Food Control 129 (2021) 108242

Contents lists available at ScienceDirect

Food Control

journal homepage: www.elsevier.com/locate/foodcont

Detection of foreign materials in cocoa beans by hyperspectral imaging technology

Ali Saeidan [a], Mehdi Khojastehpour [a,*], Mahmood Reza Golzarian [a], Marziye Mooenfard [b], Haris Ahmad Khan [c]

[a] Department of Biosystems Engineering, Ferdowsi University of Mashhad, Mashhad, Iran
[b] Department of Food Science and Technology, Ferdowsi University of Mashhad, Mashhad, Iran
[c] Farm Technology Group, Wageningen University and Research, Wageningen, the Netherlands

ARTICLE INFO

Keywords:
Cocoa beans
Foreign materials
Hyperspectral imaging
Near-infrared spectroscopy

ABSTRACT

The presence of foreign materials in a batch of cocoa beans affect its profitability, marketability and overall quality grade of the product. Therefore, the identification of these materials and their subsequent removal is very important to ensure the high quality of the final product. This study aims to investigate the feasibility of using hyperspectral imaging technology for the detection and discrimination of four categories of foreign materials (wood, plastic, stone and plant organs) that are relevant to the cocoa processing industries. The spectral image data of 250 cocoa beans and foreign material was analyzed using principal component analysis and three classification models Support Vector Machine (SVM) Linear Discriminant Analyses (LDA) and K Nearest Neighbours (KNN). Optimal wavebands, which were obtained from the second spectra graph and the first three PCs, were fed into the classification models and the performance of classifiers was compared. The results showed that SVM could reach over 89.10% accuracy in classifying cocoa beans and foreign materials. The accuracy of the SVM classifier when using optimal features as input was 86.90% for the training set and 81.28% for the test set. An external test set of data was used to test the generalization of the model. The results showed that the classification of foreign materials could be more robust when the optimal feature was used as input data.

1. Introduction

In the food industry, foreign materials such as plastic, stone and metals are defined as undesirable and unwanted pieces of solid matters present in a product. These foreign materials can end up in the final product. According to Djekic et al. (2017), in the period 1998–2015, nuts, fruits and vegetables/bakery/confectionery products were the top three categories of products with the highest physical contamination. The presence of foreign materials is considered as the main source of many consumer complaints from food manufactures and executive authorities (Edwards & Stringer, 2007). Therefore, the identification and, subsequently, the removal of foreign materials are of great importance for consumer satisfaction and health. For this reason, the presence of these kinds of materials in food products is extensively controlled by food security regulations to ensure the high quality of the final product (Djekic et al., 2017; Libânio et al., 2018; Trafialek et al., 2016). To address this issue, the European Union has established the Rapid Alert System for Food and Feed (RASFF), which provides data on food composition and safety information (RASFF, 2014).

The cocoa bean is one of the most important traded agricultural product with high demand for the industry (Sunoj et al., 2016). Cocoa is served as the main ingredient in confectionery products and chocolate manufacturing, responsible for the unique flavour and melt-in-the-mouth properties of chocolate (Astika et al., 2010). Furthermore, cocoa beans are enriched with protein, vitamins, minerals, carbohydrates, fats as well as a variety of phenolic compounds with high antioxidant activity; thus, their consumption has increased all over the world due to their benefits to human health (Teye et al., 2020).

A reliable supply of high-quality cocoa at reasonable prices is more favourable by manufacturers. Although quality is a complex attribute comprising different features, there is a direct link between the quality of cocoa beans and their price. Therefore, due to the low purchasing price by local traders, the farmers are discouraged to maintain the quality of cocoa beans. The use of improper post-harvest handling, packing the cocoa beans in an uncleaned environment and ignoring the dampness during transportation are the major sources of quality

* Corresponding author. Department of Biosystems Engineering, Ferdowsi University of Mashhad, P.O. Box 9177948978, Mashhad, Iran.
E-mail address: mkhpour@um.ac.ir (M. Khojastehpour).

https://doi.org/10.1016/j.foodcont.2021.108242
Received 15 October 2020; Received in revised form 23 March 2021; Accepted 8 May 2021
Available online 12 May 2021

deterioration. These factors affect consumer acceptance negatively (Astika et al., 2010). In Europe, the Federation of Cocoa Commerce (FCC) set contract standards for cocoa bean quality. Besides, as stated by the International Organization for Standardization (ISO), the foreign material shall not exceed 0.75% of the mass of the reference sample representing the bulk. Foreign substances can damage the manufacturer's machinery and be detrimental to the flavour as well as reducing the yield of edible material (Biscuit; ISO, 2018). Therefore, effective quality inspection systems for reducing the incidence of foreign material is extremely important from an economic point of view and consumer expectations to provide high-quality and high-safety products. The current practice for the removal of foreign materials is through manual inspection. This is a labour-intensive task, and it is difficult to identify the foreign materials that are of the same shape or colour as the cocoa beans. Machine vision has the potential of performing this task in a definite manner to save costs in terms of time and human labour.

Currently, the separation of foreign materials in chocolate factories is mostly performed with the mechanical sorters by sieving, destoning and using magnets. Despite the necessary measures to remove foreign materials in mechanical sorters, it is still observed in many cases the foreign objects pass through these sorters without being separated. This inability to separate foreign materials is more about materials that have the same shape and weight as the reference matter. There are very few innovations on the market for mechanical sorters because the technology used in these machines is timeless (Esteve Agelet, 2011). The use of X-ray is also investigated, but this technology can only identify foreign materials of higher density than the ingredients and that is why detection of plastics, plant fragments (chips, sticks and broken pallets), cardboard or insects has not been resolved so far (Mohd Khairi et al., 2018).

Recently, Lawi and Adhitya. (2018) employed multiclass ensemble least-squares support vector machine (MELS-SVM) and classified cocoa beans into four groups including normal, broken, fractured, and skin damaged beans according to their morphological and physical features. In two other studies, machine-vision technology was employed to discriminate various forms of foreign materials from soybeans (Momin et al., 2017) and walnuts (Rong et al., 2019). Most machine vision systems collect information from the images in the visible light range (Xia et al., 2019). Colour imaging techniques mimic the principle of human vision and are typically limited to the features such as colour, size and shape (Feng & Sun, 2012). Since the foreign materials may be of the same shape, colour and size, the use of colour imaging to distinguish them can be a very challenging task. One way to identify the foreign materials is to determine the material properties. One such property is the surface reflectance, where each material has a distinct behaviour when there is an interaction with incoming radiation. Some part of the incoming light is absorbed and some part is reflected from the surface of materials. This reflectance and absorbance is unique for each material and is known as their "spectral signature". We can obtain the spectral signature of materials through spectroscopy. This non-destructive optical technique is of great interest for material identification and quality estimation. Application of near-infrared spectroscopy (NIR) for predicting the quality parameters of coca beans (Sunoj et al., 2016; Xu et al., 1999) or distinguish among five types of plastic resins (Masoumi et al., 2012) were discussed in the literature. Employing (infrared) IR thermographic images for detecting foreign materials has been also investigated by Ginesu et al. (2004). Similarly, an automatic identification technique for either buried defects or foreign materials in biscuits was proposed by Senni et al. (2014).

