MDPI

Article

Nondestructive Detection of Codling Moth Infestation in Apples Using Pixel-Based NIR Hyperspectral Imaging with Machine Learning and Feature Selection

Nader Ekramirad [1] , Alfadhl Y. Khaled [1], Lauren E. Doyle [1], Julia R. Loeb [1], Kevin D. Donohue [2] , Raul T. Villanueva [3] and Akinbode A. Adedeji [1,*]

1   Department of Biosystems and Agricultural Engineering, University of Kentucky, Lexington, KY 40546, USA; nader.ekramirad@uky.edu (N.E.); alfadhl.alkhaled@uky.edu (A.Y.K.); ledo235@g.uky.edu (L.E.D.); jrlo244@g.uky.edu (J.R.L.)
2   Department of Electrical and Computer Engineering, University of Kentucky, Lexington, KY 40546, USA; kevin.donohue1@uky.edu
3   Department of Entomology, University of Kentucky, Princeton, KY 42445, USA; raul.villanueva@uky.edu
*   Correspondence: akinbode.adedeji@uky.edu; Tel.: +1-859-218-4355

Abstract: Codling moth (CM) (*Cydia pomonella* L.), a devastating pest, creates a serious issue for apple production and marketing in apple-producing countries. Therefore, effective nondestructive early detection of external and internal defects in CM-infested apples could remarkably prevent postharvest losses and improve the quality of the final product. In this study, near-infrared (NIR) hyperspectral reflectance imaging in the wavelength range of 900–1700 nm was applied to detect CM infestation at the pixel level for three organic apple cultivars, namely Gala, Fuji and Granny Smith. An effective region of interest (ROI) acquisition procedure along with different machine learning and data processing methods were used to build robust and high accuracy classification models. Optimal wavelength selection was implemented using sequential stepwise selection methods to build multispectral imaging models for fast and effective classification purposes. The results showed that the infested and healthy samples were classified at pixel level with up to 97.4% total accuracy for validation dataset using a gradient tree boosting (GTB) ensemble classifier, among others. The feature selection algorithm obtained a maximum accuracy of 91.6% with only 22 selected wavelengths. These findings indicate the high potential of NIR hyperspectral imaging (HSI) in detecting and classifying latent CM infestation in apples of different cultivars.

Keywords: apples; codling moth; hyperspectral imaging; near-infrared; machine learning; feature selection

1. Introduction

Apples are a very important fruit in the global produce market and industry. The United States of America is the second largest producer of apples, producing about 4.5 million tons of apples in 2020 [1], exporting 1 out of 3 apples grown, and averaging $1 billion annually on apple exports [2]. Additionally, apples are the most consumed fruit in the US, with the market value of about $5 billion in 2018 [2]. Because apple marketing is such a big business worldwide, preserving their quality to meet the ever increasing demands of consumers is essential. Codling moth (CM) *Cydia pomonella* (Lepidoptera: Tortricidae) is known to be the most devastating pest that infects apples [3,4]. It causes direct damage to the fruit's skin and pulp. CM is known to infest pome fruits, with a special preference for apples in almost every country the fruit is grown [3]. This larva enters the apple by feeding through the skin of the fruit, burrowing into the fruit's core to cause major damage [4]. If untreated, CM can result in up to a 50% loss in pre- and post-harvest apples [3]. Furthermore, production will only tolerate 1% of affected fruit [4], where if any apple infestation is found in some of the US' top importing countries, the whole shipment is rejected [5].

Detection of infestation, therefore, is very critical but the current manual random methods are inefficient.

Presently, apple quality assessment, including testing for possible insect infestation, is done at random, manually, and in a destructive manner. When assessing apples for packaging, inspectors visually examine the external qualities, scoring the apple surface to comply with certain specifications and tolerances for defects. To determine the internal quality, apples are cut in half to visually inspect their cross-sectional areas [6]. After testing, the used apples are discarded, wasting about three percent of the product [7]. In this way, detection of infestation is time-consuming, costly, subjective, and laborious, and yet does not assure that the batch will be pest free. Currently, machine vision has been implemented to monitor the outside of the apple at low-cost and rapid speed, but issues arise in interference of the sample's color and the presence of stem and calyx [8], coupled with the inability to inspect the internal qualities of the apple where most pest infestation damage resides. To correct this issue and increase detection efficiency, a better method is needed to identify both internal and external damages to apples by a pest such as CM.

Whereas current techniques are very wasteful, nondestructive techniques preserve the fruit, give a definitive result, and can easily look at the whole batch to ensure no bad apple gets through to the supply chain. Using forms of nondestructive testing to assess certain qualities of apples is not new. Some of the nondestructive techniques used on apples include hyperspectral imaging (HSI) [9,10], vibro-acoustic signaling [11,12], ultrasonic acoustic detection [13], delta absorbance meter [14], machine vision [15,16], and spectroscopy [10,17]. Each technique has its advantages and disadvantages; however, the one that stands out is HSI.

While some of these techniques have been used to detect and classify infested apples, such as the work done with acoustic emissions [18] and vibro-acoustic signaling [12], HSI is ideal because it conveys additional useful information for nondestructive applications. HSI combines the capability of spectroscopy and machine vision techniques. Spectroscopy is used to create a spectrum of data based on light absorbance at different wavelengths [19,20]. This is useful in finding specific chemical components but lacks a sense of location or direction since the device scans at a single point [20–22]. Additionally, machine vision converts photographic scanning of 3D objects into 2D images by capturing and documenting the reflected light into grayscale and RGB color [23]. Machine vision is great at scanning objects quickly and acquiring a sense of location, allowing for analysis of spatial qualities such as size, shape, and color [24,25]. However, it only looks at the surface of the object in primary color [20]. HSI uses the best parts of both techniques, looking at the reflectance at every point of the image showing a spectrum of reflectance for each pixel in the spatial image while still retaining the analytical benefits of the two techniques [26]. Each hyperspectral image is a three-dimensional data cube (3D hypercube) with X and Y coordinates as the spatial information and λ as the spectral data. HSI not only has the capability to detect infestation on and in the sample under test, but also is used to find the exact location of infestation due to the spatial information and the ability to evaluate the different levels of pixels in the images [27,28].

HSI has been investigated as a rapid and relatively low-cost nondestructive technique in the quality assessment of apples. This application mainly falls into three categories including external quality, internal quality, and pest detection [29,30]. Regarding external quality of apples, HSI was used to evaluate defects (e.g., surface defects and bruising) because of its ability to penetrate beneath the apple's skin. For example, bruises on apples are detectable in the range of shortwave-near-infrared (SNIR), particularly around 675 nm and 960 nm, which represent the region for carotenoids, chlorophyll pigments, sugar, and water content [31]. This reveals that the bruise on the apple causes large imbibition of water and total sugar contents at the early stage of bruising, and then causes assimilation of chlorophyll pigments and carotenoids in the subsequent stage [32]. Thus, the application of HSI to detect bruises on apples can reduce or prevent further losses from cross-contamination of others by damaged apples. HSI also has been used for nondestructive prediction of

internal quality of apples such as the nutritional value, texture, and flavor components, and in estimating physiochemical parameters such as vitamin and sugar content [33]. Additionally, HSI has been tested for safety assurance of apples through effective detection of pests [30,34]. In 2021, Ekramirad et al. [30] determined the best classification result for CM detection in apples using NIR HSI and the mean spectra extraction method on dataset consisting of three apple orientations (clayx, sides and bottom). Applying partial least squares-discriminant analysis (PLS-DA) classifier, they obtained an overall validation accuracy of 81.04%. While the calyx and side orientations had similar classification rates of about 80%, the stem orientation gave the lowest classification accuracies. These results are better than the findings of Rady et al. [34] who achieved a maximum classification rate of 74% using the side orientation of apples, using all the spectral wavelengths in the Vis-NIR range. However, they reported that by reducing the data dimensions using the sequential forward selection method, their classification accuracy was enhanced to 82%. In the same study conducted by Ekramirad et al. [30], they applied a second pixel-wise method instead of the mean spectra extraction and found an accuracy as high as 98.2% in classifying the infested and healthy pixels using the random forest (RF) classifier. However, they had manually segmented a rectangular ROI around the calyx end of the apple sample to extract pixels data by spectra for infested and healthy, which is a cumbersome subjective task that increases the processing time, hence there is a need to develop a new method for automatic extraction of target pixels. Additionally challenges to be considered in developing an automatic algorithm based on HSI for the classification of CM-infested apples include the following. First, since the shape and size of ROI in HSI affect the measurement performance [35], a proper geometry should be selected for the ROI. Second, the infested region should be well localized for accurate labeling as an infested class. Therefore, to address these issues a special procedure was developed in this study to automatically extract pixel-wise ROI around the calyx of the apples, which is the usual point of entry for the CM larvae.

