Algorithm for Mango Classification Using Image Processing and Naive Bayes Classifier

อัลกอริทึมสำหรับการคัดแยกมะม่วงด้วยการประมวลผลภาพและนาอีฟเบย์

Phantida Limsripraphan[1*], Peerapol Kumpan[2],
Nanthaphon Sathongpan[3] and Chitnarong Phengtaeng[4]
[1*,2,3,4]Department of Computer Engineering, Faculty of Industrial Technology, Pibulsongkram Rajabhat University, Phitsanulok 65000, Thailand Tel. +668 1888 9330 Fax. +66 5526 7124 E-mail: phuntida_nk@hotmail.com

วันที่รับบทความ 17 กรกฎาคม 2561 **วันที่รับแก้ไขบทความ** 28 กุมภาพันธ์ 2562 **วันที่ตอบรับบทความ** 14 พฤษภาคม 2562
Received: July. 17, 2018 **Revised:** Feb. 28, 2019 **Accepted:** May. 14, 2019

Abstract

Mangoes are one of the most preferred fruits in the world, and one of the most crucial to the economy of the Thai fruit industry. The Thai government has therefore set a policy to encourage mango growers to improve the quality of their mango fruits. Mango fruit inspection is important for the quality control system of mango fruit production. This paper presents a Bayesian approach for mango classification based on digital image processing. The algorithm is designed to identify a defect on the mango skin and classify it into one of two groups: 'unripe mango' or 'ripe mango'. To determine the defect on the mango skin, image thresholding and image labeling are applied. Colour features are extracted from RGB mango images using statistical calculations. The naive Bayes classifier is then applied to classify the colour-based feature of the mango images. The performance of the Bayesian approach was evaluated against another classification technique-the Support Vector Machine (SVM). The method was implemented and tested on 100 mango images. The surface defect detection achieved 85% accuracy. The experimental results show the superiority of the proposed method, with an accuracy of 90%, as compared to SVM-based methods, with an accuracy of 83%.

Keywords: mango classification, mango ripeness, naive bayes classifier, classification by image processing

1. Introduction

Mangoes are widely marketed both in Thailand and abroad. Phitsanulok Province has the highest mango plantation rate in the lower northern region with a total area of 40,805 acres or 55% of the entire area of 74,479 acres (Thairat news, 2016). The main economy in the Thai fruit sector is generated through the production of mango and its products. Therefore, to encourage the mango growers, the Thai government has set a policy to improve the quality of the mangoes. Mango fruit inspection is useful for the quality control system for mango fruit production. Due to the extensive production of mangoes per year, the manual way of inspection and classification processes may result in delays and poor quality which may affect the customer demand. The quick inspection and classification process of mangoes are essential for better harvesting. Factors such as the ripe mango, unripe mango and defects on the mango skin are considered. Researchers have designed an algorithm using a feature-based method to inspect and classify mango into ripe and unripe classes. This algorithm is useful to check the quality of the mangoes and in the classification process.

Research to assist the quality assurance in the agriculture sector were done using digital image processing. According to Cubero et al (2014), he developed an automatic fruit monitoring system which can measure fruit size and weight estimation using features of the fruits.

Research about mango grading and classification such as; Vyas et al. (2014) studied mango quality assessment by color separation in LAB color space. Fuzzy logic was applied to color space models to measure the ripeness of the mango (Othman et al., 2014) and (Othman et al., 2016). Mim et al. (2018) proposed an image processing technique based on RGB and HSI color feature to identify the different maturity stages of mangoes. Rungpichayapichet et al. (2016) applied near-infrared spectroscopy (NIRS) prediction models to determine the postharvest quality of mango and used second derivative pretreated spectra to classify the ripeness of mango which provides an accuracy of more than 80%. Nagle et al. (2016) developed an optical system for inspecting surface color of two Thai mango cultivars; Nam Dokmai, and Maha Chanok. The developed system applied an image processing technique to estimate the colour parameters in the LAB color space. Then the color parameters were used to predict the ripeness of mango. Eyarkai Nambi et al. (2016) developed a machine vision application to predict ripening quality in mangoes using RGB images. The developed application applied the Hierarchical clustering method to classify the ripening period into five stages based on quality parameters (physicochemical, color and textural properties). Huang et al. (2018) applied colorimetric sensor array (CSA) and image processing with principal component analysis to evaluate the mango quality. Support vector classification (SVC) model was used to classify mango into three grades, and to predict the Hardness and Total Soluble Solid (TSS) of the mango.