The spatial features of the product are also important and provide valuable information. For heterogeneous materials, traditional spectroscopic methods cannot be considered as an effective tool for evaluating food quality because they provide only information about one point on the surface of the material (Elmasry et al., 2012). Hyperspectral imaging (HSI) is a technology that combines near-infrared spectroscopy and two-dimensional imaging to obtain spatial and spectral information

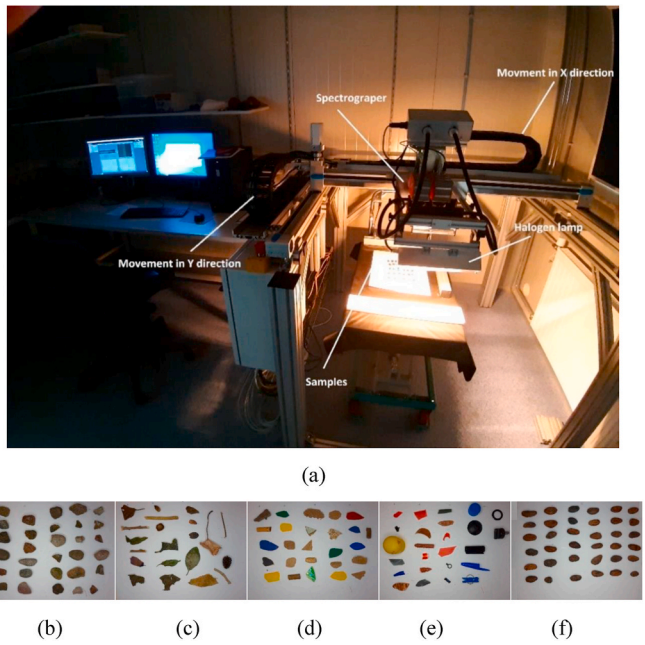


Fig. 1. Shortwave infrared hyperspectral imaging system (a), four types of foreign matter and cocoa bean samples. Stones (b), plant organs (c), Wood & paper (d), plastic (e) and cocoa bean (f).

at the same time (Elmasry et al., 2012; Gowen et al., 2007). HSI can obtain data in wavelengths from the visible to near-infrared (Xia et al., 2019). HSI comprises three-dimensional multivariate data structures with two spatial (X–Y) and one wavelength (λ) dimensions (Burger & Gowen, 2011).

HSI is used for the detection of foreign materials in blueberry (Tsuta et al., 2006) and cotton lint (Zhang et al., 2016; Guo et al., 2012a,b; Jiang & Li, 2015). Zhang et al. (2016) performed the classification of 16 types of foreign materials in cotton lint with 95% accuracy through linear discriminant analysis. Caporaso et al. (2018) used NIR-HSI to y estimate the quality features of individual cocoa beans.

Although spectroscopic techniques based on HSI were investigated previously for classification, authentication, or quality control purposes of cocoa beans (Cruz-Tirado et al., 2020; Sunoj et al., 2016; Teye et al., 2020), no study was found in the literature regarding the detection of unwanted objects by this method to the best of our knowledge. Therefore, employing HSI as a non-destructive method, with good performance is of great interest to identify foreign objects in cocoa beans. This study aims to identify and classify foreign materials in the batch of cocoa beans with help of imaging spectroscopy.

2. Material and method

2.1. Sample preparation

Four types of foreign materials that are found mostly in batches of cocoa beans were collected from a local cocoa processing company in Tabriz-Iran. This foreign material included: plastic base materials, plant organs, wood and paper-based material and small pieces of stones (Fig. 1). Around 1 kg of non-fermented cocoa beans (a mixture of 5 varieties), with uniform sizes and without obvious defects, were purchased from a local company in Tabriz-Iran. In total, 250 samples (including 50 samples for each category) were selected to do calibration and test experiments and 30 samples (including 6 samples for each category) were kept out for independent test experiment. To maintain the quality and appearance of the beans during the experiment, all samples were placed in sealed bottles and stored at 4 °C until analysis.

