Food Analytical Methods (2022) 15:917–939
https://doi.org/10.1007/s12161-021-02161-7

Comparative Analysis of Statistical and Supervised Learning Models for Freshness Assessment of Oyster Mushrooms

Tanmay Sarkar[1] · Alok Mukherjee[2] · Kingshuk Chatterjee[2] · Mohammad Ali Shariati[3,4] · Maksim Rebezov[3,4,5] · Svetlana Rodionova[3] · Denis Smirnov[3] · Ruben Dominguez[6] · Jose M. Lorenzo[6,7]

Received: 18 October 2021 / Accepted: 26 October 2021 / Published online: 18 November 2021
© The Author(s), under exclusive licence to Springer Science+Business Media, LLC, part of Springer Nature 2021

Abstract

Automatic assessment of the quality of fruits and vegetables is a growing field of research in this modern era in order to enable faster processing of good quality foods. In this work, we have analyzed ten major colour variant features of two sets of oyster mushrooms in terms of histograms of each layer of the red–green–blue colourmap, hue-saturation-vital component colourmap, luminance-chrominance colourmap and the greyscale image. Besides, texture analysis has been carried out using entropy window filtering. Apart from that, five other minor features, such as mean, standard deviation, entropy, kurtosis and skewness of each of these layers, and four other greyscale features, such as contrast, correlation, energy and homogeneity are analyzed in this work. Two different freshness assessment models employing statistical methods like principal component analysis (PCA) and supervised learning algorithms such as artificial neural network (ANN) have been used here to investigate the different features of the mushroom images and classify the same into fresh and deteriorated classes. Analysis revealed that the ANN classifier outperforms the PCA threshold classifier with almost all the features. The highest classifier accuracy is obtained as 94.4% using the ANN model and 93.3% using the PCA threshold freshness detector. Most importantly, the use of smartphones ensures portability, as well as the possibility of widespread application of the proposed models.

Keywords Oyster mushroom · Major and minor feature · Freshness class · Artificial neural network (ANN) · Food safety

Introduction

Mushrooms gain huge popularity worldwide because of the various advantages associated with like low calories, lower level of sodium, fat (2–3%) and cholesterol, making it a functional food, and it is also rich in protein (27–30%), carbohydrate (45–48%), fibre (11–13%), vitamins and essential amino acids (Dibaba and Abera 2017; Li et al. 2021). It is cultivated on large scale around the world (as it can be cultivated in different climatic conditions), hence easily available (Li et al. 2021). Continuous demand for agricultural products with higher nutritional quality is one of

Tanmay Sarkar and Alok Mukherjee contributed equally.

✉ Tanmay Sarkar
tanmays468@gmail.com

✉ Alok Mukherjee
alokmukherjee.ju@gmail.com

✉ Jose M. Lorenzo
jmlorenzo@ceteca.net

[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

[3] K. G. Razumovsky Moscow State University of Technologies and Management (The First Cossack University), 73, Zemlyanoy Val St., Moscow 109004, Russian Federation

[4] Liaocheng University, 34 Wenhua Road, Liaocheng, Shandong, China

[5] V. M. Gorbatov Federal Research Center for Food Systems of Russian Academy of Sciences, 26 Talalikhina St, Moscow 109316, Russian Federation

[6] Centro Tecnológico de La Carne de Galicia, Parque Tecnológico de Galicia, Rúa Galicia N° 4, San Cibrao das Viñas, 32900 Ourense, Spain

[7] Área de Tecnología de los Alimentos, Facultad de Ciencias de Ourense, Universidade de Vigo, 32004 Ourense, Spain

the challenges to meet the increasing food requirement of the world population. Mushroom may be a solution to meet the future food requirement. In comparison to other vegetables, mushrooms have a short shelf-life period of 3–4 days (Kumar et al. 2021). This is because unlike other vegetables, there is no cuticle in them which could protect them from any physical harm or microbial deterioration or water loss (Aguirre et al. 2009). Pathogenic bacteria and fungus can cause blemishes to the body; discolouration is very common in them due to bruising and storage period (Aguirre et al. 2009). High water activity and high respiration rate make them susceptible to microbial spoilage. Enzymatic browning is also seen in them because of high tyrosinase and phenolic content. As a result of microbial changes and browning reaction, the white colour of fresh mushroom turns brown (Aguirre et al. 2009). A wide range of product variations is the most crucial part of mushroom production. There is a huge variation in product with different stages of maturity, which cannot be maintained easily; hence, to control this, a monitoring system may be useful in this regard.

The consumer-driven vegetable market is facing a continuous increase in requirement for higher value with respect to the mushroom product quality (Li et al. 2021). The cutting-edge researches indicate adverse health issues are related to consumption of spoiled mushroom (Adebo et al. 2021). Thus, reliable scientific methods for on-site mushroom quality evaluation are a need in recent days (Przybylak et al. 2016). The conventional fruit quality evaluation is generally offline and destructive in nature. Therefore, a rapid, non-invasive, non-contact, diverse, eco-friendly and precise system is to be developed for mushroom quality assessment (Suktanarak and Teerachaichayut 2017; Caballero et al. 2017). Digital image analysis has been used to assess the quality of agricultural produce due to its proven features like non-destructive, rapid, cost-effective and accuracy (Ma et al. 2016; Hussain et al. 2018). Though the freshness of food materials is prospective, it regulates the consumer preference most. Freshness depends on the variety of the fruit, and on the experience of the evaluator (Koyama et al. 2021). Objective indices and consumer evaluation has been combined to follow the freshness concept in terms of consumer perception (Koyama et al. 2021). Quantifiable standards like colour, texture and shape perform as a critical indicator during interpretation of sensory evaluation. On contrary, physicochemical characteristics could not reflect the perception of the consumers directly. Freshness prediction with sensory evaluation based on a traditional panel-based system is expensive, time-consuming and skill-dependent whereas machine-learning-based techniques may have higher efficiency. The classification is possibly similar to human sensory assessment without the intervention of a panellist once the model is built with an adequate dataset. The quality of puffed snack (Sanahuja et al. 2018), fish (Dowlati et al. 2013; Liu et al. 2015) and olive oil (Angerosa et al. 1996) has already been evaluated with machine learning. Texture profile, sound-based sensors and image processing have been considered to construct the predictive models. To evaluate the freshness of fruit, visual perception has been considered most important (Péneau et al. 2007; Wada et al. 2007). Machine learning and digital image processing are in practice to facilitate food quality prediction. Smartphone-based digital image processing in combination with machine learning has been used in the freshness classification of squid and fish samples (Navotas et al. 2018; Hu et al. 2020).

In the case of agricultural production, one of the most crucial steps is the sorting of products. It is extremely important to sort the defective products from the superior quality ones. If there is a huge variety in products, then sorting becomes a laborious and tough job. It is seen in fresh agricultural products as various factors like cultivation environment, nutrition, planting mode affect the sorting process (Bhargava and Bansal 2020a). Grading of products during the postharvest period is essential for mushrooms. The huge workload leads to chances of less accurate grading which will ultimately affect the economic aspects in the case of manual grading. Instead of manual grading, the automatic sorting system method offers various advantages like their high speed, non-destructive evaluation and higher production capacity (Wang et al. 2018).

The covariance analysis followed by the associated eigenvalues and eigenvectors be possible with principal component analysis (PCA). Several principal directions are obtained from which the principal components (PC) with a decreasing importance order are achieved. The essential feature to identify a digital image or to reduce the dimensionality of multivariate data, PCA has been used efficiently. Defect is detected in peaches (Sun et al. 2018), apple (Bhargava and Bansal 2021) and tomatoes (Machado et al. 2020).

Biological processes have been modelled reliably and efficiently with the artificial neural network (ANN) for the evaluation of specific objectives (Gago et al. 2010). The arrangement and functioning of the human neural system are simulated in ANN for processing input information followed by decision-making. The non-linear and complex biological events could be modelled with ANN due to its inherent learning mechanism to establish the relationship between input and output features (Gago et al. 2010; Dutta Gupta and Pattanayak 2017). Internal browning of mango (Gabriëls et al. 2020), quality assessment of spray-dried powdered juice of rhubarb (Przybył et al. 2020) and chlorophyll content in apple (Pourdarbani et al. 2020), banana and avocado (Bhargava and Bansal 2020b), gradation of mulberry fruit (Azarmdel et al. 2020), detection of defective lemons (Jahanbakhshi et al. 2020) and mass estimation of *Ziziphus mauritiana* have been evaluated with ANN aided

image analysis. Table 1 shows a comparative analysis of several research works.

Methodology of Sample Preparation

Oyster Mushroom Sample Collection

White oyster mushrooms (*Pleurotus florida*) were purchased from HACCP-certified and FSSAI-registered co-operative society of Kolkata, West Bengal, India. The samples were plucked at the time of procurement. The samples were fresh and white coloured, containing 85–87% of moisture. Emphasis has been given on reducing the time gap between purchases of the mushrooms and bringing them to the laboratory (it has been set within 15 min). Room temp has been maintained at 25 ± 5 °C, with a relative humidity of 80 ± 5%. They were not exposed to direct sunlight and no other preservatives have been used so that natural decomposition can be observed.

Image Acquisition

To study the decomposition rate, pictures were captured from the very first day using Redmi Note 9 Pro (specifications: with 8-nm processor, 4-GB RAM, 128-GB memory, 48–megapixel camera, android system, and Samsung Isocell GM2 sensor) taking view from all angles. Four LED lights were installed in the room (Eveready 20-W LED Batten, 20 W, $113.5 \times 2.5 \times 3.6$ cm, luminous flux of 100 lm/W Lumen, colour temperature of 6000 K) to make sure there is proper lighting facility to capture images. Mushrooms were placed at an average distance of 20 cm from the smartphone to take images.

Evaluation of Quality Classes of the Mushrooms

The sensory qualities of the mushrooms were judged by a group of a semi-trained panel which consisted of 39 females and 30 male persons between the age group 21 and 45 years. ISO 8586–1 (1993) was followed in performing this. On the other hand, ISO 5496:2005 and ISO 10399:2004 were employed respectively to identify the colour and texture to distinguish two different mushrooms (Mukherjee et al. 2021, 2022; Sarkar et al. 2021). Whether the sample falls under good or bad class is determined by three specific sensory attributes namely shape, colour and texture. Figure 1 is a visual representation of the quality classes of the mushrooms with representative samples both from the fresh (Fig. 1(a–e)) and the deteriorated classes (Fig. 1(f–i)).

Preprocessing of Mushroom Images

The mushroom samples are pre-processed minimally with cropping the centre part of it. This allows accommodating the maximum surface of mushroom in the cropped image. We have further resized each image to 300×300 dimensions. This provides uniformity to each image sample for analyzing with the proposed freshness classifiers.

Feature Extraction from Mushroom Images

We have investigated ten major colour features of the mushroom samples and one major texture feature in this work and developed a freshness detection algorithm using two different methodologies employing PCA and ANN independently. Apart from that, we have further analyzed five minor features of each of the colourmap layers and four specific greyscale image features using the ANN classifier only. These are described as follows:

Major Features

Here, we have analyzed three major colourmaps of the mushroom sample images. These include red (R)-green (G)-blue (B) colourmap, hue (H)-saturation (S)-value components (V) colourmap and luminance (Y)-chrominance (Cb and Cr) colourmap. Each of these colour spaces contains 3 layers independently. The RGB layer is the default colourmap of the images, which are obtained just by separating the three independent layers. Each layer is analyzed to develop intensity histograms, with pixel intensity levels varying in the range from 0 to 255. The RGB colourmap is later transformed to HSV and YCbCr colourmap, from which the histograms of each of the H, S and V layers, as well as Y, Cb and Cr layers are extracted. Finally, the images are converted to greyscale form, from which the greyscale histogram is obtained. Each of these histograms is used as the key major features of analysis in the proposed work and, hence, used to develop the freshness assessment algorithms. Each colourmap is discussed in detail in the subsequent sub-sections.