Sa'ad et al. (2015) presented the use of visible imaging as a tool in grading the mangoes and applying a Fourier-descriptor method to grade the fruit by their shape. Nandi et al. (2014) described a computer vision system for the grading the mango fruit based on maturity prediction, size, and smoothness of the mango surface. The support vector machine (SVM) was used for the classification of mango into each grade.

The Naïve Bayes Classifier is a classification technique based on the Bayes theorem, with the principle of probability to assume that the value of a particular feature is independent of the value of any other feature, given the class variable. The Naïve Bayes Classifier has been applied in many classification applications. According to Ghanad et al. (2015), Naïve Bayes Classifier was used as a diagnostic tool to diagnose Parkinson's disease using Bayesian assays with the PSO Bayesian algorithm. Froese, et al. (2014) applied Bayesian classification to estimate length-weight relationships (LWR) in fishes using body shape parameter.

This study proposes an algorithm for mango fruit classification. The algorithm discovers a defect on the mango skin using an image labelling technique, then apply Naïve Bayes Classifier with color features to classify mango images into two groups; ripe and unripe.

2. Research purpose

2.1 To develop an algorithm to identify defect on for mango.

2.2 To develop an algorithm to classify the mango into ripe and unripe class.

3. Mango image data set

The mango image dataset used in the experiment were the images of the Kent mango from the COFILAB database. In this study, the images are divided into two groups; ripe and unripe. The ripe mango will have a yellow color, and the unripe mango will have a green color. Sometimes the unripe mango can both both have green and yellow color. The dataset consists of 54 ripped and 46 unripe mango images. The image is taken from a controlled image system composed of a digital camera (EOS 550D, Canon Inc., Japan) and the lighting system. The distance between mango and camera is fixed at 20 cm. Four lamps (Biolux L18W/965, 6500 K, Osram AG, Germany) with two fluorescent tubes each were used to give brightness to the image. The angle between the axis of the lens and the sources of illumination was approximately 45 degree. Polarising filters were placed in front of the lamps and camera lenses in order to reduce the impact of reflections. High-resolution images were taken with an image size of 5,184 x 3,456 pixels. The resolution of the image is 0.03 mm / pixel (Cubero et al., 2014), (Vyas et al., 2014). Figure 1 shows an image acquired from the camera.

(a) (b) (c)

Figure 1. Image acquiring system and example of mango image; a) image system, b) a sample of unripe mango, and c) a sample of a ripe mango. (Cubero et al., 2014), (Vyas et al., 2014)

4. Mango classification algorithm

The algorithm is divided into training and test process to get the classification of mango into ripe and unripe classes. The overall algorithm is shown in Figure 2.

The training process uses 26 ripe and 24 unripe mango images to train the model using Naïve Bayes and SVM. The algorithm consists of preprocessing (segmentation of mango region), surface defect detection, feature extraction and classification steps to determine whether the test image is a ripe mango or an unripe mango.