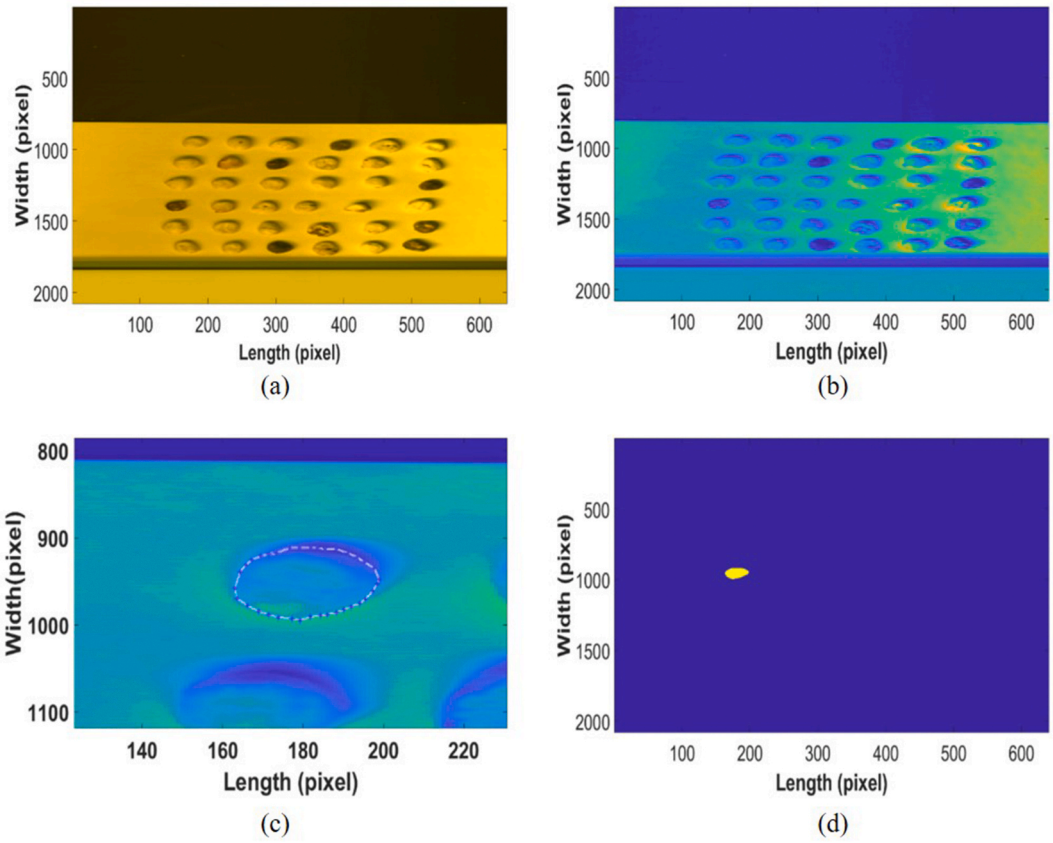


Fig. 2. Region of interest definition progress. Hyperspectral image in wave band 40 (a), Image correction by equation (1) (b), manually selected ROI (c) and creating mask image for ROI (d).

2.2. Hyperspectral image acquisition and correction

A shortwave infrared hyperspectral imaging (SWIR) system was used in the laboratory to acquire hyperspectral images from foreign materials (FMs) and cocoa beans samples. The HSI system was mainly composed of four components: a line scan imaging spectrograph (Specim FX17, Oulu, Finland), a CCD (Pixel fly QE IC × 285AL, Cooke, USA) imaging camera, a halogen-tungsten illumination source (located on top of the samples in 45° angular position respect to the samples) and a transportation mechanism with the stepper motor which moved the spectrograph in X and Y directions over the samples. A custom-made software (Isaac2, Wageningen University & Research, 2014) was used to control the hyperspectral image acquisition process. Configuration of the line scan SWIR hyperspectral imaging system is shown in Fig. 1. To take images of the samples, we first arranged them in columns and rows according to their classes and a white paper was selected as background. Spectral images were collected in the spectral wavelength range of 900–1700 nm with a resolution of 5.5 nm resulting in 112 spectral bands. The values of each three-dimensional (x, y, λ) image cube were 2080 × 656 pixels × 112 spectral levels, where (x) and (y) represent spatial dimensions and (λ) indicate the number of spectral bands, respectively.

In total, 16 scans were acquired and a 3D hyperspectral dataset was saved in the raw format for further processing. During each scan reflectance calibration was also performed. For this purpose, a white image (I_{white}) was obtained by scanning a white Teflon tile, which reflects almost 99% of received light, while the dark image (I_{dark}) was captured by fully covering the camera lenses using an opaque black cap. The corrected image (I_c) was calculated using Eq. (1) (Xia et al., 2019).

$$I_c = \frac{I_{raw} - I_{dark}}{I_{white} - I_{\text{dark}}} \quad (1)$$

where I_{raw} ,I_{white} and I_{dark} are the sample raw image, white reflectance, and dark current image, respectively.

2.3. Pixel selection and spectral data pre-processing

Chemometric data analysis was performed using Matlab (MATLAB, 2018). To choose the suitable pixels from hyperspectral images, the region of interest (ROI) for each sample was selected manually by using the *roipoly* command in the Matlab image processing toolbox as is outlined in Fig. 2(c). For each image of FM and cocoa beans, the binary mask was created on the image (Fig. 2(d)) that has less amount of noise or disorders (e.g. Image in waveband 40). The ROIs of cocoa beans and each FMs were mapped on to the original spectral images to extract the full spectra (i.e. 900–1700 nm). All samples were processed under the same conditions and data was saved as a spectral data set. The images of 250 samples (including 50 samples for each class) were implemented for pixel base classification and 1000 pixels were randomly chosen for each FM and cocoa bean group. Thus, for five different FM and cocoa bean groups, 5000 pixels were taken into consideration. A total of 5000 randomly selected pixels from five different classes were collected for spectral pixel-based classification.

The intensity of the spectrum in each waveband was normalized by their highest intensity value found in the entire spectrum (900–1700 nm) to facilitate their comparison. After normalization, the spectral values of the FMs and cocoa beans were varied within the range of −1 to 1. Each waveband was considered as a specific feature so in total there were 112 features for each pixel (Zhang et al., 2016).

For training and testing the models, firstly the total set of 5000 pixels was divided into 2 parts and 80% of pixels (4000 spectra) selected to train the classification models and the remaining 20% (1000 spectra) used as the test set. A fivefold cross-validation method was applied to

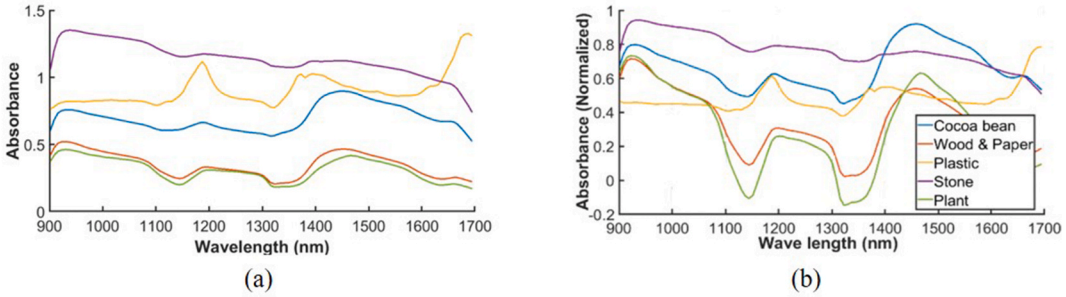


Fig. 3. Raw spectra and normalized mean spectra of cocoa bean and four foreign materials. Raw spectra (a), Normalized spectra (b).

classify FMs. Also, the term "accuracy" was utilized to compare the performance of models and terms sensitivity, specificity, recall, and precision were used to assessing the performance of selected models (Minaei et al., 2017). The mentioned assessing terms are commonly stated as:

$$\text{Accuracy} = (TP + TN)/(TP + FP + TN + FN) \quad (2)$$

$$\text{Sensitivity} = TP/(TP + FN) \quad (3)$$

$$\text{Specificity} = TN/(TN + FP) \quad (4)$$

$$\text{Precision} = TP/(TP + FP) \quad (5)$$

$$\text{Recall} = TP/(TP + FN) \quad (6)$$

where TP, TN, FP and FN are true positive, true negative, false positive and false negative, respectively.

