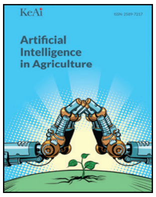
KeAi
CHINESE ROOTS
GLOBAL IMPACT

Contents lists available at ScienceDirect

Artificial Intelligence in Agriculture

journal homepage: http://www.keaipublishing.com/en/journals/artificial-intelligence-in-agriculture/



Mango internal defect detection based on optimal wavelength selection method using NIR spectroscopy

Anitha Raghavendra [a,*], D.S. Guru [b], Mahesh K. Rao [a]

[a] *Maharaja Research Foundation, Maharaja Institute of Technology, Mysore, India*
[b] *Department of Studies in Computer Science, University of Mysore, Mysore, India*

ARTICLE INFO

Article history:
Received 8 June 2020
Received in revised form 15 January 2021
Accepted 16 January 2021
Available online 10 February 2021

Keywords:
Feature selection methods
Fisher's linear discriminant analysis
Mango internal defect detection
NIR (near infrared spectroscopy)

ABSTRACT

A non-destructive technique should be developed for performance analysis of mango fruits because the spongy tissue or internal defects could lower the quality of mango fruit and incur a lack of productivity. In this study, wavelength selection methods were proposed to identify the range of wavelengths for the classification of defected and healthy mango fruits. Feature selection methods were adopted here to achieve a significant selection of wavelengths. To measure the goodness of the model, the dataset was collected using the NIR (Near Infrared) spectroscopy with wavelength ranging from 673 nm–1900 nm. The classification was performed using Euclidean distance measure both in the original feature space and in FLD (Fisher's Linear Discriminant) transformed space. The experimental results showed that the lower range wavelength (673 nm–1100 nm) was the efficient wavelength for the detection of internal defects in mangoes. Further to express the effectiveness of the model, different feature selection techniques were investigated and found that the Fisher's criterion based technique appeared to be the best method for effective wavelength selection useful for classification of defected and healthy mango fruits. The optimal wavelengths were found in the range of 702.72 nm to 752.34 nm using Fisher's criterion with a classification accuracy of 84.5%. This study showed that NIR system is a useful technology for the automatic mango fruit assessment which has the potential to be used for internal defects in online sorting, easily distinguishable by those who do not meet minimum quality requirements.

1. Introduction

With increasing urbanization, the needs of society are changing and have an impact on the overall health status of the people. This has resulted in people leaning towards more nutritional choices such as fruits and vegetables. India has emerged as the second-largest fruit producer (46 million tons), contributing almost 10% of the world production. India crops about 56% of the total world's mango. Markets in most of the countries have seen increasing demand for conventional mango varieties. Imports are restricted, in particular to developed countries, mainly due to their quarantine needs, lack of tools/techniques to meet their requirements. The lack of an automated trading program focused on non-destructive techniques to detect physiological disorders in Alphonso mango is a reason why India is still very poor in fruit export (Sandeep and Pradeep, 2014). Fruit fly, stone weevil, spongy tissue, etc., are the internal defects that occur in some mango varieties. These defects must be identified before these mangoes are exported. Human eyes cannot see the internal defects in mangoes and these defective mangoes cannot be sorted manually. The internally defected mangoes have no external symptoms either at the time of harvesting or during ripening. The affected part is visible only after slicing of the fruit (Manpreeth and Ramesh, 2010). For this reason, some countries specifically European countries have, until now, banned Indian mangoes. The causative factor for the presence of spongy tissue in the mango pulp portion has not been located so far, and there is no proven automation to monitor it during the maturation/pre-harvesting period as it is now. In such a scenario, a quick priority is to identify physiological disorders in mangoes using a non-destructive technique for the purpose of consumption and export. Fruit sorting should be done on an individual basis to ensure the quality of the product. In order to do so, an online method with a high sorting speed of at least 2 to 5 fruits per second should be established (Sang et al., 2006). Non-invasive identification of internal defects in Alphonso mango is a new area and very little research has been done to determine the internal quality of mangoes such as soluble solids, dry matter, firmness, starch and sugar content. To improve our understanding of light penetration into fruit tissues, more research involving numerical simulation techniques are needed (Bart et al., 2007).

Anitha and Mahesh (2016) have discussed mangoes with all kinds of defects such as diseases, physiological disorders and pest-related damage; and reviewed non-intrusive methods used in some fruits to detect

[*] Corresponding author.
E-mail addresses: anithbg@gmail.com (A. Raghavendra), dsg@compsci.uni-mysore.ac.in (D.S. Guru), maheshkrao_ece@mitmysore.in (M.K. Rao).

https://doi.org/10.1016/j.aiia.2021.01.005

internal defects. At Seoul National University in Korea, the model was developed for the detection of internal breakdowns in Fuji apples, Korean pears and sweet potatoes using non-destructive technologies (Sang et al., 2006). Qiang et al. (2011) have developed a Vis/NIR hyperspectral imaging system covering the spectral region from 408 nm to 1117 nm for bruise detection in kiwi fruits. Several research studies have been carried out since 1990 on determining certain indices of internal quality, specifically on total soluble solid content, starch, acidity, chlorophyll (in fruits and vegetables), protein (in grains) and total nitrogen content (in leaves) using NIR spectroscopy as a non-destructive technique. Similar qualities have been evaluated on potatoes, onions, peaches, apples, oranges, melons, mandarins, papaya and dates (Dospatliev et al., 2013).

The increasing importance of NIR spectroscopy in post-harvest technology is evident from the recent increase in the number of publications, as well as from the fact that many on-line grading manufactures have now introduced NIR systems to measure various qualities attributes (Bart et al., 2007). David and Carlos (1998) have used a non-destructive optical method for determining the internal quality of Kiwi fruit and the method based on NIR spectroscopic techniques. Non-destructive technology for the "on-line" assessment of the internal quality of fruits and vegetables using NIR spectroscopy is detailed in some publications which have been reviewed in (Krivoshiev et al., 2000). Peiris et al. (1998) have successfully determined soluble solids content in processing tomatoes. NIR based technology has also been used quite successfully to determine the ripeness of melons. Meurens and Feth (2001) conducted their research for the detection of various internal defects in apple: vitrescence, mealiness and internal discoloration (brown core) using Vis/NIR spectroscopy. Peirs et al. (2001) have investigated apples of 7 varieties and 3 orchards to determine the harvest maturity using non-destructive methods. Considerable results are obtained for measurement of maturity and detection of internal defects in apples using Vis/NIR spectroscopy (Holm, 2002). A non-destructive method has been developed for the measurement of soluble solids content in peaches "Blake" by NIR spectrometry (Peiris, 1997). With this knowledge, it is understood that the NIR spectroscopy may be one potential field that can be used for internal defect tracking in mangoes thus helping with the export of the same for India.