Red (R)-Green (G)-Blue (B) Colourmap Features

The R, G and B layers are the layers of the default RGB colourmap of the images. These are altogether denoted as a colour histogram. Minor changes in the tiny off-axis, observation axis, sometimes rotation of the image, or even small occlusion in the camera lens do not majorly affect these features. Hence, these RGB features are considered some of the most significant and useful features. This is especially important for the classifiers, which use smartphone camera-based image capturing method of the food samples due to the unavoidable difference in the smartphone camera specifications and the captured image quality. The histograms of the RGB colourmap are developed in bunches, using the number of required bins as follows:

Table 1 Performance analysis of different research methods

Mushroom variety	Aim of the research	Technique used	Dataset	Classes	Key findings	Reference
White button mushroom (*Agaricus bisporus*)	Modelling of browning in mushrooms	Greyscale transformation; segmentation followed by mathematical modelling	128 mushroom samples	-	Grey value is maintained up to 4 days at controlled atmospheric condition	Aguirre et al. (2009)
White button mushroom (*Agaricus bisporus*)	Sorting system designing	Watershed method; edge detection by Canny operator	460 observations	Grade 1 (>45 mm), 2 (25–45 mm) and 3 (<25 mm) depending on the cap size	Accuracy = 97.42%	Wang et al. (2018)
Oyster mushroom (*Pleurotus ostreatus*)	Harvest time monitoring system designing	Edge detection	10 samples	Size-dependent: ideal, almost ideal, medium Harvesting based: ready to harvest, not ready to harvest	Accuracy = 80% Precision = 60%	Rahmawati et al. (2018)
White button mushroom (*Agaricus bisporus*)	Optimisation of freeze-drying in terms of colour	Adobe Photoshop; L, a, b measurement	29 samples	-	Drying time prediction for the desired colour	Tarafdar et al. (2020)
Shiitake mushrooms	Grading system development	Binary image transformation; HSB value	24 images	10 classes based on cap opening and size	Size and area of lamella can be measured	Hazisawa et al. (2013)
White button mushroom (*Agaricus bisporus*)	Measurement of mushroom growth	Convolutional Neural Network (CNN)	500 images	-	Automatic growth rate measurement	Lu et al. (2019)
White button mushroom (*Agaricus bisporus*)	Mushroom cap measurement	CNN	500 images		Circle diameter assessment of the cap	Lu and Liaw (2020)
White button mushroom (*Agaricus bisporus*)	Quality evaluation	Preprocessing, feature extraction (area, volume, weight, colour); K-NN, fuzzy logic	250 images	3 classes (defective, healthy and moderate)	Accuracy = 95.6%	Nadim et al. (2019)
Oyster mushroom	Localisation and detection of mushroom in real-time	Single Shot Multi-Box Detector (SSD) algorithm	4600 images	-	F1 value = 0.951	Qian et al. (2020)
45 different mushroom species	Detection of the types of mushrooms	CNN	8,556 images	45 classes based on types of mushrooms	Precision = 0.78 recall = 0.73 F1 value = 0.74	Preechasuk et al. (2019)
-	Edible and poisonous mushroom classification	Preprocessing, PCA, decision tree	8124 images	2 classes (edible and poisonous)	Accuracy = 100%	Ismail et al. (2018)
Button mushroom (*Agaricus bisporus*)	Quality gradation system development	PCA, partial least square-discriminant analysis (PLS-DA)	120 mushrooms	3 classes (damaged, undamaged for 30 s and undamaged for 60 s)	PLS-DA model efficiently detect the undamaged for 30 s	Gowen et al. (2009)
Button mushroom (*Agaricus bisporus*)	Healthy and unhealthy (microbial spoilage) mushroom classification	Preprocessing, feature extraction, support vector machine	150 images	2 classes (healthy and unhealthy)	Accuracy = 90%	Masoudian and Mcisaac (2013)
Button mushroom (*Agaricus bisporus*)	Non-destructive browning monitoring	Segmentation, *K*-means clustering	10 mushroom	-	Enzymatic browning is observed	Anil and Arora (2018)

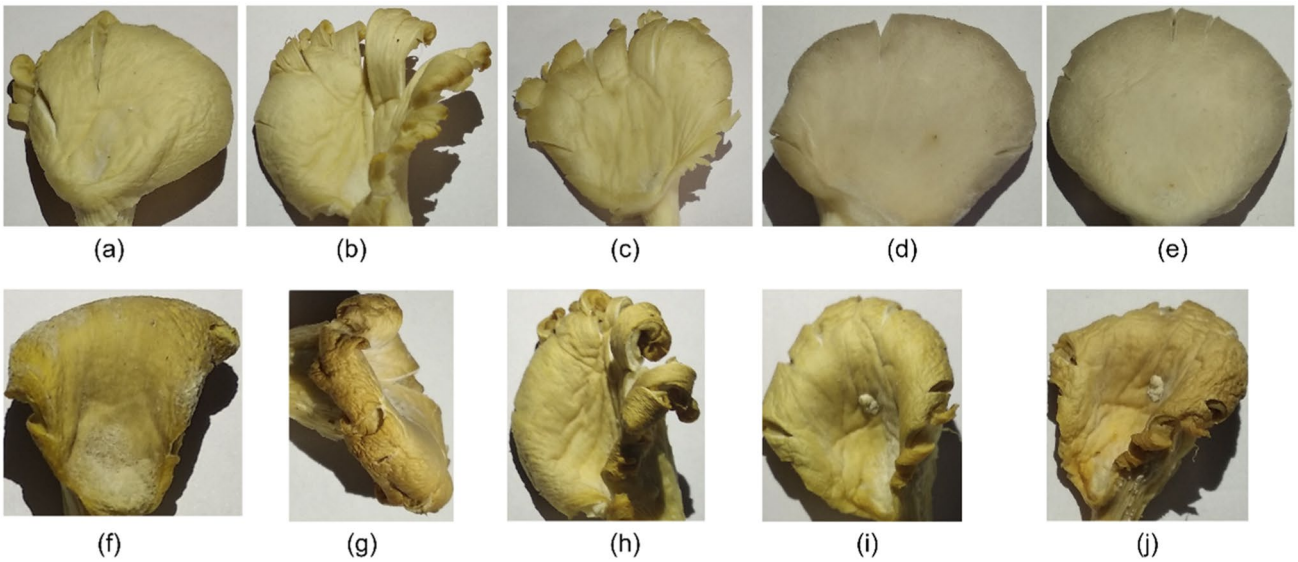

Fig. 1 Raw images of the Oyster mushrooms (*Pleurotus florida*) samples (**a**–**e**) fresh class and (**f**–**j**) deteriorated class

$$\mathrm{Bin}_a(\mathrm{R},\ \mathrm{G},\ \mathrm{B}) = [(f_{\mathrm{R1}}, f_{\mathrm{R2}}, \ldots, f_{\mathrm{Ra}})(f_{\mathrm{G1}}, f_{\mathrm{G2}}, \ldots f_{\mathrm{Ga}})(f_{\mathrm{B1}}, f_{\mathrm{B2}}, \ldots f_{\mathrm{Ba}})]$$

Here, the range of intensity values are segmented into 'a' number bunches. Further decomposition is done to segregate the R, G and B layers. The histogram is thus developed as:

$$\text{Histogram}\ (H_i) = \frac{\text{Number of pixels in} H_i}{\text{Height} \times \text{Width}} \tag{1}$$

where i is one of the layers amongst R, G or B.

Hue-Saturation-Value Component Colourmap Features

The most important use of the HSV colourmap is that it separates '*luma*' or the light intensity of the image from '*chroma*', which is the colour information of the same. That helps in histogram equalization. Hence, these features are extremely insensitive against variation in lighting conditions, thus, helping in developing robust classifiers, especially considering all weather conditions where the natural light varies, or the variation in artificial lighting or flashlight exists. Most importantly, these features help immensely in developing smartphone-based image processing algorithms, disregarding the ambient lightings or the inbuilt flash. These features are represented as follows:

$$H = \left\{ \begin{array}{c} \theta B \le G \\ 360^\circ - \theta B \ge G \end{array} \right\} \tag{2}$$

$$\text{where} \theta = \cos^{-1}\left[\sqrt{\frac{(R-G)+(R-B)}{(R-G)^2 + (R-B)(G-B)^{1/2}}}\right] \tag{3}$$

$$S = 1 - \frac{3}{R+G+B} \tag{4}$$

where R, G and B represent the red (R), green (G) and blue (B) layers.

Luminance-Chrominance Colourmap Features

The luminance can be separated effectively from the chrominance with the help of the YCbCr colour channel. The extent of light in the entire spectrum of the image is described by luminance. The colour information of a digital image in terms of cyan blue and cyan red is conveyed through Cb and Cr respectively (Binti Zaidi et al. 2015).

$$\begin{bmatrix} Y \\ Cb \\ Cr \end{bmatrix} = \begin{bmatrix} R \\ G \\ B \end{bmatrix} \begin{bmatrix} 0.257 & 0.504 & 0.098 \\ -0.148 & -0.291 & 0.439 \\ 0.439 & -0.3678 & -0.0714 \end{bmatrix} + \begin{bmatrix} 16 \\ 128 \\ 128 \end{bmatrix} \tag{5}$$

Minor Features

Apart from these major features, five other minor or secondary features of each layer of the RBG, HSV and YCbCr colourmap and the greyscale image are analyzed using the proposed ANN freshness detector. These features are mean, standard deviation (SD), entropy, kurtosis and skewness. Apart from these, four other minor features, namely,

contrast, correlation, energy and homogeneity are also extracted using the grey-level co-occurrence matrix (GLCM) of the greyscale image. A detailed discussion of some of these features is done in the following sections.

Mean

The mean of each layer is computed by computing the average or mean of matrix elements of the corresponding layer. Here, mean is computed for each layer image of the RGB, HSV and YCbCr colourmap, which are individual images, hence, matrices.

$$\text{Mean} = \text{average of (matrix elements of the image layer)} = \frac{1}{N^2}\sum_{k=0}^{N^2-1} P_k \quad (6)$$

where P_k is the pixel intensity of the $N \times N$ image corresponding to all layers independently.

Standard Deviation

The standard deviation (SD) of each layer is computed similarly by computing the standard deviation of matrix elements of the corresponding layer of the RGB, HSV and YCbCr colourmap. The generalized form of standard deviation is found as:

$$\text{SD} = \left(\frac{1}{N^2}\sum_{k=0}^{N^2-1}(P_k - \overline{P})^2\right)^{\frac{1}{2}} \quad (7)$$

where P_k is the pixel intensity of the $N \times N$ image corresponding to all layers independently and $\overline{P}$ is the mean of the same.

Entropy

The entropy of an image is the representation of the statistical measure of randomness of the pixel intensity values. Thus, it characterizes the texture of an image very efficiently. The entropy of a layer of the image, which essentially becomes a grey-level image, is measured by:

$$\text{Entropy (En)} = \sum_{x=0}^{GL-1} (\text{Probability}_x)\log_B(\text{Probability}_x) \quad (8)$$

where GL is the grey-level number and Probability_x is the probability of a pixel to have a GL value of x, and B is the base of the logarithm. The Probability_x truly contains the histogram counts of each image.

Kurtosis

Kurtosis is a statistical parameter that measures the difference in the distribution of tails of any distribution from that of a normal distribution, for which the value of kurtosis is 3. In the distributions, where there are more outliers compared to the normal distribution, the yield kurtosis exceeds 3, and for the distributions where the number of outliers is less, it has kurtosis value less than 3. It is computed as below.

$$\text{Kurtosis } (K) = \frac{1}{\text{SD}^4}E(P_k - \overline{P})^4 \quad (9)$$

where $\overline{P}$ is the mean value of the pixel intensity variable P, SD is the standard deviation of P, and $E(f(n))$ denotes the expected value of the function $f(n)$.

Skewness

Skewness is a parameter that measures the asymmetry of the probability distribution; also, it measures the extent by which any pixel is lighter or darker compared to the average level, i.e. the sample mean. A negative value of skewness denotes the distribution is shifted more towards the left of the mean level, and a positive value of skewness indicates the distribution is spread out or skewed more towards the right. mean level. Thus, it is obvious that the skewness of a perfectly symmetric distribution, which may be regarded as a normal distribution, is zero. The skewness is represented using the following equation:

$$\text{Skewness } (S) = \frac{1}{\text{SD}^3}E(P_k - \overline{P})^3 \quad (10)$$

where $\overline{P}$ is the mean value of the pixel intensity variable P, SD is the standard deviation of P, and $E(f(n))$ denotes the expected value of the function $f(n)$.

Grey-Level Co-Occurrence Matrix Features

Grey-level co-occurrence matrix or GLCM is a statistical method that examines the texture of an image by considering the spatial relationship of the pixels. Hence, GLCM is also known as the grey-level spatial dependence matrix. The GLCM functions estimate the frequency at which pairs of the pixel with specific values and in a specified spatial relationship occur by inspecting the image; thereby, it creates a GLCM. Finally, statistical measures like contrast, correlation, energy and homogeneity are computed from this matrix. Thus, GLCM is essentially the computational characteristics of the probability density function or PDF of the grey-level matrix of any particular image. Thus, GLCM is given by:

$$\text{GLCM}(p_1, p_2); p_1, p_2 = 0,1,2,3,\ldots \text{P} - 1 \quad (11)$$

where P is the numbers of pixels in the image and p_1 and p_2 are the frequency value of pixel pair. The four GLCM features, as mentioned here, are computed as described in the subsequent sub-sections.

Contrast

Contrast essentially measures the local variations in the GLCM and it is estimated by the following expression:

$$\text{Contrast } (C) = \sum_{p_{1=0}}^{P-1} \sum_{p_{2=0}}^{P-1} (\text{GLCM}(p_1 p_2))^2 \tag{12}$$

Correlation

Correlation is a measure of the joint probability of occurrence of any particular pixel pair. This is generally used to measure the deformation, displacement, optical flow and strain in an image. Correlation is estimated using the following expression.

$$\text{Correlation (Corr)} = \sum_{p_{1=0}}^{P-1} \sum_{p_{2=0}}^{P-1} \frac{p_1 p_2 GLCM(p_1 p_2) - (\overline{p}_1 \overline{p}_2)}{SD_1^2 \times SD_2^2} \tag{13}$$

where $\overline{p}_1$ and $\overline{p}_2$ are the means and $SD_1^2$and SD_2^2 are the standard deviation for p_1 and p_2 respectively.

Energy

Energy is estimated by computing the sum of squared elements in the GLCM. Energy also estimates the uniformity in an image or the angular second moment. The energy of an image is computed as follows:

$$\text{Energy } (E) = \sum_{p_{1=0}}^{P-1} p_1 \sum_{p_{2=0}}^{P-1} (p_1 - p_2)^2 \times \text{GLCM } (p_1 p_2) \tag{14}$$

Homogeneity

Homogeneity (Ho) is a parameter that measures the closeness of the distribution of elements in the GLCM to the GLCM diagonal (transverse GLCM). Homogeneity is represented by the following expression.

$$\text{Homogeneity}(Ho) = \sum_{p_{1=0}}^{P-1} p_1 \sum_{p_{2=0}}^{P-1} \frac{\text{GLCM}(p_1 p_2)}{1 + \text{mod}(p_1 - p_2)} \tag{15}$$

Analysis Methods

In this work, we have developed two classifier models employing multivariate statistical methods like principal component analysis (PCA) and supervised learning model like artificial neural network (ANN) to develop the proposed freshness detection models using the oyster mushroom images. These methods are discussed in detail in the following sub-sections.

