties, effectively distinguishing them from female parent seeds. Compared to this result, Zhang *et al.*(2022) achieved better result when detecting seed purity using hyperspectral imaging and partial least squares-discriminant analysis (PLS-DA), rather than NIR spectroscopy and LDA analysis. It highlights the effectiveness of transmittance spectra in distinguishing hybrid and female parent seeds. The accuracy rates in the testing sets of three hybrid wheat varieties were 95.69%, 98.25%, and 97.25% respectively. study was carried out to improve the semen quality of rabbits by removing dead, immotile and morphologically abnormal sperm by filtering ejaculated extended semen through five different filters Sephadex-G15, Albumin, Cotton, Synthetic Fiber, Sand and Sperm Swim-up procedure.

CONCLUSION

The findings of this study highlight the effectiveness of near-infrared spectroscopy (NIR) as a rapid, non-destructive, and reliable analytical tool for assessing seed quality and detecting impurities in faba bean seeds. Traditional methods of impurity detection, though precise, are often limited by labor intensity and time consumption. In contrast, NIR, combined with advanced multivariate techniques such as Partial Least Squares (PLS)and Principal Component Analysis (PCA) has demonstrated strong potential for accurately predicting compositional attributes, identifying foreign materials. The use of chemometric modeling has enhanced the robustness of calibration models and improved the classification accuracy for impurity detection, even under diverse genotypes and environmental conditions. Furthermore, cross-validation techniques have validated the predictive reliability of these models, making them applicable for large-scale screening in breeding programs and quality control. Overall, the integration of NIR spectroscopy with multivariate analysis represents a valuable advancement

for seed science and food safety. The methodology is not only scalable and efficient but also adaptable for use across various legumes and cereal crops. Future studies are recommended to explore the combination of NIR with complementary spectroscopic methods, such as MIR and FTIR, to further expand the range and precision of impurity detection in seed systems.

REFERENCES

Bhatti, S. K. (2018). Using Near-Infrared Reflectance Spectroscopy (NIRS) for Qualitative determination of undesirable chemical component of high nitrogen content in protein raw material used for fish feed (Master's thesis, The University of Bergen).

Boccia, P., Meconi, C., Mecozzi, M., & Sturchio, E. (2013). Molecular modifications induced by inorganic arsenic in *Vicia faba* investigated by FTIR, FTNIR spectroscopy and genotoxicity testing. *Journal of Toxicology and Environmental Health, Part A, 76*(4-5), 281-290.

Carbas, B., Machado, N., Oppolzer, D., Ferreira, L., Brites, C., Rosa, E. A., & Barros, A. I. (2020). Comparison of near-infrared (NIR) and mid-infrared (MIR) spectroscopy for the determination of nutritional and antinutritional parameters in common beans. *Food Chemistry, 306*, 125509.

Chen, M., Song, J., He, H., Yu, Y., Wang, R., Huang, Y., & Li, Z. (2024). Quantitative analysis of high-price rice adulteration based on near-infrared spectroscopy combined with chemometrics. *Foods, 13*(20), 3241.

FAO. (2018). *Seeds toolkit*. In Food and Agriculture Organization of the United Nations (Issue November).

FAO Statistical Programme of Work 2020–2021. (2020). *FAO Statistical Programme of Work 2020–2021*.

Fortier, C., Rodgers, J., Cintron, M. S., Cui, X., & Foulk, J. (2010, January). Preliminary studies of cotton non-lint content identification by near-infrared spectroscopy. In *Proc. Beltwide Cotton Conf., New Orleans, LA* (pp. 4-7).

Han, X., et al. (2023). JsiPLS model for impurity detection in seed cotton. *Infrared Physics & Technology, 134*, 104732.

Haughey, S. A., Graham, S. F., Cancouët, E., & Elliott, C. T. (2013). The application of near-infrared reflectance spectroscopy (NIRS) to detect melamine adulteration of soya bean meal. *Food Chemistry, 136*(3-4), 1557-1561.

Hernández-Hernández, C., Fernández-Cabanás, V. M., Rodríguez-Gutiérrez, G., Fernández-Prior, Á., & Morales-Sillero, A. (2022). Rapid screening of unground cocoa beans based on their content of bioactive compounds by NIR spectroscopy. *Food Control, 131*, 108347.

Johnson, J. B., Walsh, K., & Naiker, M. (2020). Application of infrared spectroscopy for the prediction of nutritional content and quality assessment of faba bean (*Vicia faba* L.). *Legume Science, 2*(3), e40.

Johnson, J. B., Walsh, K. B., & Naiker, M. (2023). Assessment of bioactive compounds in faba bean using infrared spectroscopy. *Legume Science, 5*(4), e203.

Li, D., Yang, W., & Wang, S. (2010). Classification of foreign fibers in cotton lint using machine vision and multi-class support vector machine. *Computers and Electronics in Agriculture, 74*(2), 274-279.

Lippolis, A., Polo, P. V., de Sousa, G., Dechesne, A., Pouvreau, L., & Trindade, L. M. (2024). High-throughput seed quality analysis in faba bean: leveraging Near-InfraRed spectroscopy (NIRS) data and statistical methods. *Food Chemistry: X, 23*, 101583.

Liu, N., Liu, C., Chen, L., Yu, J., Sun, X., Zhang, S., & Wu, J. (2024). Research on model transfer strategies based on the fusion of NIR-MIR spectral data. *Infrared Physics & Technology, 142*, 105545.

