Smart Fruit Grading System Using K-Means Clustering and SVM Classifier

[1] Mr. Suhas Jadhav, [2] Prof. M. U. Inamdar
[1]Student, [2]Proffesor
[1] P.G. Student, Department of E&TC, SCOE, Pune, Maharashtra, India,
[2] Assistant Professor, Department of E&TC, SCOE, Pune, Maharashtra, India.

Abstract : In this paper, computer vision-based algorithm for tomato grading is projected which workings in four steps. First is to apply preprocessing for uploaded image or images, such preprocessing step involves different algorithm methods like gray scale conversion, histogram calculation and binarization process. Discussed such all process deeply in below section. Second step is k-means segmentation to segment input image into 3 region i.e, Background, damaged and fresh area. Then, geometric, statistical and textural features from refined defected regions are extracted. Finally, for tomato grading, valued performance of SVM (Support Vector Machine) classifiers is done. Classification is computed in two approaches which in the first one, an input tomato is classified into two categories which of healthy grade and defected grade. In the second approch, the input tomato is classified into three sectors of first grade, second grade and damaged grade. In both grading steps, SVM classifier works as the finest one with detection rate of 96.8% and 98.1% for 2 categories (healthy grade and defected grade) & 3 quality categories (first grade, second grade and damaged grade), among 90 different golden delicious apple images, respectively.

IndexTerms **- SVM Classifier, K-means Clustering, Machine learning.**

I. INTRODUCTION

Image processing is a region that continues developing, with new applications being produced with each agonizing moment. It is enthralling zone which has advanced with applications in different zones like space program, media outlet and so on. The most concerning prospect of this information change is its capacity to transmit and get mind boggling information that exceeds typical composed content. Visual information sent in the method for images, has advanced as one of the significant method of impart on in the 21st century. Portrayal of image processing in detail is depicted as below.

Pistachio nut is the best non-oil fare of Iran that accounted up to 60% of worldwide pistachio advertises. It has an essential part in Iran's fares and economy as a non-oil great [1]. Because of disadvantages of manual grading, for example, monotony, subjectivity, accessibility, work prerequisites, irregularity and cost, a computerized grading framework should be produced. As of late, a few analysts have directed their endeavors on building up the grading framework. The techniques stretch out from manual-machine grading, where the highlights are resolved physically, under research facility conditions to machine vision frameworks for computerized rapid natural product sort in.

Unaya and Gosselin [13] presented a framework that utilized diverse classifiers. They utilized an ANN (artificial neural network) to section apple imperfections and afterward tried and looked at five supervised classifiers. The outcomes demonstrated that the Adaboost and SVM (support vector machine) were the best ones with around 90% precision. Zou et al. [14] presented a framework that utilized different shading cameras to filter tomato surface. They ordered tomato into gatherings of sound grade and absconded review by thresholding and accomplished 96% exactness. During this work we introduce a tomato grading computer vision algorithm that can be used in an ordinary machine vision system. computer vision-based algorithm for tomato grading is projected which workings in four steps. First is to apply preprocessing for uploaded image or images, such preprocessing step involves different algorithm methods like gray scale conversion, histogram calculation and binarization process. Discussed such all process deeply in below section. Second step is k-means segmentation to segment input image into 3 region i.e, Background, damaged and fresh area. Then, geometric, statistical and textural features from refined defected regions are extracted. Finally, for tomato grading, valued performance of SVM (Support Vector Machine) classifiers is done. Classification is computed in two approaches which in the first one, an input tomato is classified into two categories which of healthy grade and defected grade. In the second approch, the input tomato is classified into three sectors of first grade, second grade and damaged grade. The remainder of this paper is organized as follows: In next Section, the Literature Survey is explained in detail. And after, next section presents proposed method in which performance of the proposed method is evaluated among 90 different tomatoes images in Result Section, and finally, the paper ends with a section on conclusion.

