Detection of Saffron Adulteration with *Crocus sativus* Style using NIR-Hyperspectral Imaging and Chemometrics

Derick Malavi [1, 2], Amin Nikkhah [1, 2, 3], Pejman Alighaleh [4], Soodabeh Einafshar [5], Katleen Raes [1], Sam Van Haute [1, 2, *]

[1] Department of Food Technology, Safety and Health, Faculty of Bioscience Engineering, Ghent University, Coupure Links 653, 9000 Ghent, Belgium
[2] Food Chemistry and Technology Research Centre, Department of Molecular Biotechnology, Environmental Technology, and Food Technology, Ghent University Global Campus, 119, Songdomunhwa-Ro, Yeonsu-Gu, Incheon, South Korea, 21985
[3] Friedman School of Nutrition Science and Policy, Tufts University, Boston, MA, USA
[4] Department of Biosystems Engineering, Ferdowsi University of Mashhad, Iran
[5] Department of Agricultural Engineering Institute, Khorasan Razavi Agricultural and Natural Resources Research and Education Center, AREEO, Mashhad, Iran
* Correspondence: Sam.VanHaute@ghent.ac.kr; Tel.: +82 (0) 32 6264212

Abstract

Saffron is a valuable spice that is often adulterated. This study proposes using near-infrared hyperspectral imaging (NIR-HSI) and chemometrics as a fast and cost-effective method for detecting and quantifying adulteration in saffron stigmas. Adulterated saffron samples were prepared by adding *Crocus sativus* style to pure saffron stigmas in varying concentrations (20-90%). The spectral data were pre-treated using standard normal variate (SNV), and multiplicative scatter correction (MSC), while variable reduction was performed by Principal Component Analysis (PCA) and Partial Least Squares (PLS). Classification was done using Linear Discriminant Analysis (LDA), PLS-DA, Support Vector Machine (SVM), and Multi-layer Perceptron (MLP) models, while quantification was achieved by PLS, PCA, SVM, and MLP-based regression models. The HSI technique achieved correct classification rates of 95.6% to 100% in discriminating authentic saffron from plant adulterants and adulterated saffron across all the models. Regression models to quantify the percentage style adulteration in saffron demonstrated excellent prediction abilities with almost all models achieving RPD (Residual Predictive Deviation) values of 3.0-5.4. The MLP model (1 hidden layer with 3 neurons) built from SNV pre-processed and PLS reduced data (15 LVs), showed exceptional predictive capabilities, with an R^2p of 0.97, a Root Mean Squared Error of Prediction (RMSEP) of 4.3%, and an RPD of 5.4. The results demonstrate the potential of NIR-HSI and chemometrics for rapid and nondestructive detection and quantification of style in saffron stigmas.

Keywords: Food fraud; Authentication; Machine learning; Variable Reduction; Spectral Preprocessing

1 Introduction

Saffron, commonly known as 'Red Gold,' is the most expensive spice globally. It is derived from the dehydrated stigma of *Crocus sativus* L., a sterile triploid plant from the Iridaceae family. It is cultivated worldwide, with an estimated global production of 300 tons annually (Morozzi et al., 2019). Iran dominates as the largest producer, contributing to 76% of the annual saffron output. Other notable producers include Spain, Greece, Italy, Turkey, and India (Kumar et al., 2009; Hagh-Nazari & Keifi, 2007). The cultivation and harvesting of saffron are meticulous and labor-intensive, relying on manual methods, which result in relatively low yields. This, combined with its distinctive sensory properties, significantly contributes to its premium market price, which often exceeds $2,000 per kilogram (Shahnoushi et al., 2020).

Saffron is renowned for its peculiar quality, exceptional sensory attributes, and biological properties. It imparts a unique color, flavor, and aroma to foods through its secondary metabolites: crocetin, picrocrocin, and safranal (Moratalla-López et al., 2019). Crocins (trans-crocetin ester, di-β-gentiobiosyl) and crocetin sugar esters (8, 8'-diapocarotene-8, and 8'-dioic acid) are water-soluble apocarotenoids that produce yellow color hues in food. The bitter taste of the spice is derived from picrocrocin, a monoterpene glucoside [4-(-D-glucopyranosyloxy)-2,6,6-trimethyl-1-cyclohexane-1-carboxaldehyde], while its aroma is derived from the volatile fraction, safranal (2,6,6-trimethyl-1,3-cyclohexadiene-1-carboxaldehyde) (Kyriakoudi et al., 2015).

Unfortunately, saffron is frequently subject to adulteration, primarily driven by its high production costs, premium price, and perceived value (Kumar et al., 2009; Mohiuddin, 2019; Moore et al., 2012). Adulteration poses significant challenges for both consumers and traders and can manifest in various forms. These include incorporating other parts of the saffron flower, substituting it with visually similar plant materials, adding artificial colors, immersing fibers in honey or oils to increase mass, and illicitly mixing cheaper saffron categories with premium ones (Asili & Jaroszewski, 2010; Guijarro-díez et al., 2017; Jiang et al., 2014; Karimi et al., 2016; Moore et al., 2012; Moras et al., 2018; Hagh-Nazari & Keifi, 2007; Sabatino et al., 2011). These fraudulent practices are employed to enhance the spice's appearance, weight, and volume for financial gain, emphasizing the need for caution when purchasing saffron. Assessing the quality and origin of saffron presents challenges, as its appearance can be deceptively similar, and market grading is often inconsistent (Lu et al., 2019). This inconsistency further compounds the difficulty in identifying genuine saffron, making it imperative for consumers and traders to exercise caution when making purchases.

Consequently, the issue of detecting and authenticating adulterants in saffron has gained the attention of traders, consumers, and researchers. It is imperative that saffron authentication and quality control methods be continually developed, reviewed, improved, and implemented to classify grades accurately in international markets and safeguard the interests of producers and consumers (Heidarbeigi et al., 2015). This collective effort is vital for ensuring the integrity of the saffron trade and maintaining trust within the industry.

Currently, the quality of saffron in international trade is assessed according to ISO 3632-2, which specifies UV-Vis spectrophotometric method for evaluating key quality parameters including color, taste, and aroma, all of which are associated with the concentrations of crocetin esters, picrocrocin, and safranal. These parameters are quantified at specific wavelengths: 440 nm, 257 nm, and 330 nm, respectively(ISO 3632, 2011). Nevertheless, efforts are ongoing to develop more efficient means of assessing saffron quality

and authenticity. In this pursuit, various approaches such as molecular assays (Jiang et al., 2014; Soffritti et al., 2016), stable isotope analysis (Maggi et al., 2011; Wakefield et al., 2019), and chromatography (Bononi et al., 2015; Bouhadida et al., 2017; García-rodríguez et al., 2017; Guijarro-Díez et al., 2017; Morozzi et al., 2019; Rubert et al., 2016) have been employed to authenticate saffron and detect adulterants of plant origin. However, these methods, despite their accuracy and sensitivity, have drawbacks including being destructive, time-consuming, and requiring specialized expertise.

Spectroscopy has emerged as a more rapid, time-saving, and cost-effective method for detecting fraud in premium spices, with saffron being a notable example. Several infrared spectroscopic techniques and chemometrics have been employed to authenticate, detect, and quantify adulteration in saffron (Dowlatabadi et al., 2017; Karimi et al., 2016; Ordoudi et al., 2014; Shawky et al., 2020; Varliklioz Er et al., 2017). These methods have also proven instrumental in elucidating the chemical composition of saffron and discerning its geographical origin (Biancolillo et al., 2020; D'Archivio & Maggi, 2017; Li et al., 2018a; Liu et al., 2018; Zalacain et al., 2005).

Spectroscopic methods, despite relying on the precision and accuracy of the reference method, particularly in the case of NIR spectroscopy, are known for their superior reproducibility in measurements and predictions. However, spectroscopic techniques inherently lack the capability to provide spatial information about objects, rendering them less efficient when applied to heterogeneous food samples. (Huang et al., 2014; Prieto et al., 2009). For instance, NIR spectroscopy generates a mean spectrum, representing an averaged measurement of the entire sample, irrespective of the scanned area. Consequently, this results in the loss of spatial information regarding the distribution of constituents within the sample (Manley, 2014). Other methods, such as computer vision (Alighaleh et al., 2022; Kiani & Minaei, 2016), are limited to working with visible bands (RGB) and cannot provide information on spectral characteristics (Kiani et al., 2018). In response to these challenges, alternative techniques, such as hyperspectral imaging, have gained attention as promising tools for detecting fraud in food spices (Cruz-Tirado et al., 2023; Hashemi-Nasab & Parastar, 2022; Lima et al., 2020; Orrillo et al., 2019).

Hyperspectral imaging (HSI) offers a potential approach to overcome the limitations of traditional spectroscopic methods in detecting fraud and authenticating saffron. HSI technology seamlessly integrates conventional imaging and spectroscopy, thereby providing both spatial (localization) and spectral (identification) information on samples across the ultraviolet-visible (UV-Vis) and near-infrared (NIR) regions (ElMasry & Sun, 2010; Huang et al., 2014; Manley, 2014). With HSI, the spectrum of each pixel within the image is captured, a unique capability that enables HSI to potentially map out the distribution of constituents within a given sample. This makes it particularly well-suited for analyzing heterogeneous samples (Nobari Moghaddam et al., 2022).

While HSI technology has been explored for the authentication of spices such as nutmeg (Kiani et al., 2019), cumin powder (Florián-Huamán et al., 2022), cinnamon (Cruz-Tirado et al., 2023), and black pepper (Orrillo et al., 2019), research on its application for authenticating saffron stigmas is limited. In a study by Lu et al. (2019), hyperspectral imaging and multivariate spectral analysis were successfully employed to determine the authenticity, quality, and origin of different samples of saffron. The HSI spectral data and propagation neural network models could distinguish authentic saffron from saffron adulterated with safflower. In a recent study by Hashemi-Nasab & Parastar (2022), authentication of saffron using Vis-NIR (400–950 nm) HSI combined with mean-field independent component analysis was also reported to be successful.

Adulteration in saffron is a well-known problem, as other parts of the flower, such as the style that links the three stigmas to the rest of the plant, are often blended in due to their similar appearance. Using near-infrared hyperspectral imaging (NIR-HSI) to detect and quantify the presence of style as an adulterant in saffron stigmas is a promising solution. However, there is currently insufficient evidence to fully support its effectiveness. Furthermore, except for safflower, there is little information on using NIR-HSI to distinguish saffron from other plant-based adulterants. To address the specific research gap concerning the presence of style as an adulterant in saffron stigmas, the current study proposes employing HSI and chemometrics to rapidly authenticate and quantify this form of adulteration.

Therefore, the objective of this study were: (i) to assess the efficacy of spectral preprocessing methods and variable reduction techniques in enhancing model performance, (ii) to explore the potential of chemometric models, including Principal Component Analysis (PCA), Linear Discriminant Analysis (LDA), Partial Least Squares-Discriminant Analysis (PLS-DA), Support Vector Machines (SVM), and Multi-Layer Perceptron (MLP), when utilizing hyperspectral imaging (HSI) for the rapid authentication and classification of saffron, its adulterants, and adulterated saffron, and (iii) to develop predictive models, using HSI and regression techniques based on PLS, PCA, SVM, and MLP, to estimate the concentration of the style (adulterant) within saffron stigmas.

