

Article

A Novel Hyperspectral Method to Detect Moldy Core in Apple Fruits

Andrea Genangeli ¹, Giorgio Allasia ², Marco Bindi ¹, Claudio Cantini ³, Alice Cavaliere ³,
Lorenzo Genesio ³, Giovanni Giannotta ², Franco Miglietta ³ and Beniamino Gioli ^{3,*}

¹ Department of Agronomy and Land Management, University of Florence, P.le delle Cascine 18, 50144 Florence, Italy; andrea.genangeli@unifi.it (A.G.); marco.bindi@unifi.it (M.B.)

² Gruppo FOS, Via Enrico Meloni, 77/ed.A, 16152 Genova, Italy; giorgio.allasia@fos.it (G.A.); giovanni.giannotta@fos.it (G.G.)

³ Institute of Bioeconomy (IBE), National Research Council (CNR), Via Caproni 8, 50145 Florence, Italy; claudio.cantini@ibe.cnr.it (C.C.); alice.cavaliere@ibe.cnr.it (A.C.); lorenzo.genesio@ibe.cnr.it (L.G.); franco.miglietta@ibe.cnr.it (F.M.)

* Correspondence: beniamino.gioli@cnr.it

Abstract: An innovative low-cost device based on hyperspectral spectroscopy in the near infrared (NIR) spectral region is proposed for the non-invasive detection of moldy core (MC) in apples. The system, based on light collection by an integrating sphere, was tested on 70 apples cultivar (cv) Golden Delicious infected by *Alternaria alternata*, one of the main pathogens responsible for MC disease. Apples were sampled in vertical and horizontal positions during five measurement rounds in 13 days' time, and 700 spectral signatures were collected. Spectral correlation together with transmittance temporal patterns and ANOVA showed that the spectral region from 863.38 to 877.69 nm was most linked to MC presence. Then, two binary classification models based on Artificial Neural Network Pattern Recognition (ANN-AP) and Bagging Classifier (BC) with decision trees were developed, revealing a better detection capability by ANN-AP, especially in the early stage of infection, where the predictive accuracy was 100% at round 1 and 97.15% at round 2. In subsequent rounds, the classification results were similar in ANN-AP and BC models. The system proposed surpassed previous MC detection methods, needing only one measurement per fruit, while further research is needed to extend it to different cultivars or fruits.

Keywords: moldy core; internal browning; hyperspectral classification; near-infrared spectroscopy; ANN binary classification



Citation: Genangeli, A.; Allasia, G.; Bindi, M.; Cantini, C.; Cavaliere, A.; Genesio, L.; Giannotta, G.; Miglietta, F.; Gioli, B. A Novel Hyperspectral Method to Detect Moldy Core in Apple Fruits. *Sensors* **2022**, *22*, 4479. <https://doi.org/10.3390/s22124479>

Academic Editor: Ivan Andonovic

Received: 26 April 2022

Accepted: 10 June 2022

Published: 14 June 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

The apples production and post-harvest industry is among the largest fruit markets at the global scale, given the large diffusion and consumptions of apple fruits [1]. Final consumers are showing an increasing attention towards food quality and sustainability of food supply chains, aiming to consume products with low environmental impact and homogeneous organoleptic characteristics without internal or external alterations [2]. Satisfying these requirements represents one of the main objectives for farmers and food-companies involved in the production and marketing of apples [3–5]. One of the most relevant causes of quality loss is represented by the internal browning in apple post-harvest phases caused by *Alternaria* sp. (Asp), a ubiquitous genus fungorum widely present in all apple-growing areas.

The pathology caused by Asp is called mold core, or simply moldy core (MC) and initiates and produces its damage effects in the interior part of the fruits [6,7]. Previous studies have shown that the principal MC susceptible apple cultivars (cv) are represented by Fuji, Red Delicious, and Granny Smith, but other significant varieties, such as Golden Delicious, can still present relevant internal damages by MC [8]. Indeed, apple cv Golden

Delicious plays a fundamental role in the world apple industry, especially in Italy, where it is the most widely cultivated apple cv, with a total production volume of 858,423 tons [9,10]. The damage by MC is an internal injury, with dry-brown areas developed in the inner part of the fruit. The fruit does not present any external sign of damage, making MC detection with classical non-invasive methods very challenging or practically impossible [11].

Therefore, the development of non-invasive analytical methods to detect MC damages throughout the production process, especially in post-harvest phases deserves great attention from producers.

Novel non-invasive approaches deriving from different technological sectors have been successfully tested and applied in recent years to detect internal injury or internal browning in apples, such as time-frequency images of vibro-acoustic signals [12,13]; magnetic resonance techniques [14,15] and X-ray analysis [16,17]. Thereby, these techniques deliver a meaningful occasion to explore innovative analysis methods capable to detect qualitative parameters in apples, but at the same time they show limitations due to their cost, size of equipment, and operating time. Another innovative and promising solution with non-invasive methods for quality control in apple production has derived from recent developments by spectroscopy application from field to post-harvest phases [18,19]. Especially, the exploitation of light properties around the near infrared (NIR) region of the electromagnetic spectrum has captured the interest of researchers and industry in recent years, since it provides a valid alternative compared to invasive analysis methods [20,21]. Near infrared spectroscopy (NIRS) technics are based on the collection of spectral information such as absorption (ABS), reflectance (REF), and transmittance (TR) of electromagnetic signals in the spectral region spanning from 700 to 1200 nm [22].

REF based methods measure the reflected spectral signature under controlled illumination conditions and is typically used to retrieve parameters or compounds that are present on the fruit surface. REF was successfully used in several different applications, e.g., real-time quantification of biophysical and biochemical parameters through non-destructive method in citrus [23] and detection of oil palm maturity in bunches of fruits [24]. Techniques based on light reflectance in VIS/NIR have been successfully applied in apples quality control to detect a wide number of biophysical and biochemical parameters, such as external decay in apples [25], degrees brix [26], postharvest storage periods [27], and chlorophyll content [28]. ABS/TR techniques are based on generating a convenient light source at one side of the fruit, which propagates across the interior of the fruit, being finally collected at a convenient escape sensing surface. In recent years, the improvement of efficient machine learning techniques (MLT) has permitted the development of innovative analysis models in different application areas, e.g., engineering science, modeling in geology, and data reduction [29–31]. Moreover, MLT associated with TR analysis was successfully used in developing calibration models to quantify brix and pol at various stages of an industrial sugar production process [32]; therefore, the application of these methods is a relatively new field in the agricultural post-harvest science. TR and MLT were applied sporadically on apples and other commercial fruits. Nevertheless, such methods were applied sporadically on apples while they were already successfully tested in different commercial fruits; e.g., TR techniques and convolutional neural network (CNN) were successfully applied in blueberry internal damage detection with a classification accuracy over 80% [33]. Moreover, TR techniques represent an important analysis tool in high-income fruits, e.g., the grade of ripeness in nectarine (*Prunus Persica*) was evaluated with an accuracy of 88% by TR analysis in combination with Partial Least Square Regression (PLS-R) [34].