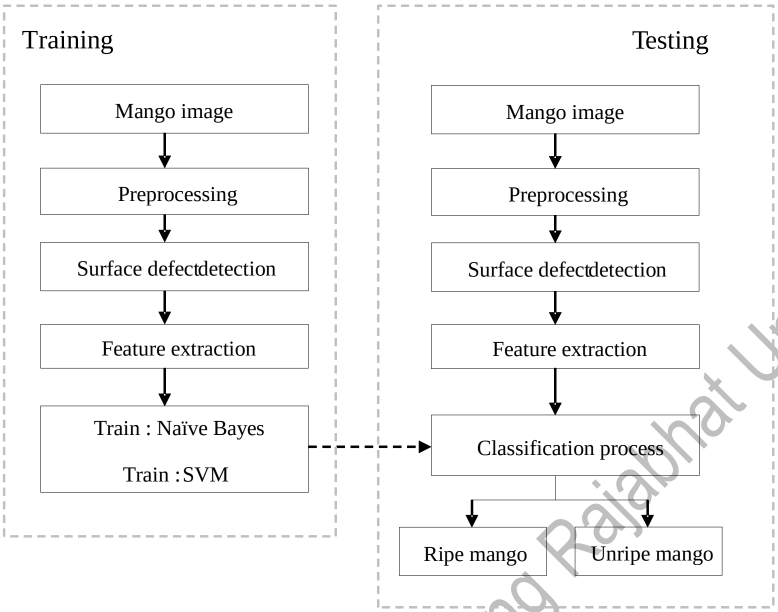
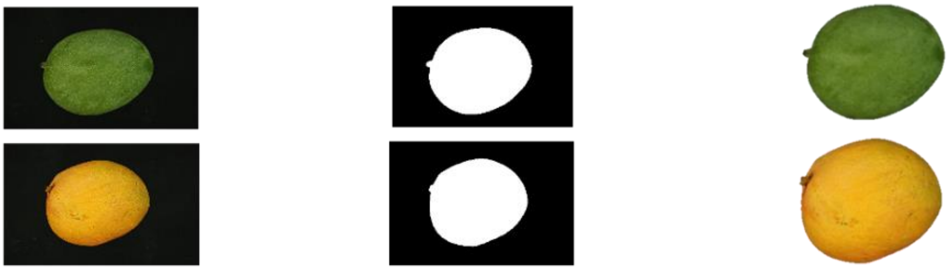


Figure 2. Overall algorithm

4.1 Preprocessing

The goal of image preprocessing is to segment mango region (remove the background and other unnecessary details) and transform into a binary image (change a color image to a black and white image). The algorithm begins with converting RGB images to grayscale images, and then enhance them using a median filter followed by adaptive thresholding to convert the test image to binary. The white mask in the binary image shows the actual mango region whereas the black area shows the background of an image. Examples of binary images and mango fruits separated from the background image are shown in Figure 3.

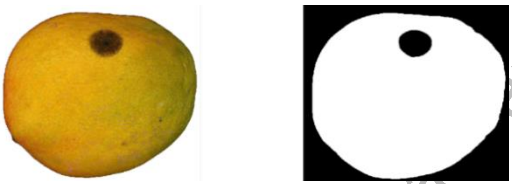


(a) (b) (c)

Figure 3. Mango region segmentation, a) original images, b) binary images, and c) segmented images.

4.2 Surface defect detection

After the mango region segmentation, an image labeling technique is applied to detect the spots of the mango. In Figure 4(b), the white region is the mango pixels, and the black region within the white region is labeled as a spot or defect surface.



(a) (b)

Figure 4. Example of surface defect detection on the mango image, a) original image, and b) defect on mango skin in binary image.

4.3 Feature extraction

This method calculates the statistical feature of color intensity to classify mango image into two groups; ripe and unripe mango. Each channel of the RBG color model of an image was extracted to analyze the color intensity of each plane; red plane, green plane and blue plane. Example of each color plane intensity is shown in Figure 5.

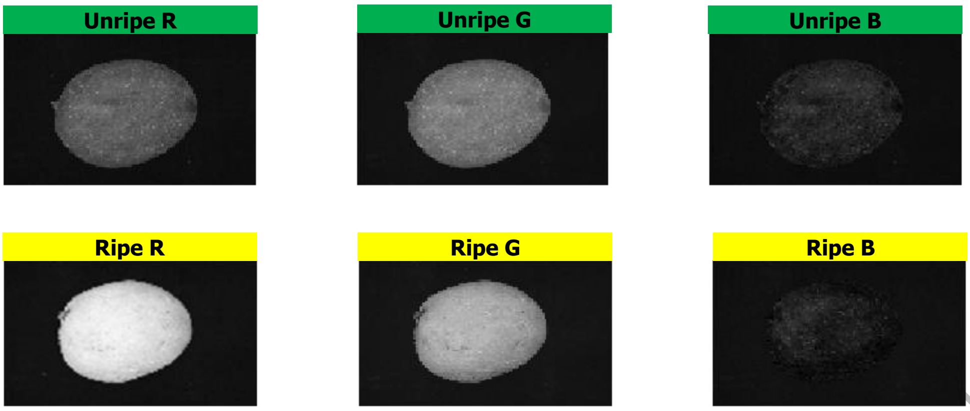


Figure 5. Example of intensity image of red plane, green plane, and blue plane.