II. LITERATURE SURVEY

Momin, M. A., et al. [1] presented, "Geometry-based mass grading of mango fruits using image processing." Developed an image (picture) acquirement and preparing framework to remove anticipated zone, border, and roundness highlights. In this framework, pictures were obtained utilizing a XGA arrange color camera of 8-bit dim levels utilizing fluorescent lighting. A picture preparing calculation in view of area based worldwide thresholding color binarization, joined with middle channel and morphological investigation was produced to group mangos into one of three mass evaluations, e.g., huge, medium, and little. To accomplish a better reviewing, two diverse evaluating highlights could be utilized as a part of succession. The picture evaluating framework is straightforward and effective and can be viewed as an appropriate first stage to motorizing the business reviewing of mangos in Bangladesh. In addition, the technique can possibly be connected to different products with reasonable changes.

Nouri-Ahmadabadi, Hosein, et al. [2] presented, "Design, development and evaluation of an online grading system for peeled pistachios equipped with machine vision technology and support vector machine." In this survey, an astute framework in view of joined machine vision (MV) and SVM (Support Vector Machine) was created for sorting of peeled pistachio kernels and shells. The framework was made out of lighting box, conveyor belt, processing unit, sorting unit and camera. A color CCD camera was

utilized to catch pictures. The pictures were digitalized by a catch card and exchanged to a PC for further investigation. At first, pictures were changed over from RGB color space to HSV color ones. For segmentation of the gained pictures, H-component in the HSV color space and Otsu thresholding method were connected. An element vector containing 30 color highlights was removed from the caught pictures. A component choice method in light of affectability examination was completed to choose unrivaled highlights. The chose highlights were introduced to SVM classifier. Different SVM models having an alternate kernel work were created and tried.

Baigvand, Mehrdad, et al. [3] presented, "Machine vision system for grading of dried figs." The framework hardware was made out of a CCD camera, a belt conveyor, a feeder, a partition unit, and a lighting framework. Three quality lists, to be specific part size, color, and size, were first classified by fig-processing specialists into the five classes. Then, the pictures of the fig tests were caught utilizing a machine vision framework. In the first place, the length of pixels in each picture and longitudinal directions of the focal point of gravity of fig pixels were separated for computing the spout discharge time. For removing the three quality indices of each class, a machine vision algorithm was created. This algorithm decided color intensity and width of each fig as the pointers of its color and size, individually. For computing the split region, the pictures were first binarized by utilizing the color intensity distinction between the split and other parts of the organic product with a specific end goal to decide the region of the split area. A grading algorithm was additionally coded in Lab-VIEW for sorting figs in view of their quality indices extricated by the picture processing algorithm into five subjective evaluations. In the grading algorithm, the estimations of these highlights were contrasted and the threshold esteem that was foreordained by a specialist. Results showed that the created framework enhanced the sorting precision for every one of the classes up to 95.2%. The framework's mean rate was 90 kg/h for processing and grading figs.

Moallem, Payman, Alireza Serajoddin, and Hossein Pourghassem. [4] presented, "Computer vision-based apple grading for golden delicious apples based on surface features." Shows, a C vision-based algorithm for brilliant delectable apple grading is proposed which works in six stages. Non-apple pixels as foundation are initially expelled from input pictures. Then, stem end is recognized by mix of morphological methods and Mahalanobis removed classifier. Calyx area is additionally distinguished by applying K-implies bunching on the Cb component in YCbCr color space. From that point forward, deserts segmentation is achieved utilizing MLP(Multi-Layer Perceptron) neural system. In the subsequent stage, stem end and calyx areas are expelled from absconded locales to refine and enhance apple grading process. Then, measurable, textural and geometric highlights from refined surrendered locales are removed. At long last, for apple grading, a correlation between execution of MLP, SVM (Support Vector Machine), and KNN (K-Nearest Neighbor) classifiers is finished. Classification is done in two conduct which in the first, an information apple is classified into two classes of healthy grade and defected grade. In the second way, the info apple is classified into three classifications of first rank grade, second rank grade and harmed grade ones. In both grading steps, SVM classifier works as the best one with acknowledgment rate of 92.5% and 89.2% for two classifications (healthy and defected) and three quality classifications (first rank, second rank and rejected ones), among 120 distinctive brilliant scrumptious apple pictures, individually, considering K-folding with K = 5. In addition, the exactness of the proposed segmentation algorithms including stem end discovery and calyx identification are assessed for two diverse apple picture databases.