In the last few years, interest for MC detection by NIRS has increased in research activities and some different NIRS applications have been proposed. The main studies concerning the assessment of the capability to detect MC presence by VIS/NIR transmittance spectra retrieval were conducted mainly in cv Fuji. Zhaoyong et al. [35] verified MC presence in cv Fuji through an acquisition method based on multiple measurements per fruit; results obtained through the application of classifications algorithms based on back propagation artificial neural network (BP-NN) and support vector machine (SVM) have

shown a classification accuracy of infected and healthy apples larger than 83%. Similarly, Shenderoy et al. [36] obtained a classification accuracy of healthy and infected apples larger than 90% by PLS-R in cv Fuji with a moldy area larger than 5%. Tian et al. [37] investigated the relationship between the orientation of fruit trough light source position, achieving best results with fruit stem-calyx axis horizontal and perpendicular to transmission belt in apples. These studies highlighted the high complexity of transmittance-based hyperspectral NIR measurements, which are strongly affected by fruit and illumination geometry and by cv specific traits and chromatic characters that strongly affect the measured spectral signatures. Results obtained on a specific cv are, therefore, likely not transposable to different ones.

The objectives of this work were:

- (i) To develop and validate an innovative and low-cost application of NIRS to detect and monitor MC presence and growth in cv Golden Delicious through a novel measurement system based on a light source—light transmission—light collection architecture. An integrating sphere (IS) with homogeneous light reflectance properties [38] was adopted to compensate the geometrical variability in each fruit and toward the illumination geometry, and a low-cost VIS-NIR commercial spectroradiometer was used to measure the transmitted radiance inside the integrating sphere.
- (ii) To develop spectral based algorithms capable of detecting the MC and classifying the fruits in a binary classification framework (e.g., classifying a fruit as healthy or moldy), based on several state of the art machine learning techniques: pattern recognition neural networks (ANN-AP), Logistic Regression (LR), Linear Support Vector Classification (SVC), Random Forest (RF), Naive Bayes (NB), K-Nearest Neighbor (KNN), and Bagging Classifier based on Decision tree (BC).
- (iii) To assess the temporal performance of the detection algorithms, i.e., to assess the amount of time after the inoculus at which it becomes detectable.
- (iv) To assess the sensitivity of the algorithms, i.e., the minimum amounts of infected tissues that can be detected.
- (v) To determine the most important spectral bands responsible for the MC detection, and the minimum number of bands that can be used to further develop low-cost-multispectral rather than hyperspectral detectors.

Finally, the technological and industrial implications of the proposed sensing technology are discussed.

2. Materials and Methods

2.1. Instrument Setup

The prototype measurement system to detect MC in apples presented here is called Apple Light Transmittance System (ALT-S) and is shown in Figure 1. It consists of a box (1) where on its top was inserted a polystyrene integrating sphere (2) with 160 mm diameter and 50 mm thickness. The sphere was externally covered with an aluminum foil to remove external light noise. The inner part of the sphere is made of polystyrene and was assumed to have white-body-like properties such as an integral spectral signature that is independent from the light geometry. The fruit (3) is placed on the base of the IS and it is detained by a neoprene gasket with 60 mm diameter (4) between the sphere and a vacuum chamber (5). The vacuum chamber is a negative pressure space that has the function of sucking the fruit to obtain a complete adherence to the gasket, thus avoiding the possibility of having photons that escape and reach the sphere and the detector without passing through the fruit core. The chamber is made of a connecting pipe (6), which connects the vacuum chamber to a vacuum pump (7), in turn connected to the outside of the box through a gasket (8). The fruit is illuminated by a NIR 40 W light source (9) placed at the base of the vacuum chamber. The NIR light source spectral range starts from 770 nm to 920 nm. Spectral data were collected by an Ocean Optics USB2000 spectrometer (Ocean Insight, Rochester, NY, USA) with a spectral sampling interval ranging from 350 to 1000 nm and a 0.2 nm spectral resolution, and internally based on a Sony ILX511 linear silicon CCD array (11, 12). The

The operational analysis time (positioning of the fruit in the sphere and spectral acquisition) was approximately 90 s for each sample. Apples were sampled every time in two different positions, vertically (T1) and horizontally (T2) with respect to the NIR light source location. A total amount of 700 spectra signatures was collected in this way (70 samples \times 5 times \times 2 positions). After the final spectral acquisition, all fruits have been cut to check the growth of MC and the degree of its development. Fruits were positioned in an image acquisition platform to acquire RGB images to retrieve information about the amount of rotten versus healthy surface. The rotten area expressed as a percentage of the total cut area was retrieved with a threshold-based segmentation method operated in MatlabR2021a (MahtWorks, Natick, MA, USA) using Image Processing Toolbox. Biometrical data such as weight, volume, height, maximum and minimum diameter at the beginning and end of the experiment were collected. The MC data consisted of 70 values determined at the end of the experiment.

2.3. Data Analysis

2.3.1. Transmittance Retrieval

The entire spectral dataset was composed of 700 spectral signatures of 2048 wavelengths (WL) ranging from 350 to 1000 nm. Only the spectral wavelengths within the range of the NIR light source were selected for the analysis, resulting in 750 bands from 770 nm to 920 nm. The set of 70 fruits was sampled in T1 and T2 positions for 5 measurement rounds. The spectral data were collected in absolute irradiance ($\mu\text{W}/\text{nm}/\text{cm}^2$) and the spectral transmittance (TR) was computed as the ratio:

$$\text{TR} = \text{FRad}/\text{RRad} \quad (1)$$

where FRad is the radiance obtained by the photons transmitted through the fruit and RRad the radiance obtained by the photons transmitted through the Delrin sphere reference target.

2.3.2. Band Ratios and Average Transmittance

A preliminary analysis was made to explore the acquired spectral dataset and verify the existence of a significant relation between MC measured at the end of the experiment and the spectral data measured at the 5-time steps along the experiment duration. The presence of such a relation can be considered as a prerequisite for the application of more complex machine learning classification methods, and may serve as an indirect estimate of the timing of the MC development by exploring different time steps. A simple band ratio index was computed for all possible combinations of couples of spectral bands transmittance ranging between 800 and 880 nm, separately for each fruit, each position, and each time step. Then the correlation coefficient between MC data (70 samples) and the multiple band ratios for the 70 fruits was computed for each measurement round and position, obtaining a correlation map for each round.

Furthermore, the average transmittance was computed for all the fruits belonging to the two classification categories (healthy and moldy), for each round, to derive additional qualitative indicators of the presence of spectral features associated to moldy state.

2.3.3. Binary Classification

The overall objective of this study was to develop and test different binary classification models. The label-encoding was adopted assigning label 1 to moldy samples in relationship with the MC presence and label 0 to healthy samples. The spectral dataset was divided in one training dataset and four test datasets, where round 5 was used to train the model and rounds 1, 2, 3, and 4 were used as test. In all the classification models developed here, each WL represented an independent variable (IV) while the MC state represented the target variable to be predicted. Different binary classifiers were evaluated by computing performance metrics on accuracy, precision, and recall commonly used in machine learning models [39]. The performance metric was estimated by computing the confusion matrix on the training dataset [40]. By the definition of the confusion matrix, C is such that $C_{i,j}$ is

equal to the number of observations known to be in group i and predicted to be in group j . Thus, in binary classification, the count of true negatives is $C_{0,0}$, false negatives is $C_{1,0}$, true positives is $C_{1,1}$ and false positives is $C_{0,1}$. Therefore, the performance metrics were obtained to evaluate the best classifiers and to compare them.