From Figure 5, it was found that the difference in color of the ripe mango and the unripe mango was significantly different in the red plane, so the red planes were selected for the extraction of mango fruit. Figure 6 shows a histogram of each color plane; red, green and blue. It is also validating the histogram of the red plane which has a wider distribution than the histogram of a green and blue plane. Therefore, the red plane can be applied to separate between ripe mango and unripe mango.

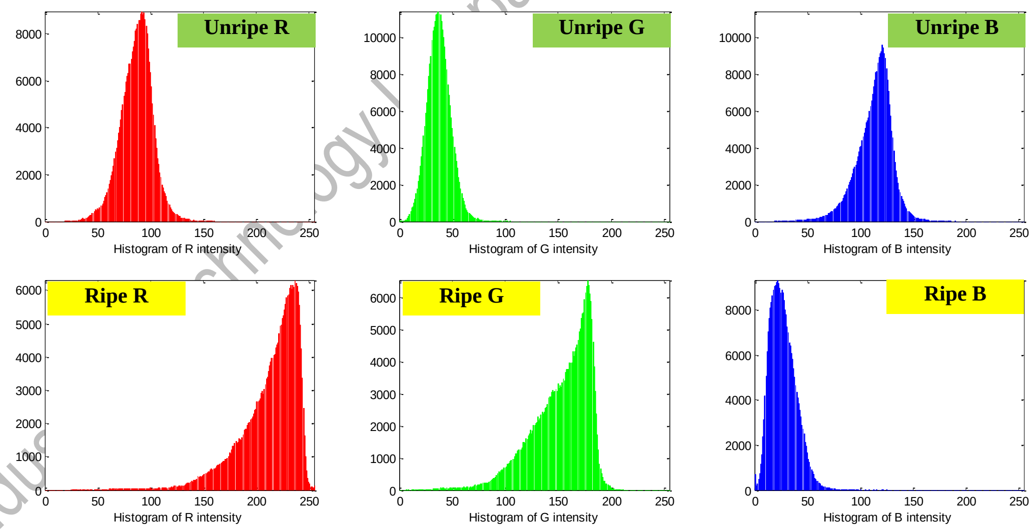


Figure 6. Histogram of intensity of red plane, green plane, and blue plane

1) Mean value feature

The mean value of an intensity of the mango region in the red plane is calculated from the images of 26 ripe and 24 unripe mangoes. Figure 7 shows a histogram of collected mean values. The histogram shows the mean values are divided into two groups, but still it has some overlapping in the values. Therefore, the mean value color could, potentially, be used for analysis but it is still necessary to improve the performance and reduce false classification produced by the overlapping of mean values as shown in Figure7.

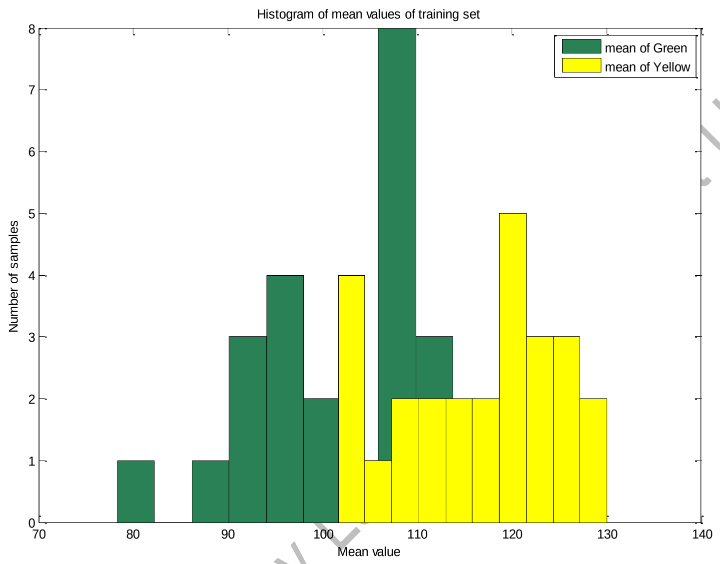


Figure 7. Histogram of the mean values of the mango region of the red plane.

2) Mean value in selected region

The second feature selected is the mean value of red intensity concentrated in the area of interest. The area of interest is chosen by framing the area around the center of the mango and creating a large rectangular box. The program then creates a small rectangular area in every corner of the rectangular box as an area of interest. If the area of interest is labeled as a spot or skin defect, then the small rectangle will be shifted to another location nearby, where there is no defect. Examples of frames, areas of interest and the average concentration of color values are displayed in Figure 8.