2 Materials and Methods

2.1 Plant materials

The saffron samples, which had undergone microwave drying, and the adulterants (safflower, saffron style, fiber, Citrus aurantium, and a mixture of style and konj) (**Figure 1**) used in this study were sourced from Zaveh County in Khorasan Razavi province, Iran. These specific samples were chosen to represent commonly encountered plant materials that are often fraudulently marketed as genuine saffron. To maintain their integrity, all samples were carefully stored at 25 °C, shielded from light and moisture, until analysis.

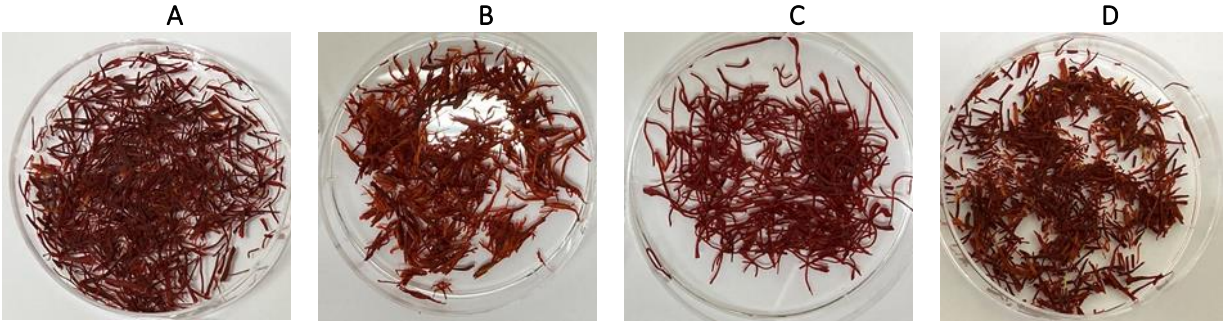


Figure 1*. RGB images of A: Authentic saffron stigmas, B: Saffron Styles; C: Style + Konj, and D: Adulterated Saffron stigmas (20% saffron stigmas + 80% saffron styles)*

2.2 Preparation of adulteration mixtures

Saffron stigmas were mixed with style to create adulterated mixtures. Admixtures were made by adding various levels of style to saffron stigmas at eight different percentages: 20%, 30%, 40%, 50%, 60%, 70%, 80%, and 90% of the total weight of the sample. Notably, 20% was selected as the lowest level of adulteration, in alignment with the market principle that suggests adding less than 20% of adulterants to

saffron is not economically feasible (Hashemi-Nasab & Parastar, 2022; Petrakis et al., 2015). A total of 84 samples, including 6 pure saffron, 30 adulterants (6 samples each for safflower, *Citrus aurantium*, saffron-style, fiber, and konj mix), and 48 adulterated saffron samples (with adulterations ranging from 20% to 90%), were prepared for hyperspectral image analysis.

2.3 Hyperspectral imaging system

A hyperspectral imaging system was employed for data acquisition. This system operated in the near-infrared region, spanning from 900 to 1700 nm, and consisted of a computer, Fx17e Specim spectral camera, light sources, an InGAs detector, and an electric displacement platform measuring 40 cm by 20 cm (Specim Lab Scanner, Spectral Imaging Oy Ltd, Finland). The acquisition, normalization, and processing of hyperspectral images were performed using Lumo Scanner, Classic ENVI (IDL 8.7.2), and ENVI (version 5.5.2) software, respectively.

2.4 Acquisition of HSI images and spectral data

The samples were evenly spread on a 60 mm diameter dish. Subsequently, the images were acquired for each sample using hyperspectral imaging. The optimal parameters for acquiring clear and undistorted spectral images were determined through repeated experiments. These parameters included a camera exposure time of 7.00 ms, a frame rate of 19.50 Hz, a platform speed of 2.6 mm/s, a platform distance of 40 cm, a distance between the sample and the lens of 15 cm, and an image resolution of 672 x 512 pixels. The operation was performed in a dark chamber to eliminate external light interference. The sample was positioned on the platform before capturing the image across the NIR spectral range from 900 to 1700 nm, resulting in an average spectral interval of 3.5 nm, producing 224 reflectance bands. The hyperspectral images were calibrated using black and white reference images to remove noise caused by uneven dark and illumination distribution by the hyperspectral camera, as per equation 1.

$$R = (I - B)/(W - B) \quad (1)$$

The calibrated hyperspectral image of the sample (R) was calculated using the raw hyperspectral image of the sample (I), the white reference image from the standard white calibration board (W, ~ 99.9% reflectance), and the black reference image obtained by closing the lens of the camera (B, ~ 0% reflectance). The images were masked to eliminate the background and other redundant parts due to Petri dishes. The average reflectance value was obtained by selecting the whole image as the region of interest (ROI).

2.5 Spectral processing

Hyperspectral imaging generates a hypercube that contains a large number of variables. Chemometrics is essential for extracting and interpreting relevant information from this spectral data (Nobari Moghaddam et al., 2022). Nonetheless, the original spectra obtained from hyperspectral imaging can be affected by issues such as scattering effects, random noise, and system noise, all of which can weaken the spectral signal and reduce the performance of models. Solid samples often feature non-uniform surfaces and can scatter light during diffuse reflection, resulting in additive and multiplicative effects. These light scattering effects can be corrected by the Standard Normal Variate (SNV) and Multiplicative Scatter Correction (MSC) algorithms (Kiani et al., 2019).

In this study, the SNV and MSC filters were applied individually to the HSI spectra. By employing the SNV and MSC algorithms, which are chemometric techniques, the quality of the spectral data can be improved, thereby enhancing the performance of classification and regression models.

2.6 Statistical Modeling

The samples were randomly divided into two sets using a train-test split method to facilitate model training and testing. Specifically, the calibration model was constructed using 70% of the total samples (n = 61 for classification and n = 44 for regression models). The remaining 30% of the samples (n = 23 for classification and n = 16 for regression models) were set aside for evaluating the model's accuracy in predicting the characteristics of randomly selected samples, serving as the test set. It's important to note that classification models were assessed using all available samples (n = 84), while the regression models were specifically evaluated using authentic saffron samples and saffron adulterated with saffron-style (n = 60).

Validation processes are essential for building chemometric models (Nicolaï et al., 2007). In our present study, all the calibration models were constructed using 10-fold cross-validation and ten repetitions, with samples randomly assigned without replacement. The calibration set samples were divided into 10 folds, employing 9 folds (k-1) for model training, and reserving 1-fold for cross-validation. This cross-validation process was iterated 10 times. Its purpose was twofold: (i) allow estimation of the number of latent variables (LVs)/factors for optimal models, which require estimation of the standard error of prediction (SE), and (ii) improve the estimation of the prediction error classification and regression models. The optimal-performing model was chosen based on the fewest factors possible factors and minimizing the error for k-fold cross-validation to prevent overfitting. The cross-validated model was then used to predict the external test set.

The process of extracting chemical information from spectral data presents challenges, primarily due to many input variables (NIR wavelengths) that exhibit multi-collinearity. This scenario is not suitable for traditional Multilinear Regression (MLR). To address this issue, statistical methods that incorporate latent variables (LVs)/factors, such as Principal Component Analysis (PCA) and Partial Least Squares (PLS), are employed for dimension reduction. Striking a balance between the number of components used is crucial since too few components may lead to the loss of valuable information (underfitting), while an excessive number of factors may introduce noise (overfitting) (Leardi, 2018).

Our study employed a variety of chemometric techniques to classify and quantify adulteration in saffron stigmas. The analysis was initiated by Principal Component Analysis (PCA) for data exploration. Supervised classification was performed using Linear Discriminant Analysis (LDA), Partial Least Square-Discriminant Analysis (PLS-DA), Support Vector Machines (SVM), and Multi-Layer Perceptron (MLP). Additionally, regression models were developed employing PCA, PLS, SVM, and MLP to predict the extent of adulteration in saffron stigmas. These techniques are fundamental tools in multivariate data analysis for food fraud detection (Nobari Moghaddam et al., 2022). It is common practice to employ multiple classification techniques and assess their performance, as highlighted by Callao & Ruisánchez, (2018). This approach does not significantly increase the overall research expenses once the research problem is well defined.

2.6.1 *Principal Component Analysis (PCA) and Principal Component Regression (PCR)*

PCA, a valuable data projection technique for sample clustering, feature selection and dimension reduction (Nobari Moghaddam et al., 2022), involves transforming the original variables into a new collection of uncorrelated variables known as principal components (PCs). These PCs capture the maximum variation in the data (Härdle & Simar, 2013). PCR takes this concept a step further by reducing the number of predictor variables, using the first few PCs instead of the original variables. This approach simplifies the regression model, enhances its interpretability, and helps prevent overfitting (Leardi, 2018).

To determine the optimal number of PCs for constructing the PCR model, the *Forward selection* approach was employed. In this technique, the regression equation is initially built with one predictor (PC), and each new predictor is individually added to the model until there is no further improvement in the error, i.e., RMSEP (Root Mean Square Error of Prediction) (Lima et al., 2020). The best model results were obtained by determining the optimum number of factors (PCs) that resulted in the lowest RMSECV.

2.6.2 *Linear Discriminant Analysis (LDA)*

LDA, a statistical method for classification and dimensionality reduction, serves two primary purposes. In classification, LDA identifies linear combinations of features that effectively separate classes within a dataset (Sánchez-López et al., 2016). n dimensionality reduction, it projects high-dimensional data onto a lower-dimensional space while preserving as much class-discriminatory information as possible (Hastie, 2009).

In this study, a combination of PCA and LDA, known as PCA-LDA, was utilized for classification. Initially, PCA was employed to reduce the dimensionality of the data. Subsequently, the optimal number of principal components (PCs) was used to construct the LDA model, determining linear combinations of features referred to as discriminant functions (DFs) that best distinguished the classes. Similarly, the selection of the best LDA model was based on the fewest DFs that resulted in the highest classification accuracy (CC) and minimized standard error (SE).

2.6.3 *Partial Least Squares (PLS)*

In the present study, PLS regression was employed to predict the concentration of style in saffron stigmas. PLS, a statistical technique widely used in chemometrics for regression, proves particularly valuable when dealing with datasets where the number of independent variables (predictors) exceeds the number of samples, as observed in this study. PLS achieves this by generating latent variables that maximize the covariance between predictor variables (spectral data) and response variables (Uncu & Ozen, 2019). By prioritizing latent variables based on their contribution to the predictive quality of the regression model, PLS facilitates the selection of a simplified model without the risk of overfitting.

In addition to PLS regression, a PLS-DA model (Partial Least Squares Discriminant Analysis) was employed to classify authentic saffron from adulterated samples. PLS-DA, a variant of PLS, is specifically designed for classification problems, particularly when dealing with datasets containing numerous predictor variables and a limited number of observations. Unlike PLSR, PLS-DA is suited for cases where the dependent variable is categorical, as exemplified in our study with saffron and plant adulterants.

The optimum number of latent variables were selected based on the "one standard error rule" approach as described by Hastie (2009). This method helps in selecting the best-performing model while simultaneously minimizing the number of latent variables. It takes into account either achieving a low RMSEP or attaining the highest percentage of classification accuracy (CC).

2.6.4 Support Vector Machines (SVM)

In the current study, Support Vector Machines (SVM) as a machine learning algorithm, employed for classification and regression tasks. SVM is renowned for its robustness and its ability to handle high-dimensional data without undue sensitivity. It's particularly well-suited for classification and regression tasks with limited training samples in high-dimensional spaces. This algorithm seeks to find a hyperplane that maximizes the distance between itself and the closest samples from each of the two classes (Deng et al., 2013).

