Expert Systems with Applications 42 (2015) 1892–1905

Contents lists available at ScienceDirect

Expert Systems with Applications

journal homepage: www.elsevier.com/locate/eswa

Using machine learning techniques for evaluating tomato ripeness

Nashwa El-Bendary [a,d,*], Esraa El Hariri [b,d], Aboul Ella Hassanien [c,d], Amr Badr [c]

[a] Arab Academy for Science, Technology, and Maritime Transport, Cairo, Egypt
[b] Faculty of Computers and Information, Fayoum University, Egypt
[c] Faculty of Computers and Information, Cairo University, Egypt
[d] Scientific Research Group in Egypt, Egypt

ARTICLE INFO

Article history:
Available online 13 October 2014

Keywords:
Image classification
Features extraction
Ripeness
Principal Component Analysis (PCA)
Support Vector Machines (SVMs)
Linear Discriminant Analysis (LDA)

ABSTRACT

Tomato quality is one of the most important factors that helps ensuring a consistent marketing of tomato fruit. As ripeness is the main indicator for tomato quality from customers perspective, the determination of tomato ripeness stages is a basic industrial concern regarding tomato production in order to get high quality product. Automatic ripeness evaluation of tomato is an essential research topic as it may prove benefits in ensuring optimum yield of high quality product, this will increase the income because tomato is one of the most important crops in the world. This article presents an automated multi-class classification approach for tomato ripeness measurement and evaluation via investigating and classifying the different maturity/ripeness stages. The proposed approach uses color features for classifying tomato ripeness stages. The approach proposed in this article uses Principal Components Analysis (PCA) in addition to Support Vector Machines (SVMs) and Linear Discriminant Analysis (LDA) algorithms for feature extraction and classification, respectively. Experiments have been conducted on a dataset of total 250 images that has been used for both training and testing datasets with 10-fold cross validation. Experimental results showed that the proposed classification approach has obtained ripeness classification accuracy of 90.80%, using one-against-one (OAO) multi-class SVMs algorithm with linear kernel function, ripeness classification accuracy of 84.80% using one-against-all (OAA) multi-class SVMs algorithm with linear kernel function, and ripeness classification accuracy of 84% using LDA algorithm.

1. Introduction

Fruits and vegetables development is characterized by a short period of cell division followed by a longer period of cell elongation by water uptake. The final fruit size mainly depends on initial cell number, rather than cell size (Cowan, Cripps, Richings, & Taylor, 2001). Fruit ripening, on the other hand, is characterized by the development of color, flavor, texture and aroma. The actual time from anthesis until full maturity can vary tremendously among species/cultivars due to genetic and environmental differences. Even between fruit on the same plant, fruit development and ripening can take more or less time depending on local microclimate conditions and differences in sink/source relations within the plant. In addition, when a fruit is harvested, the time of anthesis of a particular fruit is generally unknown, as is its full history (El Hariri, El-Bendary, Hassanien, & Badr, 2014; Lang & Hübert, 2012; Wei, Liu, Qiu, Shao, & He, 2014).

According to (FAOSTAT Database) Food and of the United Nations (FAO-UN) (2012), tomatoes world production was about 162 million tons fresh fruits produced in the year 2012 with income about 592 trillion dollars. Tomatoes are taking an important place among the fruits and vegetables all over the world, due to their continuously prevailing daily nutrition, dietary value and production income. Moreover, tomato is the fourth most important crop next to soybeans at world production (Camelo, 2004). Tomato production has been reported for 144 countries and Egypt occupies the fifth place in these countries at tomato production in both income and weight of fruit produced. Furthermore, tomato is the first most important crop in Egypt. Based on the previously stated facts, in this article we selected to focus on monitoring the ripeness stages of tomato crop, especially in Egypt. Another matter of fact is, on the categorization of crops into the two types of *climacteric* that are able to continue ripening after picking from the mother plant and *non-climacteric* that can ripe only when it is

* Corresponding author at: Arab Academy for Science, Technology, and Maritime Transport, Cairo, Egypt.

E-mail addresses: nashwa.elbendary@ieee.org, nashwa.elbendary@aast.edu (N. El-Bendary), eng.esraa.elhariri@gmail.com (E. El Hariri), aboitcairo@gmail.com (A.E. Hassanien), a.badr.fci@gmail.com (A. Badr).

URLs: http://www.egyptscience.net (N. El-Bendary), http://www.egyptscience.net (E. El Hariri), http://www.egyptscience.net (A.E. Hassanien).

http://dx.doi.org/10.1016/j.eswa.2014.09.057
0957-4174/

attached to the mother plant (Camelo, 2004; Coates & Johnson, 1997), tomato belongs to climacteric category as it can reach over-ripening stage after being harvested. Also, tomatoes have many different ripeness stages, which are (1) Green, (2) Breaker, (3) Turning, (4) Pink, (5) Light red and (6) Red stages, so they reach full red color even when harvested green. Red stage is the most preferred ripeness stage commercially (Camelo, 2004; El Hariri et al., 2014).

As has been noted, monitoring and controlling produce (fruits and vegetables) ripeness has become a very important issue in the crops industry, since ripeness is perceived by customers as the main quality indicator. Also, the product's appearance is one of the most worrying issues for producers as it has a high influence on product's quality and consumer preferences. However, up to this day, optimal harvest dates and prediction of storage life are still mainly based on subjective interpretation and practical experience (El Hariri et al., 2014). Hence, automation of that process is of a great gain for agriculture and industry fields. For agriculture, it may be used to develop automatic harvest systems and saving crops from damages caused by environmental changes. On the other hand, for industry, it is used to develop automatic sorting system or checking the quality of fruits to increase customer satisfaction level (Brezmes, Llobet, Vilanova, Saiz, & Correig, 2000; Elhariri et al., 2014). Accordingly, an objective and accurate ripeness assessment of agricultural crops is important in ensuring optimum yield of high quality products. Moreover, identifying physiological and harvest maturity of agricultural crops correctly, will ensure timely harvest to avoid cutting of either under-ripe and over-ripe agricultural crops (El Hariri et al., 2014, 2014; May & Amaran, 2011).

To put it briefly, the main research motivation of the approach proposed in this article is providing an automated multi-class classification approach for tomato ripeness measurement and evaluation via investigating and classifying the different maturity/ripeness stages based on the color features. As previously stated, among the 144 countries reported for tomato production, Egypt occupied the fifth place in both income and weight of fruit produced as well as the fact that tomato is the first most important crop in Egypt. Thus, another research motivation in this article is the selection to focus on monitoring the ripeness stages of tomato crop, especially in Egypt. The dataset used for experiments were constructed based on real sample images for tomato at different stages, which were collected from different farms in Minya city, Upper Egypt. Dataset of total 250 images was used for both training and testing datasets with 10-fold cross-validation. Training dataset is divided into 5 classes representing the different stages of tomato ripeness. The proposed approach consists of three phases; namely *pre-processing*, *feature extraction*, and *classification* phases. During pre-processing phase, the proposed approach resizes images to 250×250 pixels, in order to reduce their color index, and the background of each image has been removed using background subtraction technique. Also, each image has been converted from RGB to HSV color space. For feature extraction phase, Principal component analysis (PCA) algorithm was applied in order to generate a feature vector for each image in the dataset. Finally, for classification phase, the proposed approach applied Support Vector Machines (SVMs) and Linear Discriminant Analysis (LDA) algorithms for classification of ripeness stages.

Another basic research motivation is that, to the best of our knowledge, none of the recent ripeness classification related research works have addressed the dependency of the classification approach performance on statistics of the experimented dataset(s).

So, another contribution of this article is that it highlights the most appropriate classification algorithm considering the dependency of the classification approach performance on statistics of the experimented dataset. That has been achieved via adopting the utilization of principal component analysis (PCA) in addition to Support Vector Machines (SVMs) and Linear Discriminant Analysis (LDA) algorithms for feature extraction and classification, respectively, for tomato ripeness stages evaluation and classification considering the color features. Also, both training and testing datasets have been generated via employing the 10-fold cross validation.

An essential finding is that the performance of LDA and SVMs was highly dependent on statistics of the dataset. That is, on datasets with fewer classes (ripeness categories), and many training examples per class, SVMs had an advantage over the LDA classification approach.

The selection of both SVMs classification algorithm depended on the facts that the application of SVMs classification algorithm has may advantages such as, it deliver a unique solution, it does not need any assumptions about the functional form of the transformation, because the kernel implicitly contains a non-linear transformation. Also, if an appropriate generalization grade was chosen, even when the training sample has some bias, SVMs can be robust. Moreover, by choosing an appropriate kernel, one can put more stress on the similarity between samples. However, there is as well some limitations for using SVMs algorithm, that is the lack of transparency of results and the need for very large training time when using large datasets (Auria & Moro, 2008). On the other hand, the selection of both LDA classification algorithm depended on its advantages that are LDA has some advantages such as, the employment of projection that solves the problem of illumination by maximizing between-class scatter and minimizing within-class scatter and it need less samples in order to obtain a reliable classifier. However, one common disadvantage of LDA is the singularity problem as well as it fails when all scatter matrix are singular (Kumar & Kaur, 2012).

In general, the limitations we faced in this research are the dataset size that's needed to be larger, as the accuracy of SVMs increases by increasing the number of images per training class, and accordingly a maximum accuracy of 90.2% has been achieved.

The rest of this article is organized as follows. Section 2 introduces some recent research work related to monitoring and classification of maturity stages for tomatoes and other fruits/vegetables. Section 3 presents the core concepts of SVMs, LDA and PCA algorithms. Section 4 describes the different phases of the proposed content-based classification system; namely pre-processing, feature extraction, and classification phases. Section 5 discusses the tested image dataset and presented the obtained experimental results. Finally, Section 6 presents conclusions and addresses a number of future research suggestions.

2. Related work

This section reviews a number of current research approaches that tackle the problem of ripeness monitoring and classification for tomatoes and other fruits/vegetables.

First of all, for tomato ripeness classification, various research works have been proposed. In Zhang and McCarthy (2012), authors offered tomato maturity evaluation approach using magnetic resonance imaging (MRI). For the proposed approach, MR images were captured for tomatoes that were harvested from the field at different maturity stages. Then, for each of the MR images, the mean and histogram features of the voxel intensities in the region of interest (RoI) were calculated. Finally, partial least square discriminant analysis (PLS-DA) algorithm was applied using both the calculated features and maturity classes variables in order to deduce a maturity classification model showing that different maturity stages are embedded in MR images signal intensity.