Using NIR spectroscopy, three different measurement setups can be made for obtaining NIR spectra such as in the mode of reflectance, transmittance and interactance. When choosing the measurement setup, it is important to be aware that the NIR radiation penetration into fruit tissue decreases exponentially with the depth (Lammertyn et al., 2000). Bart et al. (2007) found a penetration depth up to 4 mm in the wavelength range of 700–900 nm and between 2 mm to 3 mm in the range of 900–1900 nm for apple. In a different optical configuration, Fraser et al. (2000) showed that the penetration depth was at least 25 mm in the range 700–900 nm in apple, while it became less than 1 mm in the range 1400–1600 nm. The limited penetration depth restricts the potential of reflectance or interactance measurements for detecting internal defects and decreases the performance of NIR based measurements of thick-skinned fruit, such as internal quality attributes of citrus. Hence, the selection of wavelength range is a very important task to provide good coverage. However, the wavelength useful for an application will often be unique to the domain because different materials will have different spectra. It is impossible to have a set of generic wavelengths for all the applications (Mayank et al., 2015). As detection of internal defects in mangoes has not yet been attempted in any previous works, the task of classifying healthy and internally defected mangoes has been undertaken in this study. In order to express the effectiveness of the classification model, pattern recognition methods were introduced newly for the analysis of NIR spectroscopic data and also for appropriate wavelength selection. In view of the above challenges, the main objective of this study was to develop a non-destructive technique that can be used to detect the internal defects in mangoes. The specific objective was to select the optimal wavelength based on the feature extraction and feature selection techniques. Ultimately, this proposed system can be applied to offline mango post-harvest procedures, because precision agriculture in horticulture aims to have efficient utilization of resources for achieving targeted production of fruits.

2. Materials and methods

2.1. Samples and spectral acquisition

First and foremost was to understand if it was possible to identify the internal defects in the Mangoes. For this purpose, experiments were conducted on mangoes of Alphonso variety using the Ocean Optics NIR Spectroscopic instrument. The specifications of the NIR-Quest are the following:

- Detector- Hamamatsu G9208-512 W InGaAs linear array.
- Wavelength range: 700 nm–2500 nm w/Grating NIR.
- Integration time: 1 ms–200 ms.
- Signal to noise ratio: 10000:1 @ 100 ms integration.

Initially, 76 mangoes were randomly picked from the market and later the surface of the mango was wiped. This study was independent of post-harvesting time and temperature; hence these parameters were not used for the purpose of classification. In this study, the measurement setup was made in reflection mode. In reflection mode, the light source and detector were mounted under a specific angle, e.g., 45 degree to avoid specular reflection (Bart et al., 2007). The schematic of STS NIR miniature spectrometer is shown in Fig. 1. The spectral range of the optical fiber was 673 nm to 1900 nm. Two optical fiber cables were used. One cable generates the spectral range between 673 nm to 1100 nm and another cable provides 1100 nm to 1900 nm range of wavelength. 1024 values of wavelength were obtained from the range of 673 nm to 1100 nm with the approximate period of 0.5 nm and 512 values of wavelength were obtained from the range of 1100 nm to 1900 nm with the approximate period of 0.56 nm. There were no clues of internal defect on the mango surfaces detected as inspected visually. Therefore, after obtaining the reflectance values, each of the mangoes had to be cut individually to confirm the internal defects as shown in Fig. 2. Based on the above procedure, samples were segregated as 43 defective and 33 healthy samples. Out of curiosity, an attempt was made to understand the reflectance spectrum of mangoes of depth penetration ability, though the research objective was only to classify internally defected and non-defected samples. Only the area and depth of the defects were measured but analysis was not done on that. The recorded data indicates that out of a limited number of samples the defected area arranges from 7 sqcm to 22 sqcm and depth ranging from 0.5 cm to

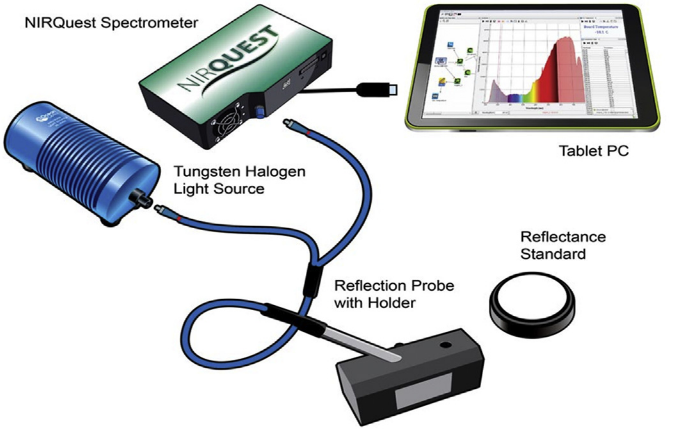
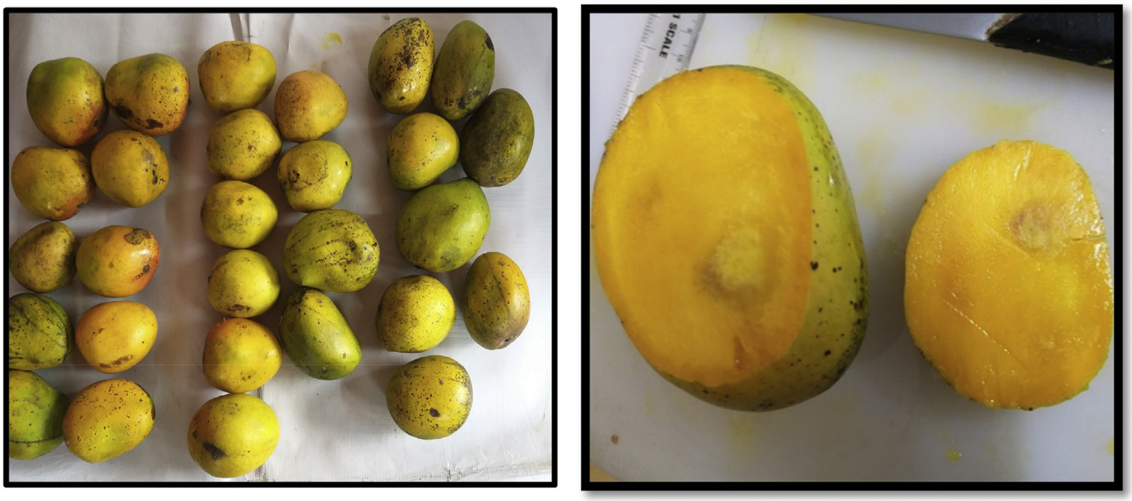


