



Supervised Learning Aided Multiple Feature Analysis for Freshness Class Detection of Indian Gooseberry (*Phyllanthus emblica*)

Tanmay Sarkar¹ · Alok Mukherjee² · Kingshuk Chatterjee²

Received: 15 May 2021 / Accepted: 22 September 2021 / Published online: 12 October 2021
© The Institution of Engineers (India) 2021

Abstract A supervised learning-based simple three class freshness detection algorithm is presented in this paper for prediction of freshness of Amla samples. Six major features from the red–green–blue (RGB) and hue–saturation–vital component (HSV) colourspace and ten other minor features have been studied here with the proposed artificial neural network (ANN) model. The proposed freshness classifier is computationally light due to the use of only ANN as the major classifying tool. More importantly, the analysis is based on images captured on smart phone only, which enables portability and hence, wide acceptability of the scheme. Accuracy of classification higher than 96.5% is achieved using the hue histogram of the image, followed by green layer histogram and others. All the major features were able to produce more than 83% efficiency in freshness class determination; whereas, minor features could achieve a highest classification accuracy of about 77%; clearly suggesting advantage of the major set of features. High efficient freshness categorization, combined with ease of computation, simple feature analysis and use of smart phones for image acquisition ensures its high possibility of real life implementation, especially incorporating within mobile application-based software development.

Keywords Major and minor feature · Feature extraction · Freshness class · Artificial neural network (ANN) · Accuracy of classification

Introduction

An intense relationship can be established between food quality and customer satisfaction. Customers always play a key role while determining the food quality. Customer satisfaction about each quality attribute of a product contributes to the overall quality, in order to make assurance about the food product quality and also to safeguard the seafood market quality maintenance is now extremely important to importing countries [1]. Sensory attributes like texture, colour, size, contribute to the exterior quality of agricultural products [2].

The erroneous output resulting from improper evaluation method is a major problem while determining the quality measuring factors visually because the process is expensive and time-consuming “best-if-used-before date” and such factors decide the market price of the product. Quality can never be maintained properly if a manual inspection method is employed for the analysis of fruits and vegetables. On the other hand, the machine vision analysis method may be more appropriate because of precision in result. Recently for agricultural industries image processing and computer vision system is an area of interest to the researches [3].

Currently, agro-food sectors, healthcare and environmental monitoring systems are using mobile diagnostics which produces rapid results based on precise primary analysis techniques. This also helps to avoid the time-consuming analysis which requires expensive instruments and also trained manpower, thus the mobile diagnostic

✉ Tanmay Sarkar
tanmays468@gmail.com

✉ Kingshuk Chatterjee

¹ Malda Polytechnic, West Bengal State Council of Technical Education, Government of West Bengal, Malda 732102, India

² Government College of Engineering and Ceramic Technology, Kolkata, India

method is a cost-effective and time-saving method. The non-invasive and rapid food scanning system is one of the emerging sectors in food technology. The low-cost, rapid, portable smartphone-based application may be used on-site for a decisive conclusion and able to replace the routine tests, trained personnel-based laboratory practices [1].

Amlaki or amla or Indian gooseberry (*Phyllanthus emblica*) is a pharmaceutically important fruit known for its immense health benefits and recognized in traditional Indian (Ayurvedic), Tibetan and Chinese folk medicines [4, 5]. It is used as a raw material in different food product formulations like candy, toffee and bar [6]. Both in the production as well in the export of the fruit, India possessed the first position [7]. It is a rich source of ascorbic acid (500–1500 mg/100 g) along with aspartic acid, proline, cysteine, alanine, and glutamic acid (essential amino acids). Phosphorus, calcium and iron are the main minerals present in amla. Due to the presence of beneficial bioactive components and several essential nutrients, the fruit possess anti-inflammatory, anti-microbial and anti-oxidant activities [8]. The fruit is seasonal in nature and available from December to March, it is shelf-stable for around 6–9 days. The alteration in the textural feature, changes in shape, generation of prominent black and/or brown spots along with bruises are the visual attributes to measure the extent of freshness, though the visual attributes depend on the perception of the expert. The texture, colour and shape are the most significant parameters that contribute towards consumer acceptance [9].

Fruit rotting is a natural phenomenon. Initially, the thick and hard cell wall maintain the quality or freshness, but in due course, the outer cell walls become soft and ultimately broken down, thus the fruit is exposed to the ambient environment. The softer outer cell wall is easy to invade by microbes and pests, thus the softening of the cell wall is the key regulator for the shelf-stability of fruits [10]. A series of biochemical reactions take place to alter the shape, texture, colour and other characteristics of the fruit.

Fruits are different from each other by means of colour, texture and shape. There are large variations with respect to colour even for the same batch of fruits. This is one of the main challenges to develop smart phone-based fruit quality detection system. The seasonal variation, cultivar, maturity level, climate condition, and the position of the fruit in trees affect the pigment development in fruits, thus responsible for specific colour attributes of the fruit. Expert opinion with manual verification is practiced therefore in the industry as well as in market place by a retail consumer. The dependency on expert opinion during the on-site procurement of the fruit by common people (or by industry) may be the reason for compromising the quality of the fruit [11]. Supervised or semi-supervised learning followed by the development of a robust classification model may

resolve the problem. In this work, a supervised learning-based classifier scheme has been developed for assessing the edibility of amla (*Phyllanthus emblica*) samples. The proposed model is simple in the analysis as it considers only an artificial neural network for designing the proposed three class freshness detection model. Six major features, viz., the Hue (H), Saturation (S) and Value (V) components from the HSV colour space and Red (R), Green (G) and Blue (B) layers of the RGB colour space are analysed here; along with ten minor features such as mean, standard deviation, entropy, kurtosis, skewness, root mean square (RMS), contrast, correlation, energy and homogeneity using RGB colour space and Gray-Level Co-Occurrence Matrix (GLCM) colour feature. The proposed model is found extremely efficient in detecting the freshness level, especially using the Hue and Green layer histograms of the HSV and RGB colour space, respectively. Most importantly, the proposed scheme is designed to analyse the images of amla fruits captured using smartphones, which is an absolute common tool among the common people. The use of smart phone-captured image-based freshness detection model aids this field of research in a bidirectional manner. First of all, the portability and ease of accessibility of this device help in capturing images in markets or any other places to aid easy freshness detection.

The flexibility of the artificial neural network (ANN) in comparison with the classical methods as well as its compatibility with complex and unstructured data from the natural world makes ANN a useful tool for analysing different aspects of food technology [12]. Image analysis along with ANN is in practice for the development of the non-invasive tool in food quality control and food safety-related challenges. The ability of the neural classifier to search patterns with a supervised learning approach is another advantageous characteristic of the ANN classifier. Potentially active molecules against *Escherichia coli* are predicted with the ANN model [13]. The spray-dried powdered juice is classified based on its quality with the ANN model [14]. Researchers have investigated on internal and external features of oranges are detected with ANN [15].

Finally, and most importantly, since the images are captured in the smartphones itself and the artificial neural network (ANN)-based algorithm used here is extremely simple in analysis; it demonstrates all practical possibilities of a probable mobile application-based design of the proposed scheme. This would help people immensely to capture the images directly and receive an approximate measure of the freshness of the samples instantly.

The aim of the work is to develop supervised learning aided multiple feature analysis model for on-site freshness determination of Indian gooseberry (*Phyllanthus emblica*).

Table 1 Comparative analysis of different research works employing supervised learning schemes

Fruit/ vegetable considered	Aim	Pre-processing used	Features	Classifier considered	Accuracy	Reference
Mango, grape, beans, and chili	Powdery Mildew disease detection	Correction of Shade; artifacts removal followed by formatting	Colour and GLCM features	ANN	87.80%	[16]
Apple	Bruise detection	Noise removal, contrast stretching	GLCM features	ANN	–	[17]
Tomato	Quality analysis in terms of colour and hardness	Resize	Size, hardness, colour features	ANN, PNN, radial basis function	–	[18]
Apples (Jonagold variety)	Defect detection	Background separation	Pixel level intensity	ANN, support vector machine	95.10% for healthy 84.80% for defected	[19]
Apple	Defect detection	Calibration	Pixel gray value	ANN	95.40%	[20]
Apple	Defect detection	Background removal, filtering	–	ANN, principal component analysis	79.00%	[21]
Apple	Detection of chilling injury	–	Spectral signature	ANN	98.40%	[22]
Tangerine (<i>Citrus clementina</i>)	Detection of rottenness	–	Spectral features	ANN	98.00%	[23]
Strawberry	Bruise detection	–	Spectral features	ANN, linear discriminant analysis	100.00%	[24]
Tomato	Defective and non- defective; ripe and un-ripe classification	Noise removal and segmentation	GLCM features	ANN	96.47%	[25]
Orange, Potato	Quality and damaged product classification	Histogram equalization	GLCM, colour feature	ANN	Potato = 92.22% Orange = 82.85%	[2]
Avocado, mango	Automatic classification	Histogram equalization	Colour and texture features	BOUNDSTAR algorithm	94.00%	[26]

Related work

Table 1 lists some of the researches in this field, their method of analysis and key outcomes of their experiment; which have helped us immensely to design the proposed model of amla fruit freshness detection; as well as, in planning for the course of the experimental work.