Application of Principal Component Analysis

Principal component analysis (PCA) is a statistical model which emphasizes the variance of a data set. The computation of the principal components is done in the following way (Jolliffe 2002). For a random variable x with y number of elements, PCA tries to develop a linear function $C_1 x$ which would have the highest variation, expressed as:

$$C_1^T x = C_{11} x_1 + C_{12} x_2 + \cdots + C_{1y} x_y = \sum_{j=1}^{y} (C_{1j} x_j) \tag{16}$$

Similarly, another linear function $C^T{}_2 x$ is developed in such a way that it becomes uncorrelated with $C^T{}_1 x$ and it has a maximum variance. Similar uncorrelated linear functions such as $C^T{}_1 x$, $C^T{}_2 x$, …, and $C^T{}_\text{m}$ are continued until mth stage, and thereby obtain the mth principal component (PC). Practically, only a few PCs, up to nth PC, say, are considered for analysis, since these PCs carry major information in the decreasing sequence. Thus, the first PC is the most significant carrier of information, followed by the second PC, third PC etc. Thus, n is much less compared to m. This reduces the dimension of the data set largely. Most often, the unknown $\sum$ is substituted by the covariance matrix, denoted by Cov, and it is given by:

$$\text{Cov} = \frac{1}{y-1} X^T X \tag{17}$$

Thus, the (j, k)th element of the same covariance matrix is given by:

$$\frac{1}{y-1} \sum_{i=1}^{y} (\tilde{x}_{ij} - \overline{x}_j)(\tilde{x}_{ik} - \overline{x}_k) \tag{18}$$

Now, if we consider $C_1^T x_i$ as $\tilde{z}_{i1}$, then the coefficients C_1^T is chosen in such a way that it should maximize the variance given by the following expression.

$$\text{variance} = \frac{1}{n-1} \sum_{i=1}^{n} (\tilde{z}_{i1} - \overline{z}_1)^2 \tag{19}$$

It is obvious that $C_1^T C_1 = 1$ and $C_1^T x$, i.e. $\tilde{z}_{i1}$, is the first PC. Thus kth largest eigenvalue of the covariance matrix (Cov) is corresponding to the kth eigenvector C_k corresponding to the kth largest eigenvalues λ_k, and the kth PC is found as z_k which is the same as $C_k^T x$. Thus, $\tilde{X}$ and $\tilde{Z}$ are related with the expression as $\tilde{Z} = \tilde{X} O$, where O is an orthogonal matrix. The kth column of this orthogonal matrix O is indicated by C_k.

As mentioned earlier that PCs are obtained by maximizing the variance, i.e. maximizing $z_k = C_k^T x = C_k^T \sum C_k$, governed by the constraint $C_k^T C_k = 1$. Lagrange multipliers are

used here as a standard method for maximizing the variance, i.e. by maximizing the function given by:

$$\left(C_k^T \sum C_k\right) - \lambda(C_k^T C_k - 1) \tag{20}$$

where λ is a Lagrange multiplier. On differentiation with respect to C_k, it yields:

$$\sum C_k - \lambda C_k = 0, \text{i.e.} \left(\sum -\lambda I\right) C = 0 \tag{21}$$

where I is an identity matrix. Using the above expression, we get independent sets of eigenvalues λ and the corresponding eigenvector C respectively. The highest eigenvalue λ_1 and the corresponding eigenvector C_1 are used to obtain the highest PC, which is given by $C_1^T x$. The subsequent PCs given by $C_k^T x$ are also obtained subsequently in a similar way. However, in this work, we have made use of only the first PC to develop the proposed PCA threshold–based freshness assessment algorithm.

Application of Artificial Neural Network

We have designed a three-layered artificial neural network (ANN) architecture for developing the freshness classifier model. The three layers correspond to the input layer, hidden layer and output layer, as is found in common ANN architectures (Ali and Dildar 2020). Histograms of each of the RGB, HSV, YCbCr, greyscale colourmap and the entropy-filtered image are given as input to the input layer of the proposed ANN architecture independently. The sigmoid activating function is employed in the proposed model, which is given as (Sarkar et al. 2020):

$$y_i = \frac{1}{1 + exp^{-(\sum \text{input} + \phi_i)}} \tag{22}$$

Here, *input* is given by $w_{ki}\theta_k$ and y_i is the response of the ith node; ϕ_i is the bias at the same node; θ_k is the kth input to the input layer; and w_{ki} is the weight of the path corresponding to the kth input. Thus, the input layer response at the ith node of the hidden layer is given by modifying expression (19) as:

$$y_i = \frac{1}{1 + \exp^{-(\sum_{k=1}^{n} w_{ki}\theta_i + \phi_i)}} \tag{23}$$

Similarly, the hidden layer output from the jth node (O_j) is obtained as:

$$O_j = \sum_{i=1}^{y} w_{ij} + \phi_j$$

where w_{ij} represent the associated weight and ϕ_j is the predisposition of the jth node of the output (Tan et al. 2018).

Sample Segmentation

We have considered two sets of independent experimental data of the mushroom samples in this work. As mentioned earlier, it is observed that the mushroom samples remain *Fresh* during the first 3 days, and from the fourth day onwards, the samples deteriorate in quality, and the sample quality falls below level 6 on the Hedonic scale. Hence, the sample images from day 4 onwards are considered to be lying in the *Deteriorate* class of samples. In the first set of mushroom samples, we have accumulated 55 *Fresh* images, captured over the first 3 days, and a similar number for *Deteriorate* samples over the next 3 days has been 65. Hence, the set contains 55 and 65 sample images of the *Fresh* and *Deteriorate* classes respectively. Another set of samples has been examined with similar quality of mushrooms and finally, 65 and 55 numbers of sample images have been captured during the first 3 days and the following 3 days respectively. Thus, the number of samples belonging to the *Fresh* and *Deteriorate* class becomes 55 and 45 respectively. Thus, the total number of samples of these two classes finally become (55 + 65) i.e. 120, and (65 + 55) i.e. 120 respectively.

Samples Analyzed Using the Proposed Schemes

The first set of 55–65 samples has been used to develop the PCA threshold–based freshness detection algorithm. The second set of 65–55 similar mushroom samples has been used for testing the proposed scheme. Thus, this second set is used here to validate the proposed PCA threshold–based classifier.

But the ANN classifier has been analyzed in a minor different scheme. The whole sample set of 120–120 mushroom images are used to train, validate and test the proposed ANN freshness detection model with a 70:15:15 percentage ratio respectively; i.e. 70% or 84–84 number of samples are used to develop the training set; 18–18 number of samples are used to validate the model; another 18–18 number of samples are used for testing the same. Training has been carried out 10 times with each feature to obtain the best possible outcome, which is considered the final outcome of the model.

Result and Discussion

Analysis of Features

Two different classification models have been designed in this work using statistical analysis and supervised learning algorithms respectively. These models are analyzed separately using the 11 different major features and 54 other minor features of the mushroom samples. These major

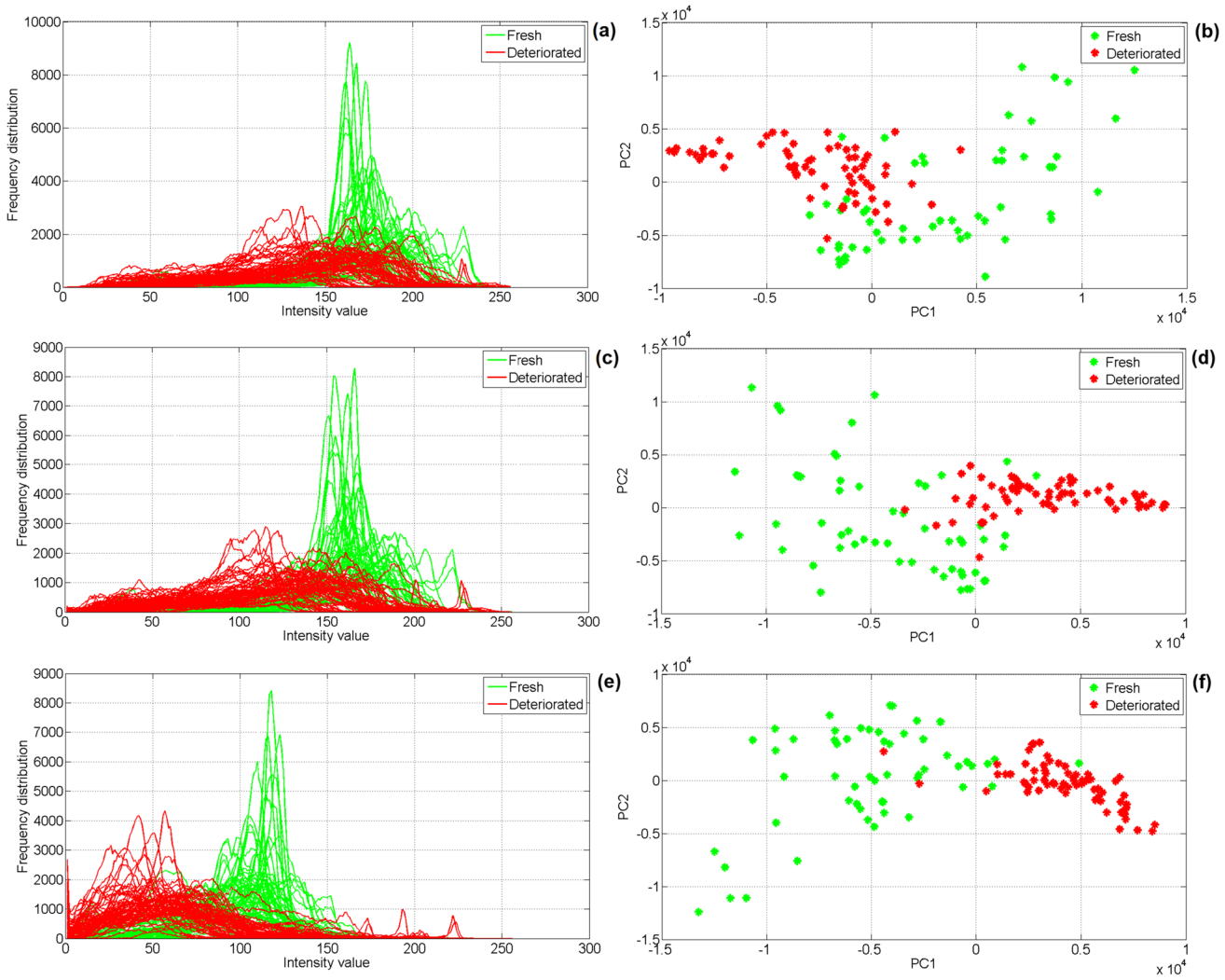

Fig. 2 Histogram and principal component scatter diagram of the RGB layers: **a** histogram of the red layer, **b** PC scatter plot of red layer histogram, **c** histogram of green layer, **d** PC scatter plot of green layer histogram, **e** histogram of blue layer, **f** PC scatter plot of blue layer histogram

features include the histograms of the three independent layers of the red (R)-green (G)-blue (B) colourmap, hue (H)-saturation (S)-vital component (V) colourmap and luminance (Y)-chrominance (Cb and Cr) colourmap, grey-scale image and texture map in the form of histograms of the entropy-filtered image. The minor features include mean, standard deviation (SD), entropy, kurtosis and skewness of each of the three layers of RGB, HSV and YCbCr colour space and the whole image itself, and the four vital grey-scale image parameters like contrast, correlation, energy and homogeneity, obtained from the grey-level co-occurrence matrix (GLCM). These major features are analyzed using the PCA threshold model, and both the major and the minor features are analyzed using the proposed ANN-based freshness level classifier.

Three features from every three colourmaps are extracted from the image samples. The histograms of the red (R), green (G) and blue (B) layers are shown in Fig. 2 along with the scatter plot of the first and the second principal components (PC1 and PC2 respectively). Similar histograms and the PC scatter plot of the hue (H), saturation (S) and vital component (V) layers of the image samples are also shown in Fig. 3. Figure 4 shows the same for the luminance (Y) and chrominance colour values (Cb and Cr layers respectively). The histogram of the greyscale image (Gr) and entropy-filtered image (En), along with the corresponding principal component scatter diagram, is shown in Figs. 5 and Fig. 6 respectively. These histograms, as well as scatter plots, correspond to the first set of sample images which are used for developing the threshold-based PCA freshness detector.