Zaborowicz, Maciej, et al. [5] presented, "Application of neural image analysis in evaluating the quality of green house tomatoes." Presents the exploration on the utilization of strategies for artificial neural modeling and computer image analysis during the time spent surveying the nature of green house tomatoes assortment Capriccio. The subject of the investigation was tomatoes of the sizes from 40 mm to 67 mm and the colors: 1– 6, incorporate middle colors. The procedure of image acquisition and acquiring observational information was directed all through the whole developing season in the period from the principal collect amidst May to the last gather toward the start of November. Attractive quality attributes were acquired on account of the RBF 37:37-39-1:1 and RBF 22:22-20-2:2 models. RBF 37:37-39-1:1 system, whose yield variable was the shade of the tomato, the preparation quality was 0.930827, the approval quality was 0.911982, and the test quality was 0.979390. The RMSE rate of system preparing for the preparation set was 0.075986, for the approval set it was 0.072194, and for the test set 0.061714.

Bhange, Manisha, and H. A. Hingoliwala. [6] presented, "Smart farming: Pomegranate disease detection using image processing." In this paper, we propose a web based apparatus that helps agriculturists for recognizing organic product infection by transferring natural product picture to the framework. The framework has an effectively trained dataset of pictures for the pomegranate organic product. Info picture given by the client experiences a few preparing ventures to distinguish the seriousness of sickness by contrasting and the trained dataset pictures. To start with the picture is resized and after that its highlights are extricated on parameters, for example, morphology, CCV, and color and clustering is finished by utilizing k-means algorithm. Next, Support vector Machine (SVM) is utilized for arrangement to group the picture as tainted or non-contaminated. A goal seek system is likewise given which is exceptionally helpful to discover the client intension. Out of three highlights separated we got best outcomes utilizing morphology. Test assessment of the proposed approach is compelling and 82% exact to distinguish pomegranate ailment.

Sofu, M. M., et al. [7] presented, "Design of an automatic apple sorting system using machine vision." This examination proposes a programmed apple arranging and quality assessment framework, which depends on real-time handling. Brilliant and Starking Delicious, and Granny Smith apple cultivars are arranged into various classes by their color, size and weight. It additionally distinguishes apples influenced by rot, scab and stain. The proposed framework comprises of a class conveyors, a roller and transporter joined with an encased lodge with machine vision, stack cell and control board units. The roller and transporter conveyors have two channels. With a specific end goal to dissect the visual properties of apples, two indistinguishable mechanical color cameras are determined to the roller transport. Four pictures of any apple moving on the transport can be caught and prepared utilizing picture handling programming in 0.52 s. Thus, the proposed machine can arranged averagely 15 apples in every second utilizing two channels, in real time.