Fig. 1. Schematic of STS NIR Miniature Spectrometer.

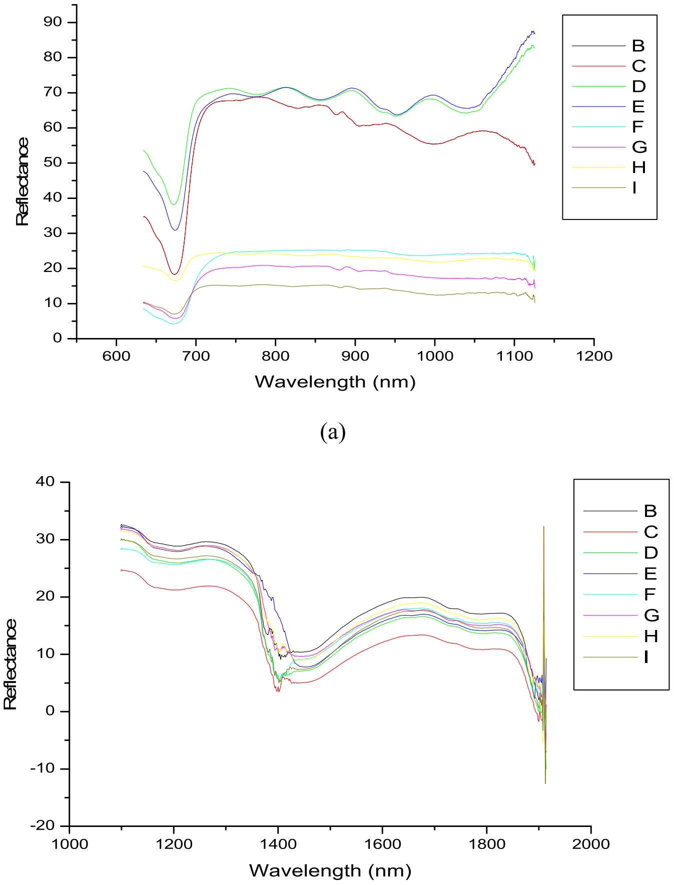


(a)
(b)

Fig. 2. (a) Mango samples (b) Internal defects found after cutting of a mango.

1.5 cm. Our target was to develop a system capable of classifying defected samples irrespective of the degree of severity which is not of great significance from a consumer point of view.

The obtained four spectroscopic data were directly plotted using Origin software which is depicted in Fig. 3. In Fig. 3(a) and (b), lines B, C, D, E are the reflectance values of healthy samples and lines F, G, H, I are the reflectance values of defective samples. From the plots, the graph appears to be separable from defective and healthy in lower range wavelength (673 nm to 1100 nm) compared to higher range wavelength (1100 nm to 1900 nm). Further, in order to express the effectiveness



(a)

(b)

Fig. 3. A plot of defective and healthy samples using origin software (a) Wavelength range 673 nm–1100 nm. (b) Wavelength range 1100 nm–1900 nm.

of the classification model, pattern recognition methods were introduced for the analysis of NIR spectroscopic data and also for the selection of appropriate wavelength.

3. Proposed wavelength extraction and selection model

In view of the above challenges, a model was proposed which selects the suitable wavelength based on the feature extraction and the feature selection techniques. The feature extraction method is highlighted in section 3.1 (Nemirko, 2016) and feature selection techniques are described in section 3.2 & 3.3 (Artur and Mário, 2012; Chulmin and Jihoon, 2007; Guru et al., 2018; Huan et al., 2010; Isabelle and Andre, 2003; Jun et al., 2016; Włodzis et al., 2003). As the mean value of the spectroscopy data for defected and healthy samples were differentiable, initially Fisher's Linear Discriminant Transformation (FLDT) was adopted, as the FLD Analysis works on the principle of inter-class variation (variance between the mean of two classes) and intra-class variation (mean of the variance of two classes).

3.1. Wavelength extraction using Fisher's linear discriminant transformation

Let *{Ai}* be a set of *P* sample column vectors of wavelength dimension *D*. The mean vector of that is given by,

$$\mu_A = \frac{1}{P}\sum_{i=1}^{P}\vec{A}_i \tag{1}$$

Let there be *K* classes$\{C_1 \ldots C_K\}$. The mean vector of the class *K* containing P_K members is

$$\mu_{\text{Ak}} = \frac{1}{P_K}\sum_{Ai\in C_K}\vec{A}_i \tag{2}$$

The inter-class scatter matrix is thus given by

$$S_B = \sum_{K=1}^{2} P_K\left(\overrightarrow{\mu_{AK}} - \overrightarrow{\mu_A}\right)\left(\overrightarrow{\mu_{AK}} - \overrightarrow{\mu_A}\right)^T \tag{3}$$

The intra-class scatter matrix is given by

$$S_W = \sum_{K=1}^{2}\sum_{Ai\in C_K}\left(\overrightarrow{A_i} - \overrightarrow{\mu_{AK}}\right)\left(\overrightarrow{A_i} - \overrightarrow{\mu_{AK}}\right)^T \tag{4}$$

The transformation matrix that repositions the data to be most separable in the matrix *W* that maximizes

$$\frac{\det\left(W^T S_B W\right)}{\det\left(W^T S_W W\right)} \tag{5}$$

$W = [W_1, W_2 \ldots\ldots\ldots W_D]$ gives a projection space of dimension *D*. The projection of vector *A* into a subspace of dimension *D* is

$$B = W_D^T A \tag{6}$$

A projection space of dimension $d < D$ can be defined by using the generalized Eigen vectors with the largest *d* Eigen values to give $W_D = [W_1, W_2, \ldots\ldots\ldots\ldots W_d]$. The Eigen vector represents the direction or components for the reduced subspace of *D*, whereas Eigen values represent the magnitudes for the directions. The generalized Eigen vectors are Eigen vectors of $S_B S_W^{-1}$.