Also, in Baltazar, Aranda, and González-Aguilar (2008), authors used 128 tomato samples that were harvested and preliminarily sorted with colorimeter choosing only those with roughly breaker color, which represent the ripeness stage where there is a definite break in color from green to tannish-yellow. So, they firstly applied data fusion to nondestructive image of fresh intact tomatoes by assessing both of colorimeter and nondestructive firmness measurements for the samples at the selected testing days using two sensors placed at different points. Then, the measurements data were normalized. Finally, a three-class Bayesian classifier was applied and the results showed that multi-sensorial data fusion is better than single sensor data and considerably reduces the classification error.

Moreover, in Polder, Van der Heijden, and Young (2002), authors proposed an approach based on spectral images analysis to measure the ripeness of tomatoes for automatic sorting. The proposed approach compared hyper-spectral images with standard RGB images for classification of tomato ripeness stages. The proposed classification approach based on individual pixels and includes gray reference in each image for obtaining automatic compensation of different light sources. The proposed approach in Polder et al. (2002) applied the Linear Discriminant Analysis (LDA) algorithm as a classification technique depending on pixels values and proved that spectral images are better than standard RGB images for measuring ripeness stages of tomatoes via offering more discriminating power.

On the other hand, for Oil palm ripeness classification, in Fadilah and Mohamad-Saleh (2014), authors proposed an automated ripeness classification system based on color feature for the problem of oil palm fresh bunch ripeness classification. They used the color of oil palm fresh bunch as a ripeness indicator. The proposed system firstly applied image segmentation using K-means clustering algorithm to separate fruits pixels from spikes ones. Then, it extracted a hue histogram of 100 bins for each image as feature vector via applying two different techniques; namely principal component analysis (PCA) and stepwise discriminant analysis (SDA), for color features reduction purpose. Finally, it applied Artificial Neural Network (ANN) as classifier to classify ripeness into four categories: un-ripe, under-ripe, ripe and over-ripe. Results showed that reducing the color features using stepwise discriminant analysis improved the performance of classification accuracy by more than 10%.

Also, in Bensaeed, Shariff, Mahmud, Shafri, and Alfatni (2014), authors proposed a hyperspectral-based classification system for the purpose of classifying the ripeness of oil palm fresh fruit bunches (FFBs). A dataset of total of 469 fruits for three types of oil palm FFBs (nigrescens, virescens, oleifera) has been collected from MPOB farm area at Kluang, Johor, Malaysia to be classified into three ripeness categories: over-ripe, ripe, and under-ripe. The proposed system firstly scanned oil palm FFBs using a hyperspectral device. Then, a pixel spectral processing step was performed by applying, background removal, followed by pixel discrimination (only reflectance data was analyzed). A Low Pass Filter was applied to data for noise reduction purpose. Finally, it applied artificial neural network (ANN) as a classifier to classify the different wavelength regions on oil palm fruit through pixel-wise processing. This system achieved an accuracy of more than 95% for all three types of oil palm fruits.

Moreover, in Jaffar, Jaafar, Jamil, Low, and Abdullah (2009), authors applied photogrammetric methodology in order to depict a relationship between the color of the palm oil fruits and their ripeness stages, then they have been sorted out physically. That proposed approach was considered as the first automation form of palm oil grading systems. Other previous grading systems faced difficulty of using the average color digital number values at RGB color space for determining ripeness, due to the fusion of palm fruit images with dirt and branches. The proposed approach applied the K-means clustering and segmented the fruit fresh bunches (FFBs) colors in an automated fashion. Then, to differentiate ripe FFBs from unripe fruits, the computed color value to R/G and R/B ratios of the digital number of the segmented images was utilized.

Also, in May and Amaran (2011), authors proposed an assessment approach for ensuring optimum yield of high quality oil in order to overcome subjectivity and inconsistency of manual human grading techniques based on experience. Palm ripeness stages were classified into under-ripe, ripe and over-ripe depending on different color intensity. The developed approach is an automated ripeness assessment using RGB and fuzzy logic feature extraction and classification model to assess the ripeness of oil palm. It depended on color intensity and achieved an accuracy of 88.74%.

Furthermore, in Zhang, Lee, Tippetts, and Lillywhite (2014), authors proposed a method for classifying harvested dates according to their color. After images capturing, a threshold segmentation method was applied to separate fruit from background. From RGB color space, they used only R–G plane, because blue channel does not give effective information for dates grading. For training stage, to generate 2D histogram (one for each maturity class), the co-occurrence of every color in R and G channel in each class was counted. After 2D histogram creation, it was normalized. Then, back projection matrix was generated. For grading stage, After background removal and extraction of R and G values, a back projection step followed by color index analysis were performed, then color grading was computed using some statistics. The proposed method in Zhang et al. (2014) achieved good results in addition to not requiring complicated training process and machine learning algorithms.

Moreover, for Hass Avocados ripeness classification, in Guerrero and Benavides (2014), authors proposed an automated classification system based on Fisher's Linear Discriminant Analysis and K-means algorithms. The RGB color space has been used and the proposed system applied some filters to minimize noises. Then, an image segmentation step was applied, using Fisher's Linear Discriminant Analysis algorithm, to separate Avocado fruit from background. Finally, K-means grouping technique was employed, in order to classify Avocados into very mature avocados from mature and green avocados categories, based on pixels percentage. The proposed system achieved an accuracy of 87.85%.

While for banana ripeness classification, a system based on artificial neural network (ANN) with image processing approach for color recognition has been designed in Paulraj, Hema, Pranesh K., and Siti Sofiah (2009) for identifying the ripeness stage of bananas. The proposed system depended on RGB color components of the captured images of banana. It used four sets of bananas with different sizes and ripeness stages. Each image of the banana was captured in four different positions and the images were captured daily until all bananas turned to be rotten. The proposed research used supervised Neural Network model with utilizing the error back propagation model. It achieved an identification accuracy of 96%.

Also, for watermelon ripeness classification, a system based on artificial neural network (ANN) with image processing approach was designed in Shah Rizam, Farah Yasmin, Ahmad Ihsan, and Shazana (2009). The proposed system measured and determined the ripeness and quality of watermelon based on its colors in YCbCr Color Space. It measured the ripeness by checking textured founded on the skin of watermelon, which was classified into ripeness index from the segmented image depending on the amount of pixels at each region. The proposed research approach achieved an accuracy of 86.51%.

For jatropha ripeness classification, in Syal, Mehta, and Darshni (2013), authors proposed an approach based on image processing techniques and fuzzy logic for fruit sorting and grading. The proposed jatropha fruit grading system based on three basic features: RGB color components, shape, and size of the fruit object. In the proposed system, there are three grades (A, B, C) for jatropha fruit ripeness depending on the selected features values. The selected features are extracted using image processing techniques, then fuzzy system was applied to decide the grade of each fruit. Experimental results showed that using fuzzy logic in the proposed system achieved accurate and very promising classification.

Furthermore, there are many research works regarding some other fruits. In Dadwal and Banga (2012), authors proposed an approach based on color image segmentation and fuzzy logic technique to classify apples into ripe, under-ripe and over-ripe categories. The proposed approach based on RGB color components, where for each fruit four images were captured from different directions. Then, segmentation algorithm was applied to these images for getting area of interest and the mean value for each color component (R, G and B) was calculated. Finally, the fuzzy logic approach was applied to decide the category of apple ripeness depending on mean values of Red, Green and Blue color components.

In Dolaty (2012), authors proposed an approach based on image processing for cherries sorting and grading. The proposed approach depended on the RGB color components of the captured cherry images. It used cherry samples from four different stages that were collected with an interval of 5 days. The sorting system of cherries according to their ripeness used color criteria and the Total Soluble Solids (TTS) in fruit to classify it into the right ripeness stage. In order to minimize the error rate in calculating the average color components, the proposed system achieved 92% accuracy in sorting cherries according to their ripeness stage.

In Damiri and Slamet (2012), authors proposed an approach based on image processing and Artificial Neural Network (ANN) to identify maturity and ripeness of lime. The proposed system depended on area, shape factor, RGB color index, and texture features to identify the ripeness of lime. These features were sent to the ANN using back propagation method for training then to implement classification. The proposed system achieved 100% accuracy in classifying the lime based on their maturity and ripeness.

Also, in one of our previous researches (Elhariri, El-Bendary, Hussein, Hassanien, & Badr, 2014), we proposed an automated ripeness classification system for bell pepper ripeness stages. The used datasets in that research were constructed based on real sample images for tomato at different stages, which were collected from different farms in Minya city, Upper Egypt and annotated by an expert. The proposed system firstly employs preprocessing step by applying image resizing, background removal and converting image from RGB to HSV color space. Then, 1D($16 \times 4 \times 4$) HSV colored histogram and color moments were computed to be used for classification purpose. Principal component analysis (PCA) algorithm was applied as feature extraction in order to generate a feature vector for each image in the dataset. Then, one-against-one Support Vector Machines (SVMs) algorithm was applied for classification of ripeness stages. The proposed system achieved an accuracy of 93.89%.

So, on the whole, many points of research concerning ripeness classification for tomatoes and other fruits/vegetables have been addressed. Most of those approaches used various classifiers, such as ANN, K-means, LDA, and fuzzy logic approaches, and reported the obtained classification accuracy of the implemented classification approach. However none of them highlighted the dependency of the classification approach performance on statistics of the experimented dataset(s). So, to the best of our knowledge, this article is the first research work aims at highlighting the most appropriate classification algorithm considering the dependency of the classification approach performance on statistics of the experimented dataset. Accordingly, in this article we adopted the utilization of principal component analysis (PCA) in addition to Support Vector Machines (SVMs) and Linear Discriminant Analysis (LDA) algorithms for feature extraction and classification, respectively, for tomato ripeness stages evaluation and classification considering the color features. In the proposed approach, we have conducted experiments on a dataset of total 250 real sample images for tomato at different stages that was used for both training and testing datasets with 10-fold cross validation. We figured out that the performance of LDA and SVMs was highly dependent on statistics of the dataset. That is, on datasets with fewer classes (ripeness categories), and many training examples per class, SVMs had an advantage over the LDA classification approach. Therefore, the obtained experimental results showed that using one-against-one (OAO) multi-class SVMs algorithm with linear kernel function outperformed using both one-against-all (OAA) multi-class SVMs algorithm with linear kernel function and LDA classification algorithms.

3. Preliminaries

This section presents a brief idea concerning the core concepts of PCA feature extraction algorithm in addition to SVMs and LDA classification algorithms.