It was mentioned before that the samples begin to deteriorate from day 4 onwards, when the Hedonic scale

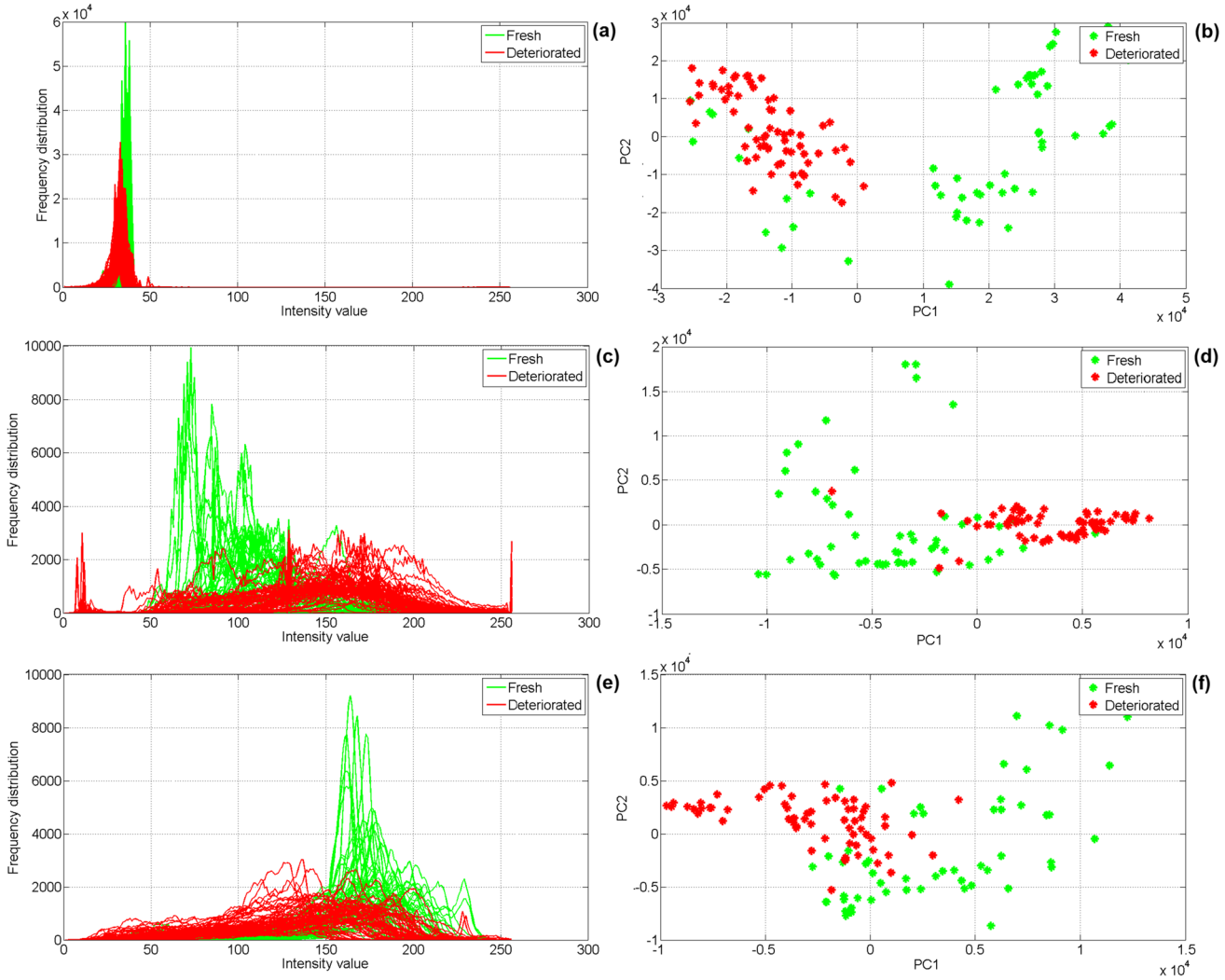

Fig. 3 Histogram and principal component scatter diagram of the HSV layers: **a** histogram of hue layer, **b** PC scatter plot of hue layer histogram, **c** histogram of saturation layer, **d** PC scatter plot of saturation layer histogram, **e** histogram of vital component layer, **f** PC scatter plot of vital component layer histogram

suggests deterioration of fewer than 6 units. Hence, we have considered samples from days 1 to day 3 as *Fresh*, and those of days 4 to 6 as a *Deteriorated* class. It is well observed from Figs. 2, 3, 4, 5 and 6 that most of the scatter plots show major overlap between the two freshness classes: *Fresh* (day 1 to day 3) and *Deteriorated* (day 4 to day 6) classes. But, in the case with few features, the scatter plots show significant separation between the two classes, and hence, the possibility of developing a distinct threshold level using the PC1 value only. In all the cases, it is observed that PC1 values are significant enough to develop classification. Hence, only the first principal component has been used here for developing the PCA threshold–based classifier scheme. This further ensures reduced computation using only one PC direction.

Classifier Outcomes Using Major Features

Careful observation of Figs. 2, 3, 4, 5 and 6 shows that green, blue, hue, saturation, luminance and entropy image histograms and PC plots are better for developing a threshold level between the *Fresh* and *Deteriorated* classes, compared to the other major features. These PCA features are hence segmented using designed thresholds to test the same model with the second set of image samples. The test results of the PCA classifier are shown in Table 2 using the major features such as histograms of colour and greyscale colourmap, as well as the texture features in terms of the histogram of the entropy image.

As mentioned earlier, the first set contains 55 numbers of the *Fresh* class image and 65 numbers of the *Deteriorated* class of images. A similar count for the second set of images

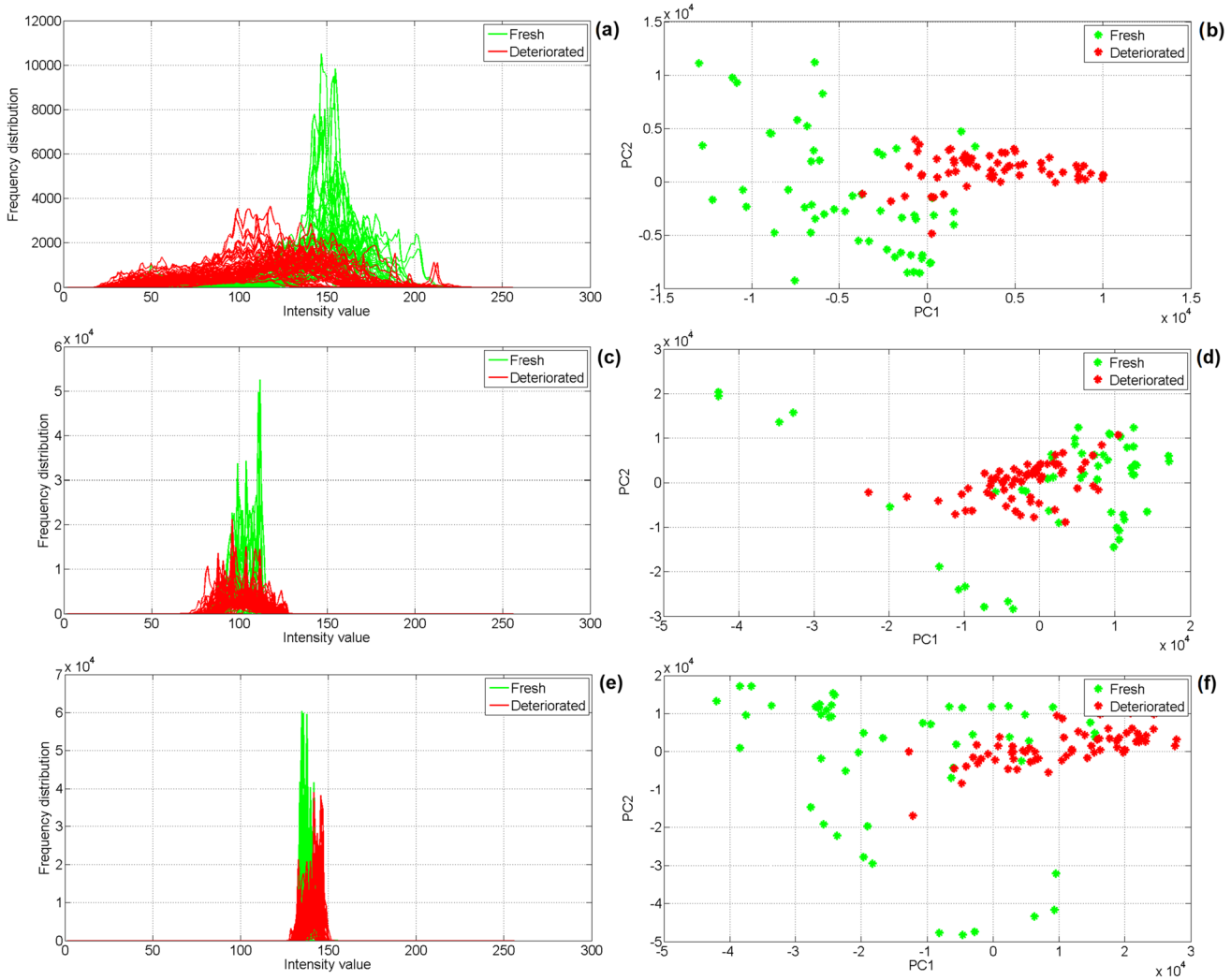

Fig. 4 Histogram and principal component scatter diagram of the YCbCr layers: **a** histogram of Luminance (Y) layer, **b** PC scatter plot of luminance (Y) layer histogram, **c** histogram of chrominance 1 (Cb) layer, **d** PC scatter plot of chrominance 1 (Cb) layer histogram, **e** histogram of chrominance 2 (Cr) layer, **f** PC scatter plot of chrominance 2 (Cr) layer histogram

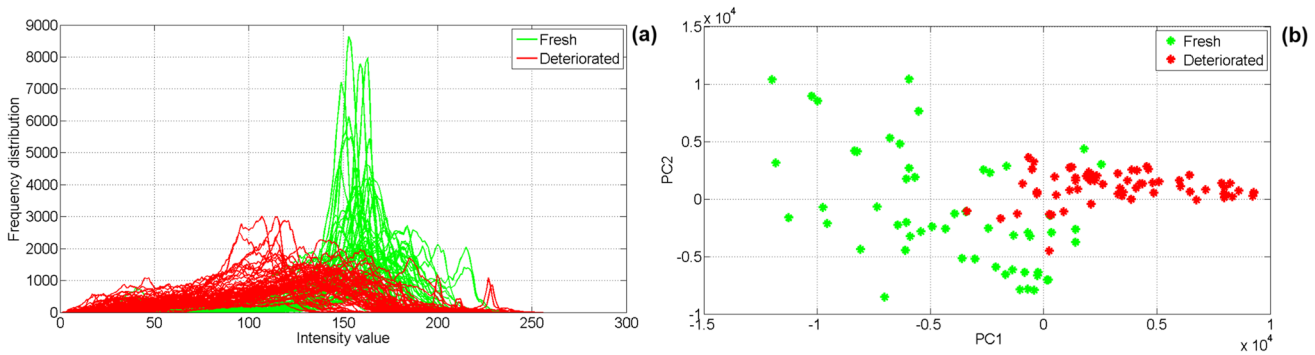

Fig. 5 Histogram and principal component scatter diagram of the greyscale layers: **a** histogram of greyscale layer, **b** PC scatter plot of greyscale layer histogram

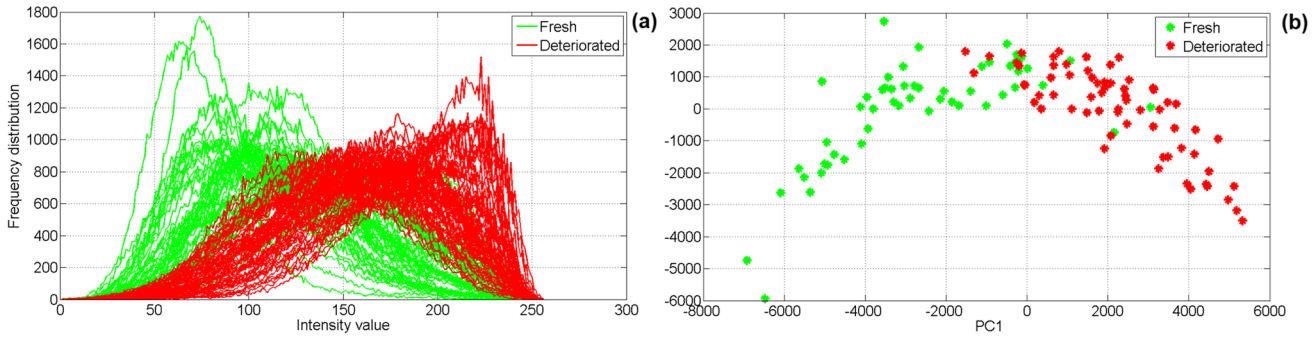

Fig. 6 Histogram and principal component scatter diagram of the entropy-filtered image: **a** histogram of entropy-filtered image, **b** PC scatter plot of entropy-filtered image histogram

Table 2 Comparative analysis of the outcomes of freshness class detection with major colour features using PCA threshold classifier and ANN classifier

Colour space	Feature	Classification accuracy (%) using PCA model	Classification accuracy (%) using ANN model
RGB	Red (R)	79.167	83.333
	Green (G)	86.667	91.667
	Blue (B)	93.333	94.444
HSV	Hue (H)	87.5	91.667
	Saturation (S)	86.667	88.889
	Vital component (V)	77.5	86.111
YCbCr	Luminance (Y)	85.833	94.444
	Chrominance 1 (Cb)	76.667	86.111
	Chrominance 2 (Cr)	79.167	91.667
Greyscale	Grey (Gr)	84.167	86.111

Table 3 Outcomes of freshness class detection with texture feature using ANN classifier

Feature	Classification accuracy (%) using PCA model	Classification accuracy (%) using ANN model
Entropy-filtered image histogram	87.5	88.889

is 65 and 55 respectively. This altogether develops 120 numbers of *Fresh* and *Deteriorated* sample images. This full set of mushroom samples are used to train, validate and test the designed ANN classifier model in 70:15:15 ratio; i.e. 70% of the entire dataset are used to train the model, 15% of the samples are used to validate the same and the remaining 15% are used for testing the classifier. The same method is repeated using a random set of training-validation-test image samples to obtain the optimal level of accuracy, using each of the 11 major features, i.e. 10 different colour features and 1 texture feature. The test results obtained using the same is shown in Table 3, along with the results of the PCA threshold freshness model.

Analysis Using Minor Features

Five minor features of each layer of RGB, HSV and YCbCr colourmap are analyzed in this work apart from the 11 major features discussed in the earlier sub-section. These features are mean, standard deviation (SD), entropy, kurtosis and skewness of the layer pixel values. Besides, four other GLCM features, namely, contrast, correlation, energy and homogeneity, are also analyzed using the greyscale sample images. We have used boxplots to identify the significance and distinctness of separation of these minor features, in order to develop distinct classification between the *Fresh* and *Deteriorated* classes. These boxplots are shown for separate layers of each of the RGB, HSV and YCbCr colourmap. The boxplots of these minor features for the red, green and blue layers are shown in Fig. 7a, b and c respectively. Similar individual layer boxplots of the HSV and YCbCr colourmap are shown in Figs. 8 and 9 respectively. Figures 10 and 11 shows the five minor features of the original RGB image and the four GLCM features of the greyscale image respectively.