The next projection vector *A* into a subspace of reduced dimension *d* is

$$B = W_d^T A \tag{7}$$

Once the projection vectors of *B* with reduced wavelength are formed, the performance of FLD transformed features can be evaluated. Inverse fisher's linear discriminant transformation can also be done from Eq. (8) to know the normal wavelengths of the transformed data space.

$$A = B{*}W_d^{-1} \tag{8}$$

3.2. Fisher's ratio for wavelength selection

For binary problems ($c_j \in \{0,1\}$), the Fisher's ratio (FiR), of the j^{th} feature is defined as.

$$\mathrm{FiR}_{\mathrm{j}} = \frac{\left|\,\overline{X}_j^{(0)} + \overline{X}_j^{(1)}\,\right|}{\sqrt{\operatorname{var}(X_j)^{(0)} + \operatorname{var}(X_j)^{(1)}}} \tag{9}$$

where $\overline{X}_j^{(0)}, \overline{X}_j^{(1)}$, $\operatorname{var}(X_j)^{(0)}$ and $\operatorname{var}(X_j)^{(1)}$, are the samples means and variances of feature *j*, for the patterns of each class.

3.3. Dispersion measures based wavelength selection

Some well-known dispersion measures that are used to compute relevance were adopted here.

(a) *The Mean Absolute Difference* (*MAD*), defined as

$$MAD_j = \frac{1}{n}\sum_{i=1}^{n}|X_{ji} - \overline{X_j}| \tag{10}$$

which computes the absolute difference from the mean value. The *MAD* is a scale-variant.

(b) *Ratio of Arithmetic Mean (AM)* to *Geometric Mean (GM)*

Another measure of dispersion applies the *Arithmetic Mean* and the *Geometric Mean*. For a given (positive) feature X_j on n patterns, the *AM* and *GM* are given by.

$$AM_j = \overline{X_j} = \frac{1}{n}\sum_{i=1}^{n}X_{ji},\quad GM_j = \left(\prod_{j=1}^{n}X_{ji}\right)^{\frac{1}{n}} \tag{11}$$

Since $AM_j \geq GM_{j,}$ with equality holding if and only if $X_{j1} = X_{j2} = X_{j3} = \ldots. = X_{jn,}$ then the ratio.

$$R_j = \frac{AM_j}{GM_j} \in (1, +\infty) \tag{12}$$

Eq. (12) can be used as a dispersion measure. Higher dispersion implies a higher value of R_j, thus a more relevant feature. Conversely, when the entire feature samples have (roughly) the same value, R_j is close to 1, indicating a low relevance feature.

(c) *Mean Median (MM) difference*

Another dispersion measure, named *Mean-Median difference* given by.

$$MM_j = \left|\,\overline{X_j} - median(X_j)\,\right| \tag{13}$$

This computes the absolute difference between the mean and median of $X_{j.}$

Once the dispersion measures described above of all the features are computed, the features are sorted in decreasing order and recommend to keep the top-m features. The computed measures are considered as relevancy scores for validation.