Methodology

The different stages of carrying out the experiment and attributes of design of the proposed ANN-based amla quality assessment model are illustrated in this section.

Collection of fruit samples

Fresh quality, sound and mature (fruits with age of 12–15 weeks after flowering with yellowish-green coloured skin and brown coloured seed, diameter 2.6–3.5 cm, weight 15–21 g, and total soluble solid 9–12°B; visual colour, firmness and size were the indices to ascertain the maturity. Shinier and lighter appearance was observed on the outer skin of the mature fruits in comparison to the immature (with creamy-white seed) one) Indian gooseberry (amla) fruits were procured (the fruits were harvested on the same day) from Agri Horticultural Society of India, Kolkata, India. In total 383 numbers of fruits were purchased, out of which 50 samples were selected randomly (the simple random sampling is considered during the sample collection with the margin of error (M_E) of ± 0.05

and 95% confidence level (C_L) at a sample proportion (S_P) of 50%). The sample size (SZ) was determined using the formula as follows:

$$SZ = \text{Population size} \times Z / (Z + \text{Population size} - 1) \quad (1)$$

where

$$Z = C_L^2 \times S_P \times (1 - S_P) / M_E^2 \quad (2)$$

The samples were brought to the laboratory immediately after purchasing (within 15 min after procurement). The samples were kept in natural ambient conditions (room temperature 25 ± 5 °C, relative humidity $80 \pm 5\%$) and away from direct sunlight. No preservatives from external sources and/or packaging materials were used to maintain the natural decay of the samples.

Acquisition of fruit images

From day 1 (the day of purchasing the fruit) images were acquired thrice daily (in an interval of 8 h) with a smartphone (Redmi Note 9 Pro) with an 8 nm processor, 4 GB RAM, 128 GB memory, 48-megapixel camera, android system and Samsung Isocell GM2 sensor in different angles. The distance between the smartphone and amla fruits (placed on a table with dimension 0.61×0.91 m) were 20 cm. The room (5×4 m) was illuminated with 4 LED lights (Eveready 20-Watt LED Batten, 20 W, $113.5 \times 2.5 \times 3.6$ cm, luminous flux of 100 lm/W Lumen, colour temperature 6000 K) (Fig. 1).

Evaluation of good, intermediate and bad fruit

To evaluate the class of the fruits (good, intermediate and bad) with respect to the colour, shape and texture a semi-trained evaluator panel was formed. During the selection of the panelists (30 males and 39 females in the age group of 21–45 years), ISO 8586-1 (1993) was maintained; while the ability of colour identification and differentiating two fruits were judged as per the ISO 5496:2005 and ISO 10399:2004, respectively [27]. The colour, shape and

texture are the three sensory attributes considered by the panel members to ascertain the class of a specific sample.

The *naïve assessors* (thrice the number of the panelists) were recruited (mixed panel) based on their willingness, availability, attitudes towards the amla fruit, aptitude and knowledge, and communication ability. They were trained to become the *initiated assessors*. Finally, the panel members were selected based on screening test like colour vision, sensitivity test, matching test, discriminating and acuity test, and descriptive ability to get the *expert assessors*. Testing and monitoring of the performance of the panelists were analysed by their reproducibility and repeatability.

Image pre-processing

A minimal pre-processing is applied over the amla images before application into the proposed model. The amla fruits are round in shape; hence, the images have been cropped from the centre part of it to accommodate the maximum amla surface in the cropped square image. Further, the image has been resized to provide uniformity to each of the sample images. Apart from this cropping and resizing, no other specific pre-processing is applied with the images which considerably reduce the computational intricacy of the proposed scheme.

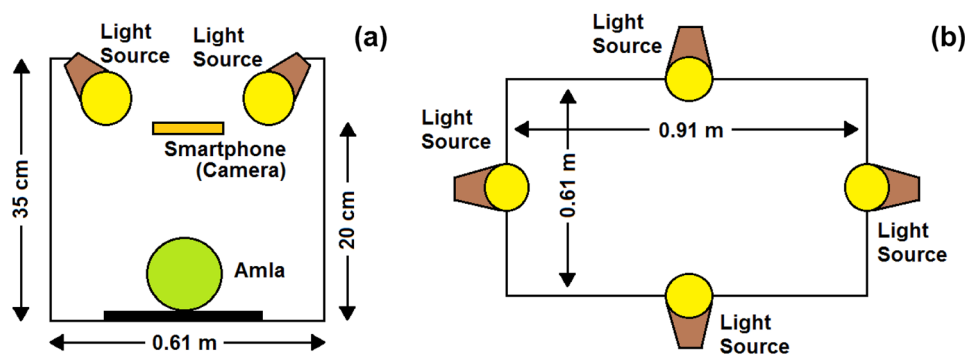
Feature extraction

The proposed model is developed using six major colour spectrum-based features including RGB and HSV colour-space and 10 minor image features as mentioned earlier. These are described as follows:

Major features

The RGB image space contains 3 layers of red, green and blue levels independently. The cropped and resized images, which are already in RGB colourspace by default, are analysed further to separate these three layers. Each layer of the images is analysed to produce the histogram map,

Fig. 1 Schematic diagram for the experimental arrangement for capturing amla images **a** side view, **b** top view (figures not to scale)



which comprises the frequency distribution of occurrence of each intensity level of the image in the range of 0–255. These red, green and blue layer histograms are used as the three of the primary features of this work.

The same cropped and resized images and now converted to HSV colourspace from the default RGB colourspace. These HSV colourspace images are again analysed to segregate the three component layers, i.e., hue, saturation and value components. Histogram maps of each of these layers are now developed to verify the pattern of change in intensity levels among the three classes.

R (red channel), G (green channel) and B (blue channel) Features The R (red channel), G (green channel) and B (blue channel) values are important features in image analysis. The red, green, and blue values of an image can be described as RGB value. Zero value implies no colour representation while the maximum concentrated colour is represented by the value 255. Each of the colours (red, green and blue) uses 8 bits each.

H (hue), S (saturation) and V (value) feature extraction HSV segregates luma (intensity of the image), from colour information (chroma). That is useful during the histogram equalization of the image. During the computer vision application, the colour components should be separated from intensity to remove the shadow or to ascertain the robustness of change in lighting, thus HSV are the key features for smartphone-based image processing. The features named H (hue), S (saturation) and V (value) are determined as follows:

$$H = \begin{cases} \theta B \leq G \\ 360^\circ - \theta B \geq G \end{cases} \tag{3}$$

$$\theta = \cos^{-1} \left[\frac{(R - G) + (R - B)}{\sqrt{(R - G)^2 + (R - B)(G - B)^{1/2}}} \right] \tag{4}$$

$$S = 1 - \frac{3}{R + G + B} \tag{5}$$

$$I = \frac{1}{3}(R + G + B) \tag{6}$$

where R = red, G = green, B = blue channel value.

Minor features

Apart from these major features, another set of minor or secondary features are also analysed using the proposed ANN architecture in order to distinguish between the three freshness classes of amla samples. These features of primary the different parameters of the image which include mean, standard deviation, entropy, kurtosis, skewness, RMS, contrast, correlation, energy and homogeneity using

Gray-Level Co-Occurrence Matrix (GLCM). These are described in detail in the following analysis.

Gray-level co-occurrence matrix (GLCM) features The texture of the sample image can be described through GLCM features by considering how often pairs of pixels with specific values and in a specified spatial relationship occur in the image, creating a GLCM, and then extracting statistical measures from this matrix. It is basically the computational characteristics of the probability density function of the grey level matrix of the concerned image and can be represented as follows:

$$GLCM(p_1, p_2); \quad p_1, p_2 = 0, 1, 2, 3, \dots, P - 1 \tag{7}$$

where P = numbers of pixels present in the image.

The GLCM features considered for this work are mean [28], standard deviation [29], entropy [2], kurtosis, skewness, RMS, contrast [2], correlation [28], energy and homogeneity [2, 16, 30, 31].