To prevent overfitting, SVM adjusts the classification decision function based on the principle of structural risk minimization, rather than simply minimizing the misclassification error on the training set (Chen et al., 2007). The optimal parameters for the SVM regression model were estimated as per the method outlined by Zeng et al. (2019).

A linear kernel function was chosen for the SVM models. The grid search was performed by varying the epsilon parameter (ε) and the "cost of constraints violation" parameter (C) to determine the best settings for the SVM models. Specifically, the epsilon parameter (ε) varied over a range of 0.9, 0.7, 0.5, 0.3, 0.2, 0.1, 0.05, 0.02, 0.01, 0.005, and 0.001. Similarly, the cost of constraints violation (C) parameter was set at 0.01, 0.1, 0.5, 0.75, 1, 3, 5, 7, 10, 25, and followed by increments of 25 to 300.

The tolerance margin, represented by the symbol ε, allows the training instances in the regression to deviate from the hyperplane by a certain value before a penalty is imposed. A high error tolerance (i.e., a large ε value) can lead to certain data patterns being disregarded and not considered in the model, resulting in underfitting. Conversely, a smaller ε value permits a smaller error tolerance but may result in overfitting.

The C parameter controls the penalty for cases outside the regression tolerance margin established based on the ε value. A high value of C results in cases outside the tolerance margin being heavily penalized, thereby reducing training bias but increasing prediction variance and potentially leading to overfitting. In contrast, low values of C may increase training bias (Zeng et al., 2019).

2.6.5 Multi-layer perceptron (MLP)

A Multilayer Perceptron is a fully connected, feedforward artificial neural network that uses layers of neurons (software nodes) to establish connections between inputs and outputs. It mimics the features of samples and predicts new tasks through training data. MLP is heavily reliant on multiple sets of parameters, such as the input layer, which passes the input vector to the network, the computation layer, and an output layer. These layers work in tandem to produce an output value (Farah et al., 2021).

In our current study, MLP was employed with two activation functions: "tansig" for the hidden layer, and "linear" for the output layer. To train the network, several training cycles (also known as epochs) were conducted, whereby the neural network was trained on all the training data for each cycle. One hidden

layer was used with the number of neurons ranging from 1 to 15. A grid search was performed to determine the optimal number of neurons. Additionally, the number of epochs was increased incrementally by 5 until a total of 200 training epochs were achieved.

2.7 Classification and prediction of adulteration by spectral angle mapper

The Spectral Angle Mapper (SAM) algorithm, integrated within the ENVI IDL software, played a pivotal role in classifying and quantifying adulteration using HSI imaging data. In our approach, the preprocessed HSI images were subjected to SAM for identification of adulterant in saffron stigmas. SAM compares the spectra of an image directly to a known target spectrum, which is commonly referred to as the endmember (Lohumi et al., 2019).

To create a spectral library, random regions of interest, each consisting of 10x10 pixels, were selected from various preprocessed images of pure saffron and saffron-style adulterant. Subsequently, the HSI images from the adulterated samples were fed into the SAM classifier. The algorithm calculated the spectral angle for each adulterant, and rule images were generated for each concentration. In the rule images, a small angle between two spectra indicated high similarity, resulting in darker regions. Conversely, pixels with high angles indicated low similarity and appeared with high intensity in the SAM rule image (Lohumi et al., 2019).

The combined use of R©Colordistance and R©Countcolors packages was then employed to select regions of pure saffron and the adulterant (saffron-style) from the SAM rule image. The process entailed randomly selecting pixels from the mapping image and treating each pixel as a point in a three-dimensional space based on its Red (R), Green (G), and Blue (B) values. The pixels were then divided into two clusters, with red indicating saffron stigmas and blue representing the adulterant. The intensity of each cluster was counted by R©Countcolors (Pessanha et al., 2023).

2.8 Assessment of model performance

A robust classification model was selected based on the average highest correct classification percentage and minimized standard error of the test set. Similarly, the performance of regression models was assessed using different parameters such as the coefficient of determination of prediction (R^2p), the root mean square of prediction (RMSEP), and the residual predictive deviation (RPD) (Shawky et al., 2020). Models with a higher R^2 value (closer to 1) are reliable and accurate for prediction. Contrarily, the lower the RMSEP, the more reliable the model is. Residual Predictive Deviation (RPD), a ratio of the standard measured deviation of the reference data to RMSEP, was also determined to compare the models in the current study with those of other studies. A model with an RPD value of < 2.5 is considered poor and can only be appropriate for very rough screening; 2.5–3.0 illustrates a good predictive model, and values > 3.0 indicate an excellent model for prediction (Florián-Huamán et al., 2022). Additionally, the slope and the intercept values of the fitted line between predicted and measured values were calculated to assess the proximity of the predicted values to the perfect fit line. The standard errors (SE) of all performance parameters for the cross-validation (R^2cv, RMESCV, and RPDcv) were estimated using ten times repeated 10-fold cross-validation.

2.9 Statistical Modelling Software

Spectral preprocessing was performed using Unscrambler X, CAMO Software AS (version 10.4, Oslo, Norway). Classification and prediction models on raw and preprocessed data were constructed and executed in RStudio version 1.4.1106. Data partitions (10 times repeated for 10-fold cross-validation) were done with "createMultiFolds" in the "hsdar" package of RStudio (Lehnert et al., 2022). PLS-DA models were built via the "plsgenomics" package (Boulesteix et al., 2018), PLS via the package "pls" (Liland et al., 2022), PCA and PCR by the "stats" package (R Core Team, 2022) while SVM and MLP models were implemented via the packages "e1071" (Meyer et al., 2022) and "monmlp" (Cannon, 2022), respectively. R packages "countcolors", "colordistance", and "pixmap" were used for analysis of HSI SAM images based on their spatial information (Pessanha et al., 2023).

One-way analysis of variance (ANOVA) was employed to ascertain whether there were any statistically significant differences in the RPD values based on the factors of preprocessing, regression tool, and variable reduction. Subsequently, post hoc analysis was conducted using the Tukey's HSD procedure. Prior to performing ANOVA, the data was tested for normality using the Shapiro–Wilk test and homogeneity of variance using Levene's test. Statistical significance was considered at $p < 0.05$. Tests for normality, homogeneity of variance, and ANOVA were also performed in R using the packages 'stats', 'car', and base R functions.

3 Results and Discussion

3.1 Visual spectral analysis

Figure 2 depicts the average SNV preprocessed spectra of pure saffron, style (adulterant), and adulterated saffron. As Figure 2 indicates, visual examination of the spectral profile makes it difficult to discern genuine saffron from contaminated saffron or saffron styles. It is, therefore, critical to use chemometrics methods to extract relevant information from HSI spectral data.

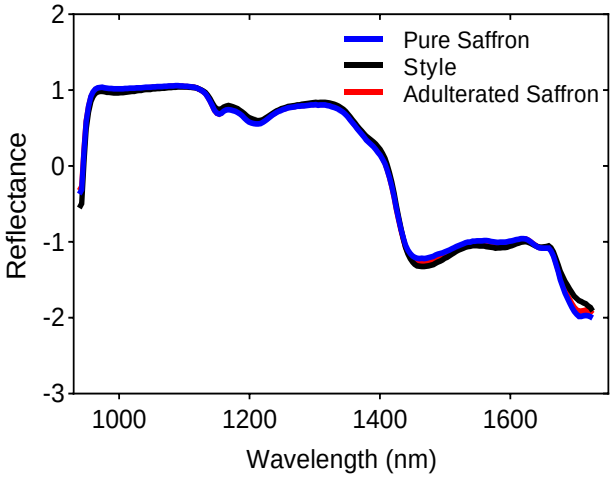


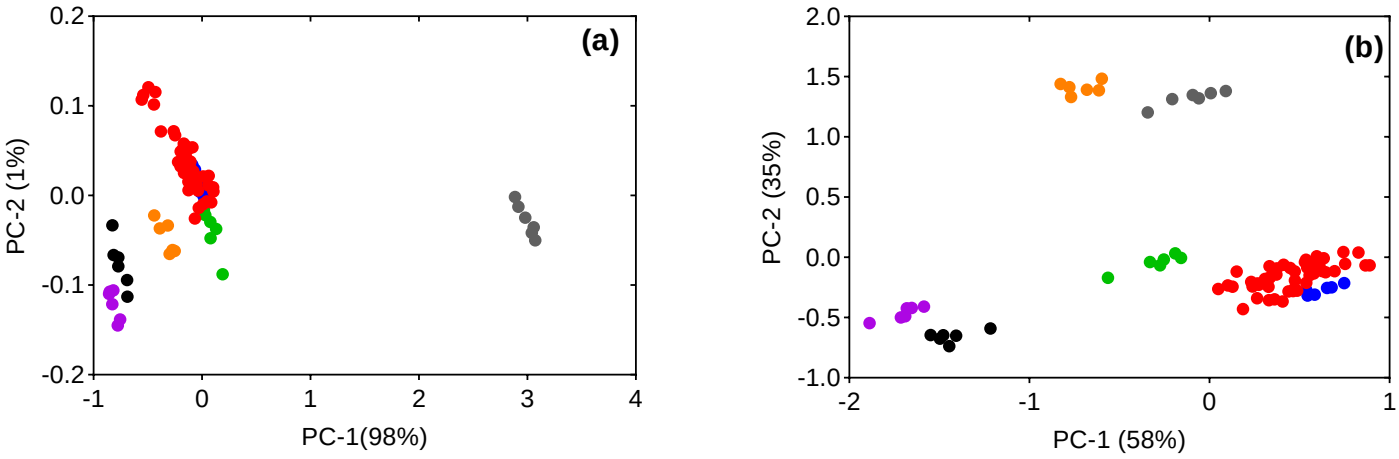
***Figure 2**. Average SNV preprocessed hyperspectral-imaging spectra of pure saffron, style (adulterant) and 90% adulterated saffron*

3.2 Explorative data analysis - Principal Component Analysis (PCA)

Although PCA cannot be used as a classification tool, it remains valuable for visualizing data patterns and clusters in reduced-dimensional spaces. PCA score plots were generated to unveil clusters of authentic saffron, adulterants, and adulterated saffron. The first two principal components, PC1 and PC2, accounted for 99% of the variation based on HSI raw spectra data. However, PCA could not sufficiently separate authentic saffron, style, and adulterated saffron from each other (**Figure 3a**).

To improve separation, MSC and SNV transformations were applied to the raw spectral data to remove artifacts and scattering effects (Dharmawan et al., 2023), as portrayed in the PC plots in **Figures 3b and 3c**. The first two principal components (PC1 & PC2) explained 93% of the variance for MSC and SNV preprocessed HSI data. Both PC score plots revealed clear discrimination between authentic saffron and the other plant adulterants without overlaps along PC1. Additionally, safflower and *Citrus aurantium* were distinguishable from other contaminants and pure saffron by PC2. These distinctions are due to spectral reflectance differences attributed to different chemical compositions among the samples. Although the separation between pure and impure saffron is subtle, it remains discernible.

Identifying the critical wavelengths that contribute to the variation in the data is essential. In this study, the effective wavelengths were identified using x-loadings results from PCA analysis. These wavelengths are concentrated in specific spectral regions such as 1120-1156 nm, 1200-1227 nm, 1375 nm, 1425 nm, and 1635 nm, corresponding to vibrations of C-H, C-C, and N-H bonds associated with various chemical compounds, including oils, carbohydrates, and proteins in the samples. **Figure 3d** illustrates this information, displaying the spectral regions that contributed to the separation of grouped samples.