2.4. Dimensionality reduction and optimum wavelength selection

Hyperspectral images consist of a large number of spectral bands and many of this information is redundant as the bands are highly correlated. Dimensionality reduction is the process of reducing the number of variables and eliminate the redundant information of the data set while maintaining meaningful data that are informative for the model (Jiang et al., 2010).

Principal components analysis (PCA) is one of the commonly used unsupervised dimensionality-reduction methods that is often used to reduce the size of large data sets by projecting the large set of variables into a smaller one using linear combinations of the variables known as principal components. The new projected variables are uncorrelated with each other and the first few components retain most of the variation present in the original variables (Xu et al., 2018). In this study, PCA was performed on all spectra to select spectral features.

Second derivative spectra (2nd spectra) is one of the spectral pretreatment techniques, which can help to select a useful feature from the spectrum. This method improves spectral resolution, repress spectral noises, and allows more specific identification of small and nearby lying absorption peaks which are not resolved in the original spectrum (Zhang et al., 2018). For selecting optimal wavelengths, the high peaks and low valleys with large differences in the second derivative spectra of the different FMs and cocoa beans were picked out.

2.5. Classification models

Three classification methods including LDA, Support Vector Machine (SVM), and K nearest neighbours (KNN) were utilized to classify the spectral images. LDA is one of the supervised classification approaches. By maximizing the variance of data, this method tries to model the difference between each class of data and find linear combinations of independent variables that separate several classes of objects or events. Briefly, LDA attempt to project a multidimensional feature space into moderate dimensional-space on which the ratio of the between-class scatter to within-class scatter is maximized (Xu et al., 2018).

The SVM is a supervised learning algorithm, which can be used for classification problems. It is robust even with a small number of input training samples. The objective of this technique is to find an optimal hyperplane in n-dimensional space as a decision surface that maximizes the distance between each class and separate data points based on margin distance. As a kernel-based classifier, SVM seems to have an especial advantage in the analysis of hyperspectral data and shows its robustness in high-dimensional data classifications (Jiang et al., 2010, pp. 79–98).

Another classifier in this study is KNN. This method is a simple non-parametric machine algorithm. In the first step, the KNN classifier identifies the k-neighbours in the training data that are closest to the test value and calculates the distance between all those categories. By the majority voting of nearest neighbours, the test class belongs to the category whose distance is minimum (Li et al., 2012).

3. Results and discussion

3.1. Pixel level classification of hyperspectral images

The average mean raw spectral and normalized spectral for five categories of FM and cocoa beans over the wavelength range of 900–1700 nm are shown in Fig. 3(a) and (b), respectively. Reflectance spectra were converted into absorption (log 1/R) spectra. The horizontal axis represents the wavelength, and the vertical axis shows the average absorption value for each feature parameter. Each piece of spectrum extracted from the ROIs was normalized by dividing the raw intensity at each wavelength by the maximum intensity value found in the entire spectrum. As shown in Fig. 3 normalization is an important step in spectral classification and without normalization, the classification does not reach correct training (Zhang et al., 2016).

To validate the effectiveness of the dimensionality reduction method, a PCA model was build based on preprocessed spectral data. Fig. 4(a) represent the individual explained variance by first 10 PCs and it can be conclude from Fig. 4(a) that the first 3 PCs have explained most of variances in our data sets. According to the PC1–PC2 and PC2-PC3 score plot presented in Fig. 4 (b, c), it can be concluded that this method is unable to differentiate between FMs and cocoa beans, suggesting the inefficiency of the PCA model for classification. In this study, pixels were selected randomly from a large set of data so it is possible to have large variance inside the class. It seems that when the number of samples per class is large and also the value of within-class variance is greater than the value of between classes variance, PCA cannot find an appropriate direction to map the original dataset into a new subspace (Aleix, 2001). Besides, this misclassification could be justified somewhat by similarity in the spectral features of some FMs, such as wood and paper group with plant group.

3.2. Selection of optimal wavelengths

As it has been presented in Fig. 5, nineteen optimal wavebands (each

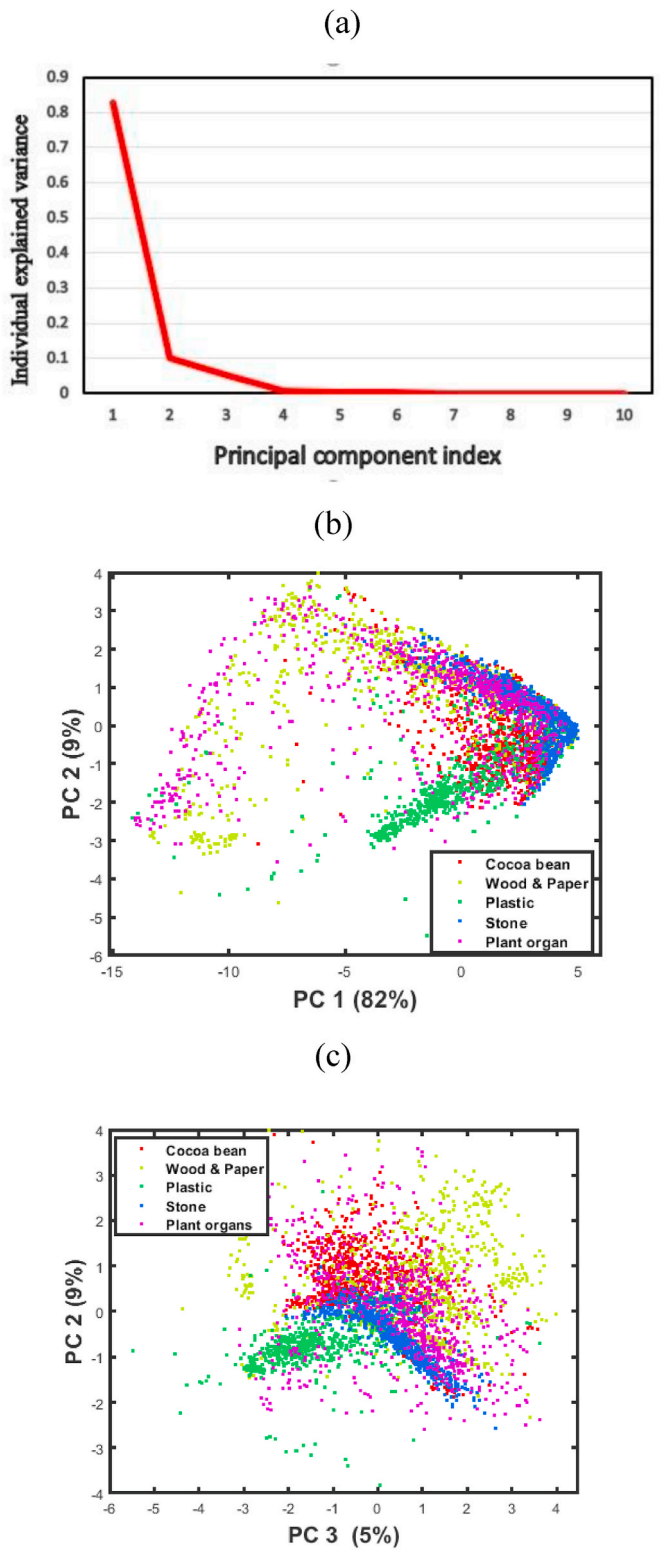


Fig. 4. Individual explained variance (a) and the score plot of PC1, PC2 (b) and PC2, PC3 (c).