Basic methodology of classifier design

Six major and ten minor features of the amla image samples have been analysed using a supervised learning scheme employing artificial neural network architecture. A gradual changes in the surface colour of the samples with days of progression have been observed. It has been observed that the Amla samples gradually develop dark patches, wrinkles all over the surface; as well as, develop a gradual variation of the base colour of the surface. Hence, the colour features of the samples have been considered only, instead of concentrating more on the textures of the samples. Thus, three major colour features of the RGB colourspace of the amla images, e.g., red, green and blue layer have been considered; as well as the image has been converted to HSV colourspace and obtained the hue, saturation and vital component features. These features are analysed independently using separate ANN models to develop the classification of the amla images. Three classes have been predefined as Good, Intermediate and Bad, as mentioned earlier. The proposed models are validated using fivefold cross-validation to remove sampling error while validating the scheme.

Apart from these six major colour features, ten other minor features, viz, Mean, Standard Deviation, Entropy, Kurtosis, Skewness, RMS, Contrast, Correlation, Energy and Homogeneity of the sample images have also been studied with independent ANN architectures. These features represent specific parameters of the images and are different in nature from the colour histograms. Thus, an effort has been made to study some of the general and diverse features of the images and tried to distinguish the difference between the three different classes. The

difference that lies in analysis methods with major and minor features is that the major features contain the histogram values of the basic colour features, whereas the minor features represent a single value of an entire image, such as mean, standard deviation etc.; all of which are representations of the parametric variation of the intensity values of the image pixels. Thus, each image is represented by an array of 256 scalar quantities in case of major features; whereas each image is denoted by a single scalar point in case of minor features. Thus, the study has been segregated into the analysis using the ANN architecture with the major and minor features. Similarly, the validation of each model for the two separate classes of features is done independently as well as the results are also described separately. The proposed freshness classification scheme is also described graphically in Fig. 2.

Ann architecture design

A simple three-layered artificial neural network architecture is designed here to implement the proposed algorithm. The three layers include an input layer, a hidden layer, and an output layer. The input layer receives the input attributes, which are the six different histogram features of the HSV and RGB colourmap, and the ten minor features obtained from the RGB image or the GLCM of the grayscale image. These features have been analysed independently, using separate networks. Next is the hidden layer, which receives weighted input values from the input layer and analyses the signals using a predefined transfer function to predict the output. Thus, the output layer provides the response of a specific input [2, 32]. In the proposed work, the output layer yields information regarding the quality grades of the amla samples. A typical sigmoidal activating function is used in this work [33, 34]. In this work, the input to the input layer is the histogram features while considering the major features; or the array of scalar values for the minor features.

The training set, test set, and validation set are distributed here in the ratio of 70:15:15. A high number, with 1000 iterations is carried out here. The learning rate is also varied from 0.5 to 0.9. The proposed neural network model for freshness assessment of amla images is developed with 1 input layer with 256 input nodes, 10 hidden layers and 1 output layer with 3 output nodes for the major feature analysis (except R) model as it contained 256 scalar points of the RGB or the HSV histograms; and 1 input layer with 1 input node, 10 hidden layers and 1 output layer with 3 output nodes for the minor feature analysis (except Kurtosis) model, as these contained single point representation of the image. For the R and Kurtosis model, the hidden layers were 5. These ANN configurations are shown in Fig. 3. The Levenberg–Marquardt ANN model is adopted

here with a nonlinear activation function and backpropagation algorithm for the proposed feed-forward ANN architecture.

Analysis of features

Each image has been analysed to extract the major and the minor features for the three classes. Six major features are considered here: Red, Green and Blue layer of the RGB colour space, and Hue, Saturation and Vital components from the HSV colour space. The total set of 50 images from each class have been studied further and constructed the histograms of each of the six major feature layers. The average levels of the 50 sample histograms are shown in Figs. 4 and 5 for the RGB and the HSV features, respectively. These figures apparently demonstrate that the hue levels of the three classes are quite distinctly separable; especially the Bad class is distinctly separable from the remaining two classes. Among the rest of the major features, the green layer histogram also apparently indicates a high probability for distinct separation of classes. The remaining features also show definite trends for classification; although, with reduced affirmation. Hence, all these major features are considered for analysis using the proposed ANN-based model.

Minor features, on the other hand, are analysed initially using a different methodology; since these features are parametric values only, inserted of an array of scalar values similar to the histograms of the major features. In total 50 samples of each class are analysed to develop boxplots for the three separate classes, from which separation of the sample images among the three classes could well be inferred. The boxplots of the ten minor features are shown from Figs. 6, 7, 8, 9 and 10, respectively.

These boxplots well represent the effectiveness of each minor feature for the freshness class separation. It is well observed from these figures that 4 minor features: Kurtosis (Fig. 7b), Contrast (Fig. 9a), Energy (Fig. 10a) and Homogeneity (Fig. 10b) are the most prominent out of the total 10 number of minor features in segregating freshness classes most effectively. This inference stands strong as these have the considerable separation in the respective class medians, as well as, in the quartile levels as observed from the respective figures; although, all of these features have some overlap with the neighbouring classes. Other minor features, except these 4, have a major overlap of values among the three freshness classes, and in some of the cases, these are very poorly separable, or even inseparable due to almost complete overlap. Hence, these less efficient features are discarded for the rest of the analysis, and only the 4 identifiable features, viz., Kurtosis, Contrast, Energy, and Homogeneity are considered for subsequent

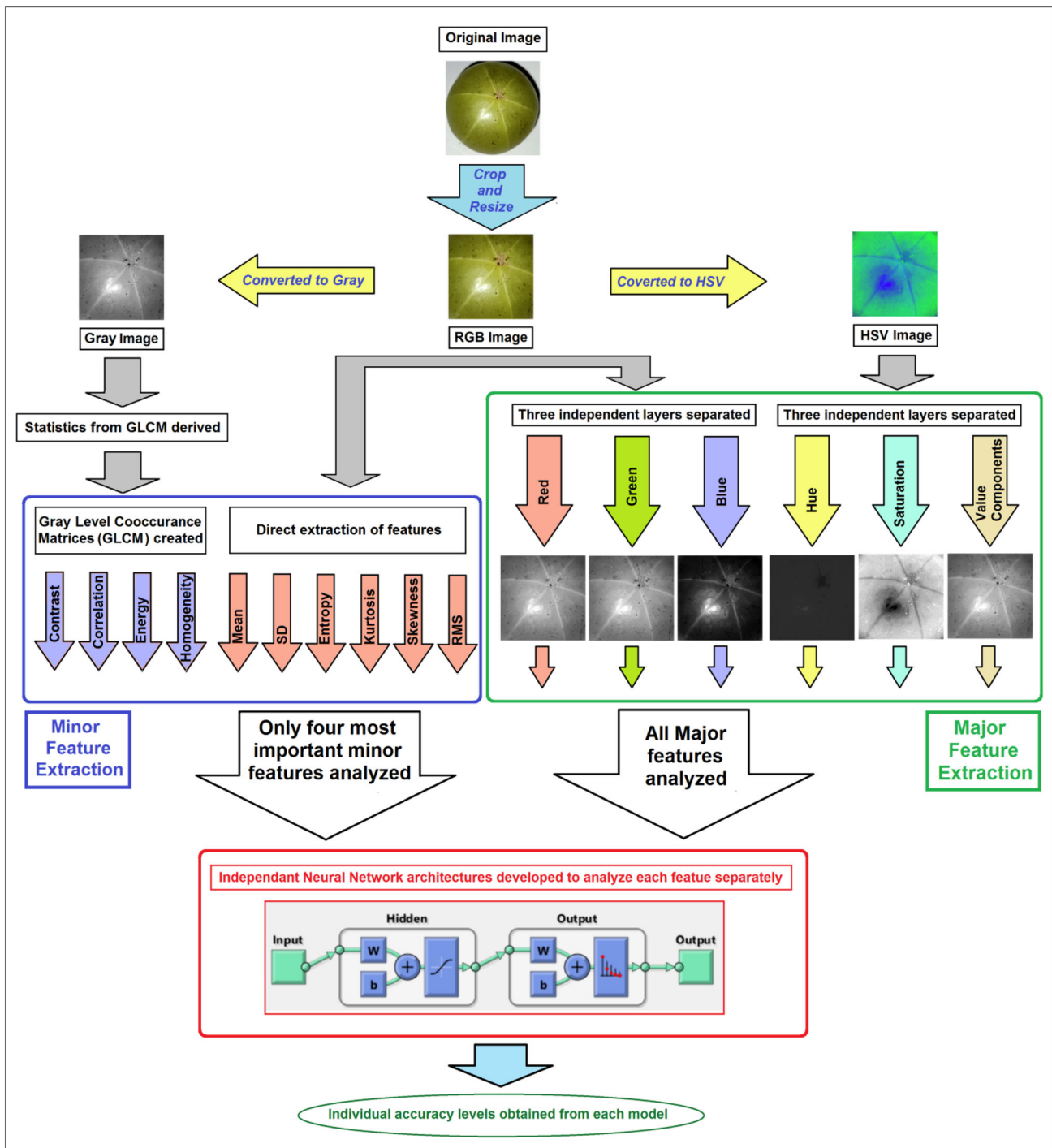


Fig. 2 Flowchart of the proposed algorithm

analysis using the proposed supervised learning algorithm employing an artificial neural network.