Supervised Classification Models

Multiple supervised classification models (MSCM) based on the scikit-learn python library [41] were evaluated in this study; respectively: LR, SVC, RF, NB, KNN, and BC [42,43]. The model parameters of the RF, NB, KNN, and BC were optimized by cross-validated grid-search over a parameter grid [44]. The best performing model was selected, and then further improved by a dimensionality reduction aimed at reducing the number of independent variables. In fact, the presence of redundant information in the spectral data can distort classification results [45]. In classification problems, some statistical techniques can be used to minimize redundant data [46]. Here, we applied a univariate feature selection, as univariate statistical test to select k features that have the strongest relationship with the output variable. To select a specific k number of features, the ANOVA F-value method [47] via the Sklearn `f_classif()` function was used, while a grid search was implemented for the tuning of k [48].

Pattern Recognition Neural Network

A binary classification model based on ANN-AP was developed using MatlabR2021a Pattern Recognition Toolbox (MathWorks, Natick, MA, USA). Backpropagation with Momentum Algorithm (BMA) represents a powerful tool to resolve non-linear problems and it was selected to train the network [49,50]. BMA is a particular class of backpropagation algorithms where the input units are propagated forward to the output layer through the connecting weights. An accurate description of BMA can be found in the work of Phansalkar et al. [51]. The network's architecture was developed in accordance with backpropagation rules and it was formed by 252 input, one layer with 252 hidden layers, one layer with two hidden layers, and two output labels. The train function 'traingdm' based on Fletcher-Powell Conjugate Gradient was used. The limit of training periods was set at 600 epochs. Other settings have been set at their default values.

3. Results and Discussion

3.1. Infection Rate

The infection rate was computed at the end of round 5 by destructive samplings and image segmentation. Results showed an infection rate of 54.2% (38 fruits over 70) against 45.8% of samples that remained healthy (33 fruits over 70). Based on the segmentation results, all the 70 samples were classified by the infection rate (Figure 2).

A relevant fraction of the infected samples (44%) had an infection level between 1% and 2% (Figure 2c). This condition made it possible to select an infection threshold of 0.51% to balance the number of samples labelled as healthy and moldy (respectively, 35 moldy and 35 healthy samples) in the binary classification training dataset. The choice of the infection threshold represents a critical issue in MC early detection [52] and has economic and industrial implications. The selection of a relatively large threshold on the one hand would facilitate the development of good classification algorithms, but on the other hand would carry the risk to classify fruits with infection rates lower than the threshold as healthy, which has relevant negative industrial impact [53]. The threshold selected here is remarkably low, improving over ten times the minimum infection level used in a similar previous study [35]. Keeping the minimum detectable infection at a low value represents a primary challenge for the performance assessment of the machine learning based methods presented here.

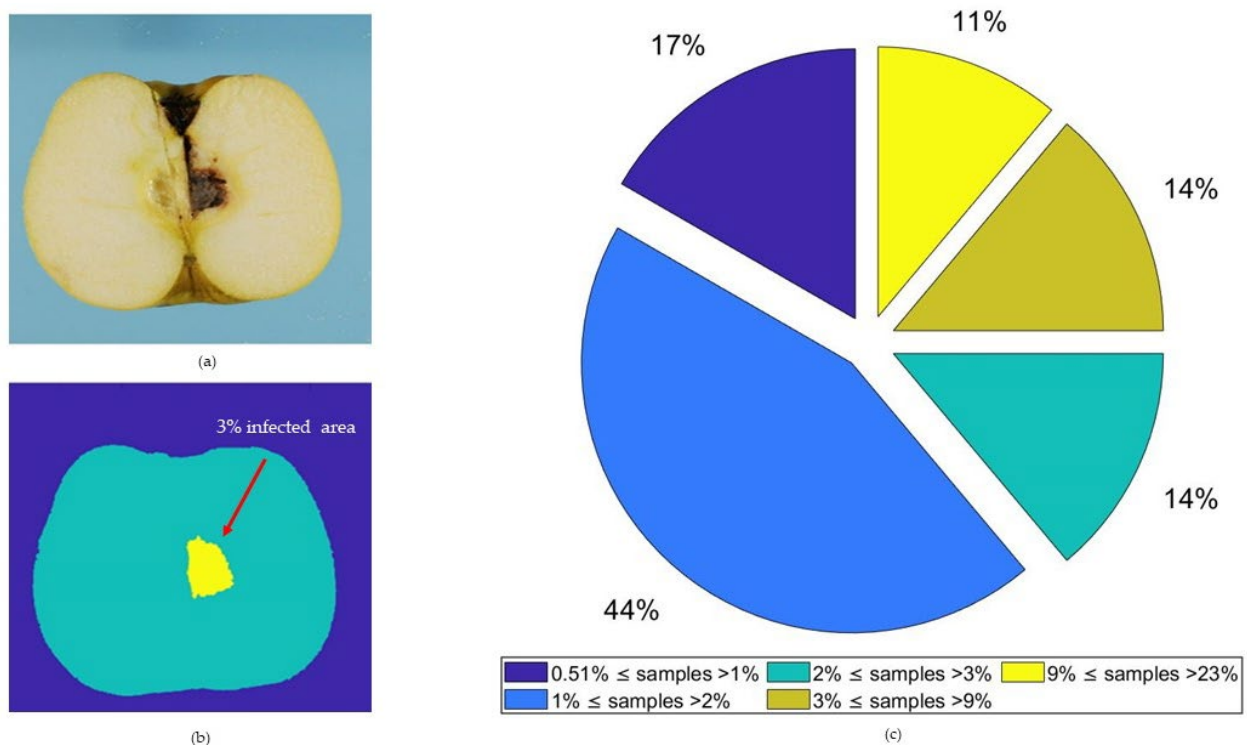


Figure 2. Example of segmentation results in a sample with a 3.0% infection rate. RGB image (a) was isolated from background, the results obtained (b) were furthermore processed to obtain necrotic area (yellow) separated from healthy area (light blue). Infection results were grouped into classes based on infection levels (c).

3.2. Spectral Correlations

The maps of correlation between transmittance band ratios and MC across all possible permutations of couples of bands was computed in T1 and T2 fruit positions and for each time step (Figure 3).

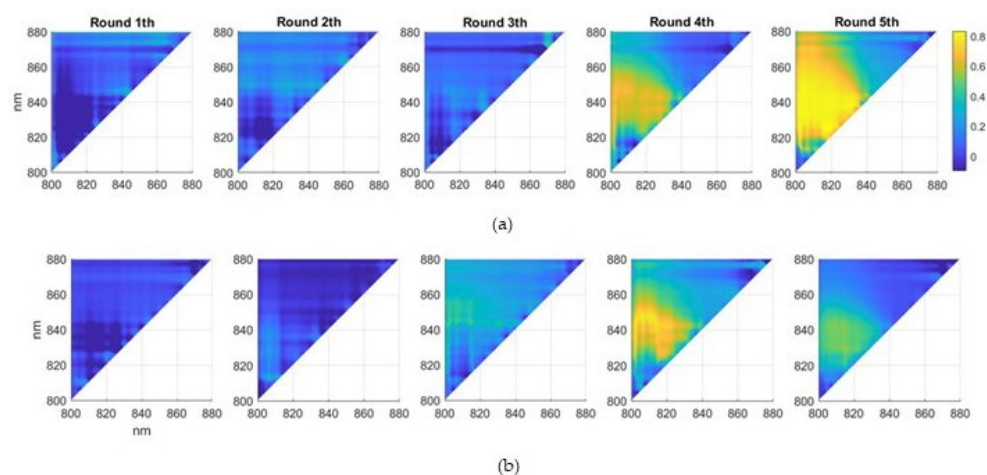


Figure 3. Maps of the correlation coefficient between MC measured at the end of the experiment (round 5), and transmittance ratios T_x/T_y based on two individual bands, computed on all combinations of x and y ranging over the sampled spectral interval (800–880 nm) for each round (1 to 5). (a,b) refer to T1 and T2 positions, respectively.