3.1. Principal component analysis (PCA)

One of the most common statistical techniques is principal component analysis (PCA). It is widely used in fields of compression for a dimensionality reduction and image recognition, data representation and features extraction tool as it guarantees better classification (Ada & Kaur, 2013; El-Bendary, Zawbaa, Hassanien, & Snasel, 2011; Suganthy & Ramamoorthy, 2012; Xiao, 2010). Basically, It reduces the dimensionality by avoiding redundant information, and transforming samples features space to features sub-space (smaller space which contains all independent variables which are needed to describe the data) via disposing all ineffectual minor components. So, it's very important to perform many preprocessing steps in order to utilize the PCA method for feature extraction. Steps of PCA algorithm are shown in Algorithm 1.

Algorithm 1. Principal component analysis (PCA) algorithm

1: Calculate the sample mean $\bar{\mu}$

$$\bar{\mu} = \frac{\sum_{i=1}^{n} X_i}{n}$$

2: Subtract sample mean from each observation X_i

$$\bar{Z}_i = X_i - \bar{\mu}$$

3: Calculate the covariance matrix C

$$C = \sum_{i=1}^{n} Z_i \cdot Z_i^t$$

4: Calculate the eigenvectors and eigenvalues of the covariance matrix C

5: Rearrange the eigenvectors and eigenvalues and select a subset as basis vectors

6: Project the data

3.2. Color features

A widely used feature in image retrieval and image classification problems is the color, which is as well an important feature for image representation (Shahbahrami & Juurlink, 2008). In this research two color descriptors are used; namely color moments and color histogram.

3.2.1. Color moments

The first three color moments, which are mean, standard deviation, and skewness (Singh & Hemachandran, 2012; Soman, Ghorpade, Sonone, & Chavan, 2012), have been proved to be efficient and effective way for representing color distribution in any image. Mean, standard deviation, and skewness for a colored image of size $N \times M$ pixels are defined by Eqs. (1)–(3).

$$\bar{x}_i = \frac{\sum_{j=1}^{M \cdot N} x_{ij}}{M \cdot N} \tag{1}$$

$$\partial_i = \sqrt{\frac{1}{M \cdot N} \sum_{j=1}^{M.N} (x_{ij} - \bar{x}_i)^2} \tag{2}$$

$$S_i = \sqrt[3]{\frac{1}{M \cdot N} \sum_{j=1}^{M.N} (x_{ij} - \bar{x}_i)^3} \tag{3}$$

where, x_{ij} is the value of image pixel j of color channel i (e.g. RGB, HSV and etc.), $\bar{x}_i$ is the mean for each channel i = (H, S and V), ∂_i is the standard deviation, and S_i is the skewness for each channel (Singh & Hemachandran, 2012; Soman et al., 2012). HSV channels can be computed for RGB channels using Eqs. (4)–(6), where R, G and B are color component of RGB color space (Meskaldji, Boucherkha, & Chikhi, 2009).

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

$$S = 1 - \frac{3[\min(R,G,B)]}{(R+G+B)} \tag{5}$$

$$V = \left(\frac{R+G+B}{3}\right) \tag{6}$$

3.2.2. Color histogram

Color histogram is a color descriptor that shows representation of the distribution of colors in an image. It represents the number of pixels that have colors in each range of colors (El-Bendary et al., 2011). Color histogram can be calculated for many color spaces (e.g. RGB, HSV, etc.). It is often used with 3-dimensional spaces like as RGB and HSV color spaces. color histogram is invariant with rotation, translation, and scale (Meskaldji et al., 2009).

3.3. Support Vector Machines (SVMs)

3.3.1. Basics

The Support Vector Machines (SVMs) is a Machine Learning (ML) algorithm that is used for classification and regression of high dimensional datasets with great results (Wu & Zhou, 2006; Zawbaa, El-Bendary, Abraham, & Hassanien, 2011, 2011). SVMs algorithm solves the classification problem via trying to find an optimal separating hyperplane between classes. it depends on the training cases which are placed on the edge of class descriptor this is called support vectors, any other cases are discarded (El Hariri et al., 2014; Suralkar, Karode, & Pawade, 2012; Tzotsos & Argialas, 2006; Zhang, Xie, & Cheng, 2010).

SVMs algorithm seeks to maximize the margin around a hyperplane that separates a positive class from a negative class (El Hariri et al., 2014; Wu & Zhou, 2006; Zawbaa et al., 2011, Zawbaa, El-Bendary, Hassanien, & Kim, 2011). Given a training dataset with n samples $(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)$, where x_i is a feature vector in a v-dimensional feature space and with labels $y_i \in -1, 1$ belonging to either of two linearly separable classes C_1 and C_2. Geometrically, the SVMs modeling algorithm finds an optimal hyperplane with the maximal margin to separate two classes, which requires to solve the optimization problem, as shown in Eqs. (7) and (8).

$$maximize \sum_{i=1}^{n} \alpha_i - \frac{1}{2} \sum_{i,j=1}^{n} \alpha_i \alpha_j y_i y_j . K(x_i, x_j) \tag{7}$$

$$Subject\text{-} \ to : \sum_{i=1}^{n} \alpha_i y_i, \quad 0 \leqslant \alpha_i \leqslant C \tag{8}$$

where, α_i is the weight assigned to the training sample x_i. If $\alpha_i > 0$, x_i is called a support vector. C is a regulation parameter used to trade-off the training accuracy and the model complexity so that a superior generalization capability can be achieved. K is a kernel function, which is used to measure the similarity between two samples. Different choices of kernel functions have been proposed and extensively used in the past and the most popular are the gaussian radial basis function (RBF), polynomial of a given degree, linear, and multi layer perceptron MLP. These kernels are in general used, independently of the problem, for both discrete and continuous data.

3.3.2. N-class Support Vector Machine Approaches

There are many problem, which are N-class classification problem not a binary one and it is well known that SVMs algorithm is a binary class classification methods SVMs algorithm is a binary class classification method and our problem is a N-class classification problem. Therefore, to use SVMs for N-class classification problem like our problem in this article, there are two different approaches can be applied in order to do that; namely *one-against-all (OAA)* and *one-against-one (OAO)* approaches (Anthony, Gregg, & Tshilidzi, 2007; Liu & Zheng, 2005).

The first approach, one-against-all (OAA), worked according to Algorithm 2.

Algorithm 2. One-against-all (OAA)

1: Construct N binary SVMs.
2: Each SVM separates one class from the rest classes.
3: Train the ith SVM with all training samples of the ith class with positive labels, and training samples of other classes with negative labels.

In the second approach, one-against-one (OAO), a SVM classifier was created for each pair of classes (for $N - class$ problem) resulting in $N(N-1)/2$ classifiers. The OAO approach worked according to Algorithm 3.

One-against-one (OAO)

1: Create $N(N-1)/2$ binary SVMs
2: Train $N(N-1)/2$ binary SVMs as follow $(1,2), (1,3), \ldots, (1,k), (2,3), (2,4), \ldots, (k-1,k)$.
3: Select the one with the largest vote (The class label that occurs the most).

3.4. Linear Discriminant Analysis (LDA)

Linear Discriminant Analysis (LDA) is a commonly used technique for data classification and dimensionality reduction. Linear Discriminant Analysis easily handles the case where the within class frequencies are unequal and their performances has been examined on randomly generated test data. it's basic idea is to find a linear transformation that best discriminate among classes, then classification can be performed in transformed space based on some metrics(Euclidean distance) (Li, Zhu, & Ogihara, 2006).

Given data $y_i,\ x_i,\ i=_1^n$, where $y_i 1, 2, \ldots, K$ is the class label, k is the number of classes and x_i is a vector of features or predictors, we seek to find the best direction in the predictor space in which the classes are separated as much as possible.

Mathematically, LDA implementation is carried out via scatter matrix analysis. For all samples of all classes, we define two measures (Fisher, 1936; Li et al., 2006):

- Within-class scatter matrix. It is defined by:

$$S_w = \sum_{j=1}^{K} \sum_{i=1}^{N_j} (x_i^j - \mu_j)(x_i^j - \mu_j)^T \tag{9}$$

  where x_i^j is the ith sample of class j, μ_j is the mean of class j, K is the number of classes, and N_j is the number of samples in class j; and
- Between-class scatter matrix. It defined by:

$$S_b = \sum_{j=1}^{K} (\mu_j - \mu)(\mu_j - \mu)^T \tag{10}$$

  where μ represents the mean of all classes. This method maximizes the ratio of between-class measure to the within-class measure in any particular data set thereby guaranteeing maximal separability. The maximization of $\frac{\delta|S_b|}{\delta|S_w|}$.

4. The proposed system

In particular, the proposed system is capable of classifying different ripeness stages of tomato. The proposed system exploits color features to identify different ripeness stages of tomato in multi-class scenario. Proposed system general structure is described at Fig. 1. It consists of three phases; namely *pre-processing*, *feature extraction*, and *classification* phases.

4.1. Pre-processing phase

During pre-processing phase, in order to reduce images color index, the proposed system resizes images to 250 × 250 pixels. Then the background of each image is removed using background subtraction technique with some morphological operations. Fig. 2. shows an example of background removal algorithm results. Also, each image is converted from RGB to HSV color space, as it is widely used in the field of color vision and close to the categories of human color perception (Yu, Li, Zhang, & Feng, 2002).

4.2. Feature extraction phase

As previously stated, the most important characteristic to assess tomato ripeness is its surface color, so this system uses HSV color histogram and color moments for ripeness stages classification. For feature extraction phase, PCA algorithm is applied as features extraction technique in order to generate a feature vector for each image in the dataset.

The proposed system transforms the input space into subspaces for dimensionality reduction. After completing the previous 1D 16 × 4 × 4 HSV histogram, 16 levels for hue and 4 levels for each of saturation and value are resulted. In addition, nine color moments, three for each channel (H, S and V channels) (mean, standard deviation, and skewness), were computed. Then, a feature vector was formed as a combination of HSV 1D histogram and the nine color moments.

4.3. Classification phase

Finally, for classification phase, the proposed system applied SVMs and LDA algorithms for classification of ripeness stages. The inputs are training dataset feature vectors and their corresponding classes, whereas the outputs are the ripeness stage of each image in the testing dataset.