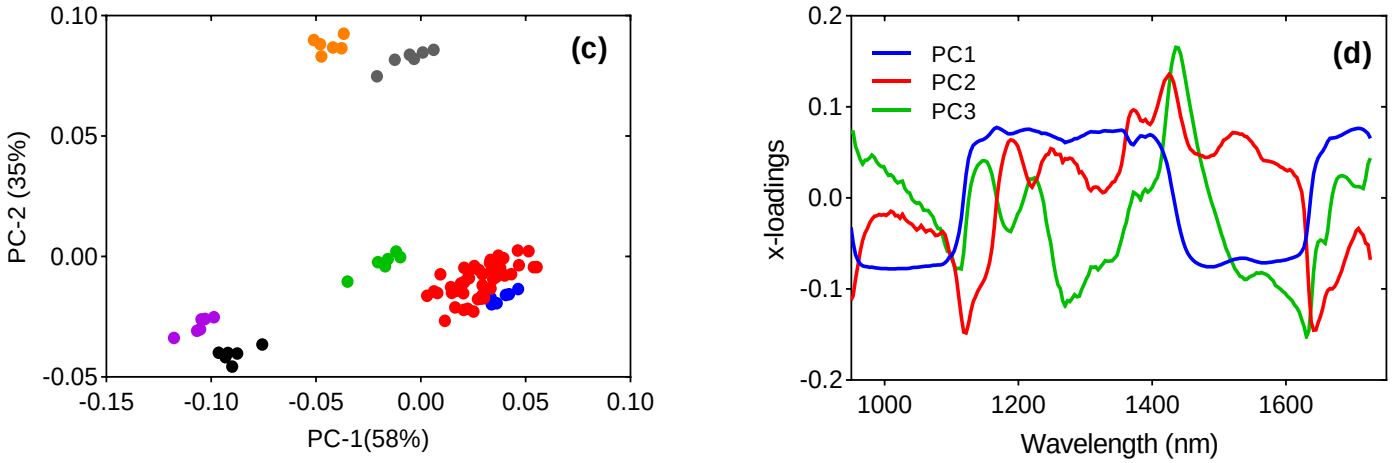


Figure 3. *Principal Component Analysis scores scatter plot for the first and second principal components (PCs). The PC scores display significant separation between saffron and other plant adulterants.* ***(a)*** *Unprocessed/raw spectra;* ***(b)*** *SNV pre-processed spectra* ***(c)*** *MSC pre-processed spectra;* ***(d)*** *PC x-loadings showing major wavelengths contributing to the separation of authentic saffron, adulterants and adulterated saffron on the PC plot. Groups on the PC plots: ● Pure Saffron; ● Adulterated saffron; ● Safflower; ● Style; ● Fiber; ● Citrus aurantium; ● Style + Konj*

According to Luo et al. (2021), agro-foods possess a complex chemical composition, leading to spectra characterized by numerous strong absorption bands originating from various constituents. Additionally, a significant number of overlapping absorption peaks exist, which presents a challenge in directly associating spectral features with individual chemical components. In the current study, specific spectral characteristics have been identified. The peaks within the wavelength ranges of 1120–1156 nm and 1200–1227 nm are attributed to the relative proportions of C-H bonds and the second overtone of C-H stretching. These features are primarily influenced by essential oils (Florián-Huamán et al., 2022). The peak at 1375 nm, associated with the second aromatic C-H combination band for CH_3, could potentially be influenced by crocins and carotenoids in saffron stigmas.

Typically, the spectral region around 1400 nm is related to the O-H first overtone of water. However, it's important to note that the moisture content of saffron is less than 12%. Therefore, the band at 1425 nm can be attributed to compounds such as cellulose and organic acids. Finally, the peak at around 1635 nm can be associated with the first overtones of C-H stretching and N-H bonds in flavones and proteins, as demonstrated in previous studies (Castro et al., 2021; Li et al., 2018).

3.3 Discrimination of genuine saffron, plant adulterants and adulterated saffron

Our study aimed to address the crucial challenge of detecting adulteration in saffron stigmas, which is vital for ensuring product integrity of this high-value spice. To achieve this objective, a rapid methodology that combines near-infrared (NIR) hyperspectral imaging and chemometrics was developed. The purpose was to discriminate between authentic saffron, plant adulterants (such as safflower, *Citrus aurantium*, saffron style, fiber, style, and konj mix), and saffron samples adulterated with saffron-style. Specifically, the efficacy of different preprocessing techniques and variable reduction methods to enhance the performance of the models were evaluated. The results of the investigation, including the optimized parameters and correct classification rates (%CC) for each model, are presented in **Table 1**.

414 **Table 1**. Average correct classification rates (%CC) for LDA, PLS-DA, MLP and SVM discriminant models for
415 cross-validation and test sets

Spectra Preprocessing	Model	Variable Reduction	Details	%CCcv	%CCp
Unprocessed	LDA	PCA	PCs=3	98.2	95.7
	PLS-DA	PLS	LVs=8	98.7	100
	MLP	none	Iterations=200, nodes=6	97.7	100
	MLP	PCA	Iterations=125, PCs=3, nodes=10	99.2	100
	MLP	PLS	Iterations=100, nodes=9, LVs=8	99.5	100
	SVM	none	kernel=linear, epsilon=0.9, cost=10, SVs=30	96.7	95.7
	SVM	PCA	kernel=linear, epsilon=0.8, cost=1, PCs=12, SVs=47	99.0	100
	SVM	PLS	kernel=linear, epsilon 0.8, cost =10, LVs=8, SVs=39	98.9	100
SNV	LDA	PCA	PCs=4	98.0	100
	PLS-DA	PLS	LVs=5	98.3	100
	MLP	none	Iterations=50, nodes=9	99.3	100
	MLP	PCA	Iterations=125, nodes=6, PCs=6	100	100
	MLP	PLS	Iterations=100, nodes=7, LVs=8	100	100
	SVM	none	kernel=linear, epsilon=0.9, cost=0.05, SVs=30	100	100
	SVM	PCA	kernel=linear, epsilon=0.8, cost=1, PCs=6, SVs=33	100	100
	SVM	PLS	kernel=linear, epsilon 0.8, cost 1, LVs=8, SVs=38	100	100
MSC	LDA	PCA	PCs=4	98.0	100
	PLS-DA	PLS	LVs=5	98.4	100
	MLP	none	Iterations=50, nodes=10	99.5	100
	MLP	PCA	Iterations=75, nodes=8, PCs=6	100	100
	MLP	PLS	Iterations=100, nodes=7, LVs=8	100	100
	SVM	none	kernel=linear, epsilon=0.9, cost=0.05, SVs=29	100	100
	SVM	PCA	kernel=linear, epsilon =0.8, cost=1, PCs=6, SVs=33	100	100
	SVM	PLS	kernel=linear, epsilon 0.8, cost =1, LVs=8, SVs=38	100	100

416 *The values represent average rate of correct classification (%CC) ± SE (Standard Error) of prediction; SNV = Standard*
417 *Normal Variate; MSC = Multiplicative Scatter Correction; PCA = Principal Component Analysis; LDA = Linear*
418 *Discriminant Analysis; PLS-DA = Partial Least Squares-discriminant analysis; SVM = Support Vector Machines; MLP =*
419 *Multilayer Perceptron; PCs = Principal Components; LVs = Latent Variables, SVs = Support Vectors; %CCcv = correct*
420 *classification for cross-validation; %CCp = correct classification for the prediction/test set*

421 Based on the findings presented in **Table 1**, it is evident that all the tested models consistently achieved
422 impressive classification accuracy rates, ranging from 95.7% to a perfect 100%. Various discriminant
423 models, including LDA, PLS-DA, MLP, and SVM, consistently demonstrate strong performance under
424 different preprocessing and variable reduction strategies. LDA consistently achieves %CCcv results of 98%
425 and %CCp of 95.7% to 100%. PLS-DA attains a perfect %CCp of 100% even with unprocessed HSI spectra
426 data. MLP excels with %CCcv and %CCp often reaching 100%, especially when appropriate preprocessing
427 and variable reduction methods are applied. SVM, is particularly effective when combined with PCA or PLS
428 for variable reduction, consistently achieves %CCcv and %CCp values of 100%. These results collectively
429 highlight the robustness of these discriminant models across different data processing approaches.

430 However, while achieving impressive classification accuracies, the data reveals occasional
431 misclassifications, particularly in cases involving unprocessed spectra. Specifically, misclassifications on the
432 test sets were observed with PCA-LDA and SVM-*no-variable-reduction* models using unprocessed spectra.
433 In both cases, one instance of adulterated saffron was falsely classified as authentic saffron. Similarly, their
434 equivalent cross-validation models misclassified adulterated saffron as pure saffron. The SVM model

erroneously classified 10 out of a total of 350 cases, while the PCA-LDA model, on the other hand, incorrectly classified 11 cases out of the same total of 350 cases. These misclassifications may arise from spectral similarities between authentic saffron and adulterated samples, highlighting the models' limitations in discerning these similarities (Mishra et al., 2018).

It was also observed that the *SVM-no-variable-cross-validation* model, when applied to unprocessed spectra, misclassified one-third of style+konj cases (10 cases) as fiber. Such misclassifications may be partially attributed to challenges related to dimensionality reduction techniques. Linear methods used during dimension reduction may not accurately estimate the intrinsic dimensionality of the data, potentially leading to the removal of critical information necessary for precise classification. This limitation likely contributed to the observed misclassifications, particularly in cases involving unprocessed spectra (Luo et al., 2021). These findings accentuate the importance of working with preprocessed data when using SVM models in cases where variable reduction techniques are not employed.

The cross-validated models on unprocessed data generally perform well, achieving %CCcv ranging from 96.7% to 99.5% and a %CCp range of 95.7% to 100%. However, the %CCp for the unprocessed spectra is slightly lower when compared to SNV and MSC preprocessed data, suggesting the potential presence of modelling noise in the unprocessed data. The application of SNV and MSC preprocessing techniques generally improved %CCp, resulting in more consistent model performance. All models using these preprocessing techniques achieved a classification accuracy of 100% on the external test set. SNV and MSC were effective in potentially eliminating random noise and light scattering effects from the HSI spectra signal, ultimately enhancing the performance of classification models. These results match those observed by Li et al. (2018c), who reported exceptional performance for SNV and MSC-NIR preprocessing when geographically classifying saffron using PLS-DA models.

Variable reduction techniques, such as PCA and PLS, were employed to reduce the number of features in the HSI spectral data and enhance the performance of discriminant models. As observed from the results, these techniques yielded parsimonious models, as indicated by a low number of PCs/LVs, especially for the linear discriminant models. PCA, through the retention of optimal number of principal components (PCs), consistently leads to outstanding results, with %CCcv and %CCp often reaching 100%. Similarly, PLS demonstrates exceptional performance when reducing latent variables (LVs), mirroring the success of PCA with %CCcv and %CCp commonly reaching 100%. For instance, MLP models achieve better results %CCcv results with either PCA or PLS reduced data as compared to the full spectra. These results suggest that use of techniques such as PCA for variable reduction improves the performance of classification models such as MLP, as demonstrated by Minaei et al. (2017). These findings also corroborate the results obtained by Dharmawan et al. (2023), which showed the superiority of MLP models integrated by PCA for verifying the origin of Arabica coffee. It can also be observed that combining LDA and PCA as a variable reduction technique resulted in a 100% correct classification rate (CCp) for SNV (3 PCs), and MSC (4 PCs) preprocessed data, as shown in Table 1. Similarly, SVM, when combined with both PCA and PLS, achieved a 100% correct classification (CC) for all the three sets of spectral data (unprocessed, SNV, and MSC). The perfect classification results can be attributed to PCA and PLS's ability to enhance the signal-to-noise ratio, eliminate redundant information (noise) by selecting optimal components/latent variables, and subsequently improve the models' performance (Leardi, 2018).