Normally, HSI technique produces large spectral data and for this, its analysis is associated with the utilization of mathematical or statistical methods to make it readable and to discover useful information about the data. However, such analyses are time consuming because of the large size of the datasets. In computational intelligence methods, dimension reduction is often used to optimize data processing time, reduce dimensionality, and enhance data generalization [36,37]. Principal component analysis (PCA) is the main dimensional reduction step used for hyperspectral data to transform its spectra into some independent features. Moreover, some approaches have been developed to optimize the HSI to perform real-time hyperspectral data reduction using the extraction of predefined features [38,39]. This is carried out by real-time multiplication of the acquired spectral data by a feature extraction operator (vector-to-scaler) consisting of the desired features (less than ten as opposed to >100 spectral features in the hypercube) predefined by experiment. The optimal feature extraction operators are usually obtained by mathematical and statistical methods, such as that applied in PCA. Some of the most frequently used classification methods include k-nearest neighbor (kNN), Linear discriminant analysis (LDA), Quadratic discriminant analysis (QDA), and Naïve Bayes (NB) [40]. In addition, PLS-DA has been proven in many studies to be a powerful classification method for HSI data analysis with high-dimensional data [41]. Additionally, the ensemble methods such as RF and gradient tree boosting (GTB) can integrate weak classifiers to achieve powerful anti-noise classifiers [42].

While HSI has been widely applied for quality assessment of agricultural products, there is no report on the application of NIR hyperspectral imaging combined with feature selection algorithm and various machine learning algorithms to detect CM latent infestation in apples. The infestation of plant tissue by pests can induce different defense mechanisms, such as hypersensitive reactions, production of metabolites and proteins, and altered plant tissue structure, leading to various reflectance spectral signatures that can be measured and localized by spectral imaging methods [43]. Thus, the main objective of the current

work was to develop and validate a robust model for the accurate detection of latent CM infestation in apples based on the NIR HSI technique. The specific objectives were to: (1) develop an automatic procedure for pixel-based extraction of infestation region on apple to address the issues related to manual segmentation of the infested area, (2) compare the results of the classification method for three major apple cultivars of Fuji, Gala, and Granny Smith, and (3) select some optimal wavebands for reducing the dimensions of the large scale HSI data leading to a multispectral imaging system.

2. Materials and Methods

2.1. Sample Preparation

The apple samples used in the experiment were USDA-certified organic Gala, Fuji, and Granny Smith cultivars purchased from a commercial market in Princeton, KY, USA in October 2020. After careful inspection, 60 sample apples similar in size, diameter, and shape, and without infestation and mechanical damage, were chosen from each cultivar. The apples were then disinfected against fungal and bacterial decay by washing in a 0.5% (*v/v*) sodium hypochlorite solution [44]. The samples were rinsed with distilled water and dried in open air at ambient conditions in the laboratory (Department of Entomology, University of Kentucky, Princeton, KY, USA). To artificially infest the apples, newly hatched neonate of CM larva was placed near the calyx end of each apple in an isolated cup (8 cm bottom diameter, 10 cm top diameter, 10 cm high) with a plastic lid. Apples of each cultivar were divided into 20 control and 40 infested groups and stored in an environmental control chamber at 27 °C and 85% relative humidity for three weeks to cause infestation to occur. The hyperspectral data acquisition was carried out in the Food Engineering lab at Biosystems and Agricultural Engineering Department, University of Kentucky, Lexington, KY, USA.

2.2. HSI System and Image Acquisition

A HSI system based on shortwave near-infrared (NIR) bands was used to acquire the hyperspectral data of all apple samples—control and infested (Figure 1). The HSI system consisted of a NIR spectrograph with a wavelength range from 900 nm to 1700 nm and a spectral resolution of 3 nm (N17E, Specim, Oulu, Finland), a moving stage driven by a stepping motor (MRC-999-031, Middleton Spectral Vision, Middleton, WI, USA), a 150 W halogen lamp (A20800, Schott, Southbridge, MA, USA), an InGaAs camera (Goldeye infrared camera: G-032, Allied Vision, Stradtroda, Germany) mounted perpendicular to the sample stage and a computer with data acquisition and analysis software (FastFrame™ Acquisition Software, Middleton Spectral Vision com, Middleton, WI, USA). Three scanning orientations of the stem, calyx, and side of each apple were captured during hyperspectral image acquisition. To acquire clear images, the parameters of the sample stage speed, the exposure time of the camera, the halogen lamp angle, and the vertical distance between the lens and the sample were set to 10 mms^{-1}, 40 ms. 45°, and 25 cm, respectively. Samples were placed on the sample stage and captured in a line scanning or pushbroom mode. The acquired hyperspectral images contained 256 wavelength bands as "*.raw" file along with a header file as "*.hdr".
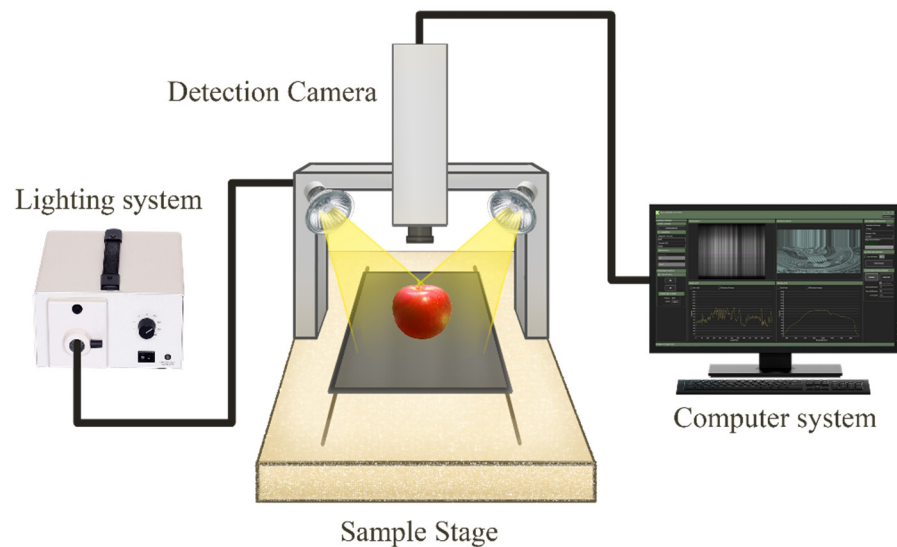
Figure 1. Schematic of the hyperspectral imaging system.

2.3. Preprocessing of Hyperspectral Images

2.3.1. Image Calibration

This is needed to correct the acquired raw images with the white and dark reference images to eliminate the influence of illumination and dark current of the camera. The reference image data were obtained after the samples were scanned every day. The dark reference images were obtained by completely covering the lens of the HSI system while turning off the lights. The white reference images were acquired using a polytetrafluoroethylene (PTFE) Teflon plate of 99% reflectivity and 10 mm thickness placed on the black sample stage. The calibration was done based on the following equation:

$$R = \frac{R_0 - R_d}{R_w - R_d} \tag{1}$$

where R_0 is the raw hyperspectral image, R_d is the dark image, and R_w is the white reflectance image [45,46].