In this phase, the classification system, previously proposed in Elhariri et al. (2014), has been utilized along with the one-against-one (OAO) approach with 10-fold cross validation for

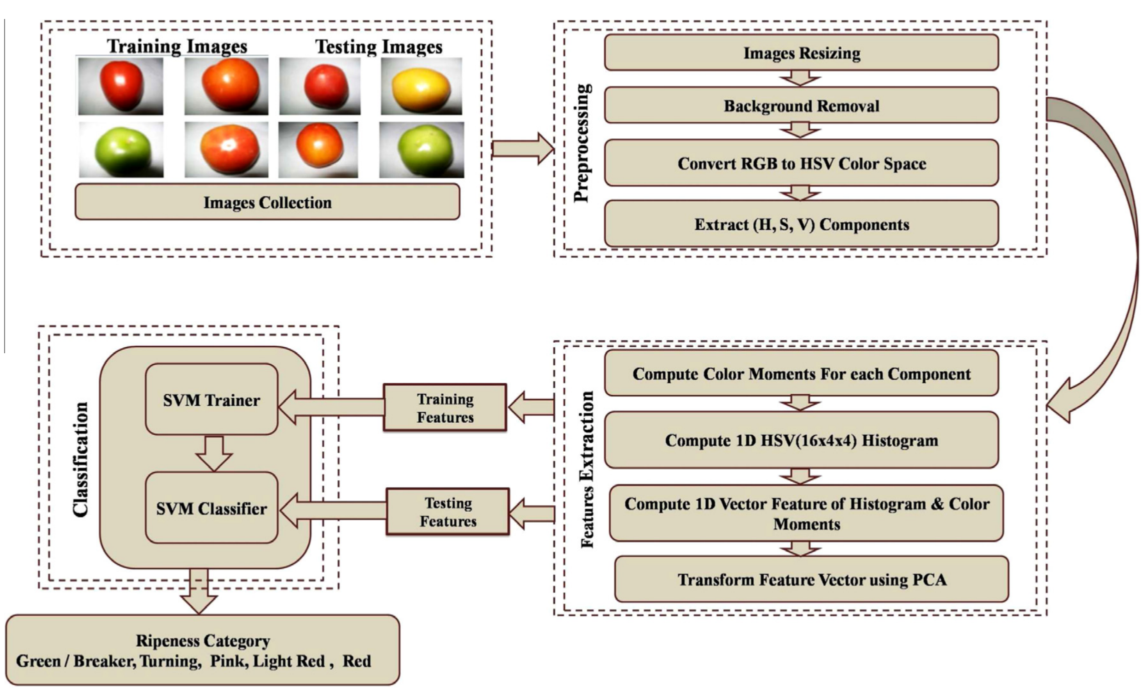


Fig. 1. Architecture of the proposed ripeness classification approach.

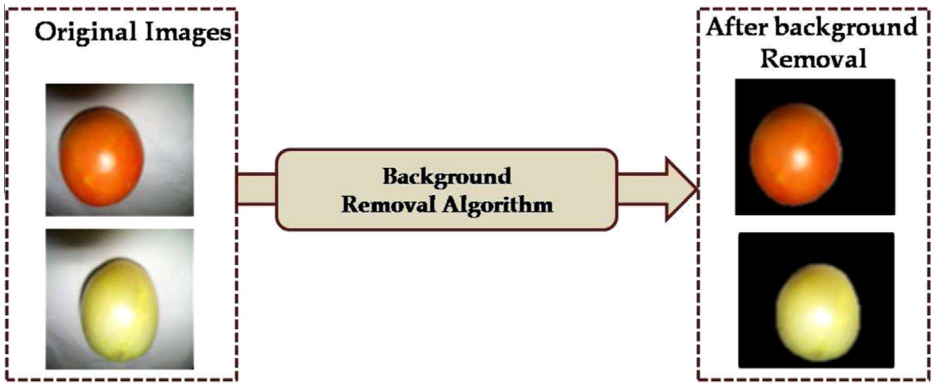


Fig. 2. Sample images before and after applying background subtraction algorithm.

multi-class SVMs problems. In addition to, another classification algorithm LDA using quadratic discriminant function with 10-fold cross validation.

5. Experimental results

Simulation experiments in this article are done on a PC with Intel Core i7 Q720@1.60 GHZ CPU and 6 GB memory. The proposed approach is designed with Matlab running on Windows 7. The datasets used for experiments were constructed based on real sample images for tomato at different ripeness stages, which were collected from different farms in Minya city. The collected datasets contained colored JPEG images of resolution 3664×2748 pixels that were captured using Kodak C1013 digital camera of 10.3 megapixels resolution. The dataset is of total 250 images were used for both training and testing datasets with 10-fold cross-validation. Training dataset is divided into 5 classes representing the different stages of tomato ripeness as shown in Fig. 3. The classes are *Green* & *Breaker*, *Turning*, *Pink*, *Light red*, and *Red* stages (U.S. Dept. Agric./AMS, 1991). For *Green* & *Breaker stage*, green represents the ripeness stage where fruit surface is completely green, however breaker represents the ripeness stage where there is a definite break in color from green to tannish-yellow, pink or red on not more than 10% of the surface. For *Turning stage*, 10–30% of the surface is not green. For *Pink stage*, 30–60% of the surface is not green. For *Light red stage*, 60–90% of the surface is not green. Finally, for *Red stage*, more than 90% of the surface is not green. Some samples of both training and testing datasets are shown in Fig. 4.

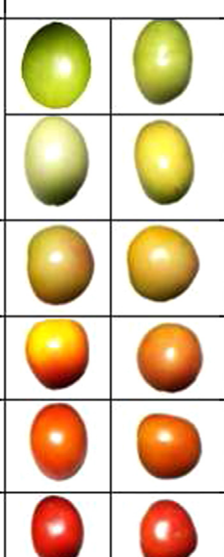
The proposed approach has been implemented considering three scenarios; namely

- Scenario I: one-against-one multi-class SVMs system using 10-fold cross. validation
- Scenario II: one-against-All multi-class SVMs system using 10-fold cross. validation
- Scenario III: LDA system using 10-fold cross validation.

5.1. Scenario I: one-against-one multi-class SVMs system using 10-fold cross validation

The first scenario presents implementing one-against-one multi-class SVMs system using10-fold cross-validation and a total of 250 images for both of training and testing datasets. The used features for classification are a combination of color HSV histogram and color moments and PCA algorithm was applied for features extraction. Moreover, SVMs algorithm was employed with different kernel functions that are: *Linear* kernel, *radial basis function (RBF)* kernel, and *Multi-Layer Perceptron (MLP)* kernel and *Polynomial* kernel (Boolchandani & Sahula, 2011; Vanschoenwinkel & Manderick, 2005) for ripeness stage classification.

Fig. 5 shows classification accuracy obtained via applying each kernel function. Figs. 6–9 show 5-class receiver operating



Stage Name	Images	Description
Green & Breaker		Green: Fruit surface is completely green. Breaker: There is a definite break in color from green to tannish-yellow, pink or red on not more than 10% of the surface.
Turning		10% to 30% of the surface is not green
Pink		30% to 60% of the surface is not green
Light Red		60% to 90% of the surface is not green
Red		More than 90% of the surface is not green

Fig. 3. Examples of tomato ripeness stages.

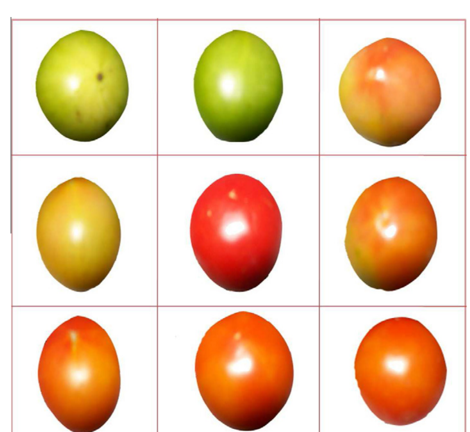


Fig. 4. Examples of training and testing samples.

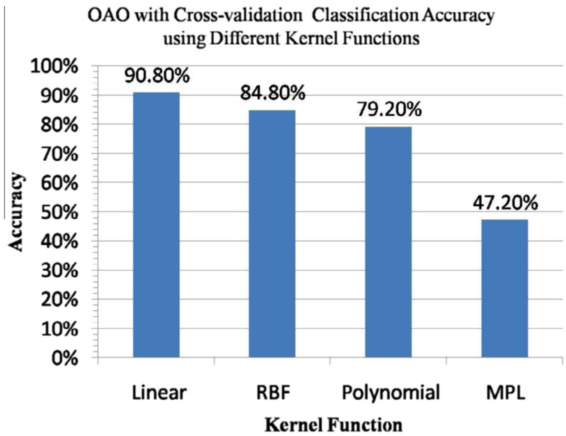


Fig. 5. Results for different SVM kernel functions using one-against-one multi-class approach and 10-fold cross-validation.

characteristic (ROC) curve and area under curve (AUC) for the best feature resulted from different kernel functions using one-against-one multi-class SVMs approach with 10-fold cross-validation and total of 250 images (used for both of training and testing). The ROC curve separates each class from other classes. From Fig. 6, showing the ROC curve for the best feature using linear kernel function for OAO multi-class SVMs with cross-validation, the applied approach separated each class from each one of the rest classes by AUCs shown at Table 1.

From Fig. 7, showing ROC curve for the best feature using MLP kernel function for OAO multi-class SVMs with cross-validation, the applied approach separated each class from each one of the rest classes by AUCs shown at Table 2.

From Fig. 8, showing ROC curve for the best feature using RBF kernel function for OAO multi-class SVMs with cross-validation, the applied approach separated each class from each one of the rest classes by AUCs shown at Table 3.