The current study’s findings demonstrate that all the models accurately classified the data, with most achieving a 100% correct classification rate (CC). This CC range aligns with the average rates reported in

previous successful authentication studies of spices such as saffron, black pepper, cinnamon, and nutmeg, which employed HSI and discriminant models like PLS-DA and SVM (Hashemi-Nasab & Parastar, 2022; Kiani et al., 2019; Lu et al., 2019; Orrillo et al., 2019). Our CC results are also comparable with those of Cruz-Tirado et al. (2023), who reported accuracies of 96.7% for PLS-DA and SVM models in their study using NIR-hyperspectral imaging for cinnamon authentication. Furthermore, our MLP–%CC results closely compare with those of Dharmawan et al. (2023), who reported %CC of 90–100% in authenticating Arabica coffee origins using PCA-MLP models based on visible and shortwave near infrared spectroscopy. This study demonstrates that multiple models and preprocessing methods can achieve high classification accuracy, consistent with the findings of Li et al. (2018).

Furthermore, the SVM models demonstrated robustness and lower sensitivity to variable reduction methods compared to other models. They consistently achieved a perfect classification rate of 100% for both cross-validation and test data with preprocessed spectra, regardless of whether PCA or PLS was used for variable reduction. This resilience can be attributed to SVM's capability to leverage the linear kernel, allowing it to maintain simplicity and interpretability by operating directly in the original feature space while still effectively classifying data with a linear decision boundary. In accordance with the present results, previous studies by Luo et al. (2021) and Mishra et al. (2018) have demonstrated regression techniques, such as SVM, have the capacity to extract more critical information for classifying complex food matrices. It is also interesting to note that SVM models achieve high classification accuracy even with the simple linear kernel in our study. These SVM models were able to maintain excellent performance even with a relatively small cost parameter, which controls the trade-off between maximizing and minimizing classification errors (Zeng et al., 2019).

Overall, the results suggest that using appropriate preprocessing techniques and variable reduction methods can enhance the performance of discriminant models. Additionally, some models, such as MLP and SVM, perform exceptionally well when the right combination of techniques is applied, achieving perfect classification rates. However, it is worth noting that all models achieved high classification rates, suggesting that using NIR-hyperspectral imaging with chemometrics is a promising method for the rapid authentication and detection of adulteration in saffron stigmas.

3.4 Prediction of adulteration levels in saffron stigmas with style based on HSI

The untapped potential of hyperspectral imaging, extending its application beyond mere classification, was also explored in the current study. Specifically, the focus was on harnessing its capabilities to accurately quantify the extent of adulteration in saffron stigmas from contamination by saffron-style (20-90%). Regression models including PCR, PLSR, SVM-R, and MLP-R, were employed to achieve this goal. The efficacy of regression models was evaluated by examining the impact of spectral preprocessing and variable reduction, as shown in **Table 2**. Most models demonstrated excellent prediction abilities with high R^2p (0.90 to 0.97), robust RPD values (3.0–5.4), and low RMSEP (4.26% to 7.26%). Only one model had an RPD value of 2.3, which was an exception.

The impact of spectral preprocessing on models designed specifically for predicting saffron adulteration with *Crocus sativus* style was investigated. Three variants were investigated: the use of Standard Normal Variate (SNV) preprocessing, Multiplicative Scatter Correction (MSC) preprocessing, and the utilization of unprocessed spectra. Contrary to the findings of previous studies by Feng & Sun (2013), Florián-Huamán et al. (2022) and Orrillo et al. (2019), which suggested that preprocessing techniques,

particularly SNV, enhanced model performance by addressing scatter effects in hyperspectral imaging (HSI) spectra, our findings revealed that these techniques did not consistently outperform the use of unprocessed data ($p = 0.684$ and F-value = 0.387). However, the MLP model without variable reduction was the best-performing model in the '*unprocessed*' category, with an RPD value of 3.38 and an RMSEP of 6.81%.

In addition to exploring spectral processing techniques, this study aimed to investigate various regression tools for predicting saffron adulteration, including linear models such as Partial Least Squares (PLS) and Principal Component Regression (PCR), as well as Support Vector Machine (SVM) and Multi-Layer Perceptron (MLP) models. Based on our findings, the choice of regression tool had an insignificant impact on model performance ($p = 0.622$ and F-value = 0.486). Notably, the MLP model that employed SNV preprocessing and PLS for dimension reduction achieved the best results. The optimized MLP model (R^2p = 0.97, RPD = 5.40) utilized 15 PLS latent variables (LVs) as input and featured a hidden layer with 3 nodes, resulting in minimal RMSEP of 4.26%. The exceptional predictive capabilities of the MLP model can be attributed to its versatility in handling both linear and non-linear data relationships. Its adaptive nature allowed it to effectively address the complexity of saffron adulteration detection. Furthermore, the MLP's architecture, specifically, one hidden layer with 3 neurons (Figure 4), was selected through a thorough evaluation process, ultimately yielding the highest level of predictive accuracy (without overfitting) among the models tested. These findings closely relate to those of Basile et al. (2022), who reported better results in the prediction of grape texture using NIR combined with PLS and ANN. Conversely, the poorest-performing model was also an MLP model, but it utilized SNV preprocessing and unprocessed spectral data. This poor performance can be attributed to the model's inability to fully extract meaningful information or remove noise from the full spectral data. These findings agree with those of Dharmawan et al. (2023), who emphasized the importance of compressing a large amount of spectral data into a smaller number of variables before feeding it into an ANN model. Overall, our findings indicate that both linear and non-linear methods can effectively be used to detect the levels of style as an adulterant in saffron stigmas.

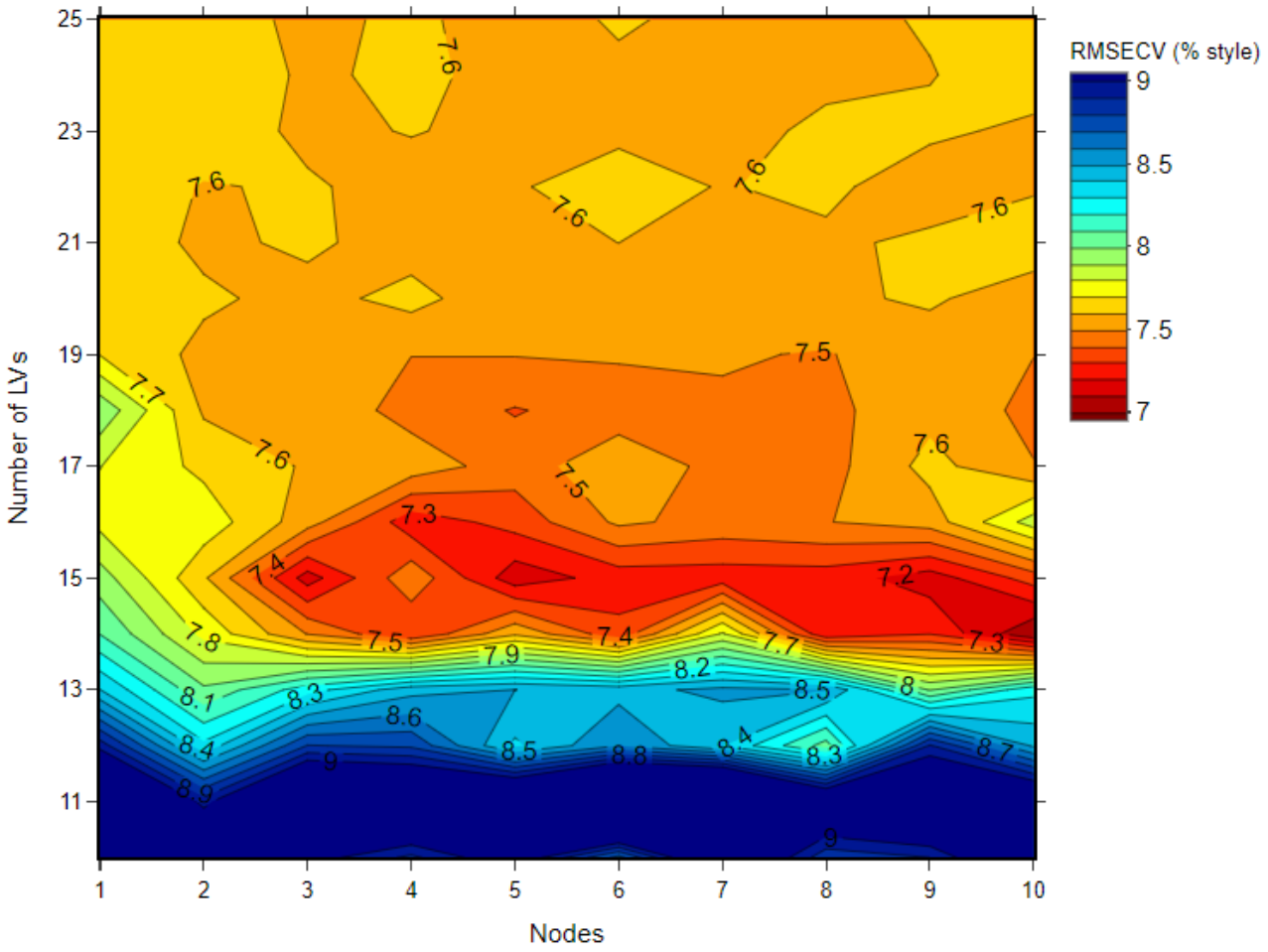


Figure 4. *RMSECV of the percentage style adulteration in saffron samples of the regression model with SNV preprocessed data, PLS dimension reduction and MLP regression, shown in function of number of nodes in the neural network and number of LVs in the PLS dimension reduction.*

Regression models that utilize the entire spectrum may unintentionally include redundant information, such as noise, collinearity, and overfitting, which can diminish their predictive capacity (Orrillo et al., 2019). To enhance predictive capabilities, it is crucial to incorporate variable reduction techniques such as Principal Component Analysis (PCA) and Partial Least Squares (PLS), which effectively eliminate redundant information from spectral data. Our findings demonstrate a significant enhancement in model performance through variable reduction ($p = 1.86\times10^{-4}$, F-value = 13.29). Specifically, PLS outperforms PCA ($p = 1.21\times10^{-3}$) and the "no variable reduction" approach (4.60×10^{-4}). PLS proves to be significantly superior in reducing spectral variables while extracting critical information for subsequent use by regression tools as compared to PCA. There was no significant difference ($p = 0.677$) in performance for models using PCA-reduced variable and full-length spectra variables. While PCA is instrumental for reduction dimensionality, it might fail to capture vital information correlated with the variable being predicted (Dharmawan et al., 2023).

As indicated by our study findings in **Table 2**, the top seven best-performing models were constructed using compressed variables from PLS. The R^2p and RMSEP for these models ranged from 0.94 to 0.97 and

4.26% to 5.57%, respectively, while their RPD values were > 4. On the contrary, the least-performing model, with an RPD value of 2.3, was built using full HSI spectra. Our findings corroborate the results by Basile et al. (2022) that emphasize on the significance of reducing the number of predictors through the PLS technique for development of more precise models and improved performance.