2.3.2. Infestation Region Acquisition

After the acquisition and correction of the hyperspectral images, the spectral information of the infested and healthy tissue was automatically extracted from ROIs using the algorithm described in Figure 2. Since the CM larvae, especially the first generation, mostly enter apples from the calyx end [47] and the initial results by Ekramirad et al. [30] showed that the highest infestation classification accuracy achieved in images from the calyx view, the ROI to extract infested pixels was segmented around the calyx end. This novel method can select the complete infested region with pixels in the healthy region as few as possible to obtain a precise infested region for subsequent classification. To do this, first the background and calyx end were segmented out using the image at 1084 nm wavelength to obtain a masked image in a binary image format using a binary thresholding method. To obtain a solid area around the calyx, the morphological image processing of erosion operation was applied. Then, the center of the calyx area was localized using mathematical operations to calculate the centroid of the eroded area. Finally, having the orientations of the center of the calyx area, a circular region with 50 pixels diameter was drawn as a mask binary region. The circular ROI fits with the spherical shape of apple fruit and it has been shown that a round ROI gave higher accuracy and predictive capability than square ROI in HSI on apples [35]. Then, the circular masked image was applied on each image of the hypercube (i.e., all the 256 wavebands) to obtain the calyx area with other

pixels equal to zero. The spectrum for each pixel in the circular ROI was then extracted and unfolded and labeled for building the dataset to develop machine learning models.
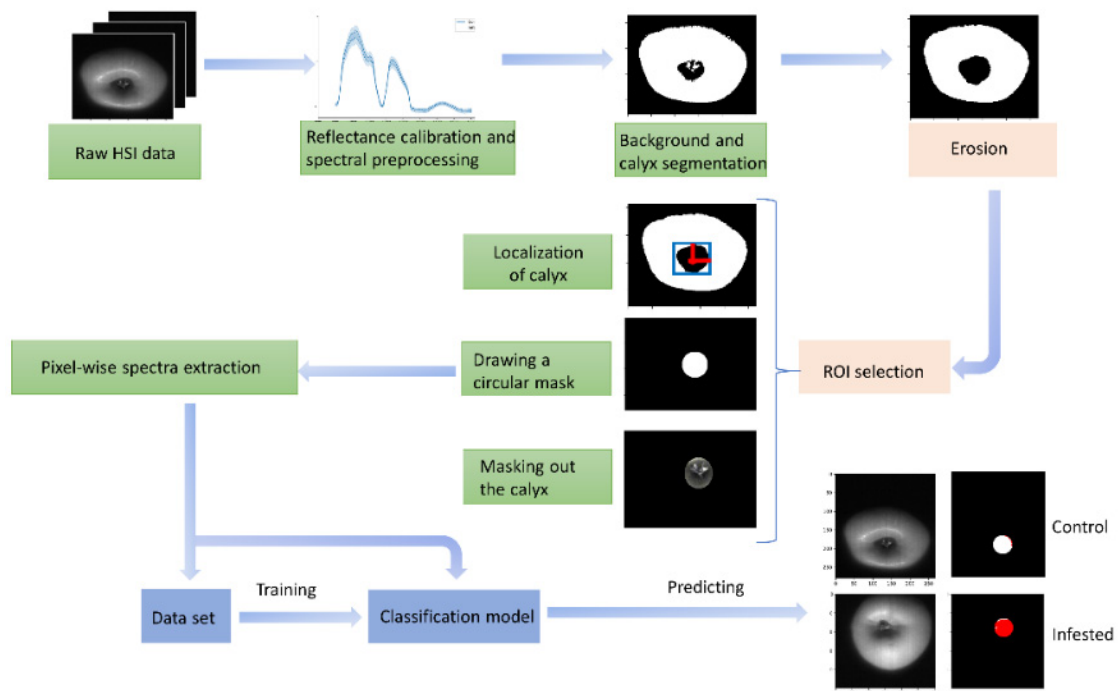
Figure 2. Flowchart of apple infestation area acquisition around calyx end for building the classification model. HIS: hyperspectral imaging; ROI: region of interest.

2.3.3. Spectral Extraction and Preprocessing

To obtain the spectral characteristics of apples, the spectral for each pixel inside the ROI was extracted in the form of reflectance intensity versus wavelength and then labeled as either infested or healthy spectral signature. After spectral data extraction, pre-processing was carried out by wavelength trimming, maximum normalization, a Savitzky-Golay smoothing filter, and mean centering to remove the noisy wavelengths at the edges of each spectrum, to get all data to the same scale, to account for particle size scattering and path length difference effects, and to keep only significant features, respectively. The maximum normalization was carried out by dividing each spectrum by the maximum value [29]. The Savitzky-Golay method involves the application of the second-order polynomial and the filter window of length 31.

2.3.4. Dimensionality Reduction

With respect to data architecture in HSI, high dimensional images with fixed training sample size can result in overfitting problems leading to degraded classification rates. It is usually the case when the size of training samples is limited in comparison to the feature space size resulting in low generalization of the results and overfitting problems [48]. The higher the dimensionality of the model, the higher the likelihood of over-fitting. Therefore, hyperspectral data size compression, especially spectral dimensionality reduction is usually required to achieve better data visualization, save storage space, eliminate redundant data, and avoid model over-fitting. Principal component analysis (PCA) was used in this study as the dimensionality reduction technique. PCA is a transformation of the data through an axis rotation, in the direction of maximum variance. Successive (principal components) PCs are the linear combinations of the variables with maximum variance, which are orthogonal to the previously computed components. The total variance can be represented in a significantly small number of components (extracted features). The PCs are the eigenvectors of the covariance matrix of the data, and the associated variance is represented

in the corresponding eigenvalues. The PCs are orthogonal and have successively ordered variances. PCA transforms multivariate data into a new coordinate system to produce new uncorrelated orthogonal variables which are called PCs or loadings. These PCs are arranged according to their eigenvalues, with the 1st PC having the highest variance, the 2nd PC containing the highest residual variance, and so on [49]. As most of the information is included in the first PCs, eliminating PCs with a small variance will remove unnecessary information. The advantages of this technique over nonlinear dimensionality reduction techniques include being easy to apply, invertible, and volume-preserving transformation.

2.3.5. Spectral Variable Selection

Acquired data from HSI systems usually have high dimensions both spatially and in spectral form. These data contain highly correlated continuous wavelengths with a lot of redundancy resulting in high complexity and computation costs. Feature selection methods through selecting optimal wavelengths can provide the informative wavelengths to build fast and simpler multispectral imaging models. In this study, the optimal wavelength selection was conducted using the sequential stepwise selection method. In this method, which is a wrapper method, a specific machine learning classifier is fitted to the dataset. It acts as a greedy search approach through evaluating all the possible combinations of features against the evaluation criterion, which is the performance measure such as accuracy, precision, sensitivity, etc. Finally, it selects the combination of optimum features that gives the best results for the specified machine learning algorithm.

*2.4. Development of Machine Learning Classifiers*

Having the spectral data for healthy and infested pixels from all the samples, the labeled dataset was built by organizing pixels (observations) as rows, and the features (spectral data) as columns. The predictor variables (features) are the spectral data while the dependent variables are the classes, namely control and infested. After building the dataset, Kennard & Stones algorithm was implemented to split 70% of the data as the training and 30% as the validation datasets. The Kennard & Stones method has been widely used in chemometrics and spectroscopy, and it has been proven to give good performance in separating spectral data into training and test sets [50]. Different machine learning classification algorithms including Linear discriminant analysis (LDA), k-Nearest Neighbors (kNN), PLS-DA, and two ensemble methods, namely Random forest (RF) and GTB, were performed and compared for their classification accuracies. For evaluating the performance of the various models in this study, five-fold cross-validation was used. The metrics used for assessing the classification performance of the models included precision, recall, and the F1-score. Precision is the positive predicted value and quantifies the correctly classified pixels as infested (the fraction of true positives out of all positive predictions whether true positive or false positive). While precision gives a quantitative measure of how exact the classifier's prediction is, recall (or sensitivity) helps avoid missing any undetected infested samples. Recall is the true positive rate that relates to the number of pixels belonging to the infested area that were predicted as positive (true positive) and those that the model incorrectly does not capture as infested (false negative). F1-score is the harmonic mean of recall (R) and precision (P), calculated as: $F1 = 2RP/(R + P)$, reflecting the balance between the classifier's precision and recall [51]. The F1-score can be used to evaluate the entire model, considering both the precision and recall, making it a sensitive metric to changes in the data distribution and ratios.

All algorithms used in this study for pre-processing and data analysis and post-processing were performed on Python 3.7 (Python Software Foundation, https://www.python.org/ (accessed on 15 October 2021)) platform and in Jupyter Editor Notebook. The open-source libraries of spectral, Numpy, Scikit-learning, and Matplotlib were used in this work.