These maps show the absence of any significant correlation at round 1 and 2 in both T1 and T2 positions, with values contained within -0.2 to 0.2 not associated with any

consistent pattern. This result is likely related to the absence of infection at these early stages of the experiment. At round 3, correlations are still very low for T1 position, while they reach values of 0.43 in T2 position, likely indicating that spectral proxies of the infection progress started to be detectable at this stage. The presence of a distinctive higher correlation region was then observed at round 4 and then round 5, with maximum correlation values in T1 position of 0.67 and 0.82 and 0.74 and 0.51 in T2 position. This evidence supports the hypothesis that MC is developing within the fruits between round 1 and round 3, it is partially developed and detectable at round 4 and fully developed at round 5, when the destructive sampling and MC determination were made. The spectral bands whose ratio was associated with the maximum correlation are very similar in round 4 and 5, at 850 nm and 805 nm in round 4, and 849 nm and 802 nm in round 5.

Correlations between band ratios and MC in T1 position (vertical) are remarkably higher than in T2 position (horizontal), revealing that the measurement position influences the MC detection capability by affecting the spectral geometry. The explanation for this difference is likely related to the geometric symmetry of apple fruits along the vertical axis that reflects in anisotropic conditions and photons homogeneously passing through the internal parts of the fruit. On the other hand, the fruit placed in a horizontal position is not symmetrical along the vertical axis aligned with the light source, likely generating fruit specific geometric conditions affecting the light penetration and the spectral sampling. Given this difference in the performance, only the spectra retrieved in T1 position were selected for the subsequent analysis and machine learning classification.

3.3. Transmittance Temporal Pattern

The average spectral transmittance pattern in T1 position was obtained for each round (Figure 4).

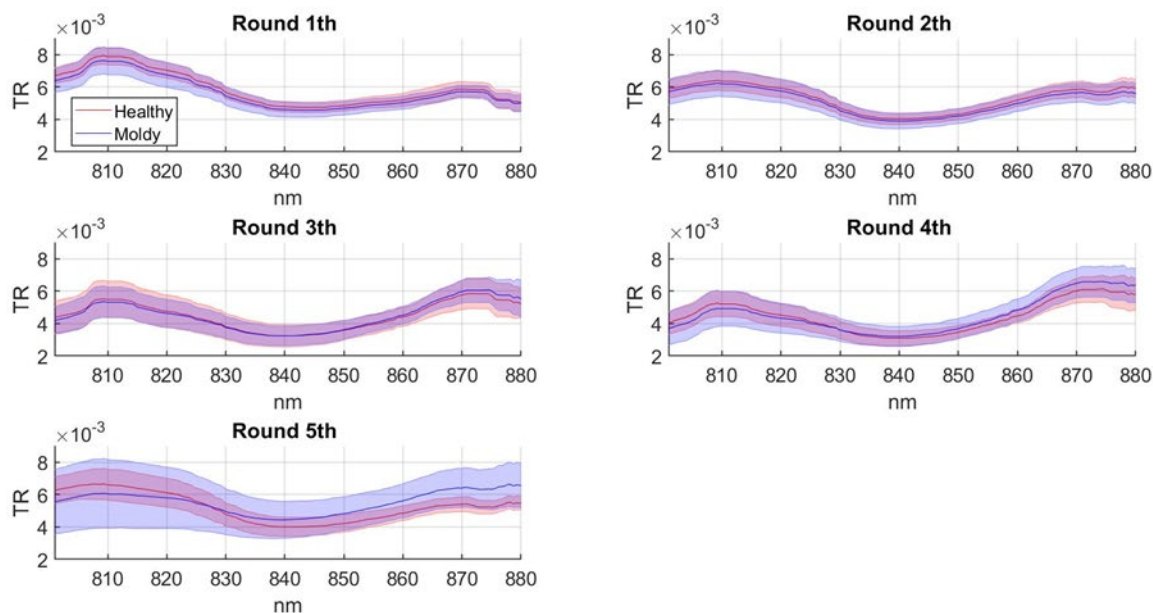


Figure 4. Temporal pattern of mean transmittance for healthy and infected samples with a threshold infection level of 2% or more. Transmittance was first averaged across all wavelengths and then mean and standard deviation were computed.

In round 1 and 2, the TR curve does not show any significant difference between healthy and moldy samples. In round 3, the magnitude of TR from 860 nm to 880 nm increased in infected samples more than in healthy ones. In round 4, the difference in TR between healthy and moldy samples is further accentuated, especially from 800 to 820 nm and from 850 to 880 nm. In round 5, results show the maximum TR difference between healthy and moldy samples. The standard deviation of the infected population increases

significantly compared to healthy samples. In addition, round 5 shows an increase of infected population's transmittance from 825 to 880 nm. Similar studies were conducted on Fuji apples by Tian et al., and Zhaoyong et al. [34,36]. Tian et al. measured in moldy samples an increase of TR in the spectral region from 775 to 830 nm and a decrease from 830 to 880 nm. In contrast, Zhaoyong et al. showed a decrease of TR in moldy samples from 700 to 820 nm. Generally, the transmittance temporal pattern obtained in our work shows a decrease of TR in moldy samples from 800 to 830 nm and an increase from 830 to 880 nm, in contrast to Tian et al.'s results and in accord to Zhaoyong et al.'s results. The difference in results might be due to the different apple varieties used in previous spectroscopy analysis. Indeed, apple varieties present morphological differences such as their texture, skin color, and chemical composition and the spectral response results are inevitably influenced by these characteristics [54]. In addition, TR differences encountered in several studies might also depend on the large number of pathogens related to MC disease [7]; therefore, a precise taxonomic classification of the detected pathogens would be recommended.

3.4. Binary Classification and ANN-AP

3.4.1. Features Reduction

Based on ANOVA univariate results, the F-value was computed in round 5 to select a specific k number of features to reduce the redundant information (Figure 5).