III. System Architecture

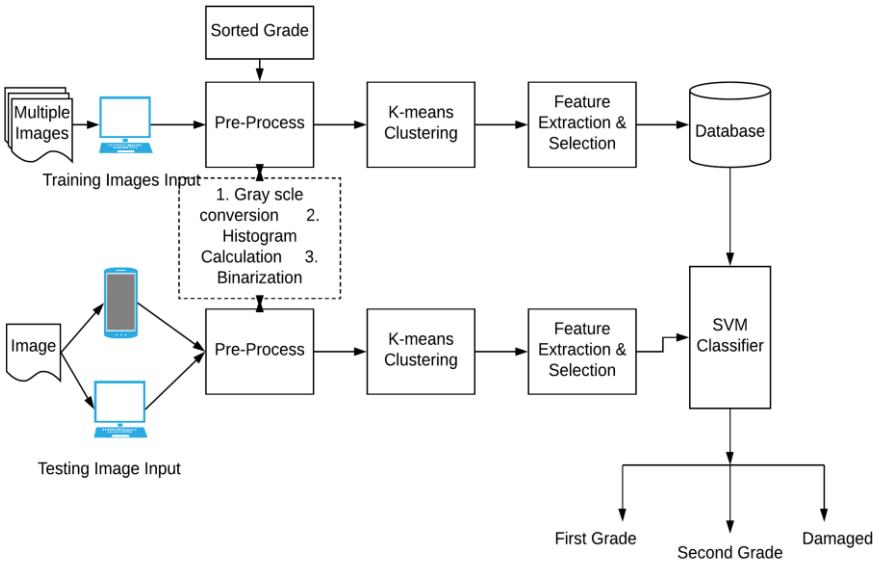

Fig.1. System Architecture for Grading System

The above fig 1 shows, proposed system architecture for tomato grading system. This system divides mainly in two parts i.e, first one is Training module and another is Testing Module. Each this module includes following 4 sub-modules as following also states its implementation process. 3 modules are common for all 2 parts i.e, preprocessing, k-means clustering and Feature Extraction/ Selection. There is only one module extend and used for feature classification by SVM classifier in testing module.

Module1 Preprocessing:

In the phase of preprocessing there are multiple preprocess algorithms are used on image. These results are used for further next process. Preprocess includes gray scale conversion, Histogram calculation, Binarization process etc. Gray scale conversion is a stage in which In photography, computing, and colorimetry a grayscale or greyscale image is one in which the estimation of every pixel is a solitary example speaking to just a measure of light, that is, it conveys just power data. Images of this sort, otherwise called monochrome or black-and-white, are composed only of shades of gray, changing from black at the weakest power to white at the most grounded. Grayscale images are particular from one-piece bi-tonal black-and-white images which, in the context of computer imaging, are images with just two colors: black and white (likewise called bilevel or paired images). Grayscale images have numerous shades of gray in the middle. In our stage gray scale is utilized to fabricate a twofold type of info image which can show and separate image in two section i.e, background part and organic product part. This is only discovery of natural product region.

The histogram plots the quantity of pixels in the image (vertical pivot) with a specific brightness esteem (even hub). Calculations in the digital manager enable the client to outwardly modify the brightness estimation of every pixel and to powerfully show the outcomes as modifications are made. Enhancements in picture brightness and contrast would thus be able to be gotten. In the field of computer vision, image histograms can be valuable devices for thresholding. Since the data contained in the diagram is a portrayal of pixel dispersion as an element of tonal variety, image histograms can be examined for peaks or potentially valleys. This limit esteem would then be able to be utilized for edge recognition, image segmentation, and co-occurrence matrices.

Binarization assumes a vital part in digital image processing, essentially in computer vision applications. Thresholding is a productive procedure in binarization. The decision of thresholding procedure is pivotal in binarization. A few thresholding calculations have been examined and proposed to define the ideal limit esteem. In this exploratory investigation, Otsu and Gaussian Otsu thresholding calculations were created and tried with a few images. The consequences of these two strategies are then compared in their execution to decide the limit esteem. Results demonstrate better execution for Gaussian Otsu's technique compared to Otsu's strategy.

Module2 K-means Clustering Segmentation:

Segmentation means dividing of picture into different gatherings or clusters of same features or having some comparability. K-means grouping strategy parcels the pictures into clusters such that atleast one a player in bunch contains a picture with real region of influenced part. In our work we can segment image in three parts which describes background region, defected region, and fresh region with the help of k-means segmentation.

Module3 Feature Extraction & Selection:

In feature extraction wanted feature vectors, for example, texture, color, structure and morphology are removed. Feature extraction is technique for including number of assets required to depict an expansive arrangement of information precisely. Factual texture features are acquired by GLCM (Gray level co-occurrence matrix) equation for texture examination and texture features are computed from measurable dissemination of watched power combinations at the predefined position with respect to others. Quantities of gray levels are imperative in GLCM additionally measurements are arranged into request of to begin with, second and higher for number of power focuses in every combination. Distinctive factual texture features of GLCM are vitality, whole entropy, covariance, data measure of correlation, entropy, contrast and backwards difference and difference entropy.