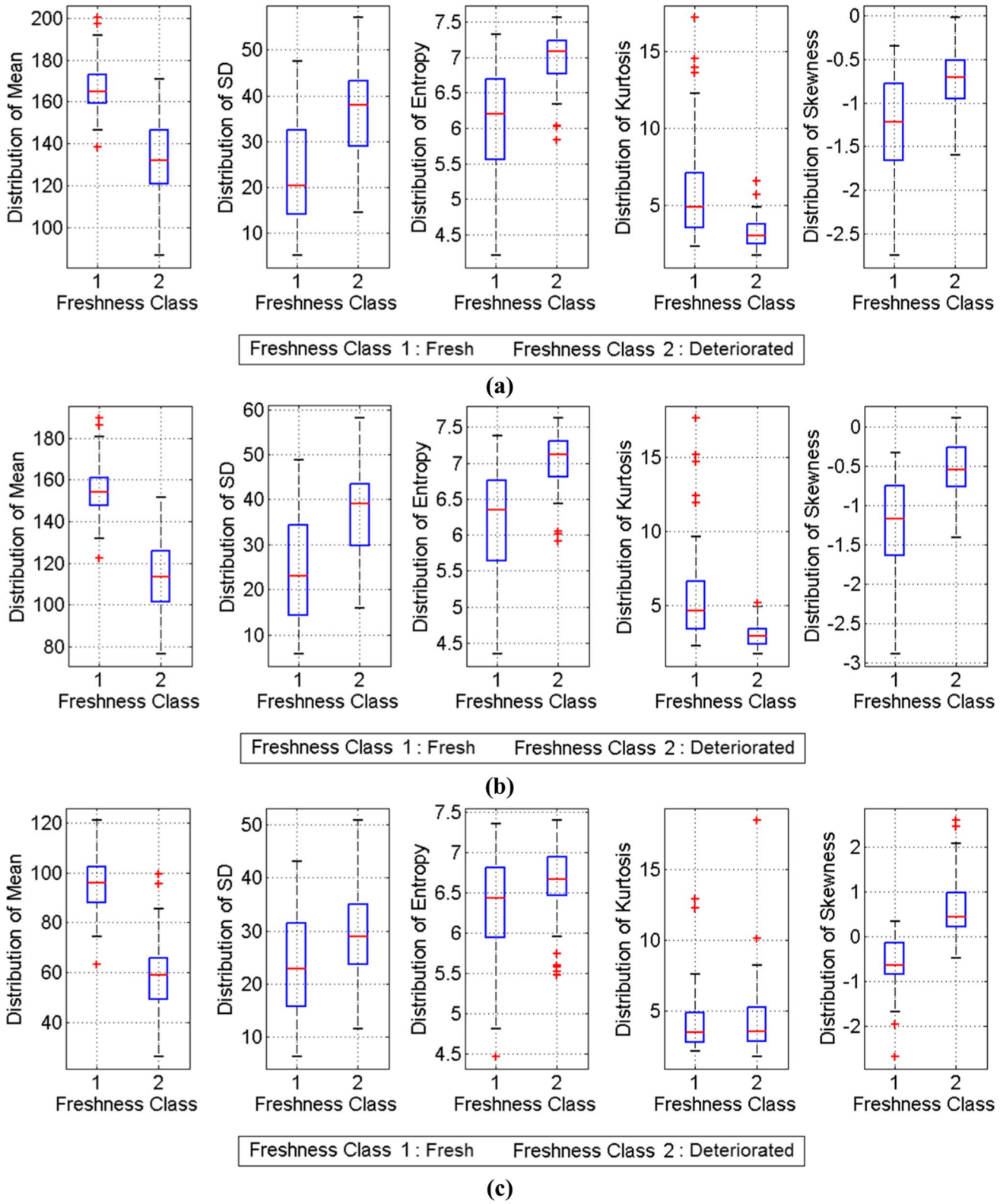

Fig. 7 Distribution of minor features such as mean, standard deviation (SD), entropy, kurtosis and skewness of the **a** red layer, **b** green layer and **c** blue layer of the RGB colourmap for *Fresh* and *Deteriorate* classes of the mushroom samples

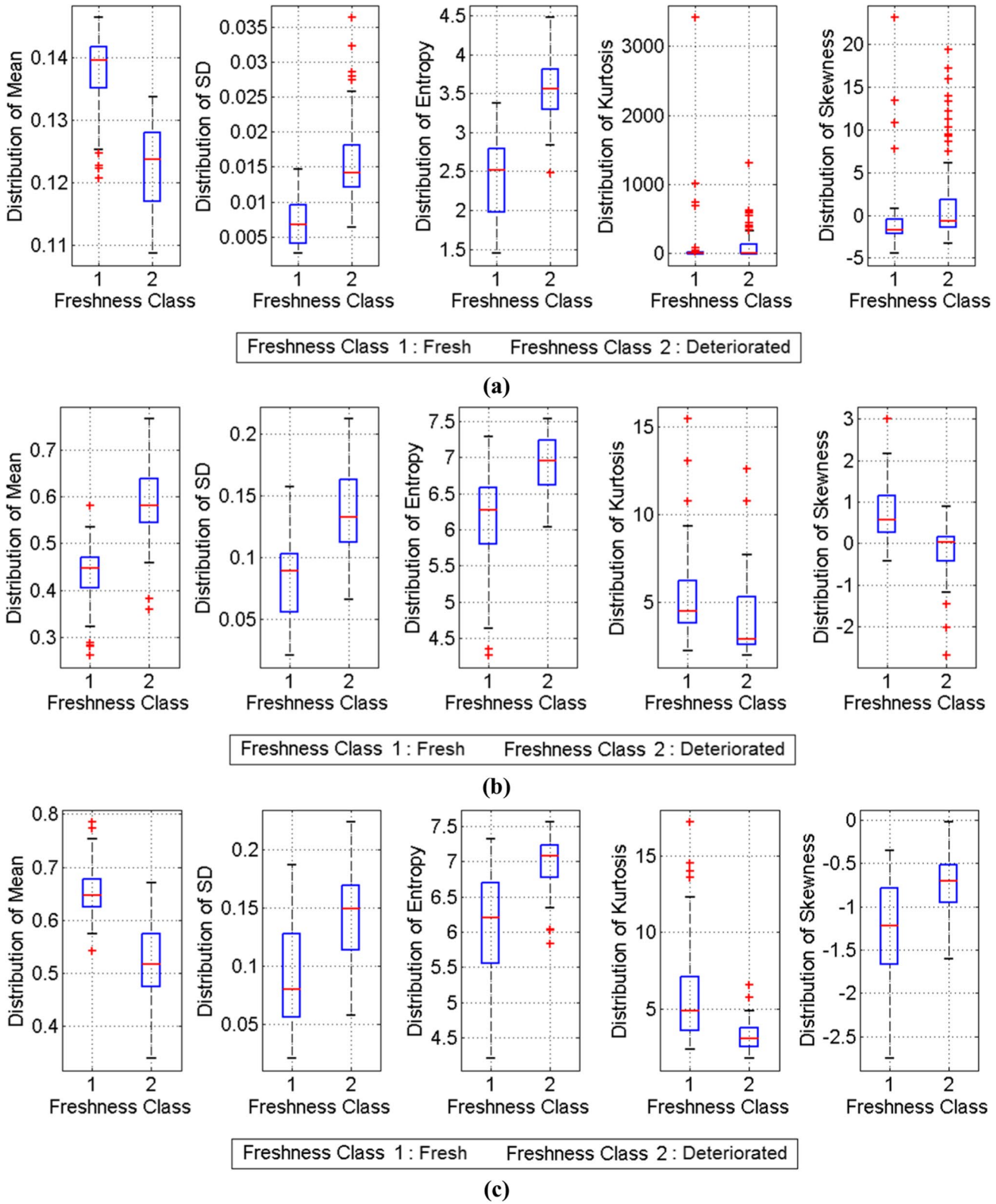

Fig. 8 Distribution of minor features such as mean, standard deviation (SD), entropy, kurtosis and skewness of the **a** hue layer, **b** saturation layer and **c** vital component layer of the HSV colourmap for *Fresh* and *Deteriorate* classes of the mushroom samples

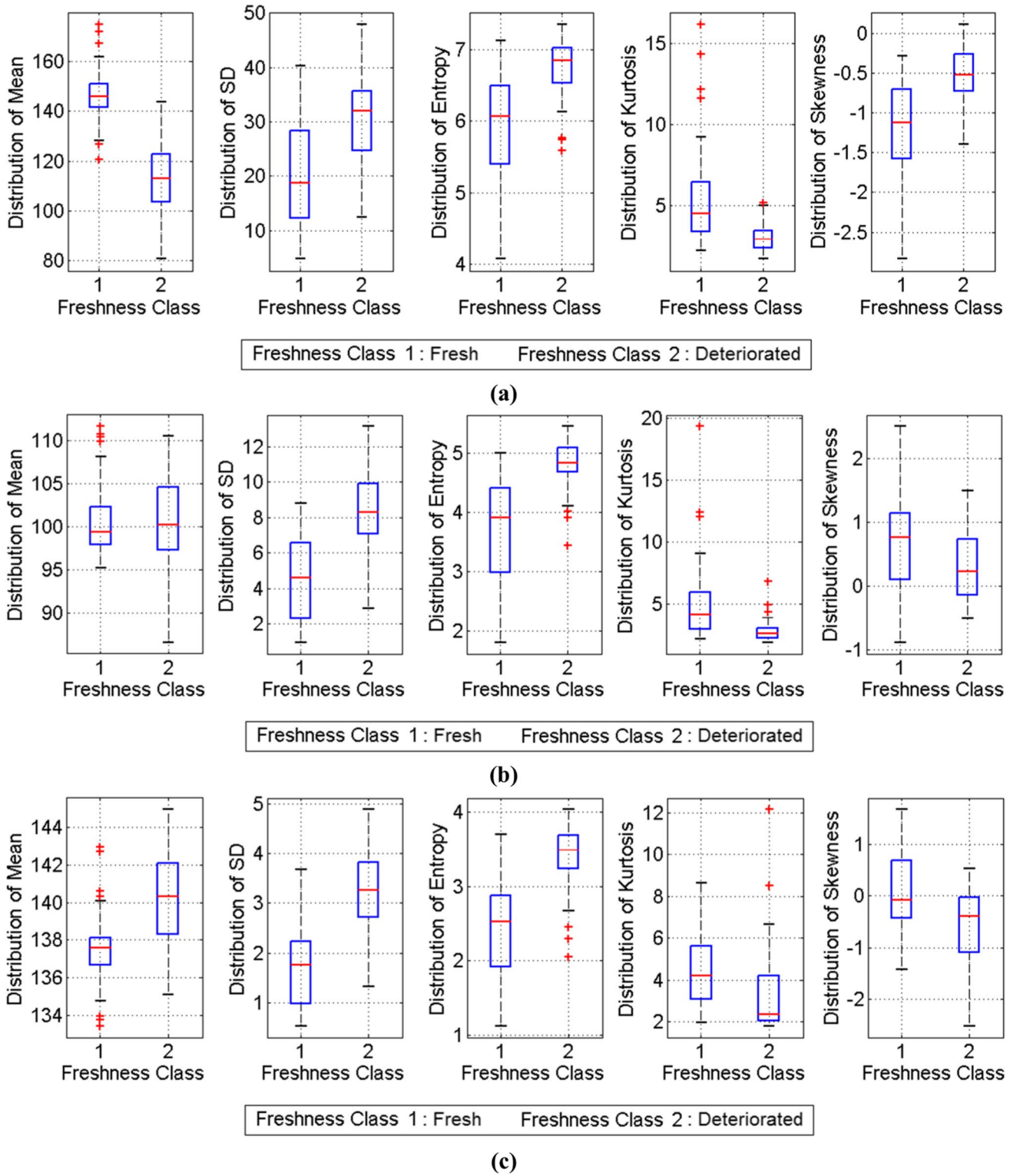

Fig. 9 Distribution of minor features such as mean, standard deviation (SD), entropy, kurtosis and skewness of the **a** luminance layer, **b** chrominance 1 layer and **c** chrominance 2 layer of the YCbCr colourmap for *Fresh* and *Deteriorate* classes of the mushroom samples

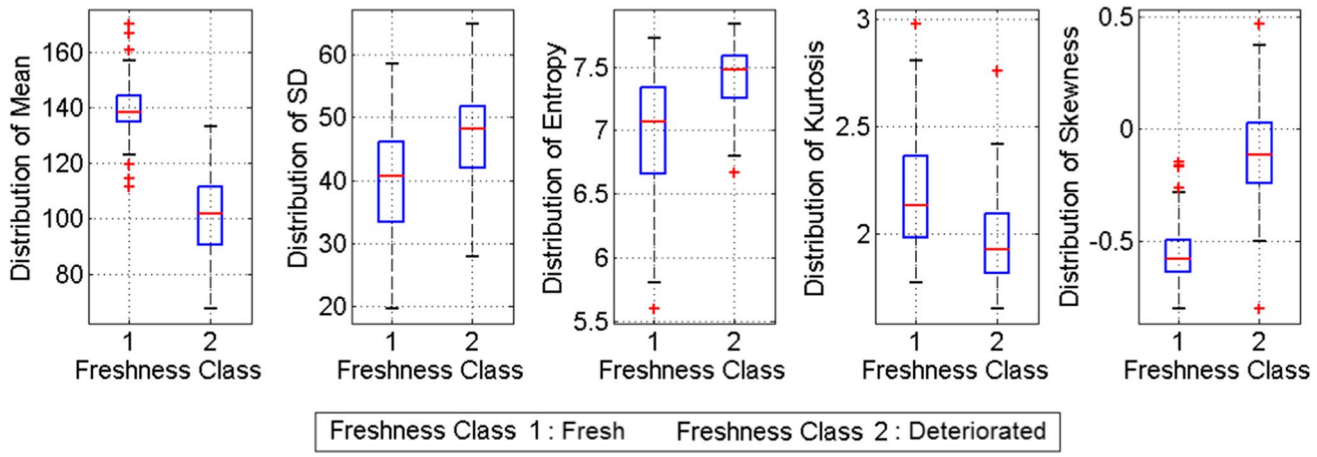

Fig. 10 Distribution of minor features such as mean, standard deviation (SD), entropy, kurtosis and skewness of the original RGB image for *Fresh* and *Deteriorate* classes of the mushroom samples