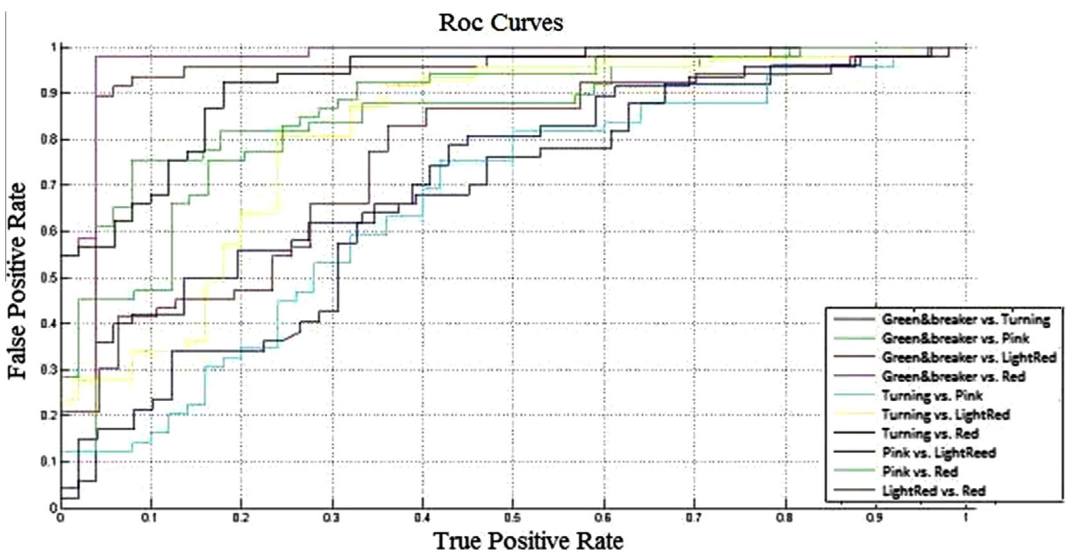


Fig. 6. ROC curve for the best feature using linear kernel function (OAO multi-class SVMs with cross-validation), AUC = 0.8219.

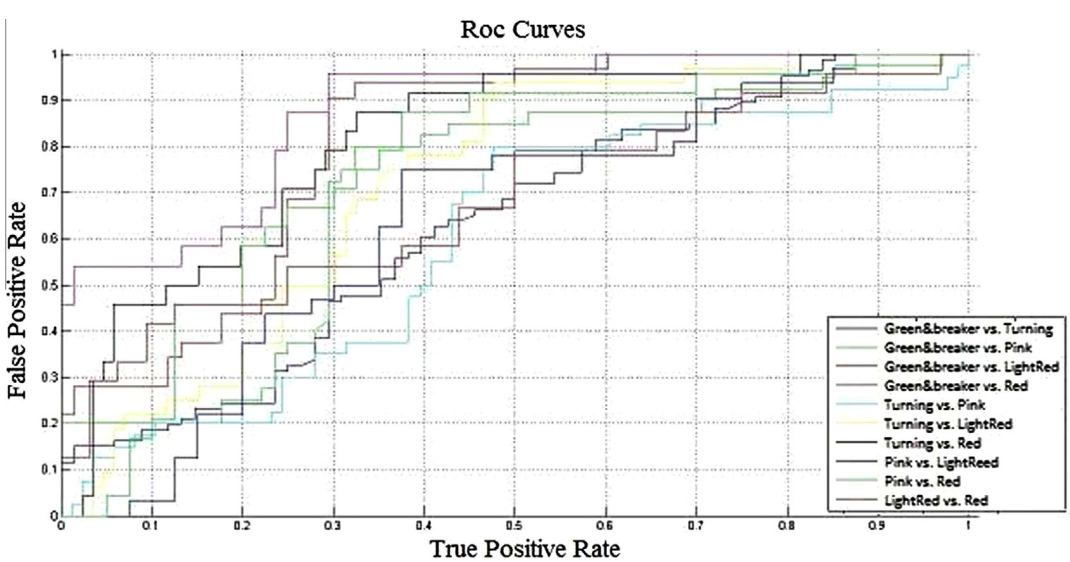


Fig. 7. ROC curve for the best feature using MLP kernel function (OAO multi-class SVMs with cross-validation), AUC = 0.7217.

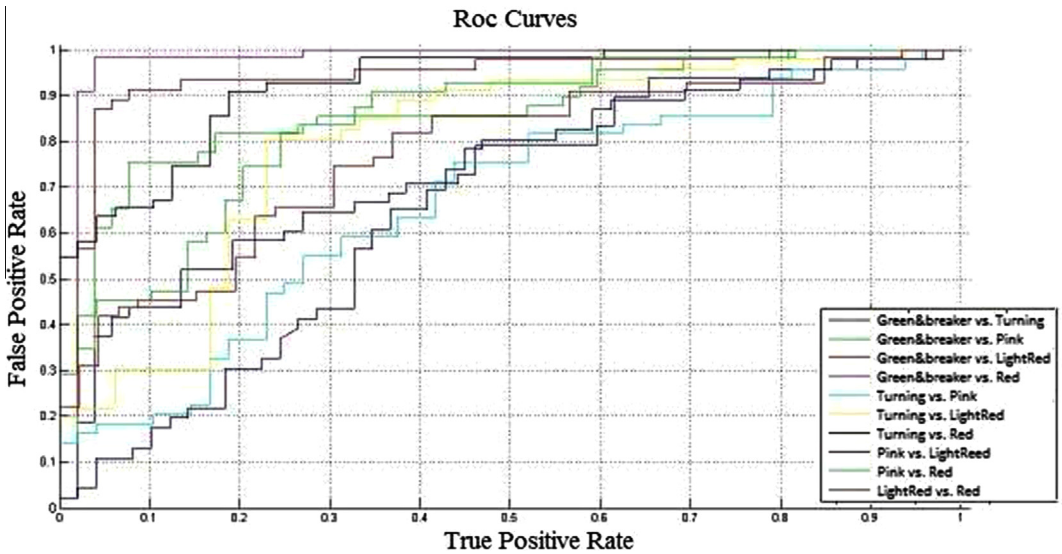


Fig. 8. ROC curve for the best feature using RBF kernel function (OAO multi-class SVMs with cross-validation), AUC = 0.8191.

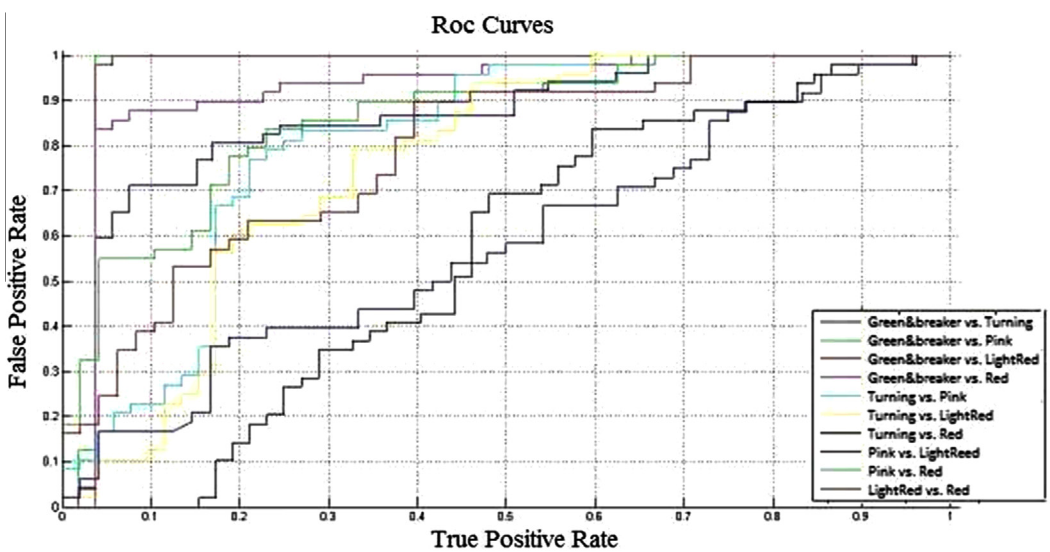


Fig. 9. ROC curve for the best feature using Polynomial kernel function (OAO multi-class SVMs with cross-validation), AUC = 0.8070.

Table 1
AUCs of OAO multiclass-SVMs using Linear kernel function and 10-fold cross validation.

	Green & Breaker	Turning	Pink	Light red	Red
Green & Breaker	–	0.7180	0.8615	0.9424	0.9785
Turning	0.7180	–	0.6649	0.8123	0.9268
Pink	0.8615	0.6649	–	0.6880	0.8652
Light red	0.9424	0.8123	0.6880	–	0.7615
Red	0.9785	0.9268	0.8652	0.7615	–

Table 2
AUCs of OAO multiclass-SVMs using MLP kernel function and 10-fold cross validation.

	Green & Breaker	Turning	Pink	Light red	Red
Green & Breaker	–	0.6299	0.6998	0.8084	0.8775
Turning	0.6299	–	0.5985	0.7197	0.8115
Pink	0.6998	0.5985	–	0.6391	0.7521
Light red	0.8084	0.7197	0.6391	–	0.6810
Red	0.8775	0.8115	0.7521	0.6810	–

Table 3
AUCs of OAO multiclass-SVMs using RBF kernel function and 10-fold cross validation.

	Green & Breaker	Turning	Pink	Light red	Red
Green & Breaker	–	0.7400	0.8646	0.9411	0.9853
Turning	0.7400	–	0.6658	0.7957	0.9231
Pink	0.8646	0.6658	–	0.6555	0.8460
Light red	0.9411	0.7957	0.6555	–	0.7735
Red	0.9853	0.9231	0.8460	0.7735	–

Table 4
AUCs of OAO multiclass-SVMs using Polynomial kernel function & 10-fold cross validation.

	Green & Breaker	Turning	Pink	Light red	Red
Green & Breaker	–	0.8621	0.9646	0.9627	0.9234
Turning	0.8621	–	0.8093	0.7620	0.5581
Pink	0.9646	0.8093	–	0.5801	0.8601
Light red	0.9627	0.7620	0.5801	–	0.7874
Red	0.9234	0.5581	0.8601	0.7874	–

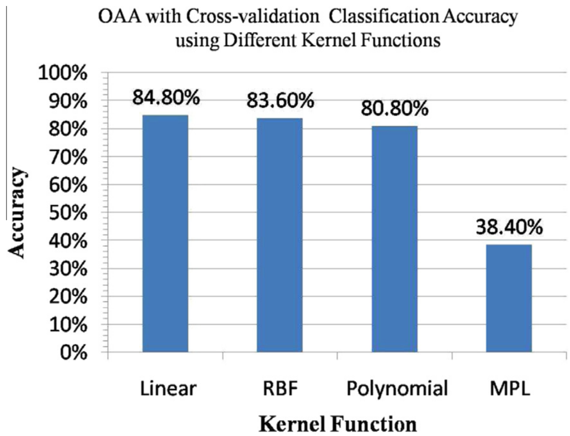


Fig. 10. Results for different kernel functions using one-against-All multi-class approach and 10-fold cross-validation.

From Fig. 9, showing ROC curve for the best feature using Polynomial kernel function for OAO multi-class SVMs with cross-validation, the applied approach separated each class from each one of the rest classes by AUCs shown at Table 4.