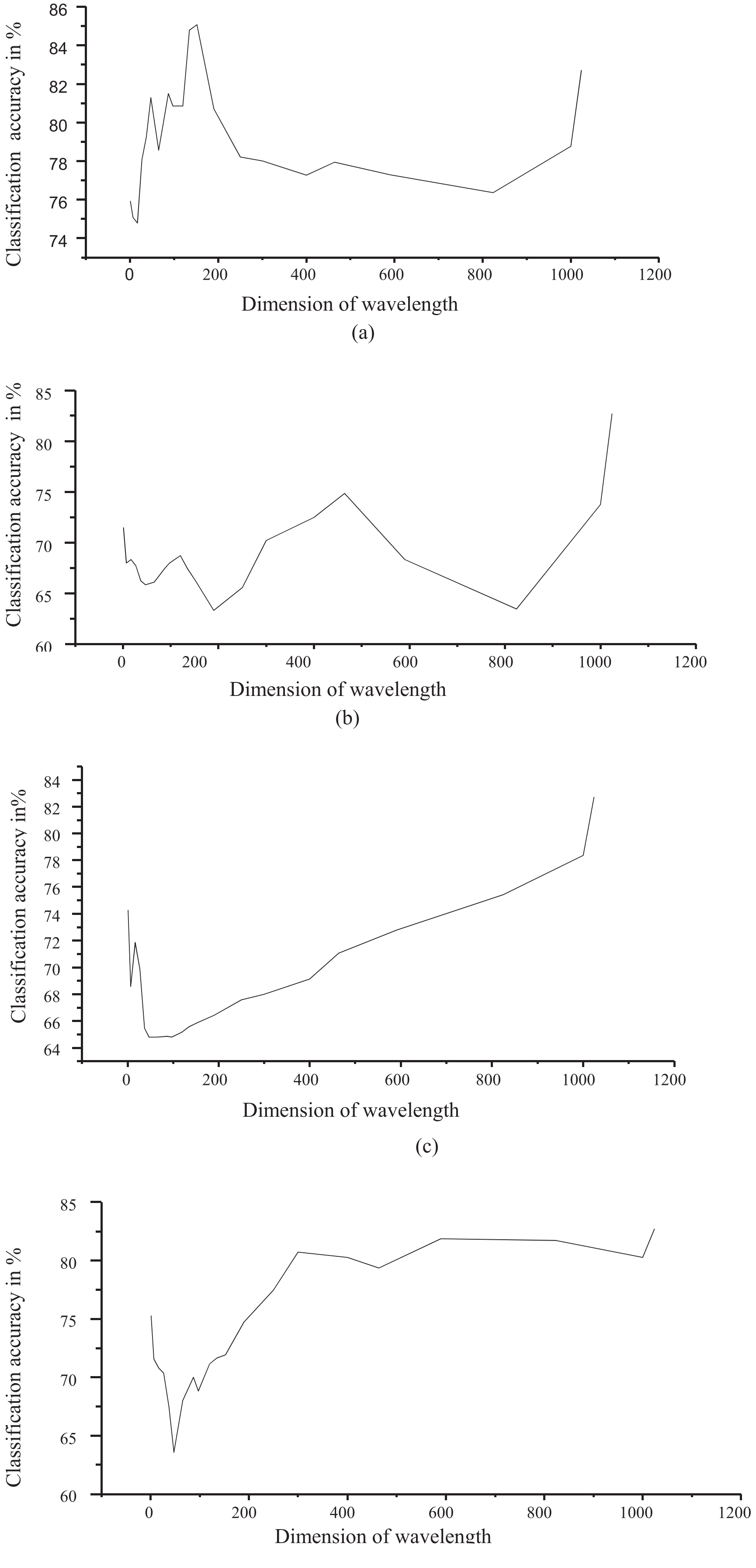


Fig. 4. Accuracy obtained using varying number of wavelength selected for the range 673 nm–1100 nm (a) Fisher's Ratio (b) AMGM (c) MM (d) MAD.

3.4. Estimation of the number of wavelengths

Once the features were sorted in decreasing order with their relevancy scores, the selection of features relies on the following estimates for efficient wavelength selection. Therefore, according to the high relevancy score, the wavelength can be selected by fixing up the dimension experimentally for better classification.

3.4.1. Cumulative relevance (CR) measure

Let $r_{i1\ldots\ldots\ldots}r_{id}$ be the sorted relevance values and be the *CR* of the top-*v* most relevant features.

$$c_f = \sum\nolimits_{f=1}^{v} r_{if} \tag{14}$$

At each step of *CR* features, evaluate the performance of the classification. It stops adding features until the best performance is obtained.

3.4.2. Relevancy with their redundancy in the score

Let $r_{i1\ldots\ldots\ldots}r_{id}$ be the sorted relevance values. If $r_{if} = r_{id}$ for *f, d* = *1, 2, ….v,* performance can be evaluated for these features. Based on the highest performance of both the features, only one feature can be considered. If $r_{if} \neq r_{id,}$ both the corresponding features can be retained.

3.4.3. Selection of ranked wavelengths

Let $r_{i1\ldots\ldots\ldots}r_{id}$ be the sorted relevance values from individual feature selection techniques. Selection of only top-1 ranked features at the beginning from all the feature selection techniques and evaluation to be carried out to verify the performance of the model. Then add the next ranked features to the top-1 ranked feature and continue to add further ranked features cumulatively until the performance drops at some point. Performance can be evaluated at each stage of feature addition.

3.4.4. Individual wavelength selection

For all the feature selection techniques, performance can be evaluated by considering individual features at a time instead of cumulative addition of features until it reaches to top- *v* most relevant features (obtained from section 3.3.1). Hence, selected features provide a peak in performance and subsequently the same evaluation can be performed for all the features.

3.5. Performance analysis

Once the number of features was estimated, samples were divided as *T1%* for training and *T2%* for testing samples. Let training samples were considered to be X_i number of samples, *i* = *1, 2, 3……N* with *d* number of wavelengths and Y_j be the testing samples, *j* = *1, 2, 3……M* with *d* number of wavelengths.

Distance measure was performed using Euclidean distance metric.

$$Z_i = \sqrt{\sum_{r=1}^{d} (X_{ri} - Y_{ri})^2}\ \text{Where}, i = 1,2,3\ldots\ldots\ldots N \tag{15}$$

The above equation involves computing the square root of the sum of squares of the differences between corresponding samples. Euclidean distance measure was performed between each testing sample *Y,* with *N* number of training samples using the above equation.

$$D = \begin{cases} C1 & \text{If} \ \ \min(Z_i)\,\epsilon\, C1 \\ C2 & \text{If} \ \min(Z_i)\ \epsilon\ C2 \end{cases} \tag{16}$$

All samples from the testing set were measured using the Euclidean distance metric from Eq. (15). The confusion matrix was built based on the result of *D* and performance was evaluated based on the confusion matrix.

4. Results

4.1. Performance of Fisher's linear discriminant transformation

Fisher's Linear Discrimination analysis projects data samples to a line that maintains direction useful for classifying data of different classes. The FLD gives a projection matrix that reshapes the scatter of a dataset to maximize class separability, defined as the ratio of the inter-class scatter matrix to the intra-class scatter matrix. This projection defines features that are optimally discriminating. First, the FLD transformation technique was applied on a lower range of 1024 wavelengths and also on a higher range of 512 wavelengths separately. Later, 525 Eigen wavelengths with non-zero Eigen values were obtained from 1024 wavelengths and 260 Eigen wavelengths with non-zero Eigen values were obtained from 512 wavelengths. Classification accuracy was calculated with FLD subspace for both lower range and higher range of wavelength by selecting non-zero Eigen wavelengths and non-negative Eigen wavelengths. However, the accuracy result was 70% for 673 nm to 1100 nm range of wavelength and 62% for 1100 nm to 1900 nm range of wavelength. Inverse transformation of FLD was also performed and mapped to original wavelengths using the obtained Eigen values from FLD transformation. Accuracy was calculated for mapped original wavelengths and achieved 80.71% for the lower range and 68% for higher range as a classification result. As the depth of penetration is higher at lower wavelengths (Fraser et al., 2000), it would be imperative to use the lower wavelengths for the analysis. By considering the successful result obtained for lower range wavelength i.e. for 673 nm–1100 nm; feature selection techniques were applied to the only lower range of wave lengths for further analysis.