Valuable insights on the effectiveness of variable reduction techniques in enhancing the accuracy and reliability of models used to detect adulteration in saffron stigmas have been demonstrated in this study. Surprisingly, it was found that neither spectral preprocessing nor regression tools influenced the model's performance. Overall, all models except one performed exceptionally well, with PLS-based models achieving higher RPD values (> 4). These findings demonstrate that HSI is a promising approach for detecting and predicting saffron stigma adulteration with style. Our research emphasizes the importance of employing variable reduction techniques, such as PLS, to enhance the development of reliable models for detecting adulteration in saffron stigma.

3.5 Classification and prediction of adulteration based on spectral angle mapper (SAM) and image analysis

As depicted in **Figure 5**, the SAM algorithm demonstrates remarkable potential in distinguishing between genuine saffron stigmas and saffron adulterated with style by utilizing spatial information. It's noteworthy that the pixel color intensity within the images increases with higher concentrations of the adulterant, as expected. Our findings are consistent with those of Lohumi et al. (2019) who reported effective visualization of adulterants in wheat flour using Raman hyperspectral imaging and SAM classification. However, it's essential to acknowledge that no classification algorithm is entirely error-free. In one instance, an authentic saffron sample was misclassified as adulterated, highlighting a limitation of the SAM classifier. This suggests that similarities in certain spectral regions between saffron and the adulterant may result in a reduced spectral angle, leading to misclassifications.

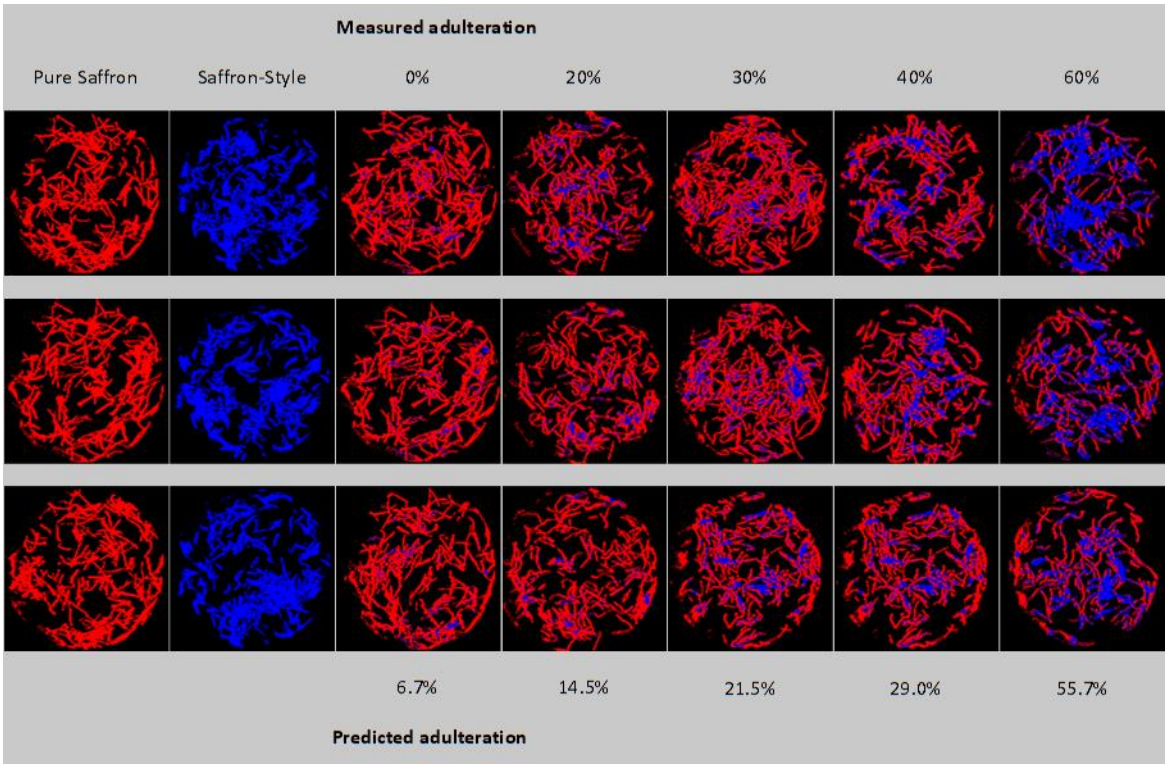


Figure 5. Resultant images for identification of saffron-style (adulterant) pixels based on spectral similarity analyses. Regions with red color intensity represent pixels of saffron stigmas, while the bluish regions indicate pixels of the adulterant (saffron-style).

Further examination of the pixel intensity from the SAM rule images was conducted to predict the levels of adulteration in saffron stigmas, as shown in Figure 5. The findings indicate that, on average, the predicted values tend to be lower compared to the measured values, except in the case of false positives. For example, a measured adulteration level of 20% was predicted as 14.5%, while an actual adulteration level of 30% was predicted as 21.5%. This discrepancy may be attributed to the potential challenges in capturing pixel information from saffron-style strands that might be concealed by saffron stigmas within the mixture. Our findings slightly deviate from those of Cheng et al. (2018), who reported superior performance of PLSR-SAM models in characterizing the degree of myofibrils' cold structural deformation in frozen pork samples using hyperspectral imaging. Therefore, it is imperative for future research endeavors to enhance SAM's performance by incorporating it with other conventional regression algorithms for the precise quantification of adulterants within saffron.

While the SAM method may not be well-suited for precise quantification, it exhibits potential for binary (yes/no) analysis in detecting adulteration. One advantage of this method is its independence from information derived from mixed adulterated samples, which reduces the labor-intensive process of constructing a model using a wide range of adulterated saffron samples. Instead, the development of a SAM model requires two sets of samples: (i) authentic saffron samples and (ii) potential adulterants. However, given the simplicity of the SAM classifier, further research is necessary to explore its applicability in authenticating saffron with other common adulterants, as well as its potential for analyzing other premium food spices. Investigating the SAM method in a broader context will contribute to an enhanced understanding of its capabilities and limitations in the realm of effective food authentication.

4 Conclusion

This study demonstrates the effectiveness of NIR-hyperspectral imaging in tandem with chemometrics for detecting adulteration in saffron stigmas. The high accuracy rates achieved in all models tested, with the majority achieving a perfect classification rate of 100%, provide compelling evidence for the potential of hyperspectral imaging as a rapid method for the authentication and detection of adulteration in saffron. Moreover, the study suggests that the choice of variable reduction and preprocessing techniques potentially affect classification accuracy at the studied adulteration levels in the saffron sample matrix. The study also highlights the resilience and lower sensitivity of SVM models to variable reduction methods, making them a reliable option in such cases.

Additionally, the potential of hyperspectral imaging for predicting the degree of saffron stigma adulteration with style was explored. The results suggest that PLS is the most effective tool for variable reduction. The MLP model coupled with SNV preprocessing and PLS for variable reduction was identified as the best-performing model for predicting adulteration in saffron. Future research with larger sample sizes is needed to validate the findings. Overall, this study provides a foundation for further research to develop hyperspectral imaging-based methods for the authentication and detection of adulteration in saffron and other high-value spices.

CRediT authorship contribution statement

Derick Malavi: Conceptualization, Methodology, Investigation, Formal analysis, Writing-Original Draft, and Visualization. **Amin Nikkah:** Writing-Review & Editing. **Pejman Alighaleh:** Writing-Review & Editing. **Soodabeh Einafshar:** Writing-Review & Editing. **Katleen Raes:** Supervision, Writing-Review & Editing. **Sam Van Haute:** Conceptualization, Formal analysis, Writing-Review & Editing, Supervision, Project administration.

Conflicts of interest

Authors declare no conflict of interest

Acknowledgements

The authors acknowledge the financial support provided by Ghent University Global Campus.

Table 2. Model Parameters for Predicting the Concentration of Saffron-Style in Adulterated Saffron Using NIR-HSI: PLS, PCR, SVM, and MLP Regression Models for cross-validation and test sets

Preprocessing	Regression tool	Variable Reduction	Details	R^2cv	R^2p	RMSECV	RMSEP	RPDcv	RPDp
Unprocessed	Linear	PCA	PCs=35	0.93	0.94	8.90	5.72	3.73	4.02
	Linear	PLS	LVs=14	0.93	0.96	8.71	5.39	3.82	4.27
	MLP	none	Iterations=200, nodes=2	0.84	0.92	13.50	6.81	2.49	3.38
	MLP	PCA	Iterations=10, nodes=4, PCs=18	0.92	0.90	9.53	7.76	3.49	2.97
	MLP	PLS	Iterations=300, nodes=2, LVs=11	0.96	0.92	6.56	6.95	5.09	3.31
	SVM	none	kernel=linear, epsilon=0.1, cost=50, SVs=35	0.93	0.91	8.56	7.12	3.89	3.23
	SVM	PCA	kernel=linear, epsilon=0.1, cost=0.1, PCs=28, SVs=33	0.92	0.91	9.12	7.03	3.65	3.27
	SVM	PLS	kernel=linear, epsilon=0.05, cost=1, LVs=16, SVs=23	0.93	0.96	8.52	5.03	3.90	4.58
SNV	Linear	PCA	PCs=23	0.93	0.93	8.76	7.26	3.80	3.17
	Linear	PLS	LVs=16	0.94	0.96	7.90	4.71	4.23	4.89
	MLP	none	Iterations=100, nodes=3	0.85	0.89	13.10	10.02	2.62	2.30
	MLP	PCA	Iterations=10, nodes=6, PCs=24	0.93	0.93	8.50	6.06	3.92	3.79
	MLP	PLS	Iterations=200, nodes=3, LVs=15	0.95	0.97	7.15	4.26	4.71	5.40
	SVM	none	kernel=linear, epsilon=0.05, cost=0.5, SVs=38	0.93	0.94	8.80	6.27	3.78	3.67
	SVM	PCA	kernel=linear, epsilon=0.1, cost=0.1, PCs=23; SVs=34	0.93	0.94	8.84	7.07	3.76	3.27
	SVM	PLS	kernel=linear, epsilon=0.03, cost=10, LVs=16, SVs=25	0.94	0.97	7.89	4.57	4.24	5.03
MSC	Linear	PCA	PCs=23	0.93	0.93	8.76	7.26	3.80	3.17
	Linear	PLS	LVs=17	0.94	0.95	7.92	5.27	4.21	4.37
	MLP	none	Iterations=100, nodes=11	0.87	0.95	12.10	7.02	2.76	3.28
	MLP	PCA	Iterations=5, nodes=10, PCs=23	0.94	0.95	8.10	5.79	4.10	3.97
	MLP	PLS	Iterations=300, nodes=3, LVs=15	0.95	0.94	7.29	5.85	4.59	3.93
	SVM	none	kernel=linear, epsilon=0.1, cost=0.5, SVs=31	0.93	0.93	8.71	6.84	3.82	3.37
	SVM	PCA	kernel=linear, epsilon=0.1, cost=0.1, PCs=22, SVs=35	0.93	0.95	8.60	7.04	3.87	3.27
	SVM	PLS	kernel=linear, epsilon=0.03, cost=10, LVs=16, SVs=26	0.94	0.94	7.96	5.57	4.19	4.13

SNV = Standard Normal Variate; MSC = Multiplicative Scatter Correction; PLS = Partial Least Squares regression; PCR = Principal Component Regression; SVM = Support Vector Machine; MLP = Multilayer Perceptron; PCs = Principal Components (PCs); LVs = Latent Variables; SVs = Support Vectors; R^2cv = coefficient of determination for cross-validation; R^2p = coefficient of determination for the test set; RMSECV = root mean square error of cross-validation; RMSEP = root mean square of prediction; RPDcv = residual predictive deviation for cross-validation; RPDp = residual predictive deviation for prediction set