Module4 SVM Classifier:

SVM (Support vector machines) are supervised learning models with related learning calculation that break down information utilized for classification and regression analysis [8]. Support vector machine depends on augmenting the base separation from the isolating hyper plane to the closest example. Just binary classification is supported in essential SVM, yet for multi-class classification case expansion can be conceivable [8]. In these expansions, extra imperatives and parameters are added to enhancement issues for dealing with the detachment of the diverse classes. SVM is a binary classifier that implies the class marks can just take two qualities ±1. We uses 2 step SVM to categorize image feature into three category, In first step SVM image is categorize in to Damaged and Fresh fruit category in 2nd step SVM image is categorized in to 1st grade and 2nd grade fruit image. 2nd step SVM is used for 2nd classified part of 1st SVM result that is after Fresh fruit grade finds.

IV. ALGORITHMS

1. K-means Clustering Algorithm

Input: Input buffered image
Output: Three (3) part clustered image
Process:
Step 1: Assign random centroid (k-centroid)
Step 2: Calculate distance from every point from every centroid.
Step 3: Assign points to those centroids with minimum distance and make clusters.
Step 4: Calculate mean for each cluster and make them new centroid.
Step 5: Repeat step 2-4 till we get same clusters.

2. SVM feature Classification for grading

Input: Three (3) part clustered image damage and fresh cluster area.
Output: Showing Actual three stage grading result (showing fruit category in the sense of 1st grade, 2nd grade, damaged grade)
Process:
Step 1: categorize Image sectors into 2grade according to ratio i.e Fresh Percentage and Damage Percentage.
Step 2: check Damaged Ratio()
Step 3: check First Grade Ratio()
Step 4: check Second Grade Ratio()
Step 5: if (OverallDamagedPercent>OverallFreshPercent) then check for damaged ratio as in step 2 if(available) then grade='damaged grade'.
Step 6: else then Grade='Second Grade'
Step 7: if(OverallDamagedPercent=<OverallFreshPercent) then check for fresh ratio as in step 3
Step 8: if (available in range) grade='First Grade'
Step 9: else check for second grade ratio if (avilable) then grade='Second grade'
Step 10: else check for damaged grade ratio if (avilable) then grade='Damaged grade'
Step 11: Check Grading By Standardized range(FreshperValue)

V. RESULTS AND DISCUSSIONS

This section is shows the result and analysis with screenshot of actual implementation. Where fig. 2 shows, gray scale and binarization process of preprocessing module phase 1. Also shows upload image process.

Where fig. 3 shows, Histogram of a gray scale image. It is also a part of preprocessing module phase in both training and testing scenario. At this section threshold point is set for binarization of image/images.