waveband including several wavelengths) were selected for spectra classification by LDA, SVM, and KNN using 2nd derivative spectra. The average amount of spectra for each class in the calibration set of samples were used to obtain the 2nd derivative spectra. As can be seen in the 2nd derivative diagram (Fig. 5), some peaks and valleys of the cocoa bean and four types of foreign materials are slightly different from each other in the corresponding wavelengths, while some other peaks and valleys have large differences. Therefore, in this research, these peaks and valleys have been selected manually as the optimal features.

The differences in chemical components cause differences in bond vibration regions between four types of FMs and cocoa beans. Wood and paper share a common source but each of them has different physical properties. Cellulose is the major carbohydrate component of wood along with hemicelluloses (20–35 per cent by weight). Lignin, extractives, and trace amounts of other materials make up the remaining portion of wood (Jones & Wegner, 2009). However, similarities in the shape of the spectrum are inherent to the wood and plant organs. This is probably due to the existence of lignin and cellulose in both of them. The chemical structure of the lignin molecule consists of several functional groups such as —CH2—OH, OH–, CH3O—, C6H6, and C— O. The C6H6-band appears at 1145 nm (Zhang et al., 2016). Because of the first overtone of O—H for cellulosic materials and the first overtone of C—H combination for both cellulose and water, the wavebands from 1350 to 1650 nm have been recognized as a combination of cellulosic materials and moisture band (Fortier et al., 2012; Rodgers et al., 2010). The 2nd overtone of C–H stretching (–CH3 or –CH2) and O–H stretching and O–H deformation are linked with an absorption band of 1263 nm due to the presence of fibres and carbohydrates (Cruz-Tirado et al., 2020) which may found in plant materials. Besides, absorption bands of 1181 and 1426 nm, corresponding to the 2nd overtone of O–H deformation and O–H stretching are also associated with water and fibre content (Cruz-Tirado et al., 2020).

The cocoa bean spectral profile represents several useful peaks where provide vital information for quantitative and qualitative analyses. These peaks attributed to functional groups such as —CH═CH 2nd overtone at around 8235 cm^{-1}(1214 nm) corresponding to protein, C—H first overtone at 7030 cm^{-1}(1422 nm) corresponding to fat, O—H combination at 5176 cm^{-1}(1931 nm) corresponding to moisture and N—H bending at around 4880 cm^{-1}(2049 nm) corresponding to some aromatic compounds. On the other hand, saturated and unsaturated fats are the most prominent component in the cocoa bean that can be identified according to spectral peaks around 7600–8000 cm^{-1}(1250–1315 nm) (Teye et al., 2020).

Polymers including Polypropylene (PP), Polyethylene (PE) and Polyethylene terephthalate (PET) are the main component of most plastics. Polyethylene contains (=CH2) (Whiteley et al., 2000) and Polypropylene consists of (=CH2), (—CH =), CH3— structures (Beswick & Dunn, 2002) where absorption of CH3— takes place at 1195 nm. The monomer of Polyethylene terephthalate (PET) is ethylene terephthalate which mainly consists of —CH2—C—O— and its absorption was seen approximately around 1395 nm (Lepoittevin & Roger, 2011). Several non-botanical bands from 1350 to 1650 nm assigned to CH 3rd overtone and O—H bands (e.g., PET) at approximately 1530–1570 nm (Fortier et al., 2012; Rodgers et al., 2010).

3.3. Classification results

The performance of the three mentioned classifiers is exhibited in Table 1. Classification using the first 3 PC as input features did not provide good classification performance. In the best model (KNN), its accuracy reached only 76.25% for the calibration set and 64.35% for the validation set. Better performance can be seen when the full spectrum was used as input and the SVM model provides the best discrimination between different types of FMs and cocoa beans representing the precision value of 89.10% for the calibration set and 83.96% for the validation set. Performance evaluation of SVM classifier using optimal wavelengths as input feature vector show rather similar classification results in which its accuracy for calibration and validation set was 86.90% and 82.20%, respectively. Among the three classification methods, LDA has the lowest performance and accuracy when using the full spectrum as input, achieved to 76.79% for the calibration set and

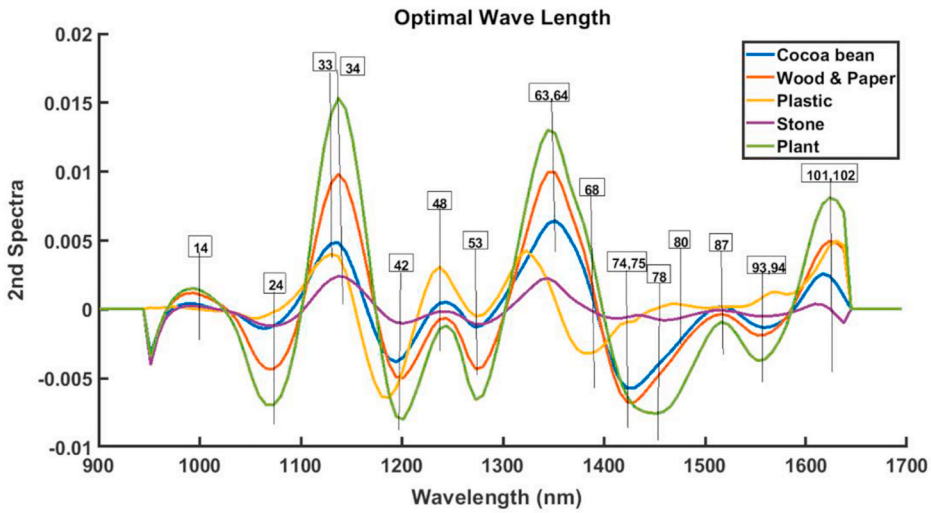


Fig. 5. Optimal wavelengths selected by 2nd spectra of preprocessed spectra.

Table 1
Classification Accuracy of three different classification models for hyperspectral images.