References

Alighaleh, P., Khosravi, H., Rohani, A., Saeidirad, M. H., & Einafshar, S. (2022). The detection of saffron adulterants using a deep neural network approach based on RGB images taken under uncontrolled conditions. *Expert Systems with Applications*, *198*(October 2021), 116890. https://doi.org/10.1016/j.eswa.2022.116890

Armona, M. A. C., La, I. N. B., Zquez, Ä., Simidou, M. A. Z. T., Lonso, G. O. L. A., Zalacain, A., Ordoudi, S. A., Díaz-Plaza, E. M., Carmona, M., Blázquez, I., Tsimidou, M. Z., & Alonso, G. L. (2005). Near-infrared spectroscopy in saffron quality control: Determination of chemical composition and geographical origin. *Journal of Agricultural and Food Chemistry*, *53*(24), 9337–9341. https://doi.org/10.1021/jf050846s

Asili, J., & Jaroszewski, J. W. (2010). *H NMR metabolic fingerprinting of saffron extracts*. 511–517. https://doi.org/10.1007/s11306-010-0221-z

Basile, T., Marsico, A. D., & Perniola, R. (2022). Use of Artificial Neural Networks and NIR Spectroscopy for Non-Destructive Grape Texture Prediction. *Foods*, *11*(3). https://doi.org/10.3390/foods11030281

Biancolillo, A., Foschi, M., & D'Archivio, A. A. (2020). Geographical Classification of Italian Saffron (Crocus sativus L.) by Multi-Block Treatments of UV-Vis and IR Spectroscopic Data. *Molecules*, *25*(10). https://doi.org/10.3390/molecules25102332

Bononi, M., Milella, P., & Tateo, F. (2015). Gas chromatography of safranal as preferable method for the commercial grading of saffron (Crocus sativus L.). *Food Chemistry*, *176*, 17–21. https://doi.org/10.1016/j.foodchem.2014.12.047

Bouhadida, N., Garcíá-Rodríguez, M. V., Moratalla-López, N., Salinas, M. R., Oueslati, S., & Alonso, G. L. (2017). Correlation between the high performance liquid chromatography with diode array detection (HPLC-DAD) and the sensory analysis based on saffron color, odor, aroma and bitterness. *Acta Horticulturae*, *1184*, 273–278. https://doi.org/10.17660/ActaHortic.2017.1184.39

Boulesteix, A.-L., Durif, G., Lambert-Lacroix, S., Peyre, J., & Strimmer., K. (2018). *plsgenomics: PLS Analyses for Genomics*. https://cran.r-project.org/package=plsgenomics

Callao, M. P., & Ruisánchez, I. (2018). An overview of multivariate qualitative methods for food fraud detection. *Food Control*, *86*, 283–293. https://doi.org/10.1016/j.foodcont.2017.11.034

Cannon, M. A. J. (2022). *monmlp: Multi-Layer Perceptron Neural Network with Optional Monotonicity Constraints*. https://doi.org/10.1007/11550907

Castro, R. C., Ribeiro, D. S. M., Santos, J. L. M., & Páscoa, R. N. M. J. (2021). Near infrared spectroscopy coupled to MCR-ALS for the identification and quantification of saffron adulterants: Application to complex mixtures. *Food Control*, *123*(November 2020). https://doi.org/10.1016/j.foodcont.2020.107776

Chen, Q., Zhao, J., Fang, C. H., & Wang, D. (2007). Feasibility study on identification of green, black and Oolong teas using near-infrared reflectance spectroscopy based on support vector machine (SVM). *Spectrochimica Acta - Part A: Molecular and Biomolecular Spectroscopy*, *66*(3), 568–574. https://doi.org/10.1016/j.saa.2006.03.038

Cheng, W., Sun, D. W., Pu, H., & Wei, Q. (2018). Characterization of myofibrils cold structural deformation degrees of frozen pork using hyperspectral imaging coupled with spectral angle mapping algorithm. *Food Chemistry*, *239*, 1001–1008. https://doi.org/10.1016/j.foodchem.2017.07.011

Cruz-Tirado, J. P., Lima Brasil, Y., Freitas Lima, A., Alva Pretel, H., Teixeira Godoy, H., Barbin, D., & Siche, R. (2023). Rapid and non-destructive cinnamon authentication by NIR-hyperspectral imaging and classification

chemometrics tools. *Spectrochimica Acta - Part A: Molecular and Biomolecular Spectroscopy*, *289*(November 2022). https://doi.org/10.1016/j.saa.2022.122226

D'Archivio, A. A., & Maggi, M. A. (2017). Geographical identification of saffron (Crocus sativus L.) by linear discriminant analysis applied to the UV–visible spectra of aqueous extracts. *Food Chemistry*, *219*, 408–413. https://doi.org/10.1016/j.foodchem.2016.09.169

Deng, S., Xu, Y., Li, L., Li, X., & He, Y. (2013). A feature-selection algorithm based on Support Vector Machine-Multiclass for hyperspectral visible spectral analysis. *Journal of Food Engineering*, *119*(1), 159–166. https://doi.org/10.1016/j.jfoodeng.2013.05.024

Dharmawan, A., Masithoh, R. E., & Amanah, H. Z. (2023). Development of PCA-MLP Model Based on Visible and Shortwave Near Infrared Spectroscopy for Authenticating Arabica Coffee Origins. *Foods*, *12*(11). https://doi.org/10.3390/foods12112112

Dowlatabadi, R., Farshidfar, F., Zare, Z., Pirali, M., Rabiei, M., Khoshayand, M. R., & Vogel, H. J. (2017). Detection of adulteration in Iranian saffron samples by 1H NMR spectroscopy and multivariate data analysis techniques. *Metabolomics*, *13*(2), 1–11. https://doi.org/10.1007/s11306-016-1155-x

ElMasry, G., & Sun, D. W. (2010). Principles of Hyperspectral Imaging Technology. *Hyperspectral Imaging for Food Quality Analysis and Control*, 3–43. https://doi.org/10.1016/B978-0-12-374753-2.10001-2

Farah, J. S., Cavalcanti, R. N., Guimarães, J. T., Balthazar, C. F., Coimbra, P. T., Pimentel, T. C., Esmerino, E. A., Duarte, M. C. K. H., Freitas, M. Q., Granato, D., Neto, R. P. C., Tavares, M. I. B., Calado, V., Silva, M. C., & Cruz, A. G. (2021). Differential scanning calorimetry coupled with machine learning technique: An effective approach to determine the milk authenticity. *Food Control*, *121*(July 2020). https://doi.org/10.1016/j.foodcont.2020.107585

Feng, Y. Z., & Sun, D. W. (2013). Near-infrared hyperspectral imaging in tandem with partial least squares regression and genetic algorithm for non-destructive determination and visualization of Pseudomonas loads in chicken fillets. *Talanta*, *109*, 74–83. https://doi.org/10.1016/j.talanta.2013.01.057

Florián-Huamán, J., Cruz-Tirado, J. P., Fernandes Barbin, D., & Siche, R. (2022). Detection of nutshells in cumin powder using NIR hyperspectral imaging and chemometrics tools. *Journal of Food Composition and Analysis*, *108*(January). https://doi.org/10.1016/j.jfca.2022.104407

García-rodríguez, M. V., López-córcoles, H., Alonso, G. L., Pappas, C. S., Polissiou, M. G., & Tarantilis, P. A. (2017). Comparative evaluation of an ISO 3632 method and an HPLC-DAD method for safranal quantity determination in saffron. *Food Chemistry*, *221*, 838–843. https://doi.org/10.1016/j.foodchem.2016.11.089

Guijarro-díez, M., Castro-puyana, M., Crego, A. L., & Marina, M. L. (2017). Journal of Food Composition and Analysis Detection of saffron adulteration with gardenia extracts through the determination of geniposide by liquid chromatography – mass spectrometry. *Journal of Food Composition and Analysis*, *55*, 30–37. https://doi.org/10.1016/j.jfca.2016.11.004

Guijarro-Díez, M., Castro-Puyana, M., Crego, A. L., & Marina, M. L. (2017). A novel method for the quality control of saffron through the simultaneous analysis of authenticity and adulteration markers by liquid chromatography-(quadrupole-time of flight)-mass spectrometry. *Food Chemistry*, *228*, 403–410. https://doi.org/10.1016/j.foodchem.2017.02.015

Härdle, W. K., & Simar, L. (2013). Applied multivariate statistical analysis. In *Applied Multivariate Statistical Analysis*. https://doi.org/10.1007/978-3-642-17229-8

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

https://doi.org/10.1016/j.foodchem.2022.133450

Hastie, T. et. all. (2009). Springer Series in Statistics The Elements of Statistical Learning. *The Mathematical Intelligencer*, *27*(2), 83–85. http://www.springerlink.com/index/D7X7KX6772HQ2135.pdf

Heidarbeigi, K., Mohtasebi, S. S., Foroughirad, A., Rafiee, S., & Rezaei, K. (2015). Detection of Adulteration in Saffron Samples Using Electronic Nose Detection of Adulteration in Saffron Samples Using Electronic Nose. *International Journal of Food Properties*, *18*(7), 1391–1401. https://doi.org/10.1080/10942912.2014.915850

Huang, H., Liu, L., & Ngadi, M. O. (2014). *Recent Developments in Hyperspectral Imaging for Assessment of Food Quality and Safety*. 7248–7276. https://doi.org/10.3390/s140407248

Jiang, C., Cao, L., Yuan, Y., Chen, M., Jin, Y., & Huang, L. (2014). Barcoding melting curve analysis for rapid, sensitive, and discriminating authentication of saffron (Crocus sativus L.) from its adulterants. *BioMed Research International*, *2014*. https://doi.org/10.1155/2014/809037

Karimi, S., Feizy, J., Mehrjo, F., & Farrokhnia, M. (2016). Detection and quantification of food colorant adulteration in saffron sample using chemometric analysis of FT-IR spectra. *RSC Advances*, *6*(27), 23085–23093. https://doi.org/10.1039/c5ra25983e

Kiani, S., & Minaei, S. (2016). Potential application of machine vision technology to saffron (Crocus sativus L.) quality characterization. *Food Chemistry*, *212*, 392–394. https://doi.org/10.1016/j.foodchem.2016.04.132

Kiani, S., Ruth, S. M. Van, & Minaei, S. (2018). Hyperspectral imaging , a non-destructive technique in medicinal and aromatic plant products industry : Current status and potential future applications. *Computers and Electronics in Agriculture*, *152*(May), 9–18. https://doi.org/10.1016/j.compag.2018.06.025

Kiani, S., van Ruth, S. M., van Raamsdonk, L. W. D., & Minaei, S. (2019). Hyperspectral imaging as a novel system for the authentication of spices: A nutmeg case study. *Lwt*, *104*(January), 61–69. https://doi.org/10.1016/j.lwt.2019.01.045

Kumar, R., Singh, V., Devi, K., Sharma, M., Singh, M. K., & Ahuja, P. S. (2009). State of art of saffron (Crocus sativus L.) agronomy: A comprehensive review. In *Food Reviews International* (Vol. 25, Issue 1). https://doi.org/10.1080/87559120802458503

Kyriakoudi, A., Ordoudi, S. A., & Tsimidou, M. Z. (2015). *Saffron , A Functional Spice*. *3*(1).