3. Results and Discussion

3.1. Spectral Analysis

Figure 3 shows the typical mean spectra for healthy and infested apple samples as normalized reflectance versus wavelength. While the average spectra of the control and the infested samples have a similar trend and curve variation tendency, the reflectance of the healthy samples is remarkably higher than the one for the infested samples (or the absorbance of the infested samples are higher than the healthy ones). Thus, this NIR reflectance difference for healthy and infested samples shows the potential to be applied for the binary classification. As reported by several authors [26,52,53], the absorbance of defective samples was higher than the healthy ones due to the cellular structure difference. As a result of plant tissue infestation, there will be biochemical, tissue structure, and pigment composition changes, leading to the different spectral signatures [54].
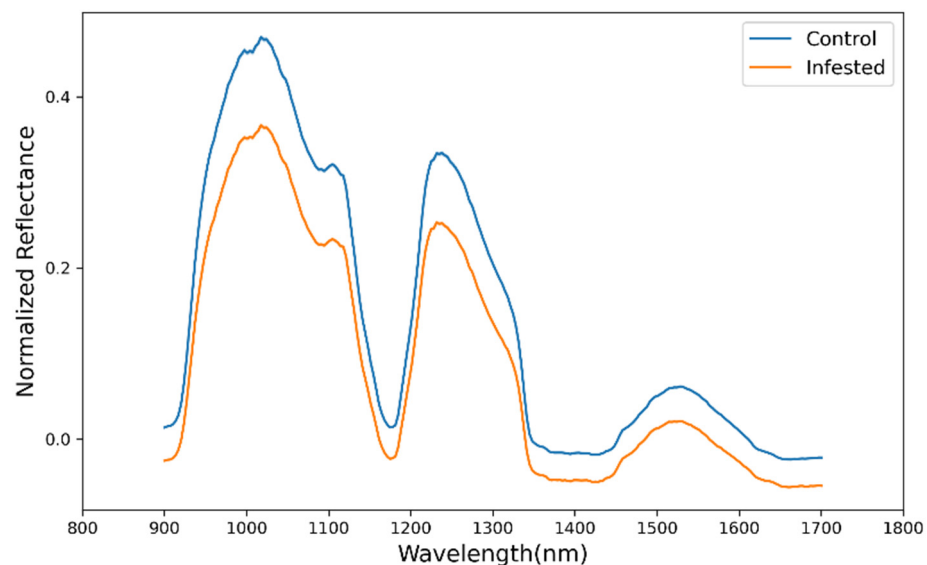
Figure 3. Mean reflectance spectra of control and CM-infested samples acquired by near-infrared hyperspectral imaging (NIR HIS). CM: codling moth.

As shown in Figure 3, there are some distinct absorption valleys around 950, 1200, and 1400 nm in the mean spectra of both sample classes. The absorption at about 950 and 1200 nm relates to the first overtones of O-H band in water molecules [55,56]. The absorption at around 1400 nm is attributed to the combination of the second overtone of C-H and the first overtone of O-H [9]. The spectral curve of samples in this study agrees with the finding of other studies on apples in the same spectral range [9].

3.2. Pre-Processing and Feature Extraction Results

For classification of infested samples of the three different cultivars of apples, PCA was performed on the preprocessed spectra before building the classification models. Results showed that the sum of the variance explained by the first three PCs for all the cultivars was more than 98% of the total variance. It means the sum of the accumulative contribution rate of the first three PCs represents 98% of the total variability of the spectral data, so it could be a reasonable way to recognize patterns in the tested samples using these limited number of PCs. As shown in Figure 4, the control and infested samples were well clustered with some minor overlaps between them. Moreover, PCA score values for infested apples were tightly clustered over the two first PCs space, while scores for control fruits were more widely scattered. Therefore, the machine learning models for classification of apples were built using the extracted PCs as features. Moscetti et al. [57] reported similar PCA score plot trends for non-infested and infested olive fruits using the NIR spectroscopy for the mean spectra of the whole fruit where the first two PCs accounted for 98.3% of

the total variance. They used the pre-processing steps of multiplicative scatter correction, a Savitzky-Golay smoothing filter, and mean centering followed by PCA dimensionally reduction and LDA, QDA, and kNN classification. Additionally, the results of Keresztes et al. [29] for pixel-based apple bruise detection using shortwave infrared HSI showed that the three first PCs represented 98.36%, 1.24%, and 0.15% of the total variance in the data. They also used the pre-processing methods of multiplicative scatter correction, Savitzky-Golay smoothing, and mean centering before PCA dimensionally reduction followed by PLS-DA classification.
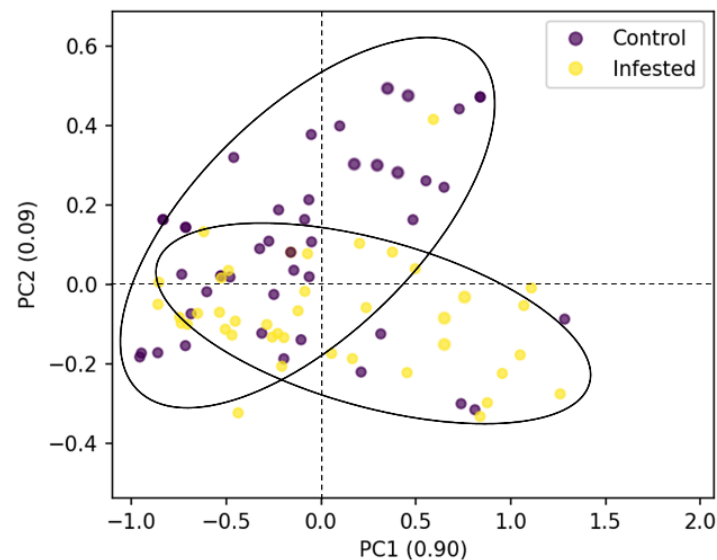
Figure 4. Principal component analysis of two types of apple sample tissues for Fuji cultivar computed from the mean spectral of the whole fruit.

3.3. The Results of Machine Learning Classification

In order to compare the results of different approaches for classification of apple samples, classification results of three approaches are presented in this section. Table 1 provides the classification results of infested and healthy Fuji apples using the mean spectra extraction method and for three orientations of stem, calyx, and side of apple along with the data for all the orientations together. In this method the reflectance spectrum for each pixel was extracted and then the average of all the pixels were calculated as the mean reflectance spectra for that sample. As shown in Table 1, the best classification result for the mean spectra extraction method was achieved using the data for all the orientations of apples and by PLS-DA and RF classifiers, with a validation accuracy of 92%. While the calyx and side orientations had similar classification rates of 88.9%, the stem orientation gave the lowest classification accuracies. These results are better than the findings of Rady et al. [34] who achieved a maximum classification rate of 74% using the mean reflectance spectra from Vis-NIR hyperspectral images of the side views of apples.

In the second approach, pixels from infested apples were localized and segmented manually using a 10×10 rectangular ROI around calyx end of samples. To do this, the ROI were selected in the images and the spectrum of each pixel in the ROI was extracted. Thus, a total of 100 spectra for each infested or control apple were extracted and labeled to build the classification dataset. Table 2 shows the classification results for the control and infested pixels in apples using the manual ROI selection method. The result gave a good performance with the accuracy of up to 99.24% for the ensemble classifiers. These findings are in good agreement with those of Munera et al. [53] who reported an overall accuracy of up to 97.5% in classifying the healthy and defective pixels in hyperspectral images of loquat fruits using the manual ROI selection method.

Table 1. Results of the PCA-based classification of control and infested samples for training and validation sets based on mean spectra for each sample.