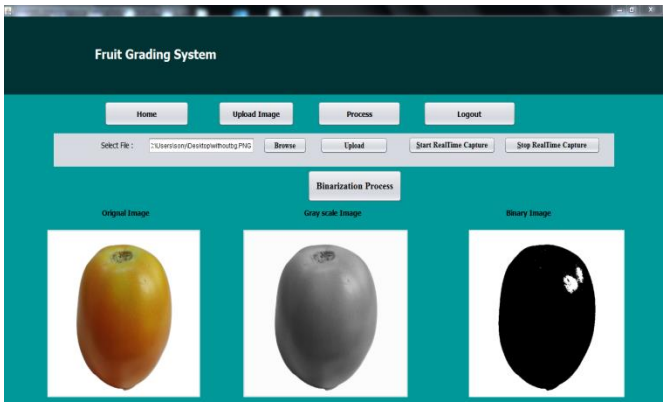

Fig. 2 Gray scale and binarization of image

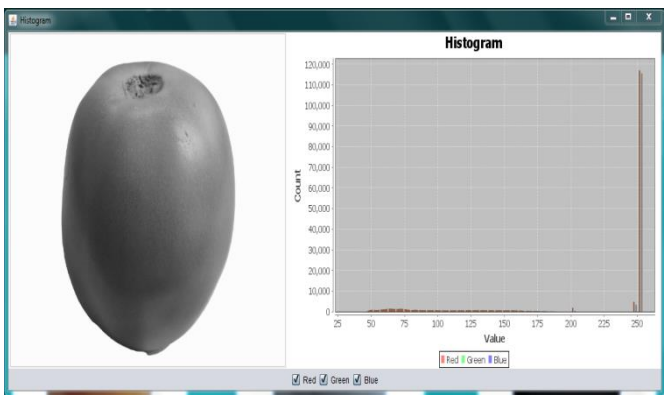

Fig. 3 Histogram calculation of image

Where fig. 4 shows, K-means clustering of an image. It is a part of segmentation module in both training and testing scenario. At this point we can divide input image into 3 parts 1st one is background, 2nd one is damaged part, 3rd one is fresh fruit area part.

Below fig. 5 shows, Result image. It is a part of result and analysis module in testing scenario. At this point after image feature extraction and selection we can apply SVM process and categorize fruit into 3 area i.e, First grade fruit, Second grade fruit, and Damaged fruit.

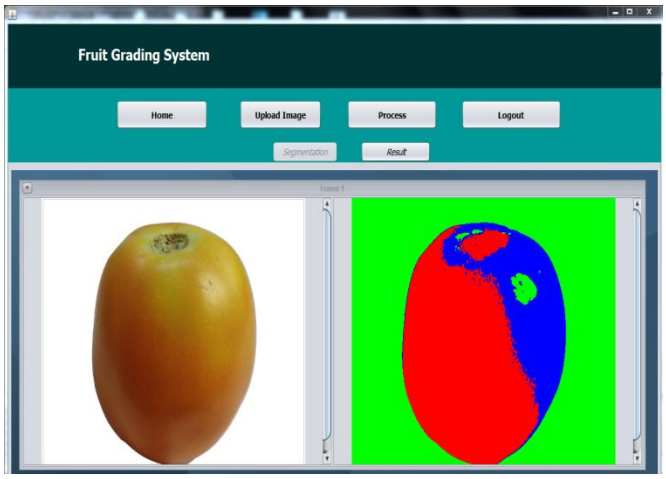

Fig. 4 K-means clustering of image

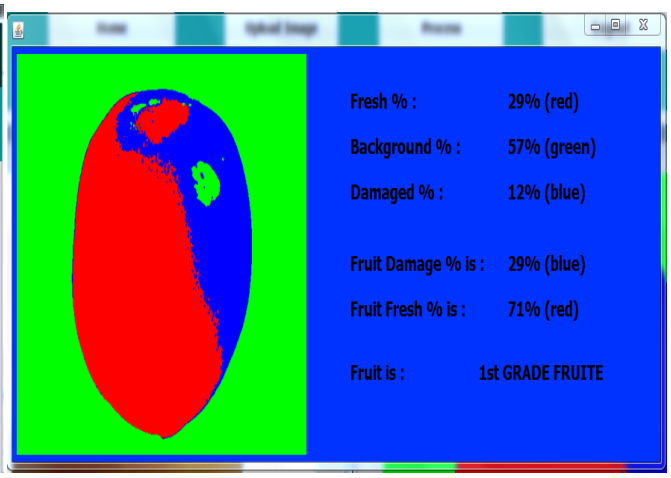

Fig. 5 After SVM Grading Result

VI. Conclusion

An image / Images processing oriented solution is projected and evaluated for the exposure and categorization of exaggerated segment of tomato fruit. Here this paper provides an proficient and precise system for exposure and categorization of tomato fruit images which is affected or damaged. The projected approach uses K-Means clustering for segmentation part and SVM classifier for classification in 3 fold classes such as First grade, Second grade, and damaged grade, as it provides high precision when check with other techniques and consumes very less time for whole processing.