Malone, A., Santi, P., Cabana, Y. C., Smith, N. M., Manning, J., Zeballos, E. Z., & Zhou, W. (2022). Cross-validation as a step toward the integration of local and scientific knowledge of geologic hazards in rural Peru. *International Journal of Disaster Risk Reduction, 67*, 102682.

Oliveira, M. M., Ferreira, M. V., Kamruzzaman, M., & Barbin, D. F. (2023). Prediction of impurities in cocoa shell powder using NIR spectroscopy. *Journal of Pharmaceutical and Biomedical Analysis Open, 2*, 100015.

Plans, M., Simó, J., Casañas, F., Sabaté, J., & Rodriguez-Saona, L. (2013). Characterization of common beans (*Phaseolus vulgaris* L.) by infrared spectroscopy: comparison of MIR, FT-NIR and dispersive NIR using portable and benchtop instruments. *Food Research International, 54*(2), 1643-1651.

Rahman, A., & Cho, B. K. (2016). Assessment of seed quality using non-destructive measurement techniques: a review. *Seed Science Research, 26*(4), 285-305.

Reddy, P., Guthridge, K. M., Panozzo, J., Ludlow, E. J., Spangenberg, G. C., & Rochfort, S. J. (2022). Near-infrared hyperspectral imaging pipelines for pasture seed quality evaluation: An overview. *Sensors, 22*(5), 1981.

Tian-xin, L., Shi-qiang, J., Xu, L., Sheng-yi, Z., Hang, R., Yan-lu, Y., & Dong, A. (2015). Maize Hybrid Seed Purity Identification Based on Near Infrared Reflectance (NIR) and Transmittance (NIT) Spectra. *Spectroscopy and Spectral Analysis, 35*(12), 3388-3392.

Williams, P., & Norris, K. (2001). *Near-infrared technology in the agricultural and food industries* (2nd ed.). American Association of Cereal Chemists.

Winkler-Moser, J. K., Singh, M., Rennick, K. A., Bakota, E. L., Jham, G., Liu, S. X., & Vaughn, S. F. (2015). Detection of corn adulteration in Brazilian coffee (*Coffea arabica*) by tocopherol profiling and near-infrared (NIR) spectroscopy. *Journal of Agricultural and Food Chemistry, 63*(49), 10662-10668.

Xue, H., Yang, Y., Xu, X., Zhang, N., & Lv, Y. (2023). Application of near infrared hyperspectral imaging

technology in purity detection of hybrid maize. *Applied Sciences, 13*(6), 3507.

Zhang, H., Hou, Q., Luo, B., Tu, K., Zhao, C., & Sun, Q. (2022). Detection of seed purity of hybrid wheat using reflectance and transmittance hyperspectral imaging technology. *Frontiers in Plant Science, 13*, 1015891.

Zhang, J., Li, M., Pan, T., Yao, L., & Chen, J. (2019). Purity analysis of multi-grain rice seeds with non-destructive visible and near-infrared spectroscopy. *Computers and Electronics in Agriculture, 164*, 104882.

الكشف عن الشوائب في بذور الفول البلدى باستخدام مطياف الأشعة القريبة من تحت الحمراء

دعاء شحاتة، رمضان الجمل، شريف رضوان، جمال المصري

قسم الهندسة الزراعية، كلية الزراعة، جامعة قناة السويس، الإسماعيلية 41522، مصر

المستخلص: يُعد الفول (*Vicia faba* L.) من المحاصيل البقولية ذات القيمة الاقتصادية والغذائية العالية على مستوى العالم، إلا أن تلوث بذوره بالشوائب مثل أجزاء الحشرات، كسور الفلقات، بقايا القشور، والمواد الغريبة الأخرى يؤدي إلى تدهور جودة البذور، وانخفاض معدل الإنبات، مما يتسبب في خسائر اقتصادية كبيرة. ولذا، فإن الحاجة تزداد إلى استخدام تقنيات حديثة تكون سريعة، دقيقة، وغير تدميرية للكشف عن هذه الشوائب وتحسين جودة البذور. تهدف هذه الدراسة إلى تقييم فعالية استخدام مطيافية الأشعة القريبة من تحت الحمراء (NIR) في النطاق الطيفي (400–1000 نانومتر) للكشف عن الشوائب في بذور الفول وتصنيفها. تم تسجيل البصمات الطيفية للبذور النقية والشوائب، وتحليلها باستخدام تقنيات التحليل الإحصائي متعدد المتغيرات، مثل تحليل المكونات الرئيسية(PCA) ، والانحدار باستخدام المربعات الصغرى الجزئية(PLS) ، وتحليل التمييز الخطي (LDA), أظهرت النتائج كفاءة عالية للنماذج المستخدمة؛ حيث بلغ معامل الارتباط في نموذج PLS نحو 0.97 مع قيمة منخفضة لجذر متوسط مربع الخطأ(RMSE ≈ 2.98) ، كما حقق نموذج LDA دقة تصنيف بلغت 84% في التمييز بين البذور النقية والشوائب. تؤكد هذه النتائج إمكانية الاعتماد على تقنيةNIR ، عند دمجها مع أدوات التحليل الكيمومتري، كأداة فعالة وسريعة وغير تدميرية لتقييم جودة بذور الفول والكشف عن الشوائب المصاحبة لها.

الكلمات المفتاحية: فول، نقاوة البذور، مطياف الأشعة القريبة من تحت الحمراء ، شوائب، تحليل متعدد المتغيرات.