Sample Orientation	Classifier [1]	Training Set (%)			Validation Set (%)		
		Precision	Recall	Total Accuracy	Precision	Recall	Total Accuracy
Stem	LDA	95.00	94.00	94.70	57.00	58.00	62.50
	kNN	58.00	57.00	57.90	36.00	42.00	62.50
	RF	**100**	**100**	**100**	**83.00**	**92.00**	**87.50**
	AdaBoost	95.00	95.00	94.70	75.00	83.00	75.00
	PLS-DA	**100**	**100**	**100**	**83.00**	**92.00**	**87.50**
Calyx	LDA	90.00	90.00	90.50	78.00	78.00	77.80
	kNN	63.00	61.00	61.90	68.00	68.00	67.00
	RF	100	100	100	88.00	83.00	83.30
	AdaBoost	**100**	**100**	**100**	**92.00**	**88.00**	**88.90**
	PLS-DA	90.00	90.00	90.50	78.00	78.00	78.00
Side	LDA	100	100	100	83.00	80.00	77.80
	kNN	86.00	80.00	80.00	75.00	60.00	55.60
	RF	100	100	100	83.00	80.00	77.80
	AdaBoost	100	100	100	83.00	80.00	77.80
	PLS-DA	**100**	**100**	**100**	**90.00**	**90.00**	**88.90**
All	LDA	80.00	80.00	79.00	71.00	73.00	72.00
	kNN	76.00	76.00	76.30	70.00	71.00	72.00
	RF	100	100	100	91.00	94.00	92.00
	AdaBoost	100	100	100	88.00	91.00	88.00
	PLS-DA	**98.00**	**98.00**	**98.00**	**91.00**	**94.00**	**92.00**

[1] LDA: Linear Discriminant Analysis, kNN: k-Nearest Neighbors, RF: Random Forest, PLS-DA: Partial Least Squares-Discriminant Analysis. Bolded line indicates the best result.

Table 2. Results of the PCA-based classification of control and infested samples for training and validation data sets based on manually selected ROI.

Classifier [1]	Training Set (%)			Validation Set (%)		
	Precision	Recall	Total Accuracy	Precision	Recall	Total Accuracy
LDA	72.20	79.20	75.24	71.60	78.40	74.64
kNN	100	99.20	99.52	99.60	98.80	99.06
RF	100	100	100	99.20	99.60	99.24
AdaBoost	100	100	100	98.00	98.4	98.20
PLS-DA	84.60	88.80	86.40	80.60	82.60	80.18

[1] LDA: Linear Discriminant Analysis, kNN: k-Nearest Neighbors, RF: Random Forest, PLS-DA: Partial Least Squares-Discriminant Analysis.

3.4. Performance of Classification Models Based on Apple Cultivar

The detailed performance of different models for classification of CM-infested and healthy samples of three apple cultivars, namely Gala, Fuji, and Granny Smith using the automatic pixel-based method developed and implemented in this study is shown in Table 3. All the classifiers gave higher results using the PCs as the input variables except for LDA and PLS-DA, which gave slightly better classification rates using raw data without dimensionality reduction. The GTB ensemble method yielded the highest classification rates for all three cultivars, reaching as high as 97.4% accuracy of the validation set for Fuji apples. Similar classification rates were achieved by Saranwong et al. [58] using HSI in the range of 400–1000 nm in reflectance mode to assess fruit fly larvae infestation in mango. Using a discriminant analysis classifier, they obtained a validation classification rate of up to 99.1% and 94.3% for infested and healthy fruits, respectively. Haff et al. [59] also researched the same insect in mango using the same method and the classification

rates reached 99% for infested samples. While they achieved a high classification rate in both studies, they artificially created pores on the fruit in a grid pattern to expose the fruit to the pest insects to have a priori knowledge of the locations of infestation. Then they extracted the spectra from the pore locations and compared them with the spectra from healthy areas to identify the spots generated in hyperspectral images of mangoes infested with fruit fly larvae. They reported that classifying the samples which were deliberately infested following a predefined pattern, and the algorithm relying on that pattern in the images would be useless in real-world applications. In a similar study to our current research, Rady et al. [34] studied the ability of Vis-NIR HSI (400–900 nm) in the reflectance mode for the detection and classification of CM infestation in GoldRush apples. Their best classification rates were obtained using decision trees at five selected wavelengths with an overall classification rate of 82%. Their relatively low classification rate can be related to the limited spectral range used to detect internal and invisible defects in samples. Moreover, they used the traditional image processing-based method combined with the mean spectral extraction for the whole sample. Thus, the broader spectral range as well as the pixel-based method for extracting the spectral signature of infested regions could be the reason behind the higher classification rates in the current study. In another study on apples for a different application, Che et al. [42] used pixel-based Vis-NIR HSI to classify the bruised Fuji fruits. They also reached their best accuracy of 99.90% with the ensemble method.

Table 3. Classification accuracy (%) for validation data set based on automatically selected pixels for three apple cultivars.

Classifier	Raw Data (No Dimensionality Reduction)				PCA-Based			
	Gala	Granny Smith	Fuji	All	Gala	Granny Smith	Fuji	All
LDA	65.38 ± 0.62	72.24 ± 0.23	70.46 ± 0.72	69.22 ± 0.10	65.38 ± 0.62	70.38 ± 0.17	66.94 ± 0.33	68.70 ± 0.14
SVM	80.18 ± 0.06	76.42 ± 0.17	81.40 ± 0.44	72.54 ± 0.36	82.60 ± 0.70	77.20 ± 0.18	81.62 ± 0.33	73.84 ± 0.39
kNN	93.72 ± 0.19	93.26 ± 0.15	95.46 ± 0.32	89.12 ± 0.12	93.80 ± 0.15	93.30 ± 0.07	95.69 ± 0.26	88.84 ± 0.11
RF	89.66 ± 0.19	89.04 ± 0.18	91.52 ± 0.27	82.82 ± 0.14	94.28 ± 0.31	93.22 ± 0.25	96.62 ± 0.13	89.74 ± 0.13
GTB	92.32 ± 0.37	91.00 ± 0.25	94.68 ± 0.39	84.66 ± 0.18	94.76 ± 0.16	93.66 ± 0.18	97.36 ± 0.28	90.00 ± 0.23
PLS-DA	62.76 ± 0.66	71.64 ± 0.24	68.56 ± 0.15	69.14 ± 0.15	62.76 ± 0.66	71.34 ± 0.16	66.92 ± 0.35	68.72 ± 0.16

PCA: principal component analysis, LDA: Linear Discriminant Analysis, SVM: support vector machine, kNN: k-Nearest Neighbors, RF: Random Forest, GTB: Gradient tree boosting, PLS-DA: Partial Least Squares-Discriminant Analysis.

Table 4 summarizes some important performance evaluation indices for the best classifier (GTB) for the three apple cultivars. The most important metric for the detection of a pest of concern such as CM is recall or sensitivity which reflects the amount of incorrectly classified infested samples (false negative) or the truly infested samples that were not detected as infested and were classified as healthy. As it is shown in Table 4, the recall values for infested samples are higher than the precision values for all the cultivars reaching as high as 0.98 (98%).

Table 4. Classification performance of gradient tree boosting for control and infested samples for three apple cultivars based on automatically selected pixels for three apple cultivars.

Cultivars	Classes	Precision	Recall	F1-Score	Overall Accuracy (%)
Fuji	Control	0.98	0.96	0.97	97.36
	Infested	0.97	0.98	0.97	
Gala	Control	0.93	0.93	0.93	94.76
	Infested	0.95	0.96	0.95	
Granny Smith	Control	0.91	0.90	0.91	93.46
	Infested	0.95	0.95	0.95	

3.5. Optimal Wavelength Selection

As mentioned above, the optimal wavelengths were selected from the whole spectra by the sequential forward selection (SFS) method to minimize variable collinearity and select the most informative variables. This algorithm started with one wavelength and then added a new one in each iteration process, and a specified number of wavelengths were selected at the end. The selections of optimal wavebands are shown in Figure 5 and Table 5. The results in Figure 5, obtained by applying SFS, illustrate a graph of classification accuracy changing with increasing the number of selected wavelengths. As shown, when 22 wavelengths variables were selected, the classification performance rate approaches an asymptote while the number of selected variables was significantly less than the raw spectral data (356 wavebands). Therefore, the optimal variable wavelength subset, which consists of 977.2, 983.9, 1050.9, 1064.3, 1081.0, 1151.28, 1184.6, 1228.0, 1248.1, 1288.1, 1351.4, 1447.9, 1530.9, 1554.2, 1574.1, 1590.7, 1627.1, 1647.0, 1653.7, 1657.0, 1663.6, and 1680.2 nm, was determined for classifying the CM infestation on Fuji apples, while the corresponding number of sampled variables was 22. The first 22 wavelengths for apples are mainly distributed around 1000, 1200, 1550, and 1650 nm.
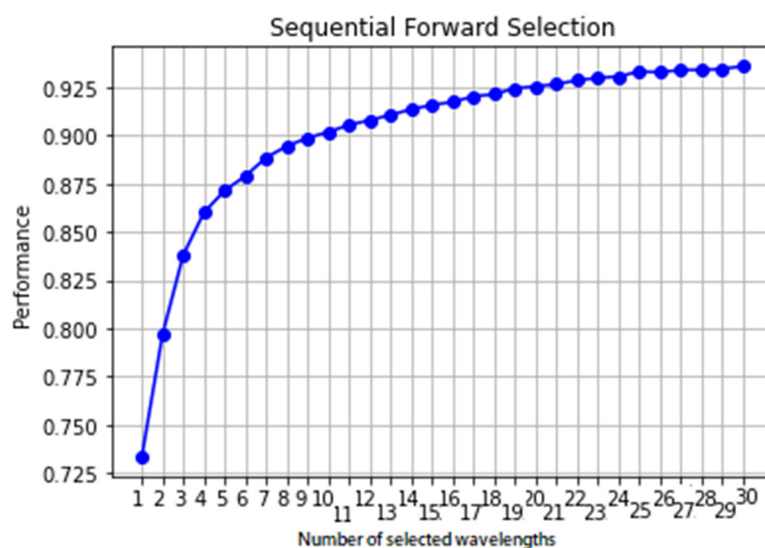
Figure 5. Classification performance (accuracy) as a function of the number of wavelengths.