Result and discussion

The proposed classifier is validated using a fivefold cross-validation scheme to classify samples into Good, Intermediate and Bad classes. This is described in the next section.

Fig. 3 The ANN model for a major and b minor features

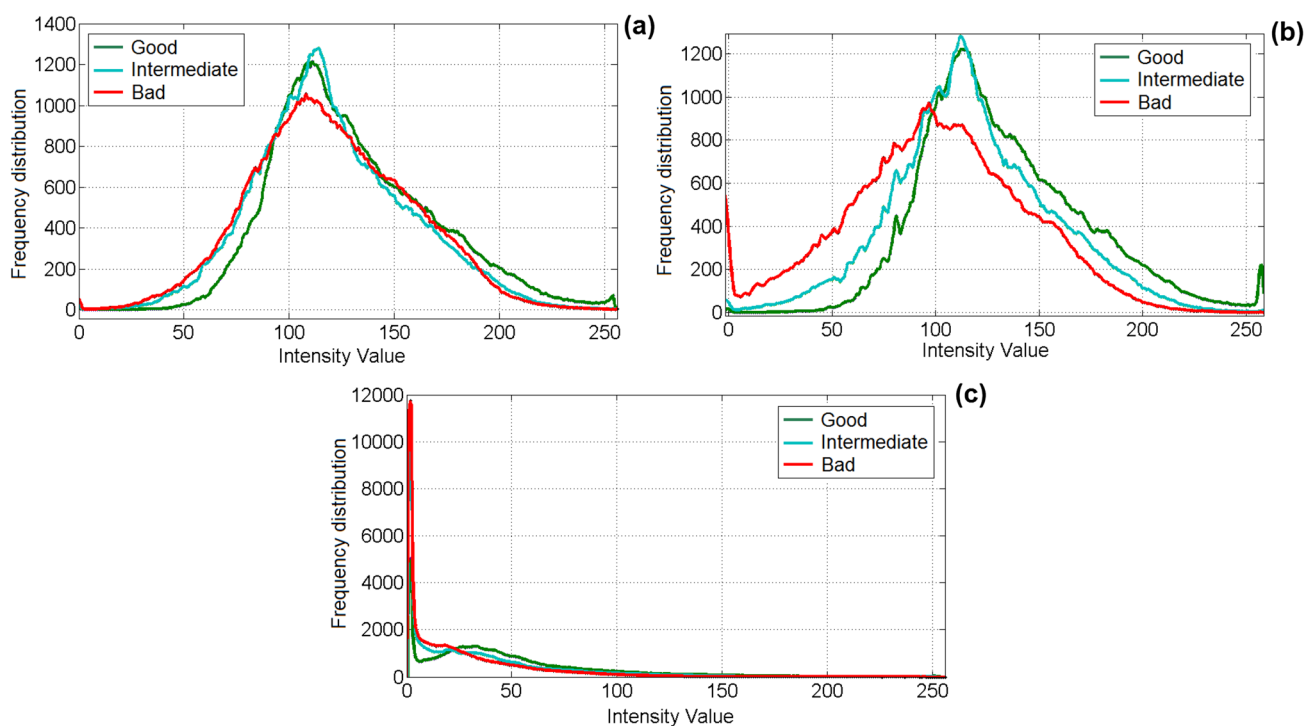
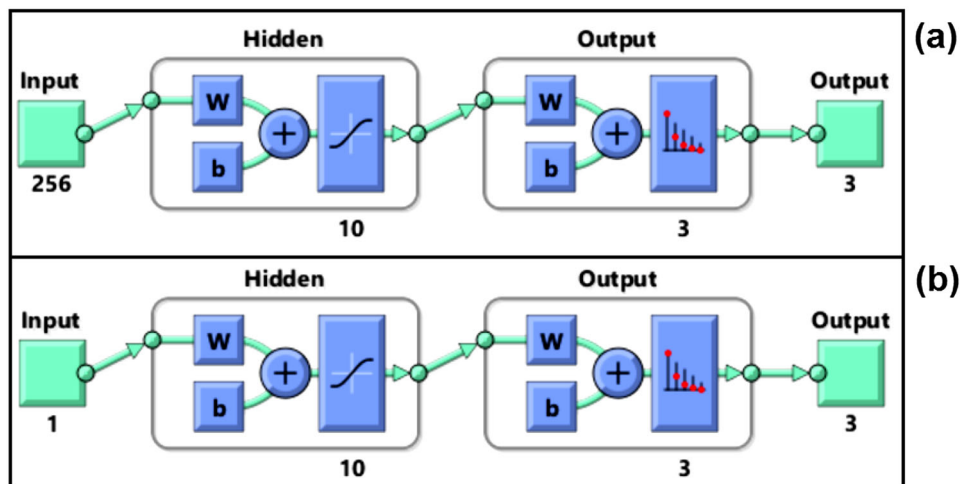


Fig. 4 Frequency distribution vs. intensity value of a red, b green and c blue layers

Method of cross-validation

The k-fold cross-validation is often considered in several researches to validate classifier models. A fivefold cross-validation model has been constructed in this proposed work to establish the proposed work due to its easy implementation characteristics along with simplicity. Fifty samples have been studied in this proposed work throughout 8 days and images were captured. Fifty image samples from each class have been studied from each of the three groups: Good, Intermediate and Bad. Hence, each

class contains 50 samples in total, which further makes the total data set length 150. Each group of the sample are further partitioned randomly into 5 sets to contain 10 random samples of each class in each set. Thus, each set contains 10 images from each three classes, thereby, making a total of 30 samples in each set. Five such sets are constructed which are used with the fivefold cross-validation model. The proposed ANN model, thus, uses four such subsets, containing 40 samples of each of the three classes, for training while the fifth subset, containing 10 samples from each class, is used as a testing dataset. Similar random

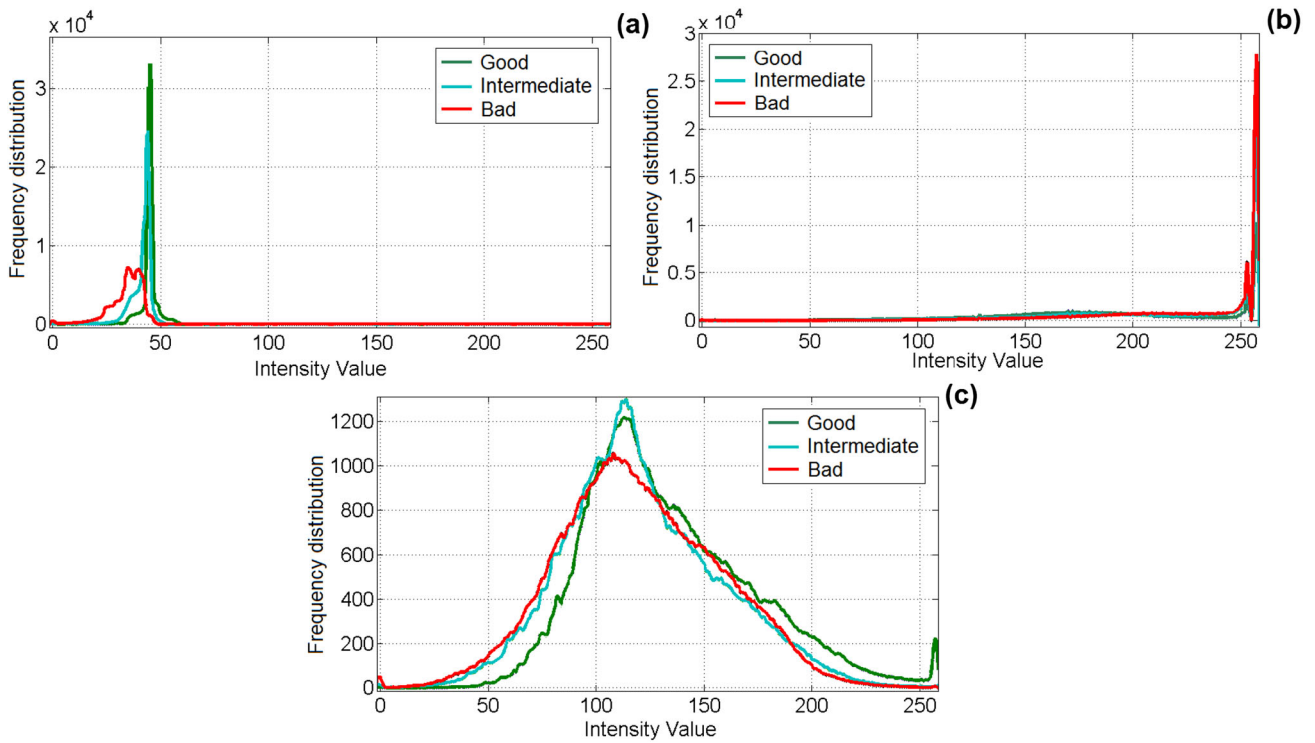


Fig. 5 Frequency distribution vs. intensity value of **a** hue, **b** saturation and **c** vital components