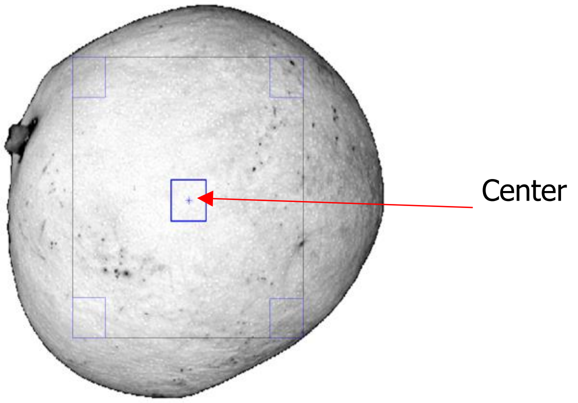


Figure 8. Selected region at center and the four areas at the corners of the black line rectangle.

4.4 Training process

In the training process, 26 ripe mango images and 24 unripe mango images were used. The statistical features which are the mean value of an intensity of color in the red plane are extracted and collected as a feature vector. The feature vectors would be analysed using a classification algorithm. Naïve Bayes classifier and SVM are applied in this step. Basic Bayesian theorem (Zhao, N. et al., 2016), (Joshi, C. et al., 2018) is:

$$P(x_i|Y) = \frac{P(Y|x)^* P(x)}{P(Y)} \quad (1)$$

Y is the feature vector of the training mango dataset which is used to calculate the probability of class x (*i* = 1 is ripe, and *i* = 2 is unripe). P(x) and P(Y) are the prior probabilities of the x and Y respectively. P(x|Y) is the probability of x when Y is given, and P(Y|x) is the probability of Y when x is given.

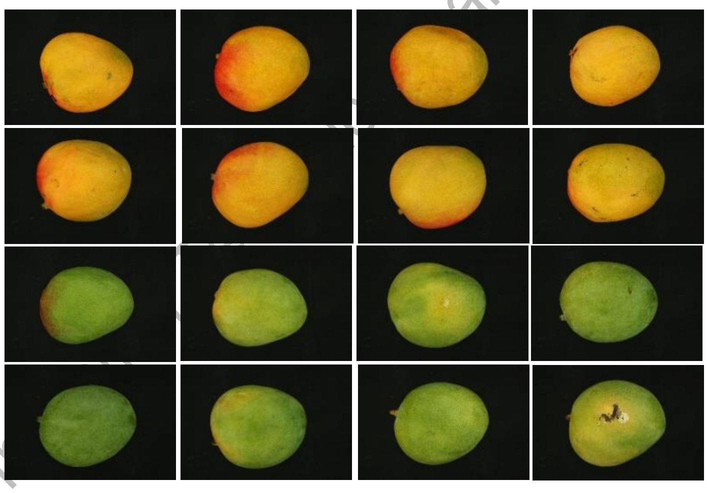
The SVM (Kotsiantis et al., 2014), (Wei, Q. et al, 2017), (Wang, J. et al., 2015), (Liao, C. H. et al., 2018) uses feature vectors of training mango images to classify ripe or unripe mangoes. Given a set of training samples, the SVM training algorithm builds a model that assigns new examples into one of the two categories. A radial basis function (RBF) kernel was used as it yields high performance in terms of accuracy compared to other kernels. The RBF kernel is defined as