After selecting optimal wavelengths by SFS, the selected optimal wavelengths carrying the most valuable information in the spectra were considered as the input variables to build the ensemble classifier model. Additionally, to further evaluate the representativeness of the chosen optimal size of the validation set, the classification results of the ensemble model based on the different number of optimal wavelengths for each of the three cultivars were compared (Table 5).

The classification accuracy of 91.6% for the validation set, which is very close to the maximum classification accuracy obtained with the whole range of wavelengths (97.4%), further validated the representativeness of the chosen optimal size of the data set. While this result is better than the results of Rady et al. [34] who achieved an accuracy of 82% in classification of CM-infested apples using Vis-NIR HSI, their best classification accuracy was obtained using only five selected wavelengths. It is worth noting that the number of wavelengths in current study reduced from 356 to 22, which only accounts for 6.17% of the total wavelengths, making the simplified model better than the model developed using the full spectra. Overall, results indicate that this is an effective way to select optimal wavelengths to build discriminant models by SFS, with a potential reduction in computational cost and relatively satisfying model performance.

Table 5. Classification performance of selected optimal wavelengths.

No. of Wavelengths	Gala		Granny Smith		Fuji	
	Selected Wavelengths (nm)	Classification Accuracy	Selected Wavelengths (nm)	Classification Accuracy	Selected Wavelengths (nm)	Classification Accuracy
30	900.1, 903.5, 920.3, 970.6, 997.4, 100.7, 1014.1, 1071.0, 1077.7, 1261.4, 1278.1, 1281.4, 1298.1, 1324.7, 1328.1, 1361.4, 1384.7, 1408.0, 1447.9, 1464.5, 1447.8, 1477.8, 1627.1, 1647.0, 1653.7, 1657.0, 1663.6, 1666.9, 1676.8, 1693.4	88.5%	900.1, 916.9, 977.2, 1010.7, 1020.8, 1030.8, 1047.6, 1074.3, 1178.0, 1181.3, 1204.7, 1274.7, 1284.7, 1294.7, 1298.1, 1304.7, 1308.1, 1371.4, 1414.6, 1471.1, 1481.1, 1494.4, 1653.7, 1660.3, 1666.9, 1673.5, 1680.2, 1683.5, 1686.8, 1693.4	87.7%	977.2, 980.6, 1044.2, 1074.3, 1077.7, 1081.0, 1137.9, 1147.9, 1151.2, 1211.3, 1264.7, 1294.7, 1314.7, 1344.7, 1348.0, 1381.3, 1421.3, 1507.7, 1530.9, 1544.2, 1560.8, 1580.7, 1623.8, 1630.5, 1647.0, 1650.3, 1653.7, 1657.0, 1663.6, 1673.5	92.4%
22	923.6, 973.9, 1000.7, 1067.6, 1081.0, 1084.4, 1127.8, 1268.1, 1281.4, 1308.1, 1351.4, 1401.3, 1411.3, 1461.2, 1491.1, 1607.3, 1643.7, 1663.6, 1670.2, 1676.8, 1690.1, 1693.4	87.8%	903.5, 916.9, 987.3, 1047.6, 1081.0, 1131.2, 1141.2,1181.3, 1204.7, 1274.7, 1288.1, 1304.7, 1371.4, 1467.8, 1471.1, 1481.1, 1643.7, 1673.5, 1680.2, 1683.5, 1686.8, 1693.4	87.5%	977.2, 983.9, 1050.9, 1064.3, 1081.0, 1151.28, 1184.6, 1228.0, 1248.1, 1288.1, 1351.4, 1447.9, 1530.9, 1554.2, 1574.1, 1590.7, 1627.1, 1647.0, 1653.7, 1657.0, 1663.6, 1680.2	91.6%
15	903.5, 990.6, 997.3, 1071.0, 1084.4, 1281.4, 1294.7, 1371.4, 1384.7, 1447.9, 1477.8, 1663.6, 1673.5, 1680.2, 1690.1	86.2%	1010.7, 1081.0, 1131.2, 1181.3, 1184.6, 1281.4, 1298.1, 1491.1, 1657.0, 1663.6, 1670.2, 1680.2, 1683.5, 1686.8, 1693.4	86.3%	977.2, 983.9, 1050.9, 1074.3, 1081.0, 1311.4, 1381.3, 1401.3, 1447.9, 1507.7, 1627.1, 1637.1, 1647.0, 1653.7, 1673.5	91.0%
5	997.3, 1084.4, 1281.4, 1663.6, 1693.4	81.5%	1014.1, 1274.7, 1494.4, 1683.5, 1693.4	80.7%	983.9, 1050.9, 1311.4, 1653.7, 1663.6	86.2%

4. Conclusions

In this study, machine learning models were developed to perform classification of CM infestation in apples using pixel-level NIR hyperspectral image data. Combined NIR HSI, machine learning and image processing methods were applied to discriminate healthy and the infested tissues in three apple cultivars. The results of three approaches were provided; the first approach was based on using image-level mean spectra extraction for the whole sample analysis, and the second and third approaches were conducted at the pixel level using manual and automatic ROI segmentation around the infested area of the sample, respectively. Furthermore, the optimal wavelengths were selected using SFS algorithm to develop multispectral models. The total classification accuracy for the infested and healthy samples are as high as 97.4% for the validation dataset using GTB ensemble classifier among others. The feature selection algorithm obtained a maximum validation accuracy of 91.6% with only 22 selected wavelengths. Therefore, the NIR HSI method

for infestation detection demonstrated the capacity to detect CM infestation in apples of different varieties with potential in post-harvest inline apple sorting applications. Overall, the good results obtained in this study represent a promising step forward for sorting technologies employed in the apple processing units especially in packinghouses and export/import inspections. Moreover, the proposed NIR HSI could be useful as a remote monitoring tool for quality control and for studying CM incidence directly in the orchard; for example, through UAV-based HSI. Finally, future research could include evaluating the computational costs and processing speed, improving hardware, and applying other machine learning methods such as deep learning, as these could improve the accuracy and the robustness of the HSI detection system.

Author Contributions: Conceptualization, N.E. and A.A.A.; methodology, N.E., A.Y.K., L.E.D., J.R.L. and R.T.V.; software, N.E.; validation, A.A.A.; formal analysis, N.E.; investigation, N.E. and A.A.A.; resources, A.A.A.; data curation, N.E. and University of Kentucky; writing—original draft preparation, N.E., A.Y.K., L.E.D. and J.R.L.; writing—review and editing, A.A.A. and N.E.; visualization, N.E.; supervision, A.A.A. and K.D.D.; project administration, A.A.A.; funding acquisition, A.A.A., K.D.D. and R.T.V. All authors have read and agreed to the published version of the manuscript.