Classification models	Calibration (Train)			Validation (Test)		
	LDA	SVM	KNN	LDA	SVM	KNN
Full spectrum	76.79	**89.10**	87.60	70.88	**83.96**	80.43
PCA (3 pc)	60.2	63.3	67	62.2	63.5	67.9
Optimal wavelength	74.51	**86.90**	82.63	69.72	**81.28**	79.72

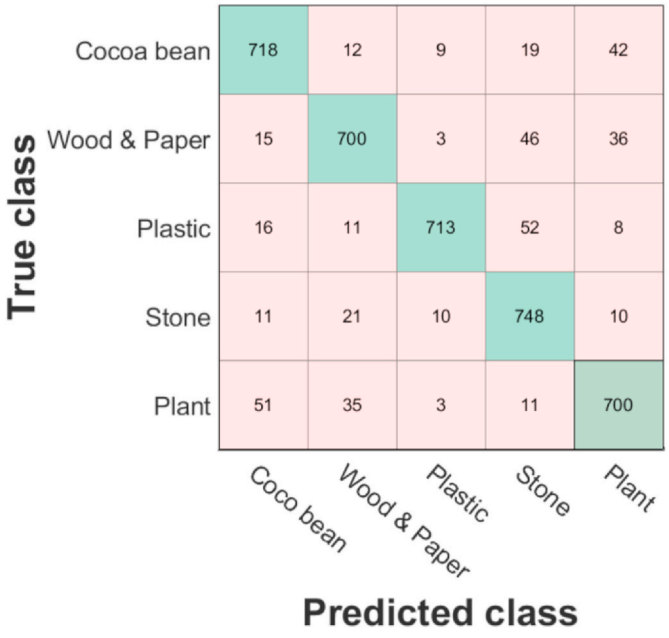


Fig. 6. Confusion matrix of the SVM model with calibration set for hyperspectral image data.

70.88% for the validation set.

The confusion matrix for the SVM model has shown in Fig. 6. This matrix illustrates the calibration accuracy of the SVM classifier when using the full spectrum as an input. According to the confusion matrix, classification reached the best separation rate for stone class, and 748 pixels from a total of 800 pixels identified correctly. It was found that the lowest separation performance occurred in the class of wood and paper as well as plant organs. For plant organs, among a total of 800 pixels, 700 pixels were classified correctly, 51 pixels were misclassified as cocoa bean, 35 pixels misclassified as wood and paper, 3 pixels assigned to plastic and 11 pixels were identified mistakenly as stone. For cocoa beans, 718 pixels were classified correctly. Also for plastic 713 pixels were truly identified.

In this study, only 5000 pixels were selected randomly from thousands of pixels so the independent test data set (including six samples for each class) that was kept out from the experiments, was utilized to validate the generalization of the selected model and its performance for an independent large data set. Fig. 7 compares SVM model performance in terms of four statistical parameters when the full-spectrum and optimal wavelength were used as model inputs. In this diagram, statistical parameters in the coloured marker were assigned for prediction results when inputs are optimal wavelengths. In this case, for each class, the precision value is greater than 67%. Compared to the case using the full-spectrum as the input, the precision of classification for each class was high. This result was reasonable because it built a model based on fewer input features and at the same time optimal features make it robust against overfitting and increased its precision.

We mapped the obtained classification results over the raw hyperspectral image to visualize the output of our classification model. We used the results from the SVM classifier that was trained on the optimal wavelengths. The resulting image is shown in Fig. 8 and according to the majority of counted coloured pixels for each object in this image, its class was determined.

4. Conclusions

In this study, we investigated the robustness and validation of a non-contact hyperspectral method to classify cocoa bean and four typical foreign materials that are relevant to the cocoa traders and cocoa processing industries. Obtained results revealed the weakness of principal component analysis as a classifier where it had no significant effect on the classification accuracy of FMs and cocoa bean samples. Whereas, the SVM classification model exhibited a significant improvement as accuracy reached over 89.1% in classifying cocoa beans and four types of foreign materials when full spectrum was used. When the optimal wavelength was used as input, the SVM model also achieved 86.9% classification accuracy for the foreign materials found in cocoa beans. The results established that the generalization and robustness of classifiers can be increased if the optimal wavelength is utilized as input features. The results in this study are promising and the next steps will be to further investigate the optimal wavelength in real situation when

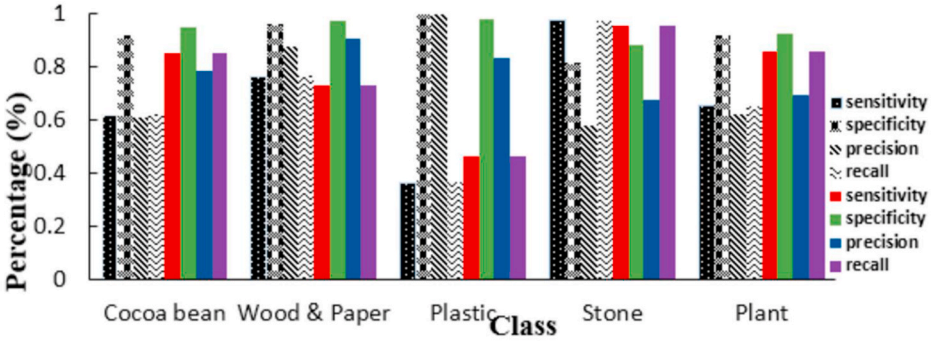
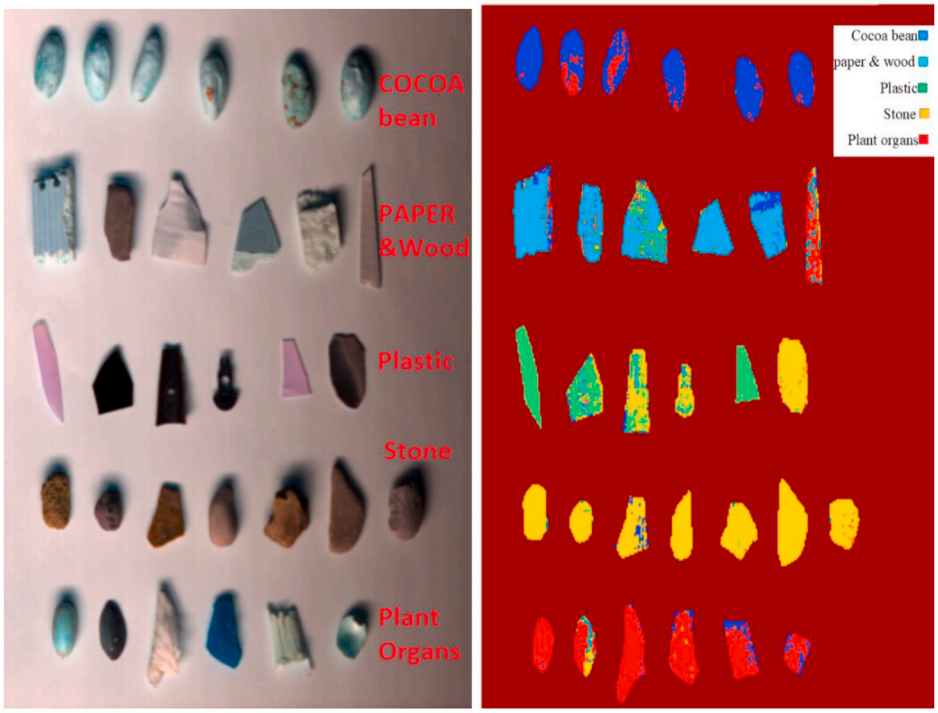