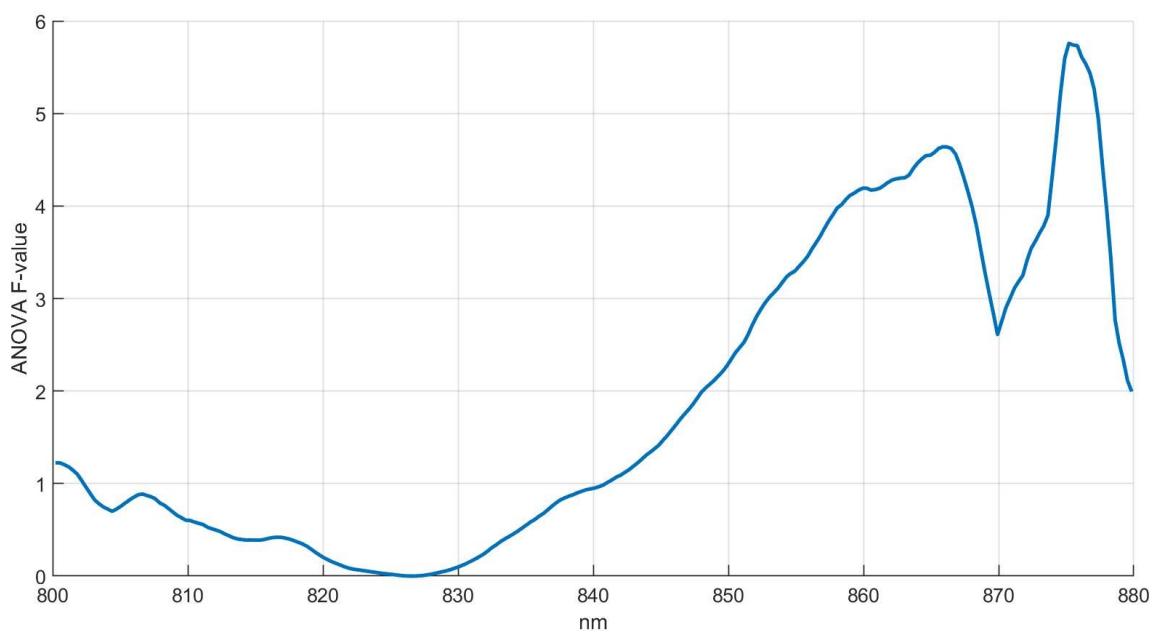


Figure 5. ANOVA F-value results from 800 to 880 nm. F-value is a ratio between two variances and a higher F-value corresponds to a significant statistical mean separation between healthy and moldy groups.

Results show a constantly increasing pattern of F-value from 830 nm to 866 nm, followed by a decrease and then a remarkable increase from 870 to 878 nm. Figure 5 shows two peaks in the spectral region from 860 to 878 nm caused by a larger variance in the spectral region from 860 to 880 nm. The F-value increases significantly from 860 to 878 nm in accordance with the increase of TR (Figure 4) obtained in temporal pattern results to the same spectral range and round. Therefore, spectral bands characterized by larger F-value indicate the spectral features most influenced by MC presence. Results obtained by ANOVA analysis allowed the reduction of the spectral range in MSCM training from 863.38 to 877.69 nm.

3.4.2. Binary Classification

Several MSCM and one ANN-AP were assessed in this study. Between MSCM, the BC reported the best training score in accuracy and precision; consequently, classification results obtained from BC and ANN-AP were compared. Training results are shown in Table 2.

Table 2. BC model and ANN-AP training results. The binary classification models were trained on round 5 and the performance metrics (accuracy, precision, and recall) were computed by confusion matrix.

Model	Accuracy	Precision	Recall
BC	0.95	0.85	0.88
ANN-AP	0.72	0.89	0.62

ANN-AP reported a higher level of precision (0.89) while BC model reported a higher level of accuracy (0.95) and recall (0.88). The accuracy measures the number of correct predictions made divided by the total number of predictions made. The precision is the number of true positives divided by the number of true positives and false positives. The recall is the ability to find all relevant instances (intuitively the ability of the classifier to find all the positive samples). The higher level of accuracy and recall reported by the BC model shows a better ability to classify correctly samples labeled (both positives and negatives) in comparison to the ANN-AP model. However, the higher level of precision reported by ANN-AP suggests a better ability to find positive instances. Indeed, a high score of precision reflects a high degree of discrimination between positive and false positive samples in training set. The precision score obtained in the training set confirms the higher capability to classify a low number of false positive by ANN-AP, as already reported in literature concerning the hyperspectral classification based on decision tree and neural network [55,56]. The importance of high precision in the training set is accentuated in a quality control industrial context, where the identification of infected samples has the greatest impact compared to the elimination of healthy samples. Therefore, failures in correctly detecting infected samples leads to a direct damage for the end consumers, which require healthy and standardized products. Differences in classification results by ANN-AP and BC models can be due to their different processing of the input variable. Indeed, ANN-AP is an algorithm inspired by biological neural networks, instead BC is based on a top-down approach of looking at the data. In ANN-AP, the value of the weights selected during the training process and the training goal is to minimize the error between values predicted by ANN-AP and true values [57]. The BC model uses a binary tree graph to assign for each data sample a target value and the target values are presented in the tree leaves. To reach the leaf the sample is propagated through nodes [58].

The two models, calibrated in round 5, were then tested on the previous rounds to assess them on completely independent datasets, also characterized by a different (e.g., lower) level of MC compared to the level used to train the models. Classification results on previous rounds are reported in Figure 6.

Overall, results obtained from both the BC and ANN-AP models exhibit an increase over time of the amount of detected infected samples; this behavior is in agreement with an expected exponential development of the infection [59]. The lowest number of positive samples was detected in round 1, respectively, 10% in BC model and 0% in ANN-AP. At round 1, a condition of complete absence of the infection was likely present, given that the inoculus was just applied at the beginning of the experiment. This condition is only met with the ANN-AP model, while the BC model reported a relatively large number of false positives (seven samples classified as positive). In round 2, the BC and the ANN-AP models detected an amount of 35.72% and 2.85% of positive samples, respectively. The temporal proximity of round 2 from to the inoculation time (six days) discourages the hypothesis that a high level of infection was reached. The correlation maps (Figure 3) and

the average transmittance patterns (Figure 4) also support the hypothesis of a complete absence of infection at round 2. Therefore, BC model is likely reporting a large number of false positives while ANN-AP is performing properly. This difference can be attributed to ANN-AP learning skills that better adapt to new data patterns and reveal a higher capacity to interpret non-linear problems, as reported in the literature from Rojas and Abiodun et al. [60,61]. In round 3, the percentage of detected positive samples resulted similarly in the two models, at 40% and 44% for BC and ANN-AP, respectively.

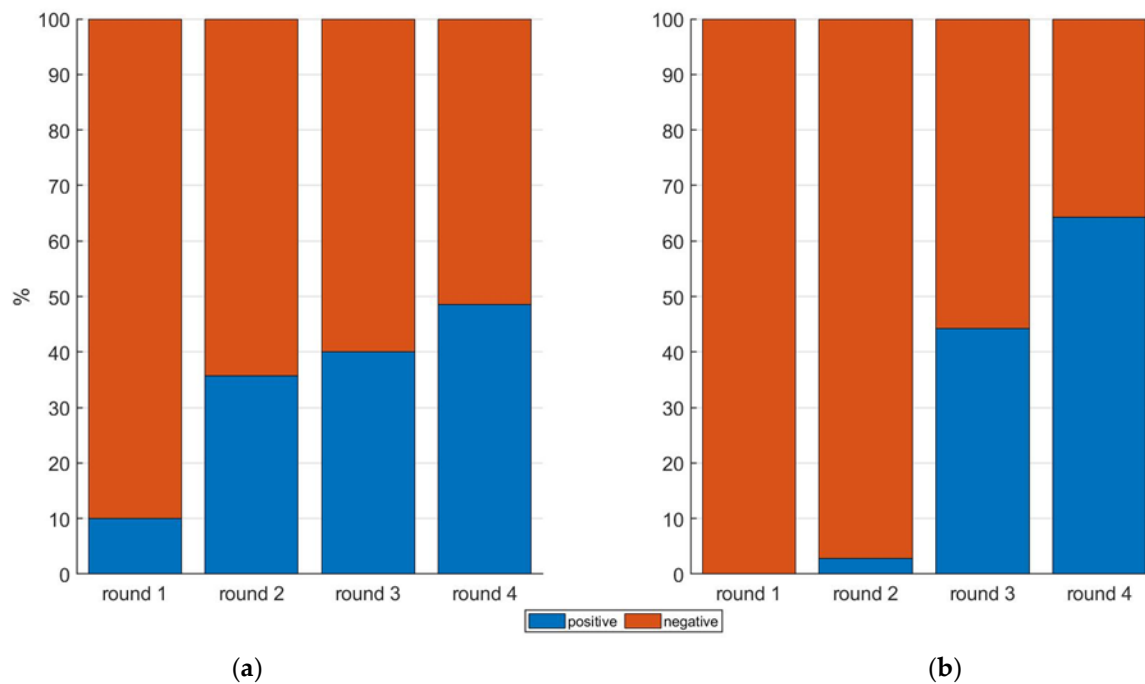


Figure 6. BC model (a) and ANN-AP (b) test results in round 1, 2, 3, 4. The binary classification results on unlabeled datasets show two classes, respectively, infected (positive) and healthy (negative). The two models were trained on round 5 with a threshold infection rate of 0.51%; therefore, the number of positive samples and negative in the training set resulted in 35 and 35 (50% positive and 50% negative), respectively.