Data Availability Statement: All data used in this project belong to the United State government and the administering institution, University of Kentucky (UK), and can be requested through the UK Library.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. USDA Foreign Agricultural Service. Available online: https://www.fas.usda.gov/data/fresh-apples-grapes-and-pears-world-markets-and-trade (accessed on 11 June 2021).
2. USApple Association. Available online: https://usapple.org/industry-at-a-glance (accessed on 2 October 2021).
3. Balaško, M.K.; Bažok, R.; Mikac, K.M.; Lemic, D.; Živković, I.P. Pest Management Challenges and Control Practices in Codling Moth: A Review. *Insects* **2020**, *11*, 38. [CrossRef]
4. Pajač, I.; Pejić, I.; Barić, B. Codling moth, Cydia pomonella (Lepidoptera: Tortricidae)–major pest in apple production: An overview of its biology, resistance, genetic structure and control strategies. *Agric. Conspec. Sci.* **2011**, *76*, 87–92.
5. Walker, J.; Lo, P.; Horner, R.; Park, N.; Hughes, J.; Fraser, T. Codling moth (Cydia pomonella) mating disruption outcomes in apple orchards. *N. Z. Plant Prot.* **2013**, *66*, 259–263. [CrossRef]
6. Lu, Y.; Lu, R. Non-Destructive Defect Detection of Apples by Spectroscopic and Imaging Technologies: A Review. *Trans. ASABE* **2017**, *60*, 1765–1790. [CrossRef]
7. United States Department of Agriculture (USDA). Plant Protection and Quarantine: USDA APHIS Annual Report 2017. 2017. Available online: https://www.aphis.usda.gov/publications/plant_health/report-ppq-2017.pdf (accessed on 15 October 2021).
8. Fan, S.; Li, J.; Zhang, Y.; Tian, X.; Wang, Q.; He, X.; Zhang, C.; Huang, W. On line detection of defective apples using computer vision system combined with deep learning methods. *J. Food Eng.* **2020**, *286*, 110102. [CrossRef]
9. Zhang, D.; Xu, Y.; Huang, W.; Tian, X.; Xia, Y.; Xu, L.; Fan, S. Nondestructive measurement of soluble solids content in apple using near infrared hyperspectral imaging coupled with wavelength selection algorithm. *Infrared Phys. Technol.* **2019**, *98*, 297–304. [CrossRef]
10. Tian, X.; Fan, S.; Li, J.; Huang, W.; Chen, L. An optimal zone combination model for on-line nondestructive prediction of soluble solids content of apple based on full-transmittance spectroscopy. *Biosyst. Eng.* **2020**, *197*, 64–75. [CrossRef]
11. Fathizadeh, Z.; Aboonajmi, M.; Beygi, S.R.H. Nondestructive firmness prediction of apple fruit using acoustic vibration response. *Sci. Hortic.* **2019**, *262*, 109073. [CrossRef]
12. Ekramirad, N.; Chadwick, A.P.; Villanueva, R.T.; Donohue, K.D.; Adedeji, A.A. Low Frequency Signal Patterns for Codling Moth Larvae Activity in Apples. In Proceedings of the 2020 ASABE Annual International Virtual Meeting, Virtual, 13–15 July 2020; p. 1.
13. Vasighi-Shojae, H.; Gholami-Parashkouhi, M.; Mohammadzamani, D.; Soheili, A. Ultrasonic based determination of apple quality as a nondestructive technology. *Sens. Bio-Sens. Res.* **2018**, *21*, 22–26. [CrossRef]
14. Cocetta, G.; Beghi, R.; Mignani, I.; Spinardi, A. Nondestructive Apple Ripening Stage Determination Using the Delta Absorbance Meter at Harvest and after Storage. *HortTechnology* **2017**, *27*, 54–64. [CrossRef]

15. Gongal, A.; Silwal, A.; Amatya, S.; Karkee, M.; Zhang, Q.; Lewis, K. Apple crop-load estimation with over-the-row machine vision system. *Comput. Electron. Agric.* **2015**, *120*, 26–35. [CrossRef]

16. Silwal, A.; Gongal, A.; Karkee, M. Apple identification in field environment with over the row machine vision system. *Agric. Eng. Int. CIGR J.* **2014**, *16*, 66–75.

17. Ma, T.; Xia, Y.; Inagaki, T.; Tsuchikawa, S. Rapid and nondestructive evaluation of soluble solids content (SSC) and firmness in apple using Vis–NIR spatially resolved spectroscopy. *Postharvest Biol. Technol.* **2020**, *173*, 111417. [CrossRef]

18. Li, M.; Ekramirad, N.; Rady, A.; Adedeji, A. Application of Acoustic Emission and Machine Learning to Detect Codling Moth Infested Apples. *Trans. ASABE* **2018**, *61*, 1157–1164. [CrossRef]

19. Qu, J.-H.; Liu, D.; Cheng, J.-H.; Sun, D.-W.; Ma, J.; Pu, H.; Zeng, X.-A. Applications of Near-infrared Spectroscopy in Food Safety Evaluation and Control: A Review of Recent Research Advances. *Crit. Rev. Food Sci. Nutr.* **2013**, *55*, 1939–1954. [CrossRef] [PubMed]

20. ElMasry, G.; Sun, D.-W. Principles of Hyperspectral Imaging Technology. In *Hyperspectral Imaging for Food Quality Analysis and Control*; Academic Press: Cambridge, MA, USA, 2010; pp. 3–43. [CrossRef]

21. Craig, A.P.; Franca, A.S.; Irudayaraj, J. Surface-Enhanced Raman Spectroscopy Applied to Food Safety. *Annu. Rev. Food Sci. Technol.* **2013**, *4*, 369–380. [CrossRef] [PubMed]

22. Li, Y.-S.; Church, J.S. Raman spectroscopy in the analysis of food and pharmaceutical nanomaterials. *J. Food Drug Anal.* **2014**, *22*, 29–48. [CrossRef]

23. Sonka, M.; Hlavac, V.; Boyle, R. *Image Processing, Analysis, and Machine Vision*; Cengage Learning: Boston, MA, USA, 2014.

24. Ma, J.; Sun, D.-W.; Qu, J.-H.; Liu, D.; Pu, H.; Gao, W.-H.; Zeng, X.-A. Applications of Computer Vision for Assessing Quality of Agri-food Products: A Review of Recent Research Advances. *Crit. Rev. Food Sci. Nutr.* **2014**, *56*, 113–127. [CrossRef] [PubMed]

25. Lorente, D.; Aleixos, N.; Gómez-Sanchis, J.; Cubero, S.; García-Navarrete, O.L.; Blasco, J. Recent Advances and Applications of Hyperspectral Imaging for Fruit and Vegetable Quality Assessment. *Food Bioprocess Technol.* **2011**, *5*, 1121–1142. [CrossRef]

26. Zhang, B.; Li, J.; Fan, S.; Huang, W.; Zhao, C.; Liu, C.; Huang, D. Hyperspectral imaging combined with multivariate analysis and band math for detection of common defects on peaches (Prunus persica). *Comput. Electron. Agric.* **2015**, *114*, 14–24. [CrossRef]

27. Peerbhay, K.; Mutanga, O.; Ismail, R. Random Forests Unsupervised Classification: The Detection and Mapping of Solanum mauritianum Infestations in Plantation Forestry Using Hyperspectral Data. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2015**, *8*, 3107–3122. [CrossRef]

28. Wu, L.; He, J.; Liu, G.; Wang, S.; He, X. Detection of common defects on jujube using Vis-NIR and NIR hyperspectral imaging. *Postharvest Biol. Technol.* **2016**, *112*, 134–142. [CrossRef]

29. Keresztes, J.C.; Goodarzi, M.; Saeys, W. Real-time pixel based early apple bruise detection using short wave infrared hyperspectral imaging in combination with calibration and glare correction techniques. *Food Control* **2016**, *66*, 215–226. [CrossRef]

30. Ekramirad, N.; Khaled, A.Y.; Doyle, L.E.; Parrish, C.A.; Villanueva, R.T.; Donohue, K.D.; Adedeji, A.A. NIR hyperspectral imaging with machine learning to detect and classify codling moth infestation in apples. In Proceedings of the 2021 ASABE Annual International Virtual Meeting, Virtual, 12–16 July 2021; p. 1. [CrossRef]

31. ElMasry, G.; Wang, N.; Vigneault, C.; Qiao, J.; ElSayed, A. Early detection of apple bruises on different background colors using hyperspectral imaging. *LWT* **2008**, *41*, 337–345. [CrossRef]

32. Wang, Z.; Sun, J.; Liao, X.; Chen, F.; Zhao, G.; Wu, J.; Hu, X. Mathematical modeling on hot air drying of thin layer apple pomace. *Food Res. Int.* **2007**, *40*, 39–46. [CrossRef]

33. Lu, Y.; Huang, Y.; Lu, R. Innovative Hyperspectral Imaging-Based Techniques for Quality Evaluation of Fruits and Vegetables: A Review. *Appl. Sci.* **2017**, *7*, 189. [CrossRef]

34. Rady, A.; Ekramirad, N.; Adedeji, A.; Li, M.; Alimardani, R. Hyperspectral imaging for detection of codling moth infestation in GoldRush apples. *Postharvest Biol. Technol.* **2017**, *129*, 37–44. [CrossRef]

35. Guo, Z.-m.; Huang, W.-q.; Peng, Y.-k.; Wang, X.; Li, J. Impact of region of interest selection for hyperspectral imaging and modeling of sugar content in apple. *Mod. Food Sci. Technol.* **2014**, *30*, 59–63.