VII. Future scope

The fundamental constraint went up against in the exploration is the dataset measure that is should have been bigger to have additionally training pictures per class, which can enhance the execution of the framework. Another confinement is that the precision rate of segmentation must be enhanced to enhance the general exactness of the classification. Different research headings and difficulties could be considered for future research. The proposed approach can be connected to other classification fields other than ripeness classification. The proposed approach can likewise be connected to mechanize the estimation of ripeness process for other climacteric harvests, for example, mango and chime pepper. Also there is another issue which related to image that is reflection or secularity and shadow effect. Due to image is confused during feature section and segmentation so there is scope to remove reflection or secularity and shadow effect by suitable technique.

References

[1] Momin, M. A., et al. "Geometry-based mass grading of mango fruits using image processing." *Information Processing in Agriculture* 4.2 (2017): 150-160.
[2] Nouri-Ahmadabadi, Hosein, et al. "Design, development and evaluation of an online grading system for peeled pistachios equipped with machine vision technology and support vector machine." *Information Processing in Agriculture* 4.4 (2017): 333-341.
[3] Baigvand, Mehrdad, et al. "Machine vision system for grading of dried figs." *Computers and Electronics in Agriculture* 119 (2015): 158-165.
[4] Moallem, Payman, Alireza Serajoddin, and Hossein Pourghassem. "Computer vision-based apple grading for golden delicious apples based on surface features." *Information Processing in Agriculture* 4.1 (2017): 33-40.
[5] Zaborowicz, Maciej, et al. "Application of neural image analysis in evaluating the quality of greenhouse tomatoes." *Scientia Horticulturae* 218 (2017): 222-229.
[6] Bhange, Manisha, and H. A. Hingoliwala. "Smart farming: Pomegranate disease detection using image processing." *Procedia Computer Science* 58 (2015): 280-288.
[7] Sofu, M. M., et al. "Design of an automatic apple sorting system using machine vision." *Computers and Electronics in Agriculture* 127 (2016): 395-405.
[8] Zhao, Yuanshen, et al. "Detecting tomatoes in greenhouse scenes by combining AdaBoost classifier and colour analysis." *biosystems engineering* 148 (2016): 127-137.
[9] Avila, Felipe, et al. "A method to construct fruit maturity color scales based on support machines for regression: Application to olives and grape seeds." *Journal of Food Engineering* 162 (2015): 9-17.
[10] Dutta, Malay Kishore, et al. "Image processing based classification of grapes after pesticide exposure." *LWT-Food Science and Technology* 72 (2016): 368-376.
[11] Cen, Haiyan, et al. "Nondestructive detection of chilling injury in cucumber fruit using hyperspectral imaging with feature selection and supervised classification." *Postharvest Biology and Technology* 111 (2016): 352-361.
[12] Guo, Chentong, et al. "Hyperspectral imaging analysis for ripeness evaluation of strawberry with support vector machine." *Journal of Food Engineering* 179 (2016): 11-18.
[13] Sanaeifar, Alireza, Adel Bakhshipour, and Miguel de la Guardia. "Prediction of banana quality indices from color features using support vector regression." *Talanta* 148 (2016): 54-61.

[14] Sa'ad, Fathinul Syahir Ahmad, et al. "Shape and weight grading of mangoes using visible imaging." *Computers and Electronics in Agriculture* 115 (2015): 51-56.
[15] Benalia, Souraya, et al. "Computer vision for automatic quality inspection of dried figs (Ficus carica L.) in real-time." *Computers and Electronics in Agriculture* 120 (2016): 17-25.
[16] Goel, Nidhi, and Priti Sehgal. "Fuzzy classification of pre-harvest tomatoes for ripeness estimation–An approach based on automatic rule learning using decision tree." *Applied Soft Computing* 36 (2015): 45-56.
[17] Yousefi, Jamileh. "Image Binarization using Otsu Thresholding Algorithm." (2011).
[18] Kader, Adel A. "Increasing food availability by reducing postharvest losses of fresh produce." *V International Postharvest Symposium 682*. 2004.