Both models suggest the presence of the infection at round 3, a finding that is supported by the correlation maps (Figure 3) and transmittance patterns (Figure 4), which reveal the first signs of the MC presence at this stage. It is possible that the machine learning methods have a higher specificity compared to a simple band ratio or an average transmittance pattern and detect an actual starting of the infection more effectively. Nevertheless, the possibility of these methods reporting false positives at this stage cannot be ruled out, and further research is needed to verify the actual infection rate at an early stage by destructive samplings that could be done only at the end of the experiment.

In round 4, 48.57% and 64.29% of samples were classified as positives in the BC and ANN-AP models, respectively; a large spread of infected samples at this stage is also confirmed by the correlation maps (Figure 3) and the average transmittance patterns that are significantly different between infected and healthy samples (Figure 5). Based on MC presence in round 5, the ANN-AP shows an overestimation of positive classifications. Overall, the classification models trained in this study might exhibit a possible uncertainty due to the limited size of the dataset [62].

This analysis gives insight on the model capability to detect MC at different development stages and it provides information about the model general applicability in real post-harvest condition. As already discussed, in an industrial context, the economic damage for primary producers due to the inputs of the infected samples in large scale retail trade is greater than the damage associated to the healthy product discard.

The BC model, while having an overall good classification capacity on the training dataset (Table 2), reported a large fraction of false positives at previous time steps while used in testing mode. This characteristic prevents its application in industrial detection given that the amount of infection and the temporal stage of the infection process are generally unknown. The ANN-AP model, on the other hand, exhibited both a good precision on the training dataset, and a consistent reproduction of the infection rates and development on the testing datasets, therefore, resulting in having a greater potential for industrial applicability. This potential should be further confirmed and assessed by means of further research deploying a significantly larger number of sampled and inoculated fruits, also encompassing infection variability across different cultivars.

4. Conclusions

This study proposes a novel early detection system based on NIRS technologies to detect MC in cv Golden Delicious apples. The measurement device, called the ALT-System, was successfully developed and tested on a set of infected samples. The ALT-System has demonstrated its capability to detect MC through one single measurement per fruit, surpassing the previous measurement systems that required repeated measurements. Therefore, the achievement of this goal represents an important starting point to develop an efficient industrial prototype. In this study, spectral features linked to the MC presence were identified in the spectral region from 863.38 to 877.69 nm, and the spectral differences in term of TR between healthy and infected samples were explained over five different measurement rounds during the development of the infection. Several binary classification methods based on decision trees and ANN methods were tested here. ANN-AP and BC models were the two best methods with a better score in terms of performance metrics such as accuracy, precision, and recall. Moreover, their performance in training and validation datasets were assessed and discussed. By comparing training results between the two methods, the ANN-AP training results exhibited the best score in precision (0.89), a parameter whose optimization is mostly associated with the industrial needs of fruits post-harvest processing. Indeed, an efficient capability to distinguish healthy samples is a primary requirement to develop a functional industrial detection system. BC and ANN-AP models were tested on independent datasets obtained at previous stages of the infection development, and the ANN-AP showed a better classification result in relation to the infection growth rate. Indeed, ANN-AP showed a predictive accuracy of 100% at round 1 and 97.15% at round 2, demonstrating a better capability compared to BC model to interpret unknown datasets. Classification errors reported here might be related to the small dataset size and a low infection threshold adopted.

In an industrial context, MC early detection through non-invasive methods remains a critical issue, and this work represents a starting point to develop an industrial-scale prototype based on NIRS. Further studies are recommended to develop a measurement system capable to overcome the limits of the ALT-System: the inability to measure multiple fruits simultaneously and the need to minimize the measurement time per fruit. In addition, further investigations are needed to integrate the ALT-System directly on a conveyor belt. Additional spectral tests on multiple apple varieties are also recommended to explore the link between skin color, texture, biochemical contents, and spectral response. Finally, the technology and methods used and developed in this work might be applied in other apple diseases detection, e.g., biotic and abiotic internal browning or post-harvest CO₂ damages, and show a high potential for application to other fruits.

Author Contributions: Conceptualization, A.G., G.A., L.G., G.G., F.M. and B.G.; methodology, A.G., C.C., L.G., F.M. and B.G.; software, A.G., A.C., F.M. and B.G.; validation, A.G., A.C., L.G., F.M. and B.G.; formal analysis, A.G., A.C., L.G., F.M. and B.G.; investigation, A.G., A.C., L.G., F.M. and B.G.; resources, A.G., A.C., F.M. and B.G.; data curation, A.G., A.C., F.M. and B.G.; writing—original draft preparation, A.G., A.C. and B.G.; writing—review and editing, A.G., G.A., M.B., C.C., A.C., L.G., G.G., F.M. and B.G.; visualization, A.G. and B.G.; supervision, F.M. and B.G.; project administration, G.A., M.B., G.G. and B.G.; funding acquisition, G.A. and B.G. All authors have read and agreed to the published version of the manuscript.

Funding: The work is supported by the project “E-crops—Technologies for Digital and Sustainable Agriculture”, funded by the Italian Ministry of University and Research (MUR) under the PON Agrifood Program (Contract ARS01_01136).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Acknowledgments: We thank Valeria Gualandri from Fondazione Edmund Mach (San Michele All’Adige, Italy) for providing the pathogen inocula.