36. Khaled, A.Y.; Aziz, S.A.; Bejo, S.K.; Nawi, N.M.; Jamaludin, D.; Ibrahim, N.U.A. A comparative study on dimensionality reduction of dielectric spectral data for the classification of basal stem rot (BSR) disease in oil palm. *Comput. Electron. Agric.* **2020**, *170*, 105288. [CrossRef]

37. Khaled, A.Y.; Aziz, S.A.; Bejo, S.K.; Nawi, N.M.; Abu Seman, I. Artificial intelligence for spectral classification to identify the basal stem rot disease in oil palm using dielectric spectroscopy measurements. *Trop. Plant Pathol.* **2021**, 1–12. [CrossRef]

38. Firtha, F. Development of data reduction function for hyperspectral imaging. *Prog. Agric. Eng. Sci.* **2007**, *3*, 67–88. [CrossRef]

39. Firtha, F.; Fekete, A.; Kaszab, T.; Gillay, B.; Nogula-Nagy, M.; Kovács, Z.; Kantor, D.B. Methods for Improving Image Quality and Reducing Data Load of NIR Hyperspectral Images. *Sensors* **2008**, *8*, 3287–3298. [CrossRef]

40. Khaled, A.Y.; Aziz, S.A.; Bejo, S.K.; Nawi, N.M.; Abu Seman, I.; Izzuddin, M.A. Development of classification models for basal stem rot (BSR) disease in oil palm using dielectric spectroscopy. *Ind. Crop. Prod.* **2018**, *124*, 99–107. [CrossRef]

41. Feng, Y.-Z.; Sun, D.-W. Application of Hyperspectral Imaging in Food Safety Inspection and Control: A Review. *Crit. Rev. Food Sci. Nutr.* **2012**, *52*, 1039–1058. [CrossRef] [PubMed]

42. Che, W.; Sun, L.; Zhang, Q.; Tan, W.; Ye, D.; Zhang, D.; Liu, Y. Pixel based bruise region extraction of apple using Vis-NIR hyperspectral imaging. *Comput. Electron. Agric.* **2018**, *146*, 12–21. [CrossRef]

43. Žibrat, U.; Stare, B.G.; Knapič, M.; Susič, N.; Lapajne, J.; Širca, S. Detection of Root-Knot Nematode *Meloidogyne luci* Infestation of Potato Tubers Using Hyperspectral Remote Sensing and Real-Time PCR Molecular Methods. *Remote Sens.* **2021**, *13*, 1996. [CrossRef]

44. Léo, R.F.L.; Miguel, F.D.S.-F.; Adalton, R.; Flávio, L.S. Relationship between fruit fly (Diptera: Tephritidae) infestation and the physicochemical changes in fresh fruits. *Afr. J. Agric. Res.* **2020**, *15*, 122–133. [CrossRef]

45. Sun, Y.; Wei, K.; Liu, Q.; Pan, L.; Tu, K. Classification and Discrimination of Different Fungal Diseases of Three Infection Levels on Peaches Using Hyperspectral Reflectance Imaging Analysis. *Sensors* **2018**, *18*, 1295. [CrossRef] [PubMed]

46. Ren, G.; Wang, Y.; Ning, J.; Zhang, Z. Using near-infrared hyperspectral imaging with multiple decision tree methods to delineate black tea quality. *Spectrochim. Acta Part A Mol. Biomol. Spectrosc.* **2020**, *237*, 118407. [CrossRef] [PubMed]

47. Howard, C. The codling moth. *Transvaal Agric. J.* **1908**, *6*, 523–526.

48. Plaza, A.; Benediktsson, J.A.; Boardman, J.W.; Brazile, J.; Bruzzone, L.; Camps-Valls, G.; Chanussot, J.; Fauvel, M.; Gamba, P.; Gualtieri, A.; et al. Recent advances in techniques for hyperspectral image processing. *Remote Sens. Environ.* **2009**, *113*, S110–S122. [CrossRef]

49. Abdi, H.; Williams, L.J. Principal component analysis. *Wiley Interdiscip. Rev. Comput. Stat.* **2010**, *2*, 433–459. [CrossRef]

50. Nawar, S.; Mouazen, A.M. Optimal sample selection for measurement of soil organic carbon using on-line vis-NIR spectroscopy. *Comput. Electron. Agric.* **2018**, *151*, 469–477. [CrossRef]

51. Garillos-Manliguez, C.A.; Chiang, J.Y. Multimodal Deep Learning and Visible-Light and Hyperspectral Imaging for Fruit Maturity Estimation. *Sensors* **2021**, *21*, 1288. [CrossRef]

52. Yu, K.-Q.; Zhao, Y.-R.; Liu, Z.-Y.; Li, X.-L.; Liu, F.; He, Y. Application of Visible and Near-Infrared Hyperspectral Imaging for Detection of Defective Features in Loquat. *Food Bioprocess. Technol.* **2014**, *7*, 3077–3087. [CrossRef]

53. Munera, S.; Gómez-Sanchís, J.; Aleixos, N.; Vila-Francés, J.; Colelli, G.; Cubero, S.; Soler, E.; Blasco, J. Discrimination of common defects in loquat fruit cv. 'Algerie' using hyperspectral imaging and machine learning techniques. *Postharvest Biol. Technol.* **2020**, *171*, 111356. [CrossRef]

54. Susič, N.; Žibrat, U.; Sinkovič, L.; Vončina, A.; Razinger, J.; Knapič, M.; Sedlar, A.; Širca, S.; Stare, B.G. From Genome to Field—Observation of the Multimodal Nematicidal and Plant Growth-Promoting Effects of Bacillus firmus I-1582 on Tomatoes Using Hyperspectral Remote Sensing. *Plants* **2020**, *9*, 592. [CrossRef]

55. Ghosh, P.K.; Jayas, D.S. Use of spectroscopic data for automation in food processing industry. *Sens. Instrum. Food Qual. Saf.* **2009**, *3*, 3–11. [CrossRef]

56. Li, J.; Luo, W.; Wang, Z.; Fan, S. Early detection of decay on apples using hyperspectral reflectance imaging combining both principal component analysis and improved watershed segmentation method. *Postharvest Biol. Technol.* **2018**, *149*, 235–246. [CrossRef]

57. Moscetti, R.; Haff, R.P.; Stella, E.; Contini, M.; Monarca, D.; Cecchini, M.; Massantini, R. Feasibility of NIR spectroscopy to detect olive fruit infested by Bactrocera oleae. *Postharvest Biol. Technol.* **2015**, *99*, 58–62. [CrossRef]

58. Saranwong, S.; Haff, R.P.; Thanapase, W.; Janhiran, A.; Kasemsumran, S.; Kawano, S. A Feasibility Study Using Simplified near Infrared Imaging to Detect Fruit Fly Larvae in Intact Fruit. *J. Near Infrared Spectrosc.* **2011**, *19*, 55–60. [CrossRef]

59. Haff, R.P.; Saranwong, S.; Thanapase, W.; Janhiran, A.; Kasemsumran, S.; Kawano, S. Automatic image analysis and spot classification for detection of fruit fly infestation in hyperspectral images of mangoes. *Postharvest Biol. Technol.* **2013**, *86*, 23–28. [CrossRef]