Leardi, R. (2018). Chemometric Methods in Food Authentication. In *Modern Techniques for Food Authentication* (2nd ed.). Elsevier Inc. https://doi.org/10.1016/b978-0-12-814264-6.00017-7

Lehnert, L. W., Meyer, H., & Bendix, J. (2022). *hsdar: Manage, analyse and simulate hyperspectral data in R.* https://cran.r-project.org/package=hsdar. Accessed 20 Sept 2022

Li, S., Shao, Q., Lu, Z., Duan, C., Yi, H., & Su, L. (2018). Rapid determination of crocins in saffron by near-infrared spectroscopy combined with chemometric techniques. *Spectrochimica Acta - Part A: Molecular and Biomolecular Spectroscopy*, *190*, 283–289. https://doi.org/10.1016/j.saa.2017.09.030

Liland, K. H., Mevik, B. H., Wehrens, R., Hiemstra, P., Liland, M. K. H., & Suggests, M. A. S. S. (2022). *{pls}: Partial Least Squares and Principal Component regression*. https://cran.ms.unimelb.edu.au/web/packages/pls/pls.pdf

Lima, A. B. S. de, Batista, A. S., Jesus, J. C. de, Silva, J. de J., Araújo, A. C. M. de, & Santos, L. S. (2020). Fast quantitative detection of black pepper and cumin adulterations by near-infrared spectroscopy and multivariate modeling. *Food Control*, *107*(July 2019), 106802. https://doi.org/10.1016/j.foodcont.2019.106802

Liu, J., Chen, N., Yang, J., Yang, B., Ouyang, Z., Wu, C., Yuan, Y., Wang, W., & Chen, M. (2018). An integrated approach combining HPLC, GC/MS, NIRS, and chemometrics for the geographical discrimination and commercial categorization of saffron. *Food Chemistry*, *253*(August 2017), 284–292. https://doi.org/10.1016/j.foodchem.2018.01.140

Lohumi, S., Lee, H., Kim, M. S., Qin, J., & Cho, B. K. (2019). Raman hyperspectral imaging and spectral similarity analysis for quantitative detection of multiple adulterants in wheat flour. *Biosystems Engineering*, *181*, 103–113. https://doi.org/10.1016/j.biosystemseng.2019.03.006

Lu, X., Xia, Z., Qu, F., Zhu, Z., & Li, S. (2019). Identification of authenticity , quality and origin of saffron using hyperspectral imaging and multivariate spectral analysis. *Spectroscopy Letters*, *0*(0), 1–10. https://doi.org/10.1080/00387010.2019.1693403

Luo, N., Yang, X., Sun, C., Xing, B., Han, J., & Zhao, C. (2021). Visualization of vibrational spectroscopy for agro-food samples using t-Distributed Stochastic Neighbor Embedding. *Food Control*, *126*(October 2020). https://doi.org/10.1016/j.foodcont.2020.107812

Maggi, L., Carmona, M., Kelly, S. D., Marigheto, N., & Alonso, G. L. (2011). Geographical origin differentiation of saffron spice (Crocus sativus L. stigmas) - Preliminary investigation using chemical and multi-element (H, C, N) stable isotope analysis. *Food Chemistry*, *128*(2), 543–548. https://doi.org/10.1016/j.foodchem.2011.03.063

Manley, M. (2014). Near-infrared spectroscopy and hyperspectral imaging: Non-destructive analysis of biological materials. *Chemical Society Reviews*, *43*(24), 8200–8214. https://doi.org/10.1039/c4cs00062e

Meyer, D., & Dimitriadou, Evgenia, Kurt Hornik, Andreas Weingessel, Friedrich Leisch, Chih-Chung Chang, C.-C. L. (2022). *Package 'e1071.'* https://cran.r-project.org/web/packages/e1071/index.html

Minaei, S., Kiani, S., Ayyari, M., & Ghasemi-Varnamkhasti, M. (2017). A portable computer-vision-based expert system for saffron color quality characterization. *Journal of Applied Research on Medicinal and Aromatic Plants*, *7*(December 2016), 124–130. https://doi.org/10.1016/j.jarmap.2017.07.004

Mishra, P., Nordon, A., Tschannerl, J., Lian, G., Redfern, S., & Marshall, S. (2018). Near-infrared hyperspectral imaging for non-destructive classification of commercial tea products. *Journal of Food Engineering*, *238*(June), 70–77. https://doi.org/10.1016/j.jfoodeng.2018.06.015

Mohiuddin, A. K. (2019). *Health Hazards with Adulterated Spices : Save t he " Onion Tears ." 1*(3), 1–6.

Moore, J. C., Spink, J., & Lipp, M. (2012). Development and Application of a Database of Food Ingredient Fraud and Economically Motivated Adulteration from 1980 to 2010. *Journal of Food Science*, *77*(4). https://doi.org/10.1111/j.1750-3841.2012.02657.x

Moras, B., Loffredo, L., & Rey, S. (2018). Quality assessment of saffron (Crocus sativus L.) extracts via UHPLC-DAD-MS analysis and detection of adulteration using gardenia fruit extract (Gardenia jasminoides Ellis). *Food Chemistry*, *257*(March), 325–332. https://doi.org/10.1016/j.foodchem.2018.03.025

Moratalla-López, N., Bagur, M. J., Lorenzo, C., Martínez-Navarro, M. E., Rosario Salinas, M., & Alonso, G. L. (2019). Bioactivity and Bioavailability of the Major Metabolites of Crocus sativus L. Flower. *Molecules*, *24*(15), 1–24. https://doi.org/10.3390/molecules24152827

Morozzi, P., Zappi, A., Gottardi, F., Locatelli, M., & Melucci, D. (2019). A quick and efficient non-targeted screening test for saffron authentication: Application of chemometrics to gas-chromatographic data. *Molecules*, *24*(14), 1–13. https://doi.org/10.3390/molecules24142602

Nicolaï, B. M., Beullens, K., Bobelyn, E., Peirs, A., Saeys, W., Theron, K. I., & Lammertyn, J. (2007). Nondestructive

measurement of fruit and vegetable quality by means of NIR spectroscopy: A review. *Postharvest Biology and Technology*, *46*(2), 99–118. https://doi.org/10.1016/j.postharvbio.2007.06.024

Nobari Moghaddam, H., Tamiji, Z., Akbari Lakeh, M., Khoshayand, M. R., & Haji Mahmoodi, M. (2022). Multivariate analysis of food fraud: A review of NIR based instruments in tandem with chemometrics. *Journal of Food Composition and Analysis*, *107*(December 2021). https://doi.org/10.1016/j.jfca.2021.104343

Ordoudi, S. A., De Los Mozos Pascual, M., & Tsimidou, M. Z. (2014). On the quality control of traded saffron by means of transmission Fourier-transform mid-infrared (FT-MIR) spectroscopy and chemometrics. *Food Chemistry*, *150*(2014), 414–421. https://doi.org/10.1016/j.foodchem.2013.11.014

Orrillo, I., Cruz-Tirado, J. P., Cardenas, A., Oruna, M., Carnero, A., Barbin, D. F., & Siche, R. (2019). Hyperspectral imaging as a powerful tool for identification of papaya seeds in black pepper. *Food Control*, *101*(February), 45–52. https://doi.org/10.1016/j.foodcont.2019.02.036

Pessanha, S., Braga, D., Ensina, A., Silva, J., Vilchez, J., Montenegro, C., Barbosa, S., Carvalho, M. L., & Dias, A. (2023). A non-destructive X-ray fluorescence method of analysis of formalin fixed-paraffin embedded biopsied samples for biomarkers for breast and colon cancer. *Talanta*, *260*(March). https://doi.org/10.1016/j.talanta.2023.124605

Petrakis, E. A., Cagliani, L. R., Polissiou, M. G., & Consonni, R. (2015). Evaluation of saffron (Crocus sativus L.) adulteration with plant adulterants by1H NMR metabolite fingerprinting. *Food Chemistry*, *173*, 890–896. https://doi.org/10.1016/j.foodchem.2014.10.107

Prieto, N., Roehe, R., Lavín, P., Batten, G., & Andrés, S. (2009). Application of near infrared reflectance spectroscopy to predict meat and meat products quality: A review. *Meat Science*, *83*(2), 175–186. https://doi.org/10.1016/j.meatsci.2009.04.016

R Core Team. (2022). *The R Stats Package*. https://stat.ethz.ch/R-manual/R-devel/library/stats/html/00Index.html

Rubert, J., Lacina, O., Zachariasova, M., & Hajslova, J. (2016). Saffron authentication based on liquid chromatography high resolution tandem mass spectrometry and multivariate data analysis. *Food Chemistry*, *204*, 201–209. https://doi.org/10.1016/j.foodchem.2016.01.003

S.Hagh-Nazari N.Keifi. (2007). *Saffron and Various Fraud Manners in its Production and Trades*. 411–416.

Sabatino, L., Scordino, M., Gargano, M., Belligno, A., Traulo, P., & Gagliano, G. (2011). HPLC/PDA/ESI-MS evaluation of saffron (Crocus sativus L.) adulteration. *Natural Product Communications*, *6*(12), 1873–1876. https://doi.org/10.1177/1934578x1100601220

Sánchez-López, E., Sánchez-Rodríguez, M. I., Marinas, A., Marinas, J. M., Urbano, F. J., Caridad, J. M., & Moalem, M. (2016). Chemometric study of Andalusian extra virgin olive oils Raman spectra: Qualitative and quantitative information. *Talanta*, *156–157*, 180–190. https://doi.org/10.1016/j.talanta.2016.05.014

Shahnoushi, N., Abolhassani, L., Kavakebi, V., Reed, M., & Saghaian, S. (2020). Economic analysis of saffron production. In *Saffron*. INC. https://doi.org/10.1016/b978-0-12-818638-1.00021-6

Shawky, E., Abu El-Khair, R. A., & Selim, D. A. (2020). NIR spectroscopy-multivariate analysis for rapid authentication, detection and quantification of common plant adulterants in saffron (Crocus sativus L.) stigmas. *Lwt*, *122*(January), 109032. https://doi.org/10.1016/j.lwt.2020.109032

Soffritti, G., Busconi, M., Sánchez, R. A., Thiercelin, J., Polissiou, M., Roldán, M., & Fernández, J. A. (2016). *Genetic and Epigenetic Approaches for the Possible Detection of Adulteration and Auto-Adulteration in Saffron (Crocus sativus L .) Spice*. 1–16. https://doi.org/10.3390/molecules21030343

Standardization, I. O. for. (2011). *Iso 3632-2*. *2010*.

Uncu, O., & Ozen, B. (2019). A comparative study of mid-infrared, UV–Visible and fluorescence spectroscopy in combination with chemometrics for the detection of adulteration of fresh olive oils with old olive oils. *Food Control*, *105*(April), 209–218. https://doi.org/10.1016/j.foodcont.2019.06.013

Varliklioz Er, S., Eksi-Kocak, H., Yetim, H., & Boyaci, I. H. (2017). Novel Spectroscopic Method for Determination and Quantification of Saffron Adulteration. *Food Analytical Methods*, *10*(5), 1547–1555. https://doi.org/10.1007/s12161-016-0710-4

Wakefield, J., McComb, K., Ehtesham, E., Van Hale, R., Barr, D., Hoogewerff, J., & Frew, R. (2019). Chemical profiling of saffron for authentication of origin. *Food Control*, *106*(June). https://doi.org/10.1016/j.foodcont.2019.06.025

Zeng, Tan, Matsunaga, & Shirai. (2019). Generalization of Parameter Selection of SVM and LS-SVM for Regression. *Machine Learning and Knowledge Extraction*, *1*(2), 745–755. https://doi.org/10.3390/make1020043