4.2. Wavelength selection using filter techniques

Further, feature selection filter techniques were implemented for lower range wavelengths, for efficient wavelength selection.

Fig. 4 shows the comparison of four feature selection techniques for the range of wavelength 673 nm–1100 nm. From Fisher's ratio, it was observed that accuracy of 84.78% was obtained for 135 wavelengths. From AMGM, 74.85% accuracy was obtained for 464 wavelengths. From MM, 82.71% accuracy was obtained for 1024 wavelengths and from MAD, 81.85% accuracy was obtained for 590 wavelengths.

There were 135 wavelengths from fisher's ratio, 464 wavelengths from AMGM, 1024 wavelengths from MM and 590 wavelengths from MAD chosen as it provided the highest accuracy in individual feature selection technique. These wavelengths were randomly chosen and plotted in Fig. 5 to know how wavelengths were distributed over the range of 633 nm–1100 nm. Once the relevancy scores were calculated, scores were arranged in descending order and it was observed that there was no redundancy in the scores as the relevancy scores were not the same.

Computing the classification accuracy by considering only top-1 ranked wavelengths from different feature selection techniques, then cumulatively adding top-2 ranked wavelengths and could be continued shown in Table 1. Classification accuracy of 79.07% for top-1 ranked wavelengths, 77.64% for the fusion of top-1 and top-2 rank wavelengths and 75.2% for the fusion of top-1, top-2 and top-3 rank wavelengths were obtained shown in Fig. 6. It was noticed that there was no overlapping of wavelengths between ranked wavelengths.

Accuracy was calculated for individual wavelength instead of cumulative wavelengths shown in Fig. 7. In the first method, 80.71% of accuracy was achieved for the wavelengths which were identified at the peaks. In the second method, 81.42% of accuracy was obtained by only selecting and combining the wavelengths which resulted in more than 80% accuracy. In the third method, 84.5% accuracy was obtained at 723.35 nm which was identified at the highest peak. Considering the

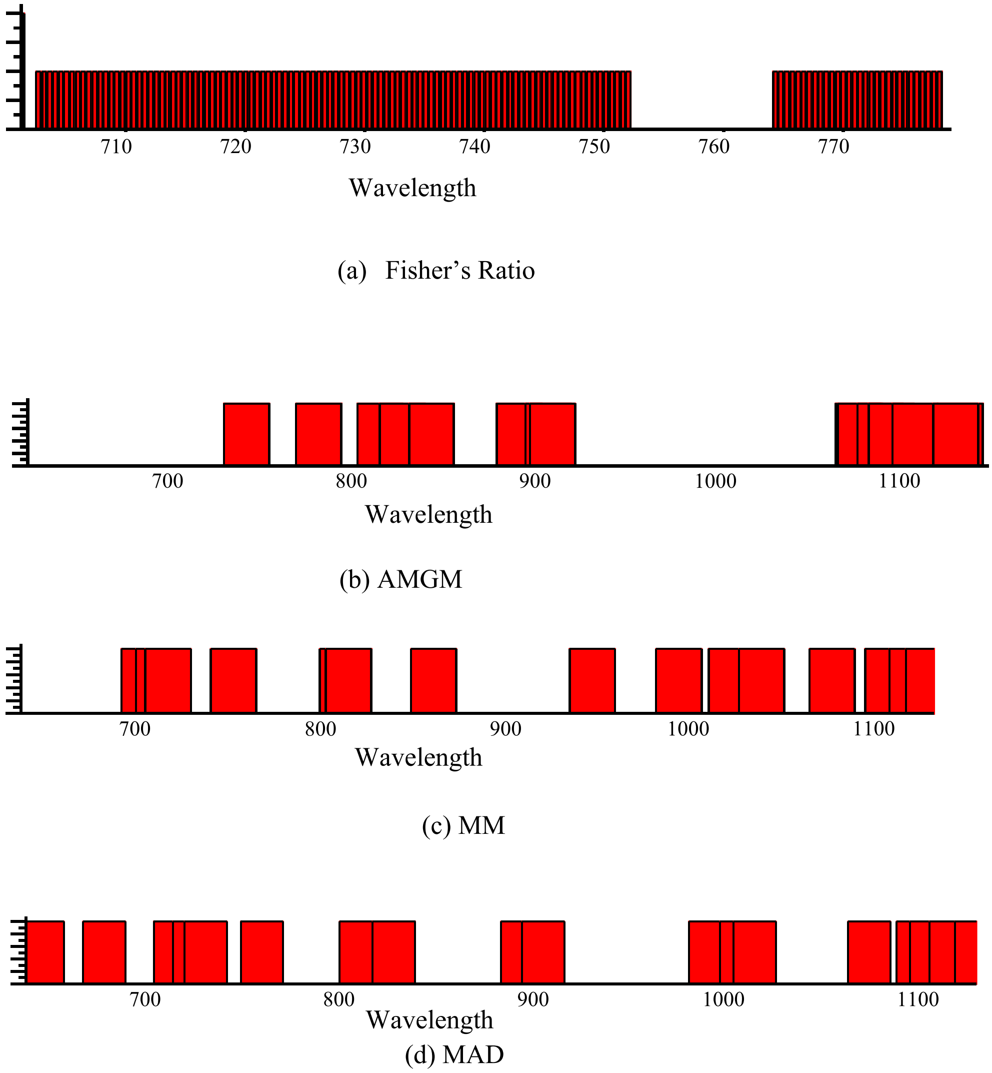


Fig. 5. Wavelength distribution over the range 633 nm–1100 nm using Fisher's ratio and different dispersion techniques.

fusion of the highest peak wavelength and their three neighbor wavelengths (722.88 nm and 723.82 nm) turned out to be 83.71% accurate.