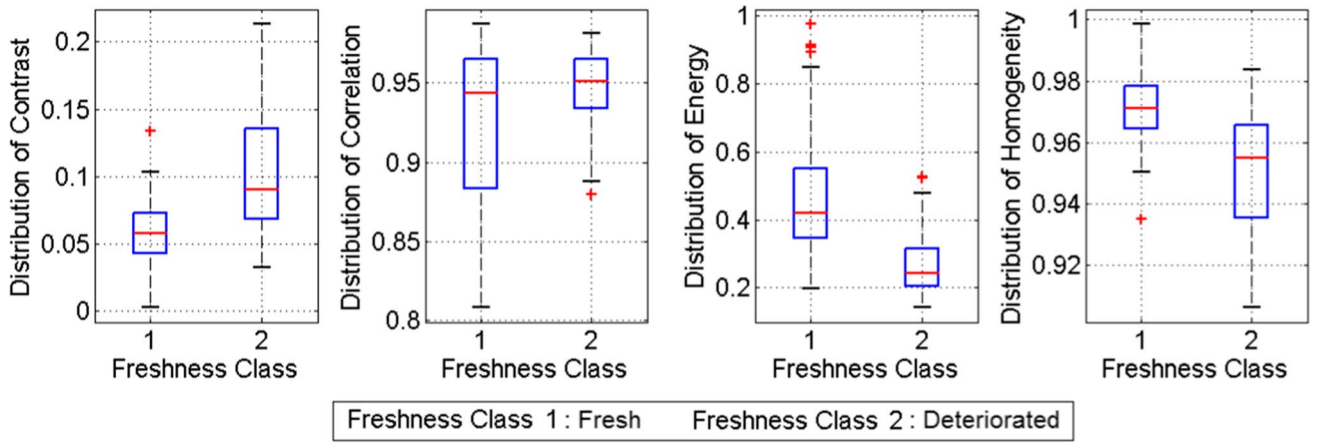

Fig. 11 Distribution of GLCM features such as contrast, correlation, energy and homogeneity of the greyscale colourmap for *Fresh* and *Deteriorate* classes of the mushroom samples

Table 4 Outcomes of freshness class detection with minor features of the RGB, HSV and YCbCr colourmap using the proposed ANN classifier

Colourmap	Layer	Accuracy of classification (%) using features such as				
		Mean	SD	Entropy	Kurtosis	Skewness
RGB	Red (R)	88.889	77.778	72.222	77.778	80.556
	Green (G)	94.444	77.778	80.556	83.333	75
	Blue (B)	94.444	61.111	63.889	44.444	91.667
HSV	Hue (H)	91.667	83.333	88.889	47.222	63.889
	Saturation (S)	83.333	75	66.667	61.111	80.556
	Vital component (V)	91.667	86.111	72.222	77.778	63.889
YCbCr	Luminance (Y)	91.667	72.223	77.778	77.778	77.778
	Chrominance 1 (Cb)	58.333	83.333	86.111	72.222	55.556
	Chrominance 2 (Cr)	80.556	80.556	83.333	66.667	69.444
Original image	Original RGB image (O)	91.667	69.444	72.223	69.444	94.444

Table 5 Outcomes of freshness class detection with GLCM features of the greyscale colourmap using the proposed ANN classifier

Colourmap	Feature	Accuracy of classification (%) using features such as			
		Contrast	Correlation	Energy	Homogeneity
Greyscale image	GLCM	75	63.889	86.111	69.444

Outcomes of the ANN Classifier Using Minor Features

Apparent observations of these boxplots show that some of the features exhibit significant separating of values between the two associated classes, whereas the other features do not show prominent significance. Hence, all these features are further analyzed only using the proposed ANN classifier model in a similar way as is done for the major features. The results so obtained are described in Table 4 and Table 5 to illustrate the classification accuracies.

The proposed methods of analysis are described in the form of a flow diagram in Fig. 12.

Performance Analysis

The performance of the model build is studied by considering the following equations.

$$\text{Accuracy } (\%) = \left(P_{\mathrm{T}} + N_T\right)/P_T + P_F + N_T + N_F \tag{24}$$

$$\text{Precision} = P_{\mathrm{T}}/\left(P_{\mathrm{T}} + P_{\mathrm{F}}\right) \tag{25}$$

$$\text{Sensitivity } \left(T_{\mathrm{PR}} \text{ or true positive rate}\right) = P_{\mathrm{T}}/\left(P_{\mathrm{T}} + P_{\mathrm{F}}\right) \tag{26}$$

$$\text{Quality measure} = 2 \times (\text{Sensitivity} \times \text{Precision})/(\text{Sensitivity} + \text{Precision}) \tag{27}$$

$$\text{Intersection of overall union} = P_{\mathrm{T}}/\left(P_T + P_F + N_F\right) \tag{28}$$

$$\text{Specificity (true negative rate)} = N_{\mathrm{T}}/\left(N_T + P_F\right) \tag{29}$$

$$\text{Negative predictive value} = N_{\mathrm{T}}/\left(N_T + N_F\right) \tag{30}$$

$$\text{False} - \text{positive rate}\left(F_{\mathrm{PR}}\right) = P_{\mathrm{F}}/\left(P_F + N_T\right) \tag{31}$$

$$\text{False} - \text{negative rate } \left(F_{\mathrm{NR}}\right) = N_{\mathrm{F}}/\left(N_F + P_T\right) \tag{32}$$

$$\text{Mathew Correlation (MC)} = \left[\left(P_{\mathrm{T}} \times N_{\mathrm{T}}\right) - \left(P_{\mathrm{F}} \times N_{\mathrm{F}}\right)\right] / \left[\left(P_{\mathrm{T}} + P_{\mathrm{F}}\right)\left(P_{\mathrm{T}} + N_{\mathrm{F}}\right)\left(N_{\mathrm{T}} + P_{\mathrm{F}}\right)\left(N_{\mathrm{T}} + N_{\mathrm{F}}\right)\right]^{1/2} \tag{33}$$

$$\text{Kappa value } (K) = \frac{A_0 - A_e}{1 - A_e} \tag{34}$$

where

$$A_0 = \left(P_{\mathrm{T}} + N_{\mathrm{T}}\right)/\left(P_{\mathrm{T}} + P_{\mathrm{F}} + N_{\mathrm{T}} + N_{\mathrm{F}}\right)$$
$$A_e = \left[\left(P_{\mathrm{T}} + P_{\mathrm{F}}\right) \times \left(P_{\mathrm{T}} + N_{\mathrm{F}}\right) + \left(N_{\mathrm{T}} + N_{\mathrm{F}}\right) \times \left(P_{\mathrm{F}} + N_{\mathrm{T}}\right)\right]/\left(P_{\mathrm{T}} + P_{\mathrm{F}} + N_{\mathrm{T}} + F_{\mathrm{F}}\right)^2$$

where P_{T} is the true positive, P_{F} is the false positive, N_{F} is the false negative, and N_{T} is the true negative.

The performance analysis of the outcomes of both the PCA threshold–based model and the ANN model are shown in Table 6. Here we have examined only the features which produce classifier accuracy higher than 90%, as these features are considered the most important ones. Other features are not elaborated here since those yields less significant outcomes.

Analysis of Results Using Major Features

It is observed that classification accuracy is higher in all the cases with ANN classifier, compared to PCA threshold freshness classifier. A comparative analysis of the outcomes of the two classifier models is shown in Fig. 13 which reveals the above fact more prominently.

It is further observed that classifier accuracy exceeds 90% level for green, blue, hue, luminance and chrominance 2 (Cr) layers using the ANN model, whereas only the blue layer is able to classify the samples with higher than 90% accuracy using the PCA threshold model. The blue layer exhibits the highest accuracy of classification with both models, which establishes the strong classification ability of the same layer. This efficiency of classification reaches the highest level of about 94.4% using the ANN model, and 93.3% with the PCA threshold model. The luminance layer also yields the same highest efficiency of 94.4% using the ANN model, although the same efficiency with the PCA threshold model is much less at about 85.8%. The red, vital component and chrominance 1 (Cb) layers, on the other hand, displays reduced accuracy of classification, especially using the PCA threshold model, when the accuracy reduces beyond 80%. The red layer produces the least efficiency of about 83% with the ANN model and the Chrominance 1 (Cb) layer shows the same with a low classification accuracy of about 77% using the PCA threshold model.

The texture feature is obtained in this work using entropy filtering, where a 9×9 windowing method is applied to obtain the entropy value of the central pixel, and thereby, forming greyscale entropy-filtered image of the same

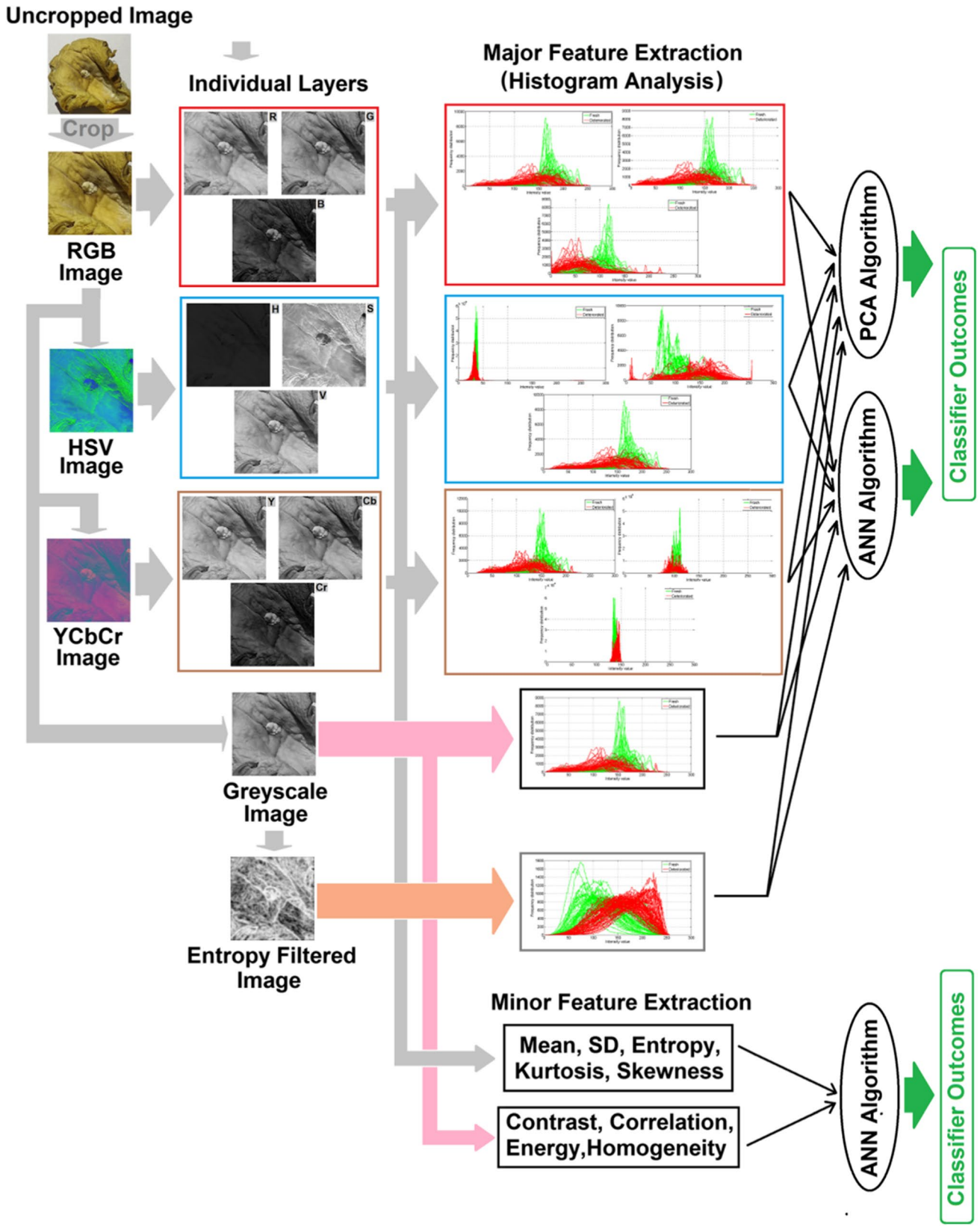

Fig. 12 Flow diagram of the proposed mushroom freshness classifier models

Table 6 Performance analysis of the obtained outcomes

Feature Type	Feature	Model	Layer	Accuracy	Precision	Sensitivity	Quality measure	Intersection of overall union	Specificity	Negative predictive value	False-positive rate	False-negative rate	Mathew Correlation	Kappa value
Major Features	Histograms	PCA	B	93.33	93.85	93.85	93.85	88.41	92.73	92.73	7.27	6.154	0.87	0.87
		ANN	G	91.67	88.89	94.12	91.43	84.21	89.47	94.44	10.5	5.882	0.83	0.83
			B	94.44	94.44	94.44	94.44	89.47	94.44	94.44	5.56	5.556	0.89	0.89
			H	91.67	94.44	89.47	91.89	85	94.12	88.89	5.88	10.53	0.83	0.83
			Y	94.44	100	90	94.74	90	100	88.89	0	10	0.89	0.89
			Cr	91.67	94.44	89.47	91.89	85	94.12	88.89	5.88	10.53	0.83	0.83
Minor Features	Mean	ANN	G	94.44	94.44	94.44	94.44	89.47	94.44	94.44	5.56	5.556	0.89	0.89
			B	94.44	94.44	94.44	94.44	89.47	94.44	94.44	5.56	5.556	0.89	0.89
			H	91.67	94.44	89.47	91.89	85	94.12	88.89	5.88	10.53	0.83	0.83
			V	91.67	88.89	94.12	91.43	84.21	89.47	94.44	10.5	5.882	0.83	0.83
			Y	91.67	88.89	94.12	91.43	84.21	89.47	94.44	10.5	5.882	0.83	0.83
			O	91.67	94.44	89.47	91.89	85	94.12	88.89	5.88	10.53	0.83	0.83
	Skewness		B	91.67	100	85.71	92.31	85.71	100	83.33	0	14.29	0.85	0.83
			O	94.44	94.44	94.44	94.44	89.47	94.44	94.44	5.56	5.556	0.89	0.89

dimension by running the window over the whole image. The efficiency of classification using the histogram of the entropy-filtered image is obtained almost equal using the two models of classifier, and the values being about 89% and 87.5% respectively using the ANN model and PCA threshold model respectively.