5.2. Scenario II: one-against-All multi-class SVMs system using 10-fold cross validation

In the second scenario, the proposed One-against-All multi-class SVMs approach was also tested using the previously stated specifications of one-against-one multi-class SVMs approach for ripeness stages classification. Fig. 10 shows classification accuracy obtained via applying each kernel function.

Figs. 11–14 show 5-class receiver operating characteristic (ROC) curve and area under curve (AUC) for the best feature for different kernel function using one-against-all multi-class SVMs approach with 10-fold cross-validation and total of 250 images (used for both of training and testing).

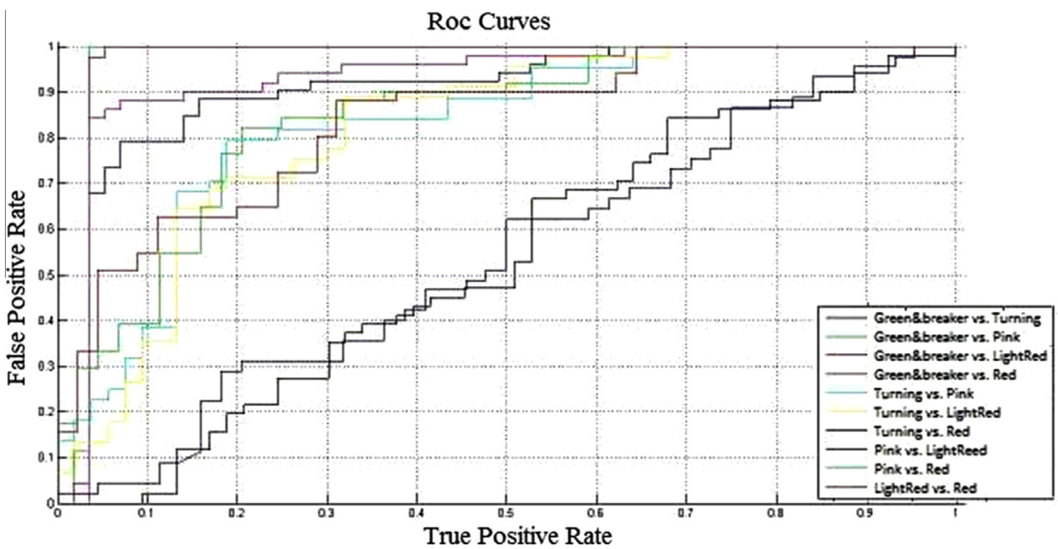


Fig. 11. ROC curve for the best feature using linear kernel function (OAA multi-class SVMs with cross-validation), AUC = 0.8154.

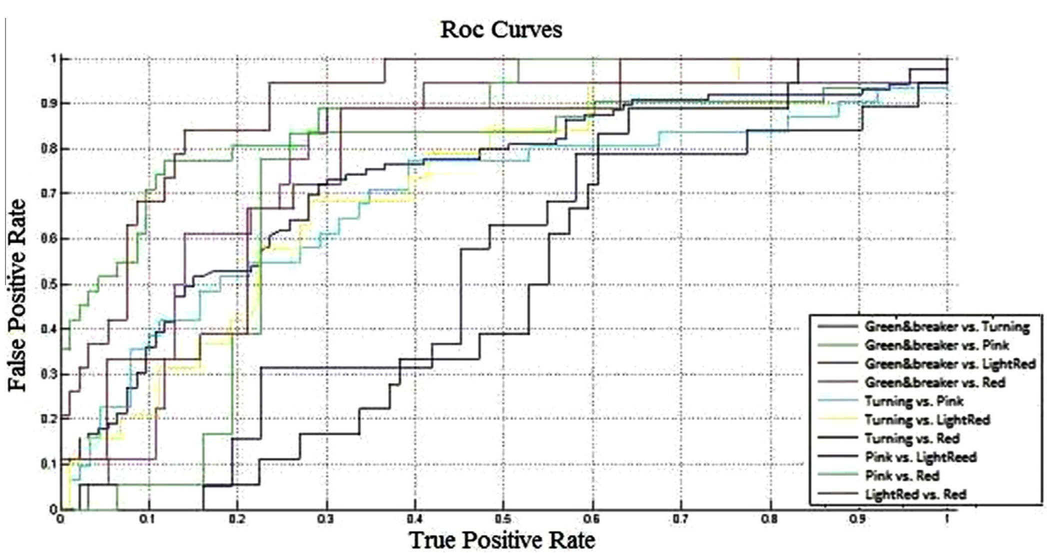


Fig. 12. ROC curve for the best feature using MLP kernel function (OAA multi-class SVMs with cross-validation), AUC = 0.7232.

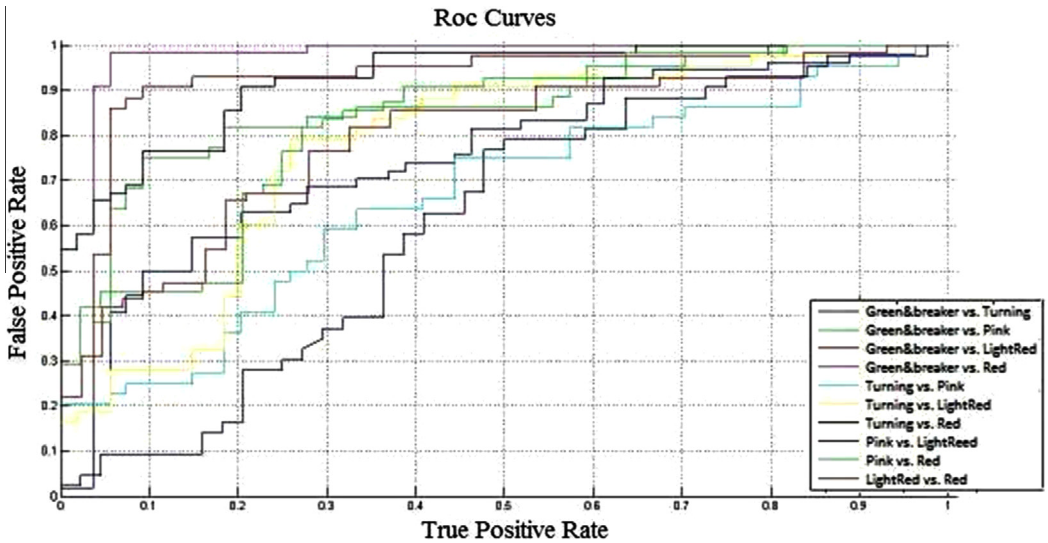


Fig. 13. ROC curve for the best feature using RBF kernel function (OAA multi-class SVMs with cross-validation), AUC = 0.8100.

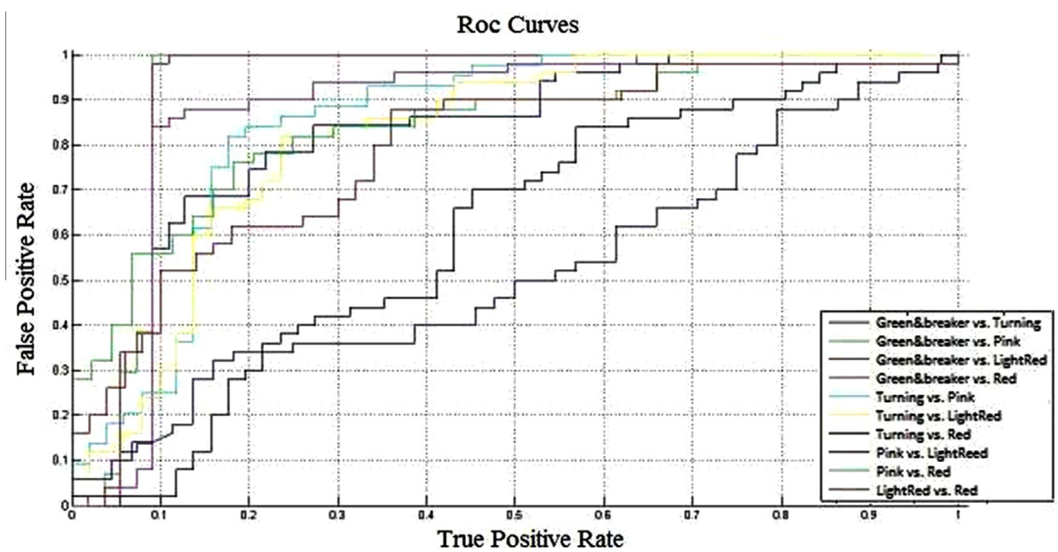


Fig. 14. ROC curve for the best feature using Polynomial kernel function (OAA multi-class SVMs with Cross-validation), AUC = 0.7967.

Table 5
AUCs of OAA multiclass-SVMs using Linear kernel function & 10-fold cross validation.

	Green & Breaker	Turning	Pink	Light red	Red
Green & Breaker	–	0.9043	0.9669	0.9653	0.9284
Turning	0.9043	–	0.8259	0.8176	0.5294
Pink	0.9669	0.8259	–	0.5361	0.8436
Light red	0.9653	0.8176	0.5361	–	0.8362
Red	0.9284	0.5294	0.8436	0.8362	–

Table 6
AUCs of OAA multiclass-SVMs using MLP kernel function & 10-fold cross validation.

	Green & Breaker	Turning	Pink	Light red	Red
Green & Breaker	–	0.7334	0.8269	0.9089	0.7826
Turning	0.7334	–	0.6901	0.7289	0.5050
Pink	0.8269	0.6901	–	0.5110	0.7616
Light red	0.9089	0.7289	0.5110	–	0.7836
Red	0.7826	0.5050	0.7616	0.7836	–

Table 7
AUCs of OAA multiclass-SVMs using RBF kernel function & 10-fold cross validation.

	Green & Breaker	Turning	Pink	Light red	Red
Green & Breaker	–	0.7555	0.8565	0.9242	0.9781
Turning	0.7555	–	0.6641	0.7730	0.9242
Pink	0.8565	0.6641	–	0.6091	0.8244
Light red	0.9242	0.7730	0.6091	–	0.7911
Red	0.9781	0.9242	0.8244	0.7911	–

Table 8
AUCs of OAA multiclass-SVMs using Polynomial kernel function & 10-fold cross validation.