Fig. 7. Comparison between SVM classification statistics (%) of FM and cocoa bean, on independent test sets (graphs in colour indicate classification statistics when inputs are optimal wavelength and graphs in hatched pattern attributed for classification statistics when inputs are full spectrum). (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)



(a) (b)

Fig. 8. Pixel-level image classification of foreign materials and cocoa beans using SVM. Hyperspectral image of Cocoa beans and four types of foreign materials (a), Pixel-level image classification (b).

the samples are randomly mixed together. Once the optimal wavelength range is determined, a customised camera can be developed that can determine the presence of foreign materials in cocoa beans in real-time in the production line. Such a system will help in reducing the labour costs and time for the removal of foreign materials, and will greatly improve the quality of the end product.

CRediT authorship contribution statement

Ali Saeidan: Conceptualization, Investigation, Data curation, Writing – original draft. **Mehdi Khojastehpour:** Supervision, Writing – review & editing. **Mahmood Reza Golzarian:** Formal analysis, Writing – review & editing. **Marziye Mooenfard:** Validation, Resources. **Haris Ahmad Khan:** Formal analysis, Writing – review & editing.

Acknowledgements

Funding for this project was provided by Ferdowsi University of Mashhad (Grant No. 49026). The authors would like to thank the Dadash Baradar Industrial co (Shoniz) for providing cocoa beans for this project.

References

Aleix, M. (2001). *Mart nez, Avinash C Kak. PCA versus LDA, 23* pp. 228–233), 2.

Astika, I., Solahudin, M., Kurniawan, A., & Wulandari, Y. (2010). *Determination of cocoa bean quality with image processing and artificial neural network. Paper presented at the the Quality Information for Competitive Agricultural Based Production System and Commerce*. Bogor, Indonesia: Proceedings of the AFITA 2010 International Conference.

Beswick, R., & Dunn, D. (2002). *Plastics in packaging: Western Europe and North America*. iSmithers Rapra Publishing.

Biscuit, C. (1996). Chocolate and confectionery Alliance (BCCCA). *Cocoa beans. Chocolate manufacturers' quality requirements* (4th ed.) United Kingdom.

Burger, J., & Gowen, A. (2011). Data handling in hyperspectral image analysis. *Chemometrics and Intelligent Laboratory Systems, 108*(1), 13–22. https://doi.org/10.1016/j.chemolab.2011.04.001

Caporaso, N., Whitworth, M. B., Fowler, M. S., & Fisk, I. D. (2018). Hyperspectral imaging for non-destructive prediction of fermentation index, polyphenol content and antioxidant activity in single cocoa beans. *Food Chemistry, 258*, 343–351. https://doi.org/10.1016/j.foodchem.2018.03.039

Cruz-Tirado, J., Pierna, J. A. F., Rogez, H., Barbin, D., & Baeten, V. (2020). Authentication of cocoa (Theobroma cacao) bean hybrids by NIR-hyperspectral imaging and chemometrics. *Food Control*, 107445. https://doi.org/10.1016/j.foodcont.2020.107445

Djekic, I., Jankovic, D., & Rajkovic, A. (2017). Analysis of foreign bodies present in European food using data from Rapid Alert System for Food and Feed (RASFF). *Food Control, 79*, 143–149. https://doi.org/10.1016/j.foodcont.2017.03.047

Edwards, M., & Stringer, M. (2007). Observations on patterns in foreign material investigations. *Food Control, 18*(7), 773–782. https://doi.org/10.1016/j.foodcont.2006.01.007

Elmasry, G., Kamruzzaman, M., Sun, D.-W., & Allen, P. (2012). Principles and applications of hyperspectral imaging in quality evaluation of agro-food products: A review. *Critical Reviews in Food Science and Nutrition, 52*(11), 999–1023. https://doi.org/10.1080/10408398.2010.543495

Esteve Agelet, L. (2011). *Single seed discriminative applications using near infrared technologies*.

Feng, Y.-Z., & Sun, D.-W. (2012). Application of hyperspectral imaging in food safety inspection and control: A review. *Critical Reviews in Food Science and Nutrition, 52*(11), 1039–1058. https://doi.org/10.1080/10408398.2011.651542

Fortier, C., Rodgers, J., Foulk, J., & Whitelock, D. (2012). *Near-infrared classification of cotton lint, botanical and field trash*.

Ginesu, G., Giusto, D. D., Margner, V., & Meinlschmidt, P. (2004). Detection of foreign bodies in food by thermal image processing. *IEEE Transactions on Industrial Electronics, 51*(2), 480–490. https://doi.org/10.1109/tie.2004.825286

Gowen, A. A., O'Donnell, C. P., Cullen, P. J., Downey, G., & Frias, J. M. (2007). Hyperspectral imaging–an emerging process analytical tool for food quality and safety control. *Trends in Food Science & Technology, 18*(12), 590–598. https://doi.org/10.1016/j.tifs.2007.06.001

Guo, J., Ying, Y., Li, J., Rao, X., Kang, Y., & Shi, Z. (2012a). Detection of foreign materials on surface of ginned cotton by hyper-spectral imaging. *Transactions of the Chinese Society of Agricultural Engineering, 28*(21), 126–134. https://doi.org/10.3969/j.issn.1002-6819.2012.21.018

Guo, J., Ying, Y., Rao, X., Li, J., Kang, Y., & Shi, Z. (2012b). Detection of trashes in combed cotton using hyper-spectral images. *Nongye Jixie Xuebao= Transactions of the Chinese Society for Agricultural Machinery, 43*(12), 197–203. https://doi.org/10.6041/j.issn.1000-1298.2012.12.036

ISO, I. (2018). 22000: Food safety management systems—requirements for any organization in the food chain. *International Standard*, 1–48.