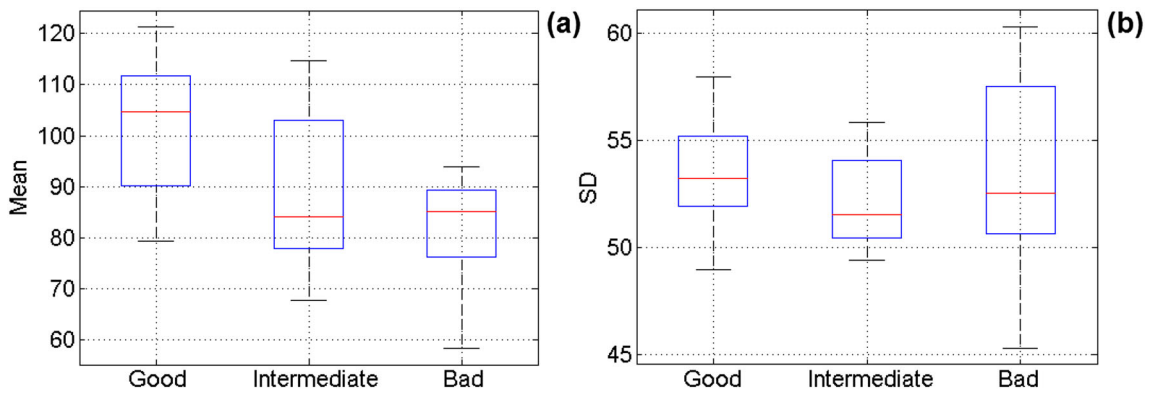


Fig. 6 Box plot separating three freshness classes using minor features **a** mean and **b** standard deviation (SD)

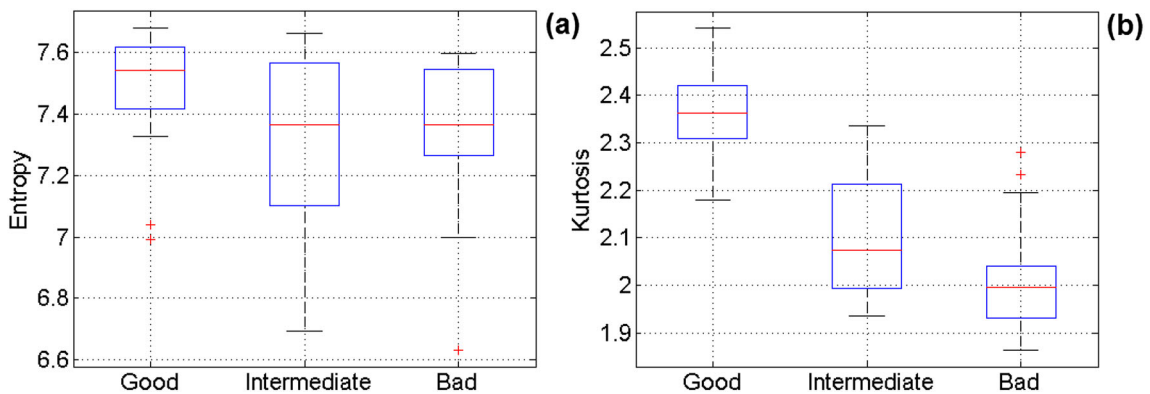


Fig. 7 Box plot separating three freshness classes using minor features **a** entropy and **b** kurtosis

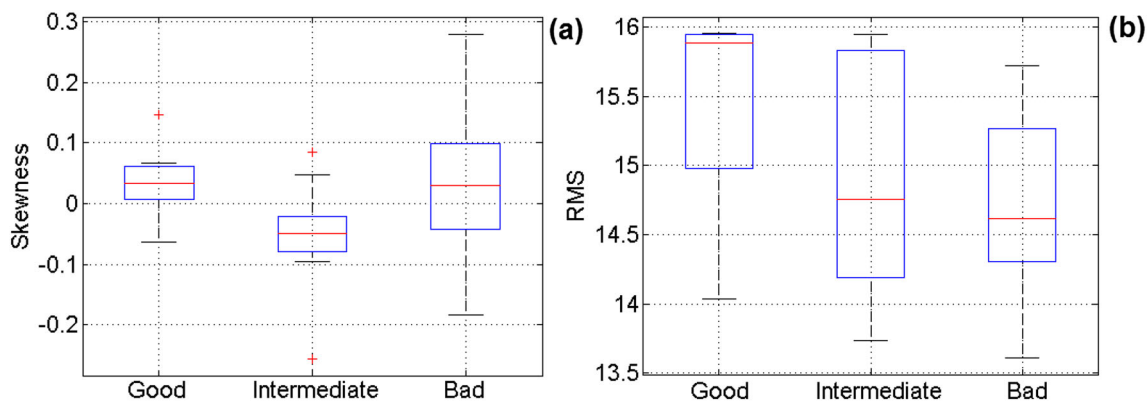


Fig. 8 Box plot separating three freshness classes using minor features **a** skewness and **b** root mean square (RMS)

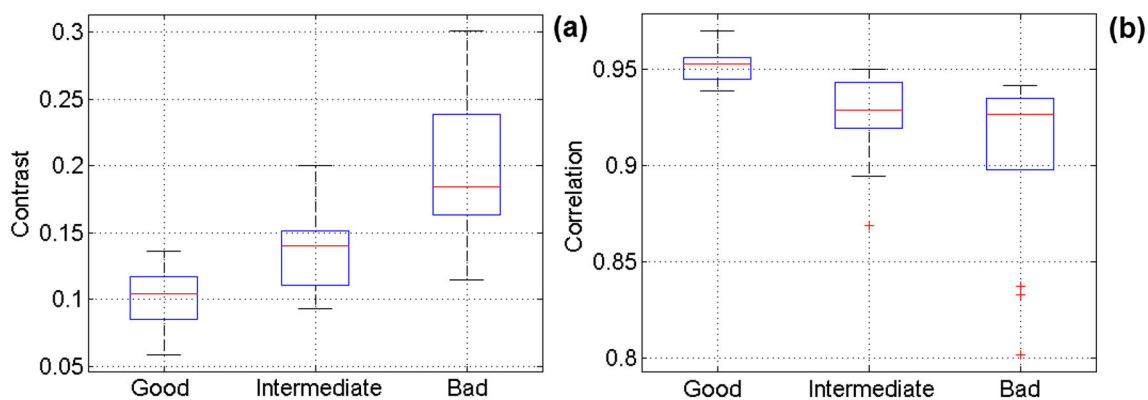


Fig. 9 Box plot separating three freshness classes using minor features **a** contrast and **b** correlation