	Green & Breaker	Turning	Pink	Light red	Red
Green & Breaker	–	0.8178	0.9227	0.9225	0.8771
Turning	0.8178	–	0.8480	0.8196	0.6075
Pink	0.9227	0.8480	–	0.5230	0.8368
Light red	0.9225	0.8196	0.5230	–	0.7924
Red	0.8771	0.6075	0.8368	0.7924	–

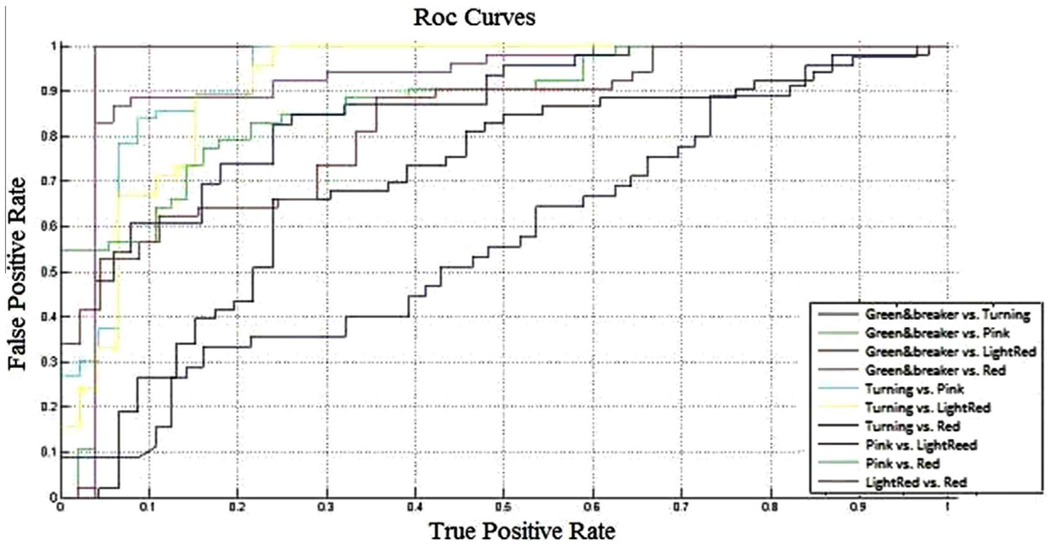


Fig. 15. ROC curve for the best feature using LDA classifier with cross-validation, AUC = 0.8531.

From Fig. 11, showing ROC curve for the best feature using linear kernel function for OAA multi-class SVMs with cross-validation, the applied approach separated each class from each one of the rest classes by AUCs shown at Table 5.

From Fig. 12, showing ROC curve for the best feature using MLP kernel function for OAA multi-class SVMs with cross-validation, the applied approach separated each class from each one of the rest classes by AUCs shown at Table 6.

From Fig. 13, showing ROC curve for the best feature using RBF kernel function for OAA multi-class SVMs with cross-validation, the applied approach separated each class from each one of the rest classes by AUCs shown at Table 7.

From Fig. 14, showing ROC curve for the best feature using Polynomial kernel function for OAA multi-class SVMs with cross-validation, the applied approach separated each class from each one of the rest classes by AUCs shown at Table 8.

5.3. Scenario III: LDA system with 10-fold cross validation

In the third scenario, the proposed LDA multi-class system was also tested using the previously stated specifications for ripeness stages classification. Classification accuracy obtained via applying LDA classifier is 84%.

Fig. 15, showing ROC curve for the best feature using LDA system with cross-validation, the applied approach separated each class from each one of the rest classes by AUCs shown at Table 9.

From the previously depicted experimental results, we found out that the one-against-one multi-class SVMs approach is better than the One-against-All multi-class SVMs and LDA approaches, when applied for ripeness stage classification. Fig. 16 shows a comparison between accuracies obtained by each of the three approaches.

The accuracy measure is calculated as shown in Eq. (11).

Table 9
AUCs of LDA classifier & 10-fold cross validation.

	Green & Breaker	Turning	Pink	Light red	Red
Green & Breaker	–	0.8478	0.9621	0.9604	0.9196
Turning	0.8478	–	0.9340	0.9159	0.7055
Pink	0.9621	0.9340	–	0.5720	0.8807
Light red	0.9604	0.9159	0.5720	–	0.8327
Red	0.9196	0.7055	0.8807	0.8327	–

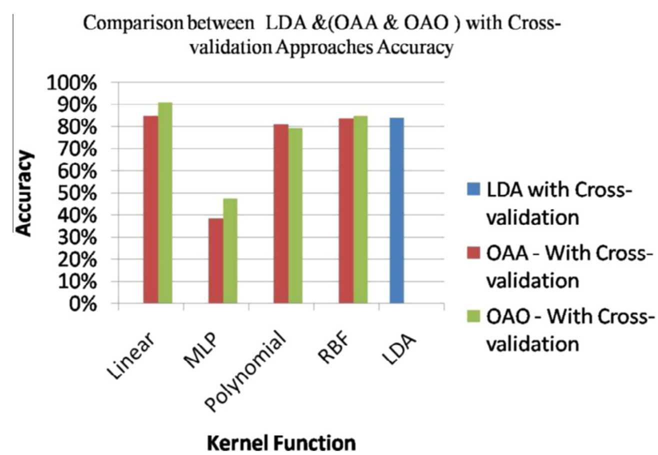


Fig. 16. Comparison between the classification accuracy of OAA, OAO multi-class SVMs and LDA approaches.

$$\text{Accuracy} = \frac{\text{Number of correctly classified images}}{\text{Total number of testing images}} * 100 \qquad (11)$$

6. Conclusions and future work

In recent years, the use of computer vision and machine learning techniques for evaluating/estimating fruits ripeness has become widely used. From another perspective, tomato maturity has been related to quantifiable parameters, that reflect the biochemical changes during ripening. Color is used as a major method in determining maturity of tomato. However, skin color of tomato varies from cultivar to another cultivar even at the same maturity stage (El Hariri et al., 2014; Molyneux, Lister, & Savage, 2004; Zhang & McCarthy, 2012). During ripening, tomatoes go through a series of highly ordered physiological and biochemical changes, such as chlorophyll degradation and increased activity of cell wall-degrading enzymes, bring on changes in color, firmness, and development of aromas and flavors (El Hariri et al., 2014; Prasanna, Prabha, & Tharanathan, 2007). For tomatoes, ripeness issue is often handled via classifying harvested produce according to discrete color classes going from immature green to mature red, as stated in some recent researches that have classified tomatoes in different maturity stages based on measurements of color (Aranda-Sanchez, Baltazar, & González-Aguilar, 2009; El Hariri et al., 2014; Hahn, 2002). Different tomato products have distinct requirements for maturity to achieve quality standards; hence, tomato maturity is one of the most important factors associated with the quality of processed tomato products.

Accordingly, automation of the ripeness stage classification process for crops is of a great use for different domains; such as agriculture, via developing automatic harvest systems and saving crops from damages caused by environmental changes and ensuring optimum yield of high quality products. Also, for industry, through developing automatic sorting systems for checking the quality of fruits to increase customer satisfaction level. Moreover, for trading, identifying physiological and harvest maturity of agricultural crops correctly, will ensure timely harvest for exporting and to avoid cutting of either under-ripe and over-ripe agricultural crops (El Hariri et al., 2014, 2014; May & Amaran, 2011).

In conclusion, the approach proposed in this article has one main research motivation that is providing an automated multi-class classification approach for tomato ripeness measurement and evaluation via investigating and classifying the different maturity/ripeness stages based on the color features. An essential finding is that the performance of classification algorithms was highly relative to statistics of the experimented datasets. That is, on datasets with fewer classes (ripeness categories), and many training examples per class, SVMs had an advantage over other classification approaches. Many points of research concerning ripeness classification for tomatoes and other fruits/vegetables have been addressed by other researchers; however none of those classification approaches addressed the dependency of the classification approach performance on statistics of the experimented dataset(s).

The proposed system has three main phases; pre-processing implemented by applying resizing, background removal, and extracting color components for each image, PCA based feature extraction applied to each pre-processed image in order to obtain HSV histogram and color moments feature vectors, and finally, SVMs and LDA models are generated for ripeness stage classification. The proposed approach has been implemented considering two scenarios via applying one-against-one multi-class SVMs system, One-against-All multi-class SVMs system and LDA system using 10-fold cross-validation. Based on the obtained the experimental results, the highest ripeness classification accuracies of 90.80% and 84.80% have been achieved by the first scenario and the second scenarios, respectively, using linear kernel function and 84% using third scenario. Thus, it can be concluded that the ripeness classification accuracy obtained by the OAO multi-class SVMs approach is better than ripeness classification accuracy obtained by the OAA multi-class SVMs and LDA approaches.

On the other hand, the limitations we faced in this research are the dataset size that's needed to be larger, as the accuracy of SVMs increases by increasing the number of images per training class, and accordingly a maximum accuracy of 90.2% has been achieved using our current experimented dataset.

For future research, variety of challenges and research directions could be considered. Some general research directions are to consider applying the approach proposed in this article to different fields of objects classification other than crops ripeness stages classification. For example, in industry, it could be applied for achieving higher accuracy for objects sorting systems or checking the quality of objects to increase customer satisfaction level. Moreover, for trading, the proposed approach can be of a great use for detecting and choosing the best ripeness stage according to production needs such as exporting, manufacturing, ... etc. Another open problem is to apply the proposed approach on different crops, other than tomatoes, in order to automate the whole process of harvesting and detect damages to save crops.

From the perspective of the utilized algorithms in the proposed approach, a number of future research could be achieved via classifying different objects or crops by involving other features (texture, shape, size, ... etc.) according to the classified objects nature. Moreover, other Machine Learning approaches could be employed in order to address the advantages and limitations of applying each of them. Another direction of research is to use nondestructive/non-invasive detection technologies of food quality/maturity such as hyperspectral imaging systems, colorimetric, Near Infrared Spectroscopy, and non-invasive smart sensing technologies.

References

Ada & Kaur, R. (2013). Feature extraction and principal component analysis for lung cancer detection in ct scan images. *International Journal of Advanced Research in Computer Science and Software Engineering, 3*(3), 187–190.

Anthony, G., Gregg, H., Tshilidzi, M., 2007. Image classification using svms: One-against-one vs one-against-all. arXiv preprint arXiv:0711.2914.

Aranda-Sanchez, J. I., Baltazar, A., & González-Aguilar, G. (2009). Implementation of a bayesian classifier using repeated measurements for discrimination of tomato fruit ripening stages. *Journal of Biosystems Engineering, 102*(3), 274–284.

Auria, L., Moro, R. A. 2008. Support vector machines (svm) as a technique for solvency analysis, Technical report, Discussion papers: German Institute for Economic Research (Deutsches Institut fr Wirtschaftsforschung, Berlin).