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

References

- Gonzalez, C.G. Global trade and competition in fresh and processed product markets. In *World Trade Review*; Aksoy, M.A., Beghin, J.C., Eds.; Cambridge University Press: Cambridge, UK, 2006; Volume 5, pp. 237–243. [\[CrossRef\]](#)
- Jaeger, S.R.; Antúnez, L.; Ares, G.; Swaney-Stueve, M.; Jin, D.; Harker, F. Quality perceptions regarding external appearance of apples: Insights from experts and consumers in four countries. *Postharvest Biol. Technol.* **2018**, *146*, 99–107. [\[CrossRef\]](#)
- El Khaled, D.; Castellano, N.; Gazquez, J.A.; Salvador, R.G.; Manzano-Agugliaro, F. Cleaner quality control system using bioimpedance methods: A review for fruits and vegetables. *J. Clean. Prod.* **2017**, *140*, 1749–1762. [\[CrossRef\]](#)
- Granatstein, D.; Kirby, E.; Ostenson, H.; Willer, H. Global situation for organic tree fruits. *Sci. Hortic.* **2016**, *208*, 3–12. [\[CrossRef\]](#)
- O’Rourke, A.D. *The World Apple Market*; Routledge: London, UK, 2018. [\[CrossRef\]](#)
- Elfar, K.; Zoffoli, J.P.; Latorre, B.A. Identification and characterization of *Alternaria* species associated with moldy core of apple in Chile. *Plant Dis.* **2018**, *102*, 2158–2169. [\[CrossRef\]](#)
- Gao, L.L.; Zhang, Q.; Sun, X.Y.; Jiang, L.; Zhang, R.; Sun, G.Y.; Zha, Y.L.; Biggs, A.R. Etiology of moldy Core, core browning, and core rot of Fuji apple in China. *Plant Dis.* **2013**, *97*, 510–516. [\[CrossRef\]](#)
- Ntasiou, P.; Myresiotis, C.; Konstantinou, S.; Papadopoulou-Mourkidou, E.; Karaoglanidis, G.S. Identification, characterization and mycotoxigenic ability of *Alternaria* spp. causing core rot of apple fruit in Greece. *Int. J. Food Microbiol.* **2015**, *197*, 22–29. [\[CrossRef\]](#)
- Italian Institute of Statistics. Available online: <https://www.istat.it/> (accessed on 18 May 2022).
- Italian Association of Apple Producers. Available online: <http://www.assomela.it> (accessed on 18 May 2022).
- Basson, E.; Meitz-Hopkins, J.C.; Lennox, C.L. Morphological and molecular identification of fungi associated with South African apple core rot. *Eur. J. Plant Pathol.* **2019**, *153*, 849–868. [\[CrossRef\]](#)
- Zhang, H.; Zha, Z.; Kulasiri, D.; Wu, J. Detection of early core browning in pears based on statistical features in vibro-acoustic signals. *Food Bioprocess Technol.* **2021**, *14*, 887–897. [\[CrossRef\]](#)
- Zhao, K.; Zha, Z.; Li, H.; Wu, J. Early detection of moldy apple core based on time-frequency images of vibro-acoustic signals. *Postharvest Biol. Technol.* **2021**, *179*, 111589. [\[CrossRef\]](#)
- Chayaprasert, W.; Stroshine, R. Rapid sensing of internal browning in whole apples using a low-cost, low-field proton magnetic resonance sensor. *Postharvest Biol. Technol.* **2005**, *36*, 291–301. [\[CrossRef\]](#)
- Cho, B.-K.; Chayaprasert, W.; Stroshine, R.L. Effects of internal browning and watercore on low field (5.4MHz) proton magnetic resonance measurements of T2 values of whole apples. *Postharvest Biol. Technol.* **2008**, *47*, 81–89. [\[CrossRef\]](#)
- van Dael, M.; Verboven, P.; Zanella, A.; Sijbers, J.; Nicolai, B. Combination of shape and X-ray inspection for apple internal quality control: In silico analysis of the methodology based on X-ray computed tomography. *Postharvest Biol. Technol.* **2018**, *148*, 218–227. [\[CrossRef\]](#)
- Hu, W.; Li, J.; Zhu, X.; Wang, Z.; Piao, D.; Zhao, L. Nondestructive detection of underlying moldy lesions of apple using frequency domain diffuse optical tomography. *Postharvest Biol. Technol.* **2019**, *153*, 31–42. [\[CrossRef\]](#)
- Cen, H.; He, Y. Theory and application of near infrared reflectance spectroscopy in determination of food quality. *Trends Food Sci. Technol.* **2007**, *18*, 72–83. [\[CrossRef\]](#)
- Sirisomboon, P. NIR spectroscopy for quality evaluation of fruits and vegetables. *Mater. Today Proc.* **2018**, *5*, 22481–22486. [\[CrossRef\]](#)

20. Cattaneo, T.M.P.; Stellari, A. Review: NIR spectroscopy as a suitable tool for the investigation of the horticultural field. *Agronomy* **2019**, *9*, 503. [CrossRef]
21. Tan, J.Y.; Ker, P.J.; Lau, K.Y.; Hannan, M.A.; Tang, S.G.H. Applications of photonics in agriculture sector: A review. *Molecules* **2019**, *24*, 2025. [CrossRef]
22. Li, L.; Zhang, Q.; Huang, D. A review of imaging techniques for plant phenotyping. *Sensors* **2014**, *14*, 20078–20111. [CrossRef]
23. Ali, A.; Imran, M. Remotely sensed real-time quantification of biophysical and biochemical traits of Citrus (*Citrus sinensis* L.) fruit orchards—A review. *Sci. Hortic.* **2021**, *282*, 110024. [CrossRef]
24. Goh, J.Q.; Shariff, A.R.M.; Nawi, N.M. Application of optical spectrometer to determine maturity level of oil palm fresh fruit bunches based on analysis of the front equatorial, front basil, back equatorial, back basil and apical parts of the oil palm bunches. *Agriculture* **2021**, *11*, 1179. [CrossRef]
25. 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]
26. Sánchez, N.H.; Lurol, S.; Roger, J.M.; Bellon-Maurel, V. Robustness of models based on NIR spectra for sugar content prediction in apples. *J. Near Infrared Spectrosc.* **2003**, *11*, 97–107. [CrossRef]
27. Giovanelli, G.; Sinelli, N.; Beghi, R.; Guidetti, R.; Casiraghi, E. NIR spectroscopy for the optimization of postharvest apple management. *Postharvest Biol. Technol.* **2014**, *87*, 13–20. [CrossRef]
28. Zude-Sasse, M.; Truppel, I.; Herold, B. An approach to non-destructive apple fruit chlorophyll determination. *Postharvest Biol. Technol.* **2002**, *25*, 123–133. [CrossRef]
29. Shariati, M.; Mafipour, M.S.; Mehrabi, P.; Shariati, A.; Toghroli, A.; Trung, N.T.; Salih, M.N.A. A novel approach to predict shear strength of tilted angle connectors using artificial intelligence techniques. *Eng. Comput.* **2020**, *37*, 2089–2109. [CrossRef]
30. Mammoliti, E.; Di Stefano, F.; Fronzi, D.; Mancini, A.; Malinverni, E.S.; Tazioli, A. A Machine learning approach to extract rock mass discontinuity orientation and spacing, from laser scanner point clouds. *Remote Sens.* **2022**, *14*, 2365. [CrossRef]
31. Junior, F.M.R.; Bianchi, R.A.; Prati, R.C.; Kolehmainen, K.; Soininen, J.-P.; Kamienski, C.A. Data reduction based on machine learning algorithms for fog computing in IoT smart agriculture. *Biosyst. Eng.* **2022**, *220*. [CrossRef]
32. Melo, B.H.D.S.; Sales, R.F.; Filho, L.D.S.B.; da Silva, J.S.P.; Sousa, A.G.C.D.A.; Peixoto, D.M.C.; Pimentel, M.F. Handheld near infrared spectrometer and machine learning methods applied to the monitoring of multiple process stages in industrial sugar production. *Food Chem.* **2021**, *369*, 130919. [CrossRef]
33. Wang, Z.; Hu, M.; Zhai, G. Application of deep learning architectures for accurate and rapid detection of internal mechanical damage of blueberry using hyperspectral transmittance data. *Sensors* **2018**, *18*, 1126. [CrossRef]
34. Munera, S.; Blasco, J.; Amigo, J.M.; Cubero, S.; Talens, P.; Aleixos, N. Use of hyperspectral transmittance imaging to evaluate the internal quality of nectarines. *Biosyst. Eng.* **2019**, *182*, 54–64. [CrossRef]
35. Zhaoyong, Z.; Lei, Y.; Su, D.; Zhang, H.; He, D.; Chenghai, Y. Detection of moldy core in apples and Its symptom types USING transmittance spectroscopy. *Int. J. Agric. Biol. Eng.* **2016**, *9*, 148–155.
36. Shenderoy, C.; Shmulevich, I.; Alchanatis, V.; Egozi, H.; Hoffman, A.; Ostrovsky, V.; Lurie, S.; Ben Arie, R.; Schmilovitch, Z. NIRS detection of MOLDY core in apples. *Food Bioprocess Technol.* **2009**, *3*, 79–86. [CrossRef]
37. Tian, X.; Wang, Q.; Huang, W.; Fan, S.; Li, J. Online detection of apples with moldy core using the Vis/NIR full-transmittance spectra. *Postharvest Biol. Technol.* **2020**, *168*, 111269. [CrossRef]
38. Jacques, J.A.; Kuppenheim, H.F. Theory of the integrating sphere. *J. Opt. Soc. Am.* **1955**, *45*, 460–466. [CrossRef]
39. Hossin, M.; Sulaiman, M.N. A review on evaluation metrics for data classification evaluations. *Int. J. Data Min. Knowl. Manag. Process* **2015**, *5*, 1–11. [CrossRef]
40. Marom, N.D.; Rokach, L.; Shmilovici, A. Using the confusion matrix for improving ensemble classifiers. In Proceedings of the 2010 IEEE 26th Convention of Electrical and Electronics Engineers in Israel, Eilat, Israel, 17–20 November 2010; pp. 000555–000559. [CrossRef]
41. Pedregosa, F.; Varoquaux, G.; Gramfort, A.; Michel, V.; Thirion, B.; Grisel, O.; Blondel, M.; Prettenhofer, P.; Weiss, R.; Dubourg, V.; et al. Scikit-learn: Machine learning in python. *J. Mach. Learn. Res.* **2011**, *12*, 2825–2830.
42. Myles, A.J.; Feudale, R.N.; Liu, Y.; Woody, N.A.; Brown, S.D. An introduction to decision tree modeling. *J. Chemom.* **2004**, *18*, 275–285. [CrossRef]
43. Breiman, L. Random forests. *Mach. Learn.* **2001**, *45*, 5–32. [CrossRef]
44. Gulzat, T.; Lyazat, N.; Siladi, V.; Gulbakyt, S.; Maksatbek, S. Research on predictive model based on classification with parameters of optimization. *Neural Netw. World* **2020**, *30*, 295–308. [CrossRef]
45. Omar, N.; Jusoh, F.B.; Binti, R.; Othman, M.S. Review of feature selection for solving classification problems. *J. Res. Innov. Inf. Syst.* **2013**, *7*, 64–70.
46. Chen, R.-C.; Dewi, C.; Huang, S.-W.; Caraka, R.E. Selecting critical features for data classification based on machine learning methods. *J. Big Data* **2020**, *7*, 52. [CrossRef]
47. Cho, B.-K.; Kim, M.S.; Baek, I.; Kim, D.-Y.; Lee, W.-H.; Kim, J.; Bae, H.; Kim, Y.-S. Detection of cuticle defects on cherry tomatoes using hyperspectral fluorescence imagery. *Postharvest Biol. Technol.* **2013**, *76*, 40–49. [CrossRef]
48. Brownlee, J. How to Perform Feature Selection with Numerical Input Data. Available online: <https://machinelearningmastery.com/feature-selection-with-numerical-input-data/> (accessed on 7 January 2022).