5. Discussion

In this study, an internal defected and healthy mango classification model was attempted based on feature extraction and feature selection techniques for efficient wavelength selection. Classification accuracy of 82.71% for lower range wavelengths and 64.86% for higher range wavelengths were obtained using the original features. Further, the FLD transformation and filter-based feature selection technique were adopted to achieve a better result. It was observed that classification accuracy was 70% for lower range (673 nm–1100 nm) with 525 non-zero Eigen wavelengths and 62% for a higher range (1100 nm to 1900 nm) with 215 non-zero Eigen wavelengths using FLD transformation. The model provides 80.71% of accuracy for the lower range wavelengths and 68% for the higher range wavelengths for the wavelengths which were identified after computing inverse FLD transformation. Hence, it was noticed that lower range NIR spectroscopy wavelength was suitable for identifying internal defects in mangoes. In order to choose an appropriate wavelength range, a lower range wavelength was considered for further feature selection techniques.

Table 1
Top-1 rank and top-2 rank wavelengths of fisher's ratio, AMGM, MM and MAD.

Feature selection techniques	Wavelength at top-1 rank	Wavelength at top-2 rank
Fisher's ratio	715.84 nm	716.31 nm
AMGM	1123.73 nm	1114.26 nm
MM	1122.24 nm	1121.74 nm
MAD	1124.75 nm	1125.73 nm

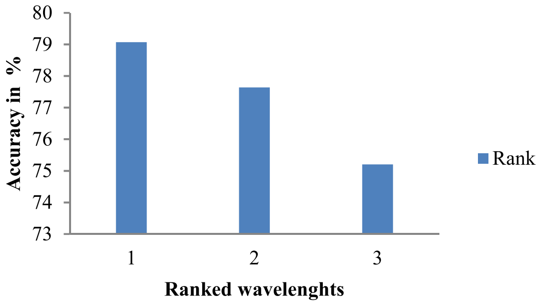


Fig. 6. Classification performance by cumulative addition of top-1, top-2 and top-3 ranked wavelengths.

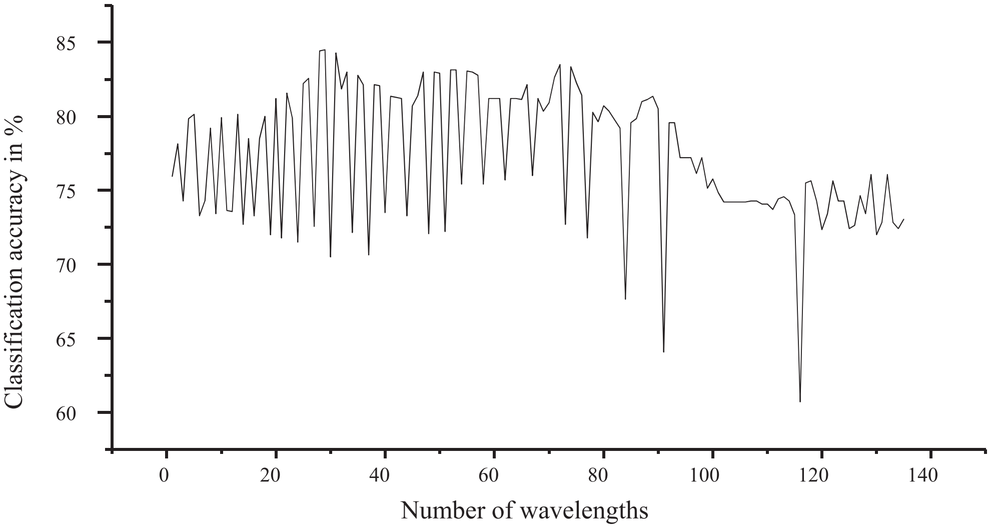


Fig. 7. Performance analysis for individual wavelengths using Fisher's criterion.

For further analysis, four filter-based feature selection techniques were used for the lower range of wavelength i.e. 673 nm–1100 nm. However it was observed from Fig. 4, Fisher's ratio technique outperforms the dispersion measure techniques. Because 84.78% accuracy was obtained using fisher's ratio for 135 cumulative wavelengths based on its relevancy score. The results showed that Fisher's criterion appeared to be the best technique and, further to have an efficient wavelength, an optimized wavelength selection method was adopted here. From Fig. 5(a), two-wavelength ranges were found based on Fisher's criterion. One with the range of 702.72 nm to 752.34 nm having 105 wavelengths and another range of wavelength is 764.34 nm to 778.07 nm having 30 wavelengths. 81.28% classification accuracy was obtained for considering 105 wavelengths and 74.14% for 30 wavelengths. Out of the entire feature selection techniques, Fisher's criterion seems to be the best method of choosing the wavelength from the NIR for the classification of healthy and defective mango. Further, to design any model like NIR camera from the NIR spectrum, an appropriate range of wavelength is needed; hence, efficient wavelength selection can be done using Fisher's criterion.

6. Conclusion

Appropriate wavelength selection methods were proposed from NIR spectroscopy, data ranging from 673 nm −1900 nm for classification of healthy and internally defected mangoes. Wavelength selection methods were proposed based on the feature extraction and feature selection techniques. Experimentation results showed that the wavelength range of 673 nm–1100 nm was the suitable range to detect the internal defects compared to the range 1100 nm–1900 nm. Filter based feature selection techniques were used and fisher's criterion was chosen to be the best wavelength selection technique compared to other selection technique. The highest accuracy obtained was 84.5% with efficient wavelengths. The optimal wavelength was found in the range of 702.72 nm to 752.34 nm using fisher's criterion. This study laid a foundation for further development of the NIR camera/model for the spongy tissue detection on mango fruits.

However, further tuning of our classification model towards predicting the degree of severity of defection would help us in further classification of defected samples into less defected, moderately defected and severely defected and this work is kept as our future research work. Future research in this direction will help to build a more effective model for a large dataset including an efficient classification algorithm and speeding up the processing of data to fulfill the goal of detecting internal defects on mango fruits in real-time.

Credit Author Statement

Anitha Raghavendra- Conceptualization, Methodologies, Software, writing original draft preparation, Writing-Reviewing and Editing. **D.S. Guru**-Supervision, Validation, Visualization, Writing- review and editing **Mahesh.K.Rao**- Supervision.

Conflicts of interest

There is no conflict of interest in our study.

Acknowledgments

Maharaja Research Foundation (MRF), Mysore, supports this research work. We appreciate the support of MRF. We also appreciate Shwetha and Jayakumar, Comtek Scientific Pvt. Ltd., Bangalore for providing NIR spectroscopy to collect the dataset.