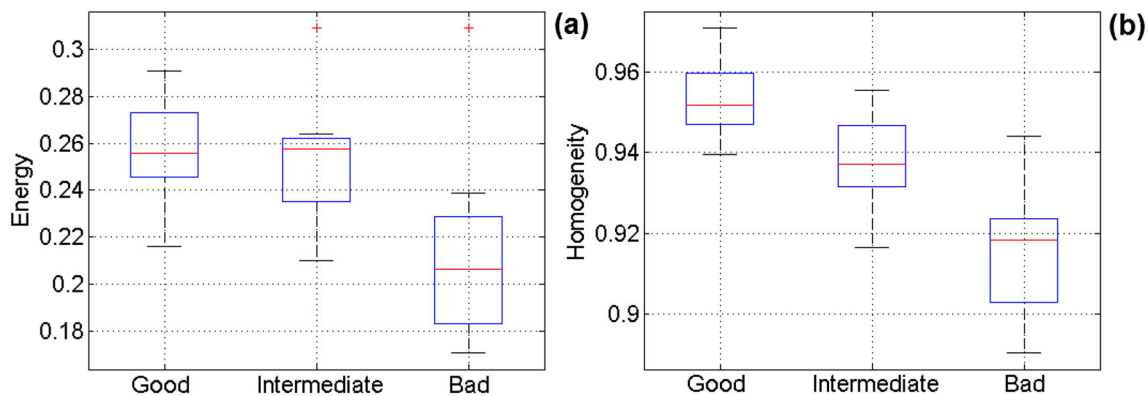
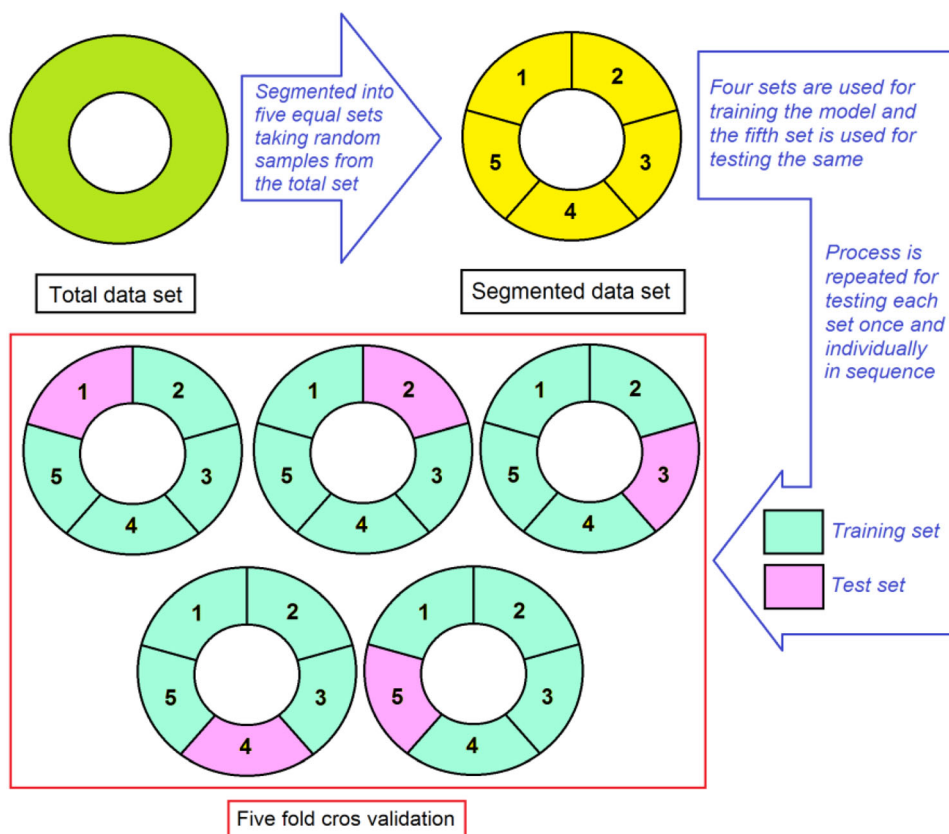


Fig. 10 Box plot separating three freshness classes using minor features **a** energy and **b** homogeneity

sets are constructed for each of the three freshness classes. The entire process is repeated five times ($k = 5$) considering each of the subsets once for testing. The illustrative depiction of the cross-validation is presented in Fig. 11.

The cross-validation method may reduce or nullify the biasness possibility for the proposed ANN model.

Fig. 11 Schematic representation of fivefold cross-validation method



Results with major features

In this work, the accuracy levels of freshness classification of the amla images using few major and minor features with the proposed ANN models have been computed. In this sub-section, the detailed results obtained from the fivefold cross-validation model using the 6 major colour features of the images have been described. These results are shown in Tables 2 and 3.

Results with minor features

The proposed ANN model is further tested for classification accuracy using the minor features obtained from the fivefold cross-validation model. These results are shown in Table 4, along with the final accuracy levels obtained using the major features, described in the earlier sub-section.

Analysis of results and key features of the work

Analysis of Tables 2, 3 and 4 reveal important qualities regarding the outcomes of the proposed classifier. Table 4 shows the summarized form of the results, which shows that Hue histograms of the images yield the highest efficiency of 96.533% for freshness classification of amla sample images, followed by 94.4% efficiency, obtained

with the green layer of the RGB colourmap. This level of efficiency of classification is considerably high compared to existing researches as referred to in Table 2. All the other remaining features such as Saturation and Vital components of the HSV colourmap and Red and Blue components of the RGB colourmap yield more than 83% classifier accuracy with the three freshness class model; which is again appreciable.

Freshness analysis has been carried out by researchers for fruits like apple [35–37], banana [38], amla [39] and mangoes [40] and vegetables like eggplant [41]. The findings of the present work is in agreement with the previous reported results. Apart from this high efficiency of classification, the proposed work is using a simple ANN architecture as the primary classifier tool, which confirms reduced computational burden. More importantly, the model uses amla images captured using smartphones only. Smartphones are portable devices and are widely accessed by a major group of people. The use of smartphones enables the widespread use of the proposed algorithm. High efficiency of freshness classification, combined with ease of computation and image capturing using smartphones enables a broad scope for applying the proposed model in developing mobile application software. This would ensure people to capture fruit images at any point in the market and simultaneously analyse the image using the

Table 2 ANN classifier outcomes using the HSV layers

Test set index	Training set index	Sample class						Total number of correct results	Total number of Wrong results	Accuracy (%)	Overall accuracy (%)
		Good		Intermediate		Bad					
		Correct	Wrong	Correct	Wrong	Correct	Wrong				
<i>Feature analysed: hue layer histogram</i>											
1	2, 3, 4, 5	49	1	47	3	49	1	145	5	96.667	96.534
2	3, 4, 5, 1	50	0	47	3	48	2	145	5	96.667	
3	4, 5, 1, 2	48	2	49	1	50	0	147	3	98	
4	5, 1, 2, 3	49	1	47	3	48	2	144	6	96	
5	1, 2, 3, 4	49	1	46	4	48	2	143	7	95.334	
<i>Feature analysed: saturation layer histogram</i>											
1	2, 3, 4, 5	42	8	44	6	42	8	128	22	85.334	83.067
2	3, 4, 5, 1	43	7	39	11	41	9	123	27	82	
3	4, 5, 1, 2	41	9	43	7	43	7	127	23	84.667	
4	5, 1, 2, 3	42	8	41	9	42	8	125	25	83.334	
5	1, 2, 3, 4	41	9	39	11	40	10	120	30	80	
<i>Feature analysed: value components layer histogram</i>											
1	2, 3, 4, 5	42	8	40	10	43	7	125	25	83.334	84.134
2	3, 4, 5, 1	42	8	41	9	42	8	125	25	83.334	
3	4, 5, 1, 2	44	6	39	11	41	9	124	26	82.667	
4	5, 1, 2, 3	41	9	43	7	44	6	128	22	85.334	
5	1, 2, 3, 4	44	6	42	8	43	7	129	21	86	

Table 3 ANN classifier outcomes using the RGB layers

Test set index	Training set index	Sample class						Total number of correct results	Total number of Wrong results	Accuracy (%)	Overall Accuracy (%)
		Good		Intermediate		Bad					
		Correct	Wrong	Correct	Wrong	Correct	Wrong				
<i>Feature analysed: red layer histogram</i>											
1	2, 3, 4, 5	45	5	40	10	43	7	128	22	85.334	87.067
2	3, 4, 5, 1	46	4	43	7	43	7	132	18	88	
3	4, 5, 1, 2	47	3	42	8	45	5	134	16	89.334	
4	5, 1, 2, 3	44	6	43	7	43	7	130	20	86.667	
5	1, 2, 3, 4	45	5	42	8	42	8	129	21	86	
<i>Feature analysed: green layer histogram</i>											
1	2, 3, 4, 5	47	3	48	2	47	3	142	8	94.667	94.401
2	3, 4, 5, 1	49	1	46	4	46	4	141	9	94	
3	4, 5, 1, 2	48	2	47	3	48	2	143	7	95.334	
4	5, 1, 2, 3	47	3	48	2	48	2	143	7	95.334	
5	1, 2, 3, 4	48	2	45	5	46	4	139	11	92.667	
<i>Feature analysed: blue layer histogram</i>											
1	2, 3, 4, 5	46	4	47	3	48	2	141	9	94	91.734
2	3, 4, 5, 1	46	4	46	4	45	5	137	13	91.334	
3	4, 5, 1, 2	48	2	44	6	44	6	136	14	90.667	
4	5, 1, 2, 3	45	5	45	5	47	3	137	13	91.334	
5	1, 2, 3, 4	49	1	44	6	44	6	137	13	91.334	

Table 4 Summarized form of ANN classifier outcomes using selected minor features

Feature class	Feature analysed	Number of correct predictions (in fivefold cross-validation)	Number of wrong predictions (in fivefold cross-validation)	Classification accuracy (%)
Major	Red	653	97	87.067
	Green	708	42	94.4
	Blue	688	62	91.733
	Hue	724	26	96.533
	Saturation	623	127	83.067
	Value components	631	119	84.133
Minor	Kurtosis	499	251	66.533
	Contrast	542	208	72.267
	Energy	568	182	75.733
	Homogeneity	583	167	77.733