Jiang, Y., & Li, C. (2015). Detection and discrimination of cotton foreign matter using push-broom based hyperspectral imaging: System design and capability. *PloS One, 10*(3), Article e0121969. https://doi.org/10.1371/journal.pone.0121969

Jiang, L., Zhu, B., & Tao, Y. (2010). *Hyperspectral image classification methods Hyperspectral imaging for food quality analysis and control*. Elsevier.

Jones, J. P. E., & Wegner, T. H. (2009). Wood and paper as materials for the 21st century. *MRS Online Proceedings Library, 1187*(1), 51–60.

Lawi, A., & Adhitya, Y. (2018). *Classifying physical morphology of cocoa beans digital images using multiclass ensemble least-squares support vector machine*. Paper presented at the Journal of Physics: Conference Series.

Lepoittevin, B., & Roger, P. (2011). Poly (ethylene terephthalate). *Handbook of engineering and speciality thermoplastics, 3*, 97–126.

Libânio, D., Garrido, M., Jácome, F., Dinis-Ribeiro, M., Pedroto, I., & Marcos-Pinto, R. (2018). Foreign body ingestion and food impaction in adults: Better to scope than to wait. *United European gastroenterology journal, 6*(7), 974–980. https://doi.org/10.1177/2050640618765804

Li, C., Zhang, S., Zhang, H., Pang, L., Lam, K., Hui, C., et al. (2012). Using the K-nearest neighbor algorithm for the classification of lymph node metastasis in gastric cancer. *Computational and mathematical methods in medicine, 2012*. https://doi.org/10.1155/2012/876545

Masoumi, H., Safavi, S. M., & Khani, Z. (2012). Identification and classification of plastic resins using near infrared reflectance. *International Journal of Mechanical and Industrial Engineering, 6*, 213–220. https://doi.org/10.5281/zenodo.1076916

MATLAB, V. (2018). *R2018a. 9.4. 0. Natick, MA, USA*: The MathWorks Inc.

Minaei, S., Shafiee, S., Polder, G., Moghadam-Charkari, N., van Ruth, S., Barzegar, M., et al. (2017). VIS/NIR imaging application for honey floral origin determination. *Infrared Physics & Technology, 86*, 218–225. https://doi.org/10.1016/j.infrared.2017.09.001

Mohd Khairi, M. T., Ibrahim, S., Md Yunus, M. A., & Faramarzi, M. (2018). Noninvasive techniques for detection of foreign bodies in food: A review. *Journal of Food Process Engineering, 41*(6), Article e12808.

Momin, M. A., Yamamoto, K., Miyamoto, M., Kondo, N., & Grift, T. (2017). Machine vision based soybean quality evaluation. *Computers and Electronics in Agriculture, 140*, 452–460. https://doi.org/10.1016/j.compag.2017.06.023

RASFF, E. (2014). *Rapid Alert system for food and feed (RASFF)* Accessed.

Rodgers, J., Fortier, C., Montalvo, J., Cui, X., Kang, S. Y., & Martin, V. (2010). Near infrared measurement of cotton fiber micronaire by portable near infrared instrumentation. *Textile Research Journal, 80*(15), 1503–1515. https://doi.org/10.1177/0040517510361799

Rong, D., Xie, L., & Ying, Y. (2019). Computer vision detection of foreign objects in walnuts using deep learning. *Computers and Electronics in Agriculture, 162*, 1001–1010. https://doi.org/10.1016/j.compag.2019.05.019

Senni, L., Ricci, M., Palazzi, A., Burrascano, P., Pennisi, P., & Ghirelli, F. (2014). On-line automatic detection of foreign bodies in biscuits by infrared thermography and image processing. *Journal of Food Engineering, 128*, 146–156. https://doi.org/10.1016/j.jfoodeng.2013.12.016

Sunoj, S., Igathinathane, C., & Visvanathan, R. (2016). Nondestructive determination of cocoa bean quality using FT-NIR spectroscopy. *Computers and Electronics in Agriculture, 124*, 234–242. https://doi.org/10.1016/j.compag.2016.04.012

Teye, E., Anyidoho, E., Agbemafle, R., Sam-Amoah, L. K., & Elliott, C. (2020). Cocoa bean and cocoa bean products quality evaluation by NIR spectroscopy and chemometrics: A review. *Infrared Physics & Technology, 104*, 103127. https://doi.org/10.1016/j.infrared.2019.103127

Trafialek, J., Kaczmarek, S., & Kolanowski, W. (2016). The risk analysis of metallic foreign bodies in food products. *Journal of Food Quality, 39*(4), 398–407. https://doi.org/10.1111/jfq.12193

Tsuta, M., Takao, T., SuGIYAMA, J., Wada, Y., & Sagara, Y. (2006). Foreign substance detection in blueberry fruits by spectral imaging. *Food Science and Technology Research, 12*(2), 96–100. https://doi.org/10.3136/fstr.12.96

Whiteley, K. S., Heggs, T. G., Koch, H., Mawer, R. L., & Immel, W. (2000). *Polyolefins. Ullmann's Encyclopedia of industrial chemistry*.

Xia, C., Yang, S., Huang, M., Zhu, Q., Guo, Y., & Qin, J. (2019). Maize seed classification using hyperspectral image coupled with multi-linear discriminant analysis. *Infrared Physics & Technology, 103*, 103077. https://doi.org/10.1016/j.infrared.2019.103077

Xu, J.-L., Esquerre, C., & Sun, D.-W. (2018). Methods for performing dimensionality reduction in hyperspectral image classification. *Journal of Near Infrared Spectroscopy, 26*(1), 61–75. https://doi.org/10.1177/0967033518756175

Xu, B., Fang, C., & Watson, M. (1999). Clustering analysis for cotton trash classification. *Textile Research Journal, 69*(9), 656–662. https://doi.org/10.1177/004051759906900906

Zhang, R., Li, C., Zhang, M., & Rodgers, J. (2016). Shortwave infrared hyperspectral reflectance imaging for cotton foreign matter classification. *Computers and Electronics in Agriculture, 127*, 260–270. https://doi.org/10.1016/j.compag.2016.06.023

Zhang, S., Zeng, X., Ding, T., Guo, L., Li, Y., Ou, S., et al. (2018). Microarray profile of circular RNAs identifies hsa_circ_0014130 as a new circular RNA biomarker in non-small cell lung cancer. *Scientific Reports, 8*(1), 1–11. https://doi.org/10.1038/s41598-018-21300-5