Baltazar, A., Aranda, J. I., & González-Aguilar, G. (2008). Bayesian classification of ripening stages of tomato fruit using acoustic impact and colorimeter sensor data. *Computers and Electronics in Agriculture, 60*(2), 113–121.

Bensaeed, O., Shariff, A., Mahmud, A., Shafri, H., & Alfatni, M. (2014). *Oil palm fruit grading using a hyperspectral device and machine learning algorithm. IOP conference series: Earth and environmental science* (Vol. 20). IOP Publishing [p. 012017].

Boolchandani, D., & Sahula, V. (2011). Exploring efficient kernel functions for support vector machine based feasibility models for analog circuits. *International Journal of Design, Analysis & Tools for Circuits and Systems, 1*(1).

Brezmes, J., Llobet, E., Vilanova, X., Saiz, G., & Correig, X. (2000). Fruit ripeness monitoring using an electronic nose. *Sensors and Actuators B: Chemical, 69*(3), 223–229.

Camelo, A. F. L. (2004). Manual for the preparation and sale of fruits and vegetables: From field to market. *Food & Agriculture Organization, 151*.

Coates, L., & Johnson, G. (1997). Postharvest diseases of fruit and vegetables. *Journal of Plant Pathogens and Plant Diseases*, 533–548.

Cowan, A. K., Cripps, R. F., Richings, E. W., & Taylor, N. J. (2001). Fruit size: Towards an understanding of the metabolic control of fruit growth using avocado as a model system. *Physiologia Plantarum, 111*(2), 127–136.

Dadwal, M., & Banga, V. (2012). Estimate ripeness level of fruits using rgb color space and fuzzy logic technique. *International Journal of Engineering and Advanced Technology, 2*(1), 225–229.

Damiri, D. J., & Slamet, C. (2012). Application of image processing and artificial neural networks to identify ripeness and maturity of the lime (citrus medica). *International Journal of Basic and Applied Science, 1*(2), 171–179.

Dolaty, H. (2012). Sorting and grading of cherries on the basis of ripeness, size and defects by using image processing techniques. *International Journal of Agriculture and Crop Sciences (IJACS), 4*(16), 1144–1149.

El-Bendary, N., Zawbaa, H. M., Hassanien, A. E., & Snasel, V. (2011). Pca-based home videos annotation system. *International Journal of Reasoning-based Intelligent Systems, 3*(2), 71–79.

El Hariri, E., El-Bendary, N., Hassanien, A. E., & Badr, A. (2014). *Automated ripeness assessment system of tomatoes using pca and svm techniques*. In M. Sarfraz (Ed.). *Computer vision and image processing in intelligent systems and multimedia technologies*. IGI global.

Elhariri, E., El-Bendary, N., Fouad, M. M. M., Platoš, J., Hassanien, A. E., & Hussein, A. M. (2014). *Multi-class svm based classification approach for tomato ripeness. Innovations in bio-inspired computing and applications*. Springer [pp. 175–186].

Elhariri, E., El-Bendary, N., Hussein, A. M. M., Hassanien, A. E., & Badr, A. (2014). *Bell pepper ripeness classification based on support vector machine. The second IEEE international conference on engineering and technology (ICET 2014)*. IEEE.

Fadilah, N., Mohamad-Saleh, J. 2014. Color feature extraction of oil palm fresh fruit bunch image for ripeness classification. In *13th International Conference on Applied Computer and Applied Computational Science (ACACOS'14)* (pp. 51–55).

Fisher, R. A. (1936). The use of multiple measurements in taxonomic problems. *Journal of Annals of Eugenics, 7*(2), 179–188.

Food and of the United Nations (FAO-UN), A.O. 2012. Fao statistical yearbook 2013-world food and agriculture. <http://faostat.fao.org/site/339/default.aspx>.

Guerrero, E. R., & Benavides, G. M. (2014). *Automated system for classifying Hass avocados based on image processing techniques. 2014 IEEE Colombian conference on communications and computing (COLCOM)*. IEEE [pp. 1–6].

Hahn, F. (2002). Ae–automation and emerging technologies: Multi-spectral prediction of unripe tomatoes. *Journal of Biosystems Engineering, 81*(2), 147–155.

Jaffar, A., Jaafar, R., Jamil, N., Low, C. Y., & Abdullah, B. (2009). Photogrammetric grading of oil palm fresh fruit bunches. *International Journal of Mechanical and Mechatronics Engineering, 9*, 18–24.

Kumar, S., & Kaur, H. (2012). Face recognition techniques: Classification and comparisons. *International Journal of Information Technology and Knowledge Management, 5*(2), 361–363.

Lang, C., & Hübert, T. (2012). A colour ripeness indicator for apples. *Food and Bioprocess Technology, 5*(8), 3244–3249.

Li, T., Zhu, S., & Ogihara, M. (2006). Using discriminant analysis for multi-class classification: An experimental investigation. *Journal of Knowledge and Information Systems, 10*(4), 453–472.

Liu, Y., & Zheng, Y. F. (2005). *One-against-all multi-class svm classification using reliability measures. IEEE international joint conference on neural networks (IJCNN'05)* (Vol. 2). IEEE [pp. 849–854].

May, Z., & Amaran, M. (2011). *Automated ripeness assessment of oil palm fruit using rgb and fuzzy logic technique. The 13th WSEAS international conference on mathematical and computational methods in science and engineering (MACMESE 2011)*. Stevens Point, Wisconsin, USA: World Scientific and Engineering Academy and Society (WSEAS) [pp. 52–59].

Meskaldji, K., Boucherkha, S., & Chikhi, S. (2009). Color quantization and its impact on color histogram based image retrieval accuracy. In *First international conference on networked digital technologies, NDT'09*. IEEE.

Molyneux, S. L., Lister, C. E., & Savage, G. P. (2004). An investigation of the antioxidant properties and colour of glasshouse grown tomatoes. *International Journal of Food Sciences and Nutrition, 55*(7), 537–545.

Paulraj, M., Hema, C. R., Pranesh K., R., & Siti Sofiah, M. R. (2009). Color recognition algorithm using a neural network model in determining the ripeness of a banana. In *The international conference on man–machine systems (ICoMMS)*. Universiti Malaysia Perlis [pp. 2B71–2B74].

Polder, G., Van der Heijden, G., & Young, I. (2002). Spectral image analysis for measuring ripeness of tomatoes. *Transactions-American Society of Agricultural Engineers, 45*(4), 1155–1162.

Prasanna, V., Prabha, T., & Tharanathan, R. (2007). Fruit ripening phenomena–An overview. *Journal of Critical Reviews in Food Science and Nutrition, 47*(1), 1–19.

Shah Rizam, M., Farah Yasmin, A., Ahmad Ihsan, M., & Shazana, K. (2009). Non-destructive watermelon ripeness determination using image processing and artificial neural network (ann). *International Journal of Electrical and Computer Engineering, 4*(6).

Shahbahrami, A., Juurlink, D. B. B. 2008. Comparison between color and texture features for image retrieval.

Singh, S. M., & Hemachandran, K. (2012). Content-based image retrieval using color moment and Gabor based image retrieval using color moment and Gabor texture feature. *International Journal of Computer Science Issues, 9*(1), 299–309.

Soman, S., Ghorpade, M., Sonone, V., Chavan, S. 2012. Content based image retrieval using advanced color and texture features, In *International Conference in Computational Intelligence (ICCIA 2012)*, Vol. ICCIA (pp. 1–5).

Suganthy, M., & Ramamoorthy, P. (2012). Principal component analysis based feature extraction, morphological edge detection and localization for fast iris recognition. *Journal of Computer Science, 8*(9), 1428–1433.

Suralkar, S., Karode, A., & Pawade, P. W. (2012). Texture image classification using support vector machine. *International Journal of Computer Technology & Applications, 3*(1).

Syal, S., Mehta, T., & Darshni, P. (2013). Design & development of intelligent system for grading of jatropha fruit by its feature value extraction using fuzzy logics. *International Journal of Advanced Research in Computer Science and Software Engineering (IJARCSSE), 3*(7), 1077–1081.

Tzotsos, A., Argialas, D. 2006. A support vector machine approach for object based image analysis. In *International conference on object-based image analysis (OBIA06)*.

U.S. Dept. Agric./AMS, Washington, D. 1991. United states standards for grades of fresh tomatoes. <http://www.ams.usda.gov/standards/vegfm.htm>.

Vanschoenwinkel, B., & Manderick, B. (2005). *Appropriate kernel functions for support vector machine learning with sequences of symbolic data. Deterministic and statistical methods in machine learning* (Vol. 3635). Springer [pp. 256–280].

Wei, X., Liu, F., Qiu, Z., Shao, Y., & He, Y. (2014). Ripeness classification of astringent persimmon using hyperspectral imaging technique. *Food and Bioprocess Technology, 7*(5), 1371–1380.

Wu, Q., & Zhou, D.-X. (2006). Analysis of support vector machine classification. *Journal of Computational Analysis & Applications, 8*(2).

Xiao, B. (2010). *Principal component analysis for feature extraction of image sequence. 2010 International conference on computer and communication technologies in agriculture engineering (CCTAE)* (Vol. 1). IEEE [pp. 250–253].

Yu, H., Li, M., Zhang, H.-J., & Feng, J. (2002). *Color texture moments for content-based image retrieval. International Conference on Image Processing* (Vol. 3). IEEE [pp. 929–932].

Zawbaa, H. M., El-Bendary, N., Abraham, A., & Hassanien, A. E. (2011). *Svm-based soccer video summarization system. 2011 Third world congress on nature and biologically inspired computing (NaBIC)*. IEEE [pp. 7–11].

Zawbaa, H. M., El-Bendary, N., Hassanien, A. E., & Kim, T.-h. (2011). *Machine learning-based soccer video summarization system, multimedia, computer graphics and broadcasting*. Springer [pp. 19–28].

Zhang, D., Lee, D.-J., Tippetts, B. J., & Lillywhite, K. D. (2014). Date maturity and quality evaluation using color distribution analysis and back projection. *Journal of Food Engineering, 131*, 161–169.

Zhang, L., & McCarthy, M. J. (2012). Measurement and evaluation of tomato maturity using magnetic resonance imaging. *Journal of Postharvest Biology and Technology, 67*, 37–43.

Zhang, Y., Xie, X., & Cheng, T. (2010). *Application of pso and svm in image classification. 2010 3rd IEEE international conference on computer science and information technology (ICCSIT)* (Vol. 6). IEEE [pp. 629–631].