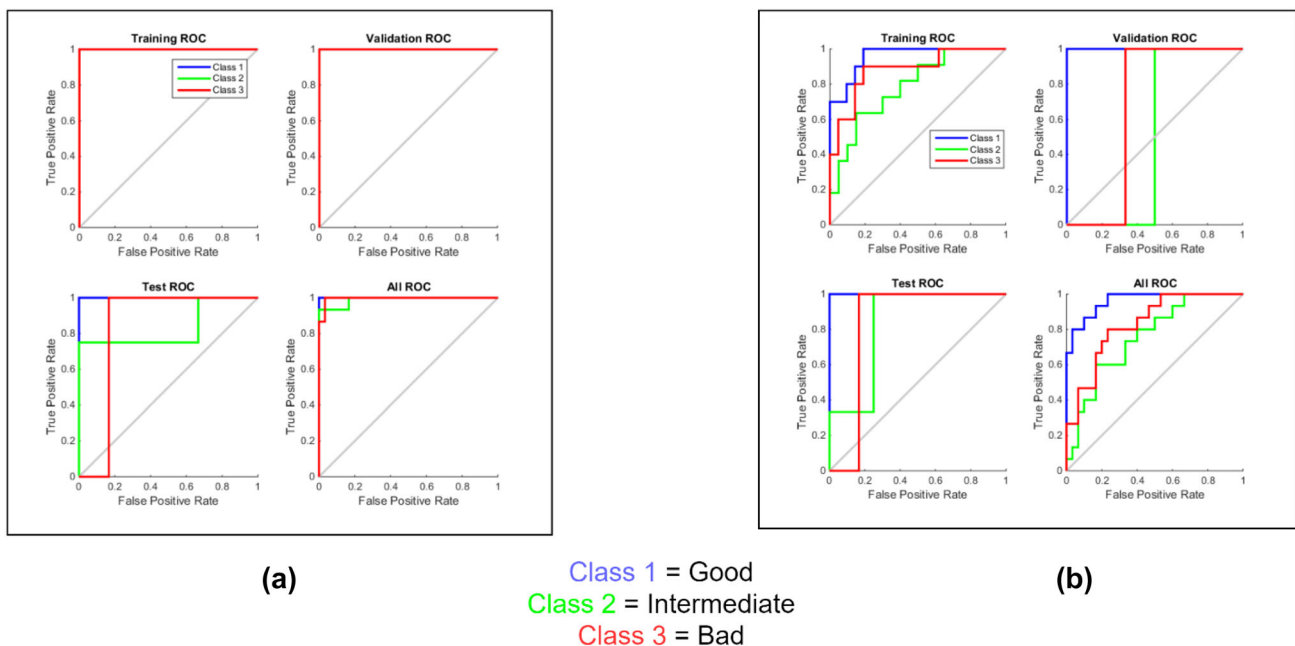


Fig. 12 ROC curve for the ANN model built with feature **a** hue and **b** kurtosis

probable application to obtain an instant prediction of the freshness of the fruit. This would enable wide and effective use of the proposed classifier and establish its practicality in this field of research.

Receiver operating characteristic (ROC) curve analysis

Upon variation of the discrimination threshold the classifier’s performance is illustrated by Fig. 10 (ROC curve). The ROC curve has been constructed by plotting sensitivity

(true positive rate) against 1-specificity (false positive rate) for each possible CO (cut-off) value.

$$\text{True positive rate}(CO_n) = P(QR \geq CO|E+)$$
 (8)

$$\text{False positive rate}(CO_n) = P(QR \geq CO|E-)$$
 (9)

where CO_n is the nth cut-off value, P is the probability density, $E+$, $E-$ = positive and negative error respectively, QR = quantitative results.

Therefore, ROC curve can be represented as follows:

$$\text{ROC} = \{\text{False positive rate}(\text{CO}_n), \text{True positive rate}(\text{CO}_n), \text{CO} \in (-\infty, +\infty)\} \quad (10)$$

In terms of QR, the ROC can be expressed as follows:

$$\text{ROC} = \{(\text{QR}, \text{ROC}(\text{QR})), \text{QR} \in (0, 1)\} \quad (11)$$

where the ROC maps QR to true positive rate (CO_n), and CO_n matching to false positive rate (CO_n) = QR.

Alteration in the scalar threshold value in the proposed ANN model is able to provide the ROC curve. The performance of the proposed ANN model (nonparametric classifier) can be analysed with the aid of the ROC curve [42]. The greater the area under the curve (AUC) better will be the performance of the model. The features with higher AUC thus build a more robust model, which is able to predict the class of the sample with higher reliability. Figure 12a, reveals that the ROC curve for the model built with hue feature possessed maximum AUC for training, testing, validation, and for the overall model, while the same for the model built with kurtosis feature possessed minimum AUC (Fig. 12b). Due to the scale-invariant nature of AUC, it measures the precision of the predictions, rather than their absolute values. It is also invariant with respect to the classification threshold. Irrespective of the classification threshold it helps to estimate the prediction quality of the proposed model.

Conclusion

The proposed work is about the prediction of the freshness of amla (*Phyllanthus emblica*) fruit images among the three predefined freshness classes. The model uses supervised learning tools such as artificial neural network (ANN) to analyse colour gradient based on major image features such as Hue, Saturation and Vital components of the HSV colourmap and Red, Green and Blue layers of the RGB colourmap; and minor features such as mean, standard deviation, entropy, kurtosis, skewness, RMS, contrast, correlation, energy and homogeneity using RGB colour-space and GLCM colour features. The proposed scheme contains reduced computational complexity as it is designed using ANN only. A high efficiency of classification of the proposed model, especially using Hue and Green layers of the amla images; ease of computation and more importantly, use of smartphones for capturing images enable its possibility of using the algorithm in developing smartphones application-based software for instant freshness detection of amla fruits.

Acknowledgements The authors acknowledge Malda polytechnic, Malda throughout the study.

Funding Not available.

Availability of data and materials Not applicable.

Declarations

Conflict of interest Authors declare no conflict of interest.

References

- G. Ratani, P. Dario, F. Cavallo, Smartphone-based food diagnostic technologies: a review. *Sensors (Basel)* (2017). <https://doi.org/10.3390/s17061453>
- S.S.E. Ali, S.A. Dildar, An efficient quality inspection of food products using neural network classification. *J Intell Syst* **29**, 1425–1440 (2020). <https://doi.org/10.1515/jisys-2018-0077>
- A. Bhargava, A. Bansal, Fruits and vegetables quality evaluation using computer vision: a review. *J King Saud Univ Comput Inf Sci* **33**, 243–257 (2021). <https://doi.org/10.1016/j.jksuci.2018.06.002>
- M.S. Uddin, M.A. Al, M.S. Hossain et al., Exploring the effect of *Phyllanthus emblica* L. on cognitive performance, brain antioxidant markers and acetylcholinesterase activity in rats: promising natural gift for the mitigation of Alzheimer's disease. *Ann Neurosci* **23**, 218–229 (2016). <https://doi.org/10.1159/000449482>
- R. Tewari, V. Kumar, H.K. Sharma, Pretreated Indian gooseberry (*Emblca officinalis*) segments: kinetic, quality and microstructural parameters. *J Inst Eng Ser A* **102**, 523–534 (2021). <https://doi.org/10.1007/s40030-021-00538-9>
- M.P. Kapoor, K. Suzuki, T. Derek et al., Clinical evaluation of *Emblca Officinalis* Gaertn (Amla) in healthy human subjects: Health benefits and safety results from a randomized, double-blind, crossover placebo-controlled study. *Contemp Clin Trials Commun* **17**, 100499 (2020). <https://doi.org/10.1016/j.conctc.2019.100499>
- S. Gantait, M. Mahanta, S. Bera, S.K. Verma, Advances in biotechnology of *Emblca officinalis* Gaertn. syn. *Phyllanthus emblica* L.: a nutraceuticals-rich fruit tree with multifaceted ethnomedicinal uses. *3 Biotech* **11**, 62 (2021). <https://doi.org/10.1007/s13205-020-02615-5>
- R. Chaphalkar, K.G. Apte, Y. Talekar et al., Antioxidants of *Phyllanthus emblica* L Bark extract provide hepatoprotection against ethanol-induced hepatic damage: a comparison with Silymarin. *Oxid Med Cell Longev* **201**, 3876040 (2017). <https://doi.org/10.1155/2017/3876040>
- T. Sarkar, S. Saha, M. Saluddin, R. Chakraborty, Drying kinetics, fourier-transform infrared spectroscopy analysis and sensory evaluation of sun, hot-air, microwave and freeze dried mango leather. *J Microbiol Biotechnol Food Sci* **10**, 1–7 (2021)
- B.N. Yao, K. Tano, H.K. Konan et al., The role of hydrolases in the loss of firmness and of the changes in sugar content during the post-harvest maturation of *Carica papaya* L. var solo 8. *J Food Sci Technol* **51**, 3309–3316 (2014). <https://doi.org/10.1007/s13197-012-0858-x>
- I.R. Fermo, T.S. Cavali, L. Bonfim-Rocha et al., Development of a low-cost digital image processing system for oranges selection using hopfield networks. *Food Bioprod Process* **125**, 181–192 (2021). <https://doi.org/10.1016/j.fbp.2020.11.012>
- T. Sarkar, M. Salauddin, S. Hazra et al., Comparative approach of artificial neural network and thin layer modelling for drying kinetics and optimization of rehydration ratio for bael (*Aegle marmelos* (L) correa) powder production. *Econ Comput Econ Cybern Stud Res* **55**, 167–184 (2021). <https://doi.org/10.24818/18423264/55.1.21.11>