Analysis of Results Using Minor Features

The classifier accuracies obtained using the minor features such as mean, standard deviation (SD), entropy, kurtosis and skewness are shown graphically in Fig. 14.

Analysis of the minor features mostly show that the average classifier accuracy is highest with the mean values of almost all the layers of each colourmap, except for the Chrominance 1 (Cb) layer, where the accuracy is exceptionally poor at below 60%. In the case of blue and green layers, the same accuracy reaches as high as about 94.4%, which is even compared to the accuracy level reached using some of the major features using the ANN model. The classifier accuracy using the same mean feature is also found higher than 90% using other layers such as hue, vital component, luminance and the original RGB image. On the other hand, the kurtosis feature produces the least average accuracy of classification of less than 70% considering all the layers of each colourmap. The standard deviation (SD) feature yields the highest efficiency with the vital component layer and it ranges in the order of 86%, and the lowest accuracy for the same parameter being 61% using the blue layer. The highest and the lowest level of accuracy of classification, using the entropy feature has been found as about 89% and 63% using hue and blue layers respectively. Analysis of the kurtosis feature has produced the worst average accuracy considering all layers, with many of the results lying below 80% and 70% and even the same hue and blue layers have yielded an efficiency level of below 50%. The skewness feature has also yielded an average accuracy level of below 80%, similar to SD and entropy features, whereas the blue layer and the original RGB image have yielded an average accuracy level exceeding 90% using the same skewness feature.

Receiver Operating Characteristic Plot Evaluation

The performance of the proposed ANN classifier is exemplified with the help of the receiver operating characteristic (ROC) curve of Fig. 15, constructed with varying discrimination threshold and plotting T_{PR} against F_{PR} for every possible CUO (cut-off) value.

$$T_{\mathrm{PR}}\left(\mathrm{CUO}_x\right) = P_d\left(\mathrm{RQ} \geq \mathrm{CUO} \middle| E_r^+\right) \tag{35}$$

$$F_{\mathrm{PR}}\left(\mathrm{CUO}_x\right) = P\left(\mathrm{RQ} \geq \mathrm{CUO} \middle| E_r^-\right) \tag{36}$$

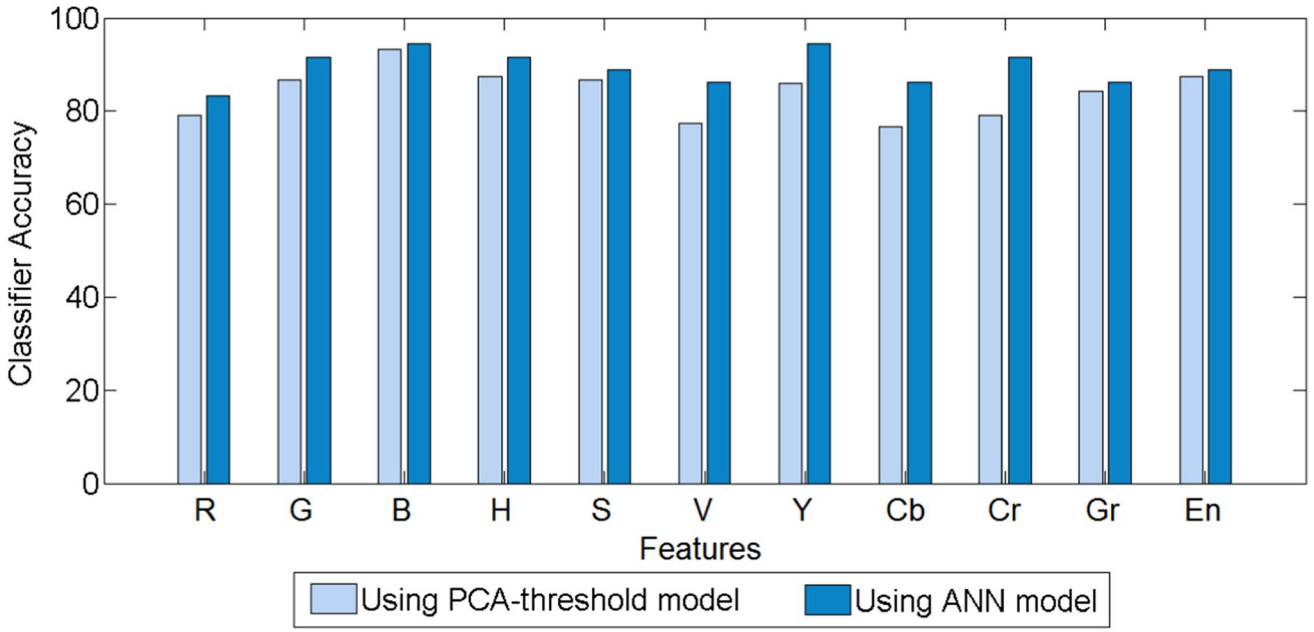

Fig. 13 Comparative analysis of PCA threshold and ANN classifier using different major colour and texture features: red (R), green (G), blue (B), hue (H), saturation (S), vital component (V), luminance (Y), chrominance 1 (Cb), chrominance 2 (Cr), greyscale image (Gr) and entropy-filtered image (En)

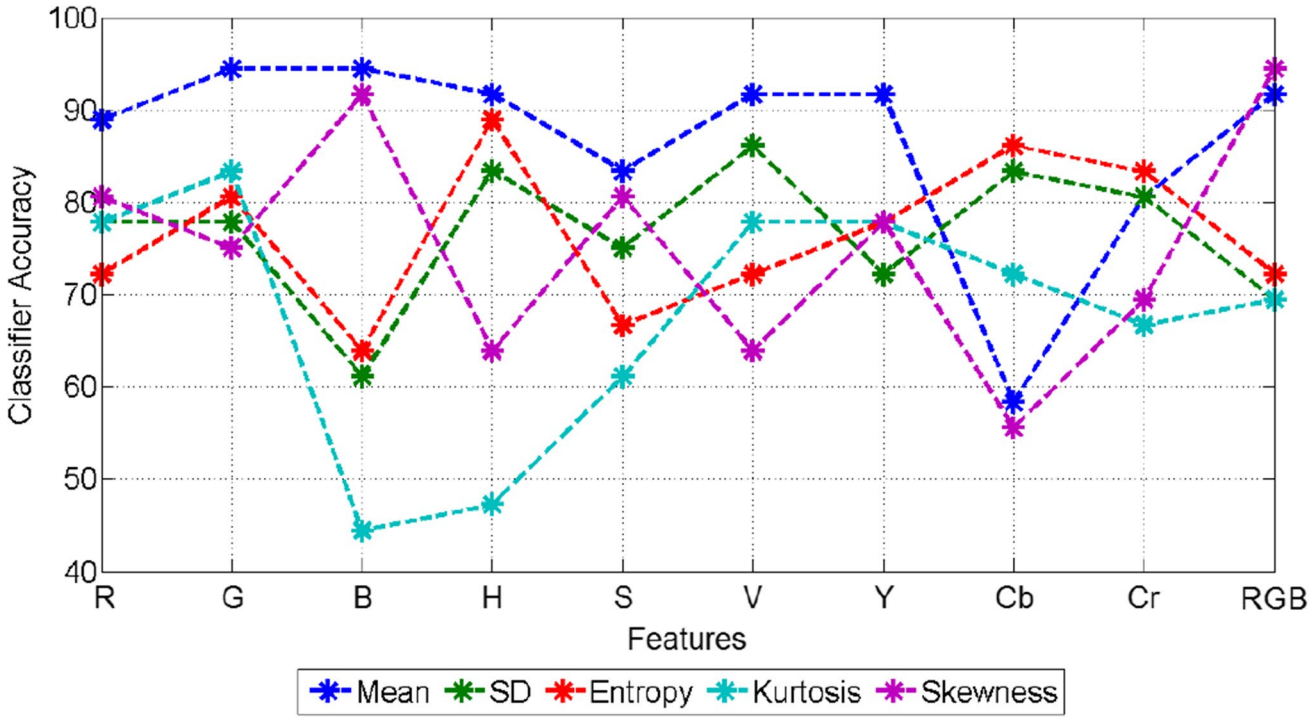

Fig. 14 Classifier accuracy obtained with ANN classifier using different minor features like mean, standard deviation (SD), entropy, kurtosis and skewness of different layers: red (R), green (G), blue (B), hue (H), saturation (S), vital component (V), luminance (Y), chrominance 1 (Cb), chrominance 2 (Cr) and complete RGB image (RGB)

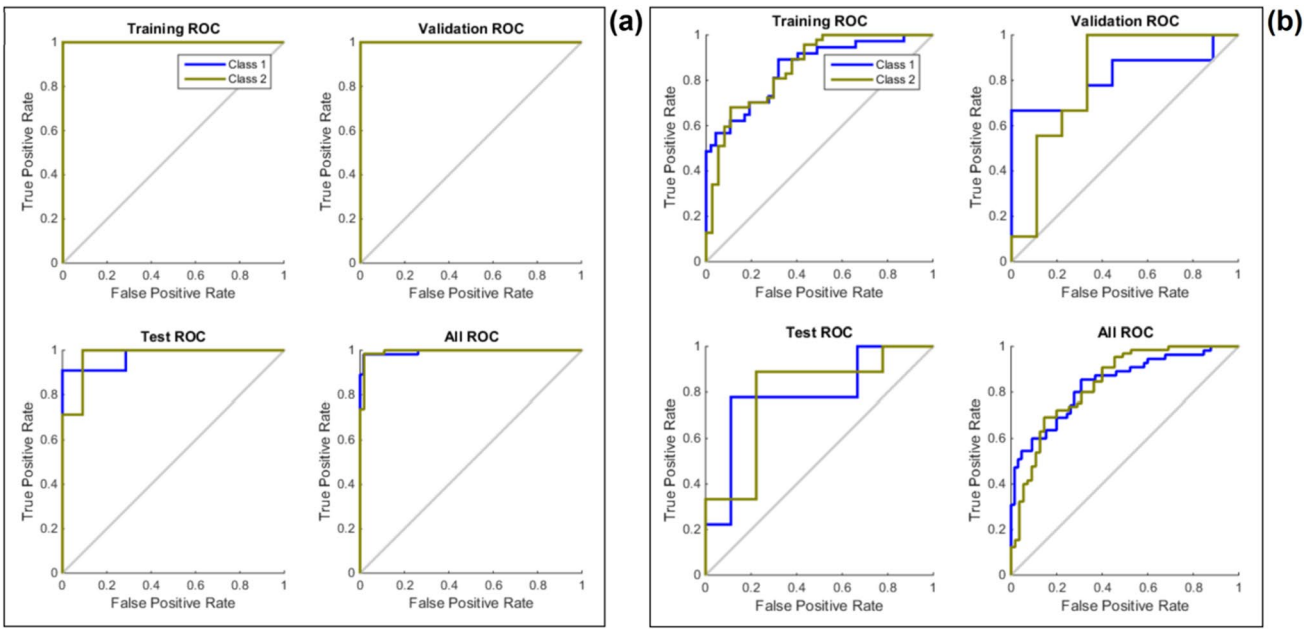

Fig. 15 ROC curve for **a** B-ANN model and **b** Cr-kurtosis-ANN model

where CUO_x is the cut-off value at the position, P_d is the probability density, E_r+ and E_r- are the positive and negative errors respectively and R_Q are the quantitative results.

Thus, the representation of the ROC curve is as follows:

$$\mathrm{ROC} = \left\{F_{\mathrm{PR}}\left(\mathrm{CUO}_x\right), T_{\mathrm{PR}}\left(\mathrm{CUO}_x\right), \mathrm{CUO} \in (-\infty, +\infty)\right\} \tag{37}$$

In terms of R_Q, the ROC is described as follows:

$$\mathrm{ROC} = \left\{\left(R_Q, \mathrm{ROC}\left(R_Q\right)\right), R_Q \in (0,1)\right\} \tag{38}$$

where the ROC maps R_Q to T_{PR} (CUO_x) and CUO_x matching to F_{PR} (CUO_x) = RQ.

The higher the area under the curve (AUC), the better the performance of the proposed classifier will be. Therefore, the model (build with the extracted features) with higher AUC is obviously the efficient one, as the ROC plot is built during the alteration in the scalar threshold value of the model (Lorente et al. 2013). The model constructed with the feature '*B*' as the input layer has the highest AUC for the overall model as well as for the testing, training and validation set (Fig. 15a), the model built with the feature named kurtosis obtained from Cr channel possessed minimum AUC (Fig. 15b). Instead of measuring the absolute value, the AUC is unaltered with scale, thus able to determine the precision of the predictions. However, it is independent of the classification threshold. The eminence of the prediction is therefore illustrated by AUC irrespective of the classification threshold.