49. Rukhaiyar, S.; Samadhiya, N. A polyaxial strength model for intact sandstone based on artificial neural network. *Int. J. Rock Mech. Min. Sci.* **2017**, *95*, 26–47. [[CrossRef](#)]
50. Scales, L.E. *Introduction to Non-Linear Optimization*. Macmillan International Higher Education, 1985. Available online: <https://books.google.com/books?hl=it&lr=&id=AEJdDwAAQBAJ&oi=fnd&pg=PR9&dq=Scales,+L.+E.+%E2%80%98Introduction+to+Non-Linear+Optimization%E2%80%99.+Macmillan+International+Higher+Education&ots=fR1BoRYYT6&sig=ZMFrOS9ZZ4a1Tx7BiuqMFeFgzFo> (accessed on 25 January 2022).
51. Phansalkar, V.; Sastry, P. Analysis of the back-propagation algorithm with momentum. *IEEE Trans. Neural Netw.* **1994**, *5*, 505–506. [[CrossRef](#)]
52. Zhang, J.; Huang, Y.; Pu, R.; Gonzalez-Moreno, P.; Yuan, L.; Wu, K.; Huang, W. Monitoring plant diseases and pests through remote sensing technology: A review. *Comput. Electron. Agric.* **2019**, *165*, 104943. [[CrossRef](#)]
53. Martinelli, F.; Scalenghe, R.; Davino, S.; Panno, S.; Scuderi, G.; Ruisi, P.; Villa, P.; Stroppiana, D.; Boschetti, M.; Goulart, R.L.; et al. Advanced methods of plant disease detection: A review. *Agron. Sustain. Dev.* **2015**, *35*, 1–25. [[CrossRef](#)]
54. Pu, Y.-Y.; Feng, Y.-Z.; Sun, D.-W. Recent progress of hyperspectral imaging on quality and safety inspection of fruits and vegetables: A review. *Compr. Rev. Food Sci. Food Saf.* **2015**, *14*, 176–188. [[CrossRef](#)]
55. Shihab, T.; Al-Hameedawi, A.; Hamza, A. Random forest (RF) and artificial neural network (ANN) algorithms for LULC mapping. *Eng. Technol. J.* **2020**, *38*, 510–514. [[CrossRef](#)]
56. Raczko, E.; Zagajewski, B. Comparison of support vector machine, random forest and neural network classifiers for tree species classification on airborne hyperspectral APEX images. *Eur. J. Remote Sens.* **2017**, *50*, 144–154. [[CrossRef](#)]
57. Hinton, G.E. 20—Connectionist learning procedures. In *Machine Learning*; Elsevier: Amsterdam, The Netherlands, 1989. [[CrossRef](#)]
58. Breiman, L.; Friedman, J.H.; Olshen, R.A.; Stone, C.J. *Classification and Regression Trees*; Routledge: Boca Raton, FL, USA, 2017. [[CrossRef](#)]
59. Trinci, A.P.J. Exponential growth of the germ tubes of fungal spores. *J. Gen. Microbiol.* **1971**, *67*, 345–348. [[CrossRef](#)]
60. Raúl, R. The backpropagation algorithm. In *Neural Networks: A Systematic Introduction*; Rojas, R., Ed.; Springer: Berlin/Heidelberg, Germany, 1996; pp. 149–182. [[CrossRef](#)]
61. Abiodun, O.I.; Jantan, A.; Omolara, A.E.; Dada, K.V.; Mohamed, N.A.; Arshad, H. State-of-the-art in artificial neural network applications: A survey. *Heliyon* **2018**, *4*, e00938. [[CrossRef](#)] [[PubMed](#)]
62. Barbedo, J.G.A. Impact of dataset size and variety on the effectiveness of deep learning and transfer learning for plant disease classification. *Comput. Electron. Agric.* **2018**, *153*, 46–53. [[CrossRef](#)]