13. A. Badura, J. Krysiński, A. Nowaczyk, A. Buciuński, Application of artificial neural networks to prediction of new substances with antimicrobial activity against *Escherichia coli*. *J Appl Microbiol* **130**, 40–49 (2021). <https://doi.org/10.1111/jam.14763>
14. K. Przybył, J. Gawalek, K. Koszela, Application of artificial neural network for the quality-based classification of spray-dried rhubarb juice powders. *J Food Sci Technol* (2020). <https://doi.org/10.1007/s13197-020-04537-9>
15. J. Gómez-Sanchis, L. Gómez-Chova, N. Aleixos et al., Hyperspectral system for early detection of rottenness caused by *Penicillium digitatum* in mandarins. *J Food Eng* **89**, 80–86 (2008). <https://doi.org/10.1016/j.jfoodeng.2008.04.009>
16. J.D. Pujari, R. Yakkundimath, A.S. Byadgi, Recognition and classification of produce affected by identically looking powdery mildew disease. *Acta Technol Agric* **17**, 29–34 (2014). <https://doi.org/10.2478/ata-2014-0007>
17. D. Jawale, M. Deshmukh, Real time automatic bruise detection in (Apple) fruits using thermal camera, in *2017 International Conference on Communication and Signal Processing (ICCSP)* (2017), p. 1080–1085
18. M. Zaborowicz, P. Boniecki, K. Koszela et al., Application of neural image analysis in evaluating the quality of greenhouse tomatoes. *Sci Hortic (Amsterdam)* **218**, 222–229 (2017). <https://doi.org/10.1016/j.scienta.2017.02.001>
19. D. Unay, B. Gosselin, Automatic defect segmentation of ‘Jonagold’ apples on multi-spectral images: a comparative study. *Postharvest Biol Technol* **42**, 271–279 (2006). <https://doi.org/10.1016/j.postharvbio.2006.06.010>
20. D. Ariana, D.E. Guyer, B. Shrestha, Integrating multispectral reflectance and fluorescence imaging for defect detection on apples. *Comput Electron Agric* **50**, 148–161 (2006). <https://doi.org/10.1016/j.compag.2005.10.002>
21. S. Bennedsen, B.D.L. Peterson, A. Tabb, Identifying apple surface defects using principal components analysis and artificial neural networks. *Trans ASABE* **50**, 2257–2265 (2007). <https://doi.org/10.13031/2013.24078>
22. G. ElMasry, N. Wang, C. Vigneault, Detecting chilling injury in Red Delicious apple using hyperspectral imaging and neural networks. *Postharvest Biol Technol* **52**, 1–8 (2009). <https://doi.org/10.1016/j.postharvbio.2008.11.008>
23. J. Gómez-Sanchis, J.D. Martín-Guerrero, E. Soria-Olivas et al., Detecting rottenness caused by *Penicillium* genus fungi in citrus fruits using machine learning techniques. *Expert Syst Appl* **39**, 780–785 (2012). <https://doi.org/10.1016/j.eswa.2011.07.073>
24. M. Nagata, J.G. Tallada, T. Kobayashi, Bruise detection using NIR hyperspectral imaging for strawberry (*Fragaria×ananassa* Duch.). *Environ Control Biol* **44**, 133–142 (2006). <https://doi.org/10.2525/ecb.44.133>
25. M.P. Arakeri, Lakshmana, Computer vision based fruit grading system for quality evaluation of tomato in agriculture industry. *Procedia Comput Sci* **79**:426–433. <https://doi.org/10.1016/j.procs.2016.03.055> (2016)
26. H. Calvo, M.A. Moreno-Armendáriz, S. Godoy-Calderón, A practical framework for automatic food products classification using computer vision and inductive characterization. *Neuro-computing* **175**, 911–923 (2016). <https://doi.org/10.1016/j.neucom.2015.06.095>
27. P. Lestringant, J. Delarue, H. Heymann, 2010–2015: How have conventional descriptive analysis methods really been used? a systematic review of publications. *Food Qual Prefer* **71**, 1–7 (2019). <https://doi.org/10.1016/j.foodqual.2018.05.011>
28. W.K. Mutlag, S.K. Ali, Z.M. Aydam, B.H. Taher, Feature extraction methods: a review. *J Phys Conf Ser* **1591**, 12028 (2020). <https://doi.org/10.1088/1742-6596/1591/1/012028>
29. G. Kumar, P. K. Bhatia, A detailed review of feature extraction in image processing systems, in: *2014 Fourth International Conference on Advanced Computing and Communication Technologies* (2014), p. 5–12
30. P. Mohanaiah, P. Sathyanarayana, L. Gurukumar, Image texture feature extraction using GLCM approach. *Int J Sci Res Publ* **3**, 1–5 (2013)
31. R.R. Waliyansyah, U.H.A. Hasbullah, Comparison of tree method, support vector machine, Naïve Bayes, and logistic regression on coffee bean image. *Emit Int J Eng Technol* (2021). <https://doi.org/10.24003/emitter.v9i1.536>
32. D. Lahiri, M. Nag, T. Sarkar et al., Antibiofilm activity of α -amylase from *Bacillus subtilis* and prediction of the optimized conditions for biofilm removal by response surface methodology (RSM) and artificial neural network (ANN). *Appl Biochem Biotechnol* (2021). <https://doi.org/10.1007/s12010-021-03509-9>
33. T. Sarkar, M. Salauddin, S. Hazra, R. Chakraborty, Artificial neural network modelling approach of drying kinetics evolution for hot air oven, microwave, microwave convective and freeze dried pineapple. *SN Appl Sci* **2**, 1621 (2020). <https://doi.org/10.1007/s42452-020-03455-x>
34. C. Tan, F. Sun, T. Kong et al., A survey on deep transfer learning, in *International Conference on Artificial Neural Networks* (Springer, 2018), p. 270–279
35. S.N. Jha, R.D. Rai, S. Gunasekaran, Visual spectroscopy and colour modelling for nondestructive quality evaluation of apple. *J Agric Eng* **47**, 9–13 (2010)
36. S.N. Jha, D.R. Rai, R. Shrama, Physico-chemical quality parameters and overall quality index of apple during storage. *J Food Sci Technol* **49**, 594–600 (2012). <https://doi.org/10.1007/s13197-011-0415-z>
37. S.N. Jha, R. Garg, Non-destructive prediction of quality of intact apple using near infrared spectroscopy. *J Food Sci Technol* **47**, 207–213 (2010). <https://doi.org/10.1007/s13197-010-0033-1>
38. P. Jaiswal, S.N. Jha, P.P. Kaur et al., Prediction of textural attributes using color values of banana (*Musa sapientum*) during ripening. *J Food Sci Technol* **51**, 1179–1184 (2014). <https://doi.org/10.1007/s13197-012-0614-2>
39. A. Mukherjee, K. Chatterjee, T. Sarkar, Entropy-aided assessment of Amla (*Emblca officinalis*) quality using principal component analysis. *Biointerface Res Appl Chem* **12**, 2162–2170 (2022). <https://doi.org/10.33263/BRIAC122.21622170>
40. S.N. Jha, P. Jaiswal, K. Narsaiah et al., Textural properties of mango cultivars during ripening. *J Food Sci Technol* **50**, 1047–1057 (2013). <https://doi.org/10.1007/s13197-011-0431-z>
41. S.N. Jha, T. Matsuoka, Development of freshness index of eggplant. *Appl Eng Agric* **18**, 555 (2002). <https://doi.org/10.13031/2013.10142>
42. D. Lorente, N. Aleixos, J. Gómez-Sanchis et al., Selection of optimal wavelength features for decay detection in citrus fruit using the ROC curve and neural networks. *Food Bioprocess Technol* **6**, 530–541 (2013). <https://doi.org/10.1007/s11947-011-0737-x>

Publisher’s Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.