Conclusion

The proposed work illustrates an analysis for assessing the edibility of oyster mushrooms, using the images of two classes of samples, e.g. *Fresh* and *Deteriorated*. The present work uses two major analysis tools for this purpose. The first one is a statistical analysis model employing principal component analysis and the second one is an artificial neural network model. These models are used to analyze image features of each layer of the red (R)-green (G)-blue (B) colourmap, hue (H)-saturation (S)-vital component (V) colourmap, luminance (Y)-chrominance (Cb and Cr) colourmap, greyscale image and the entropy-filtered image. Observations reveal that the supervised learning model exceeds the statistical model in terms of the accuracy of classification. More so, some of the features show more than 90% accuracy of classification of the samples, which is, in fact, highly accurate for all real-life applications. Most importantly, the use of smartphones for capturing the sample images exhibits the possibility of widespread application and portability of the proposed algorithm.

Funding Thanks to GAIN (Axencia Galega de Innovación) for supporting this review (grant number IN607A2019/01).

Data Availability All the data used in the manuscript are available in the tables and figures.

Code Availability Not applicable.

Declarations

Ethics Approval Not applicable.

Consent to Participate All authors have given their full consent to participate.

Consent for Publication All authors have given their full consent for publication.

Conflict of Interest Tanmay Sarkar declares that he has no conflict of interest. Alok Mukherjee declares that he has no conflict of interest. Kingshuk Chatterjee declares that he has no conflict of interest. Mohammad Ali Shariati declares that he has no conflict of interest. Maksim Rebezov declares that he has no conflict of interest. Svetlana Rodionova declares that she has no conflict of interest. Denis Smirnov declares that he has no conflict of interest. Ruben Dominguez declares that he has no conflict of interest. Jose M. Lorenzo declares that he has no conflict of interest.

References

Adebo OA, Molelekoa T, Makhuvele R et al (2021) A review on novel non-thermal food processing techniques for mycotoxin reduction. Int J Food Sci Technol 56:13–27. https://doi.org/10.1111/ijfs.14734

Aguirre L, Frias JM, Barry-Ryan C, Grogan H (2009) Modelling browning and brown spotting of mushrooms (Agaricus bisporus) stored in controlled environmental conditions using image analysis. J Food Eng 91:280–286. https://doi.org/10.1016/j.jfoodeng.2008.09.004

Ali SSE, Dildar SA (2020) An efficient quality inspection of food products using neural network classification. J Intell Syst 29:1425–1440. https://doi.org/10.1515/jisys-2018-0077

Angerosa F, Di GL, Vito R, Cumitini S (1996) Sensory evaluation of virgin olive oils by artificial neural network processing of dynamic head-space gas chromatographic data. J Sci Food Agric 72:323–328. https://doi.org/10.1002/(SICI)1097-0010(199611)72:3%3c323::AID-JSFA662%3e3.0.CO;2-A

Anil A, Arora M (2018) Image processing based non-destructive testing of mushroom sample. In: 2018 Second International Conference on Electronics, Communication and Aerospace Technology (ICECA). pp 1625–1629

Azarmdel H, Jahanbakhshi A, Mohtasebi SS, Muñoz AR (2020) Evaluation of image processing technique as an expert system in mulberry fruit grading based on ripeness level using artificial neural networks (ANNs) and support vector machine (SVM). Postharvest Biol Technol 166:111201. https://doi.org/10.1016/j.postharvbio.2020.111201

Bhargava A, Bansal A (2020a) Machine learning based quality evaluation of mono-colored apples. Multimed Tools Appl 79:22989–23006. https://doi.org/10.1007/s11042-020-09036-9

Bhargava A, Bansal A (2021) Classification and grading of multiple varieties of apple fruit. Food Anal Methods. https://doi.org/10.1007/s12161-021-01970-0

Bhargava A, Bansal A (2020b) Automatic detection and grading of multiple fruits by machine learning. Food Anal Methods 13:751–761. https://doi.org/10.1007/s12161-019-01690-6

Binti Zaidi NI, Binti Lokman NAA, Bin Daud MR et al (2015) Fire recognition using RGB and YCbCr color space. ARPN J Eng Appl Sci 10:9786–9790

Caballero D, Pérez-Palacios T, Caro A et al (2017) Prediction of pork quality parameters by applying fractals and data mining on MRI. Food Res Int 99:739–747. https://doi.org/10.1016/j.foodres.2017.06.048

Dibaba T, Abera S (2017) Nutritional quality of Oyster mushroom (Pleurotus ostreatus) as affected by osmotic pretreatments and drying methods. Food Sci Nutr 5:989–996. https://doi.org/10.1002/fsn3.484

Dowlati M, Mohtasebi SS, Omid M et al (2013) Freshness assessment of gilthead sea bream (Sparus aurata) by machine vision based on gill and eye color changes. J Food Eng 119:277–287. https://doi.org/10.1016/j.jfoodeng.2013.05.023

Dutta Gupta S, Pattanayak AK (2017) Intelligent image analysis (IIA) using artificial neural network (ANN) for non-invasive estimation of chlorophyll content in micropropagated plants of potato. Vitr Cell Dev Biol - Plant 53:520–526. https://doi.org/10.1007/s11627-017-9825-6

Gabriëls SHEJ, Mishra P, Mensink MGJ et al (2020) Non-destructive measurement of internal browning in mangoes using visible and near-infrared spectroscopy supported by artificial neural network analysis. Postharvest Biol Technol 166:111206. https://doi.org/10.1016/j.postharvbio.2020.111206

Gago J, Landín M, Gallego P (2010) Strengths of artificial neural networks in modeling complex plant processes. Plant Signal Behav 5:743–745. https://doi.org/10.4161/psb.5.6.11702

Gowen AA, O'Donnell CP, Frias JM, Downey G (2009) Hyperspectral imaging for mushroom (agaricus bisporus) quality monitoring. In: 2009 First Workshop on Hyperspectral Image and Signal Processing: evolution in Remote Sensing. pp 1–4

Hazisawa T, Toda M, Sakoil T, et al. (2013) Image analysis method for grading raw shiitake mushrooms. In: The 19th Korea-Japan Joint Workshop on Frontiers of Computer Vision. pp 46–52

Hu J, Zhou C, Zhao D et al (2020) A rapid, low-cost deep learning system to classify squid species and evaluate freshness based on digital images. Fish Res 221:105376. https://doi.org/10.1016/j.fishres.2019.105376

Hussain A, Pu H, Sun D-W (2018) Innovative nondestructive imaging techniques for ripening and maturity of fruits – a review of recent applications. Trends Food Sci Technol 72:144–152. https://doi.org/10.1016/j.tifs.2017.12.010

Ismail S, Zainal AR, Mustapha A (2018) Behavioural features for mushroom classification. In: 2018 IEEE Symposium on Computer Applications & Industrial Electronics (ISCAIE). pp 412–415

Jahanbakhshi A, Momeny M, Mahmoudi M, Zhang Y-D (2020) Classification of sour lemons based on apparent defects using stochastic pooling mechanism in deep convolutional neural networks. Sci Hortic (amsterdam) 263:109133. https://doi.org/10.1016/j.scienta.2019.109133

Jolliffe I (2002) Principal component analysis, 2nd edn. Springer. https://doi.org/10.1007/b98835

Koyama K, Tanaka M, Cho B-H, et al. (2021) Predicting sensory evaluation of spinach freshness using machine learning model and digital images. PLoS One 16:e0248769

Kumar P, Kumar V, Goala M et al (2021) Integrated use of treated dairy wastewater and agro-residue for Agaricus bisporus mushroom cultivation: experimental and kinetics studies. Biocatal Agric Biotechnol 32:101940. https://doi.org/10.1016/j.bcab.2021.101940

Li H, Tian Y, Menolli N Jr et al (2021) Reviewing the world's edible mushroom species: a new evidence-based classification system. Compr Rev Food Sci Food Saf 20:1982–2014. https://doi.org/10.1111/1541-4337.12708

Liu X, Jiang Y, Shen S et al (2015) Comparison of Arrhenius model and artificial neuronal network for the quality prediction of rainbow trout (Oncorhynchus mykiss) fillets during storage at different temperatures. LWT - Food Sci Technol 60:142–147. https://doi.org/10.1016/j.lwt.2014.09.030

Lorente D, Aleixos N, Gómez-Sanchis J et al (2013) Selection of optimal wavelength features for decay detection in citrus fruit using

the ROC curve and neural networks. Food Bioprocess Technol 6:530–541. https://doi.org/10.1007/s11947-011-0737-x

Lu C-P, Liaw J-J (2020) A novel image measurement algorithm for common mushroom caps based on convolutional neural network. Comput Electron Agric 171:105336. https://doi.org/10.1016/j.compag.2020.105336

Lu C-P, Liaw J-J, Wu T-C, Hung T-F (2019) Development of a mushroom growth measurement system applying deep learning for image recognition. Agron 9(1):32. https://doi.org/10.3390/agronomy9010032

Ma J, Sun D-W, Qu J-H et al (2016) Applications of computer vision for assessing quality of agri-food products: a review of recent research advances. Crit Rev Food Sci Nutr 56:113–127. https://doi.org/10.1080/10408398.2013.873885

de Machado T, A, Costa AG, Rodrigues RE, et al (2020) Quality of tomatoes under different transportation conditions by principal component analysis. Rev Ceres 67:448–453. https://doi.org/10.1590/0034-737x202067060004

Masoudian A, Mcisaac KA (2013) Application of support vector machine to detect microbial spoilage of mushrooms. In: 2013 International Conference on Computer and Robot Vision. pp 281–287

Mukherjee A, Chatterjee K, Sarkar T (2021) Entropy-aided assessment of Amla (Emblica officinalis) quality using principal component analysis. Biointerface Res Appl Chem 12(2):2162–2170. https://doi.org/10.33263/BRIAC122.21622170

Mukherjee A, Sakar T, Chatterjee K (2021) Freshness assessment of Indian gooseberry (Phyllanthus emblica) using probabilistic neural network. J Biosyst Eng 46(3). https://doi.org/10.1007/s42853-021-00116-8

Nadim M, Ahmadifar H, Mashkinmojeh M, yamaghani MR (2019) Application of image processing techniques for quality control of mushroom TT -. gums-cjhr 4:72–75. https://doi.org/10.29252/cjhr.4.3.72

Navotas IC, Santos CN V, Balderrama EJM, et al. (2018) Fish identification and freshness classification through image processing using artificial neural network

Péneau S, Brockhoff PB, Escher F, Nuessli J (2007) A comprehensive approach to evaluate the freshness of strawberries and carrots. Postharvest Biol Technol 45:20–29. https://doi.org/10.1016/j.postharvbio.2007.02.001

Pourdarbani R, Sabzi S, Hernández-Hernández M et al (2020) Non-destructive estimation of total chlorophyll content of apple fruit based on color feature, spectral data and the most effective wavelengths using hybrid artificial neural network-imperialist competitive algorithm. Plants (basel, Switzerland) 9https://doi.org/10.3390/plants9111547

Preechasuk J, Chaowalit O, Pensiri F, Visutsak P (2019) Image analysis of mushroom types classification by convolution neural networks. In: Proceedings of the 2019 2nd Artificial Intelligence and Cloud Computing Conference. Association for Computing Machinery, New York, NY, USA, pp 82–88

Przybył K, Gawałek J, Koszela K (2020) Application of artificial neural network for the quality-based classification of spray-dried rhubarb juice powders. J Food Sci Technol. https://doi.org/10.1007/s13197-020-04537-9

Przybylak A, Boniecki P, Koszela K et al (2016) Estimation of intramuscular level of marbling among Whiteheaded Mutton Sheep lambs. J Food Eng 168:199–204. https://doi.org/10.1016/j.jfoodeng.2015.07.035

Qian Y, Jiacheng R, Pengbo W, et al. (2020) Real-time detection and localization using SSD method for oyster mushroom picking robot*. In: 2020 IEEE International Conference on Real-time Computing and Robotics (RCAR). pp 158–163

Rahmawati D, Ibadillah AF, Ulum M, Setiawan H (2018) Design of automatic harvest system monitoring for oyster mushroom using image processing BT - Proceedings of the International Conference on Science and Technology (ICST 2018). Atlantis Press, pp 143–147

Sanahuja S, Fédou M, Briesen H (2018) Classification of puffed snacks freshness based on crispiness-related mechanical and acoustical properties. J Food Eng 226:53–64. https://doi.org/10.1016/j.jfoodeng.2017.12.013

Sarkar T, Salauddin M, Hazra S, Chakraborty R (2020) 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. https://doi.org/10.1007/s42452-020-03455-x

Sarkar T, Mukherjee A, Chatterjee K (2021) Supervised learning aided multiple feature analysis for freshness class detection of Indian gooseberry (Phyllanthus emblica). J Inst Eng (India): A. https://doi.org/10.1007/s40030-021-00585-2

Suktanarak S, Teerachaichayut S (2017) Non-destructive quality assessment of hens' eggs using hyperspectral images. J Food Eng 215:97–103. https://doi.org/10.1016/j.jfoodeng.2017.07.008

Sun Y, Wei K, Liu Q et al (2018) Classification and discrimination of different fungal diseases of three infection levels on peaches using hyperspectral reflectance imaging analysis. Sensors (basel) 18https://doi.org/10.3390/s18041295

Tan C, Sun F, Kong T, et al. (2018) A survey on deep transfer learning. In: In International conference on artificial neural networks. Springer, pp 270–279

Tarafdar A, Shahi NC, Singh A (2020) Color assessment of freeze-dried mushrooms using Photoshop and optimization with genetic algorithm. J Food Process Eng 43:e12920. https://doi.org/10.1111/jfpe.12920

Wada Y, Tsuzuki D, Kobayashi N et al (2007) Visual illusion in mass estimation of cut food. Appetite 49:183–190. https://doi.org/10.1016/j.appet.2007.01.009

Wang F, Zheng J, Tian X et al (2018) An automatic sorting system for fresh white button mushrooms based on image processing. Comput Electron Agric 151:416–425. https://doi.org/10.1016/j.compag.2018.06.022

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