Further, we are grateful to all the reviewers who have shared their critical comments which made our manuscript more readable in addition to enhancing the quality of the content.

References

Anitha, Raghavendra, Mahesh, Rao, 2016. A survey on internal defect detection in fruits by non-intrusive methods. Int. J. Latest Trends Eng. Technol. 6 (3), 343–348.

Artur, J. Ferreira, Mário, A.T. Figueiredo, 2012. Efficient feature selection filters for high-dimensional data. Pattern Recogn. Lett. 33 (13), 1794–1804.

Bart, M. Nicolai, Katrien, Beullens, Els, Bobelyn, Ann, Peirs, Saeys, Wouter, Karen, I. Theron, Jeroen, Lammertyn, 2007. Non-destructive measurement of fruit and vegetable quality by means of NIR spectroscopy: a review. Post-Harv. Biol. Technol. 46, 99–118.

Chulmin, Yun, Jihoon, Yang, 2007. Experimental comparison of feature subset selection methods. IEEE Int. Conf. Data Min. https://doi.org/10.1109/ICDMW.2007.77.

David, C. Slaughter, Carlos, H. Crisosto, 1998. Non-destructive internal quality assessment of kiwifruit using near-infrared spectroscopy. Sem. Food Analys. 3, 131–140.

Dospatliev, L., Katrandzhiev, N., Kostadinova, G., 2013. Use of near infrared spectroscopy technology for assessment of the internal quality of some fruits and vegetables-review. For. Sci. Technol. 3, 39–48.

Fraser, D.G., Kunnemeyer, R., McGlone, V.A., Jordan, R.B., 2000. Near infrared light penetration into an apple. Postharvest Biol.Technol. 22, 191–194.

Guru, D.S., Mahamad, Suhil, Lavanya, Narayana, Raju, Vinay, Kumar, N., 2018. An alternative framework for univariate filter based feature selection for text categorization. Patt. Recogn. Lett. 103, 23–31.

Holm, F., 2002. European cooperation of the development of sensors for food quality assessment. Innovation workshop flair-flow. Advanced sensor devices for on-line monitoring of food quality, Sofia http://www.ceeri.res.in/main/x-ray%20imaging.pdf.

Huan, Liu, Hiroshi, Motoda, Rudy, Setiono, Zheng, Zhao, 2010. Feature selection: an ever evolving frontier in data mining. JMLR: Workshop Conf. Proc. 10, 4–13.

Isabelle, Guyon, Andre, Elisseeff, 2003. An introduction to variable and feature selection. J. Mach. Learn. Res. 3, 1157–1182.

Jun, Chin, Ang Andri, Mirzal, Habibollah, Haron, Haza, Nuzly, Hamed, Abdull, 2016. Supervised, unsupervised, and semi-supervised feature selection: a review on gene selection. IEEE/ACM Trans. Comput Biol Bioinform. 13 (5), 971–989.

Krivoshiev, G., Chalucova, R., Moukarev, M., 2000. A possibility for elimination of the peel disturbing effect in non-destructive determination of fruit and vegetables internal quality by Vis/NIR spectroscopy. Lebensm.-Wiss. Technol. 33, 344–349.

Lammertyn, Jeroen, Ann, Peirs, Joss, De Baerdemaeker, Bart, Nicolai, 2000. Light penetration properties of NIR radiation in fruit with respect to non-destructive quality assessment. Postharvest biol. Technol. 18 (2), 121–132.

Manpreeth, Singh, Ramesh, Kumar, 2010. Quality parameters of mango and potential of non-destructive techniques for their measurement-a review. J Food Sci Technol-Springer. 47 (1), 1–14.

Mayank, Goel, Eric, Whitmire, Alex, Mariakakis, Scott, Saponas T., Neel, Joshi, Dan, Morris, Brain, Guenter, Marcel, Gavriliu, Gaetano, Borriello, Shwetak, N. Patel, 2015. HyperCam: Hyperspectral Imaging for Ubiquitous Computing Applications. pp. 145–156 UbiComp'15.

Meurens, M., Feth, F., 2001. Detection of Internal Defects Inside Defects Inside Apples by Vis/NIR Transmission Spectroscopy. www.inapg.inra.fr/ens_rech/siab/asteq/abstract_3rdagm.

Nemirko, A.P., 2016. Transformation of feature space Basedon Fisher's linear discriminant. Patt. Recogn. Image Analys. 26 (2), 257–261.

Peiris, K., 1997. Non-destructive determination of solids content of peach by near infrared spectroscopy. Sensors for Non-destructive Testing Measuring the Quality of Fresh Fruits and Vegetables. Processing from the Sensors for Non-Destructive Testing International Conference.

Peiris, K., Dull, G., Leffler, R., Kays, S., 1998. Near infrared spectroscopic technique for non destructive determination of soluble solids content in processing tomatoes. J. Amer. Soc. Hort. Sci 123, 1089–1093.

Peirs, A., Touchant, K., Schenk, A., Nicolai, B., 2001. FT-NIR spectroscopy to evaluate picking date of apples. ISHS Acta Horticult. 553, 477–480.

Qiang, Lu, Ming, Jie, Tang, Jian, Cai, Rong, Zhao, Jie-wen, Saritporn, Vittayapadung, 2011. Vis/NIR hyperspectral imaging for detection of hidden bruises on kiwifruits. Czech J. FoodSci. 29 (6), 595–602.

Sandeep, S. Musale, Pradeep, M. Patil, 2014. Database development of defective and healthy alphonso mangoes. Int. J. Adv. Agric. Environ. Eng. 1 (1), 2349–2531.

Sang, Ha, Kyu, N.O.H., CHOI, Hong, 2006. Nondestructive quality evaluation technologies for fruits and vegetables. International Seminar on Enhancing. International Seminar on Enhancing Export Competitiveness of Asian Fruits Corpus ID: 9864344.

Włodzis, Ław Duch, Tomasz, Winiarski, Jacek, Biesiada, Adam, Kachel, 2003. Feature selection and ranking filters. Artificial Neural Networks and Neural Information Processing – ICANN/ICONIP.