$$K(x_i, x_j) = exp\left(-\frac{\|x_i - x_j\|}{2\sigma^2}\right) \quad (2)$$

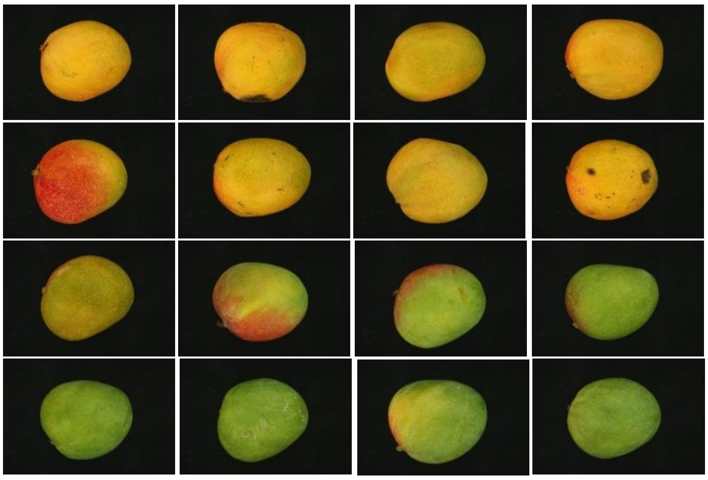
where x_i and x_j are the samples of the class ripe and unripe respectively, and $\sigma > 0$ is a constant that defines kernel width. The optimal kernel function parameter σ was found experimentally to be 0.4.

4.5 Classification

In the classification process, Bayesian theorem is used to calculate the probabilities of two classes using the feature vectors of test dataset. The test mango image is determined as ripe if the probability of class x_1 is higher than the probability class x_2; otherwise, the test mango image is determined as unripe. To evaluate the performance of the classification process, we compare the performance of Bayesian classification and SVM classification. For SVM classification, the feature vector of the test mango image is classified into two class using support vector derived from a model of RBF kernel from the training process. The test image is determined as ripe if the prediction score of class ripe is higher than the prediction score of class unripe. Example of the train images and test images are shown Figure 9.



(a)



(b)

Figure 9. Example of image data set, a) example of train image, and b) example of test image.

5. Result and discussion

In the experiment, 54 ripe mango images and 46 unripe mango images were taken as a test set. The test set was tested with the proposed algorithm to classify mangoes as ripe and unripe using the Bayesian theorem and was compared to SVM. In the training process, 26 ripe and 24 unripe mango images which are included in the testing set were used as a training set. The results are shown in Table 1.

Table 1. Experimental results

Method	Testing mango image								Accuracy (%)
	Ripe (54 images)				Unripe (46 images)				
	Correct	%	Incorrect	%	Correct	%	Incorrect	%	
Surface defect detection	45	88.33	9	16.67	40	86.96	6	13.04	85.00
Bayesian classification	50	92.59	4	7.41	40	86.96	6	13.04	90.00
SVM classification	46	85.19	8	14.81	37	80.43	9	19.57	83.00

From Table1, the surface defect detection process achieved 85% accuracy. In the classification process to classify test mangoes into 2 group; ripe and unripe, the proposed Bayesian classification provided 90% accuracy, better than SVM classification which provided 83% accuracy.

The false detection of surface defect detection was caused by the very small defect on the mango surface. The ripeness classification uses the statistical feature which is mean value of the red plane which has some overlapping in the mean values of the ripe and unripe mangoes; this is caused the classification method to the false result.

6. Conclusion

This paper proposed the use of classification algorithms to detect mango surface defects and classify mangoes into two classes: ripe and unripe. To detect mango surface defects, image thresholding and image labeling were used. Mean values are applied to extract features for the classification process using Bayesian classification. The proposed Bayesian-based method was compared to the SVM classification method. The results validate the proposed Bayesian-based method having a higher performance than the SVM classification.

The limitation of this study is that the proposed algorithm could classify mango images using color features but could not accurately classify mangoes that were ripe with a green surface color and mango that are always yellow although still unripe.

General Comments: Please check the APA referencing. I have tried to correct some of the intext citations having et al.

7. References

Cubero, S. et al. (2014). A new method for pedicel/peduncle detection and size assessment of grapevine berries and other fruits by image analysis. **Biosystems engineering,** Vol.117, 62-72.

Eyarkai Nambi, V., Thangavel, K., Shahir, S., & Thirupathi, V. (2016). Comparison of Various RGB Image Features for Nondestructive Prediction of Ripening Quality of

"Alphonso" Mangoes for Easy Adoptability in Machine Vision Applications: A Multivariate Approach. **Journal of Food Quality,** Vol.39(6), 816-825.

Froese, R., Thorson, J. T., & Reyes, R. B. (2014). A Bayesian approach for estimating length-weight relationships in fishes. **Journal of Applied Ichthyology,** Vol.30(1), 78-85.

Ghanad, N. K., & Ahmadi, S. (2015). Combination of PSO Algorithm and Naive Bayesian Classification for Parkinson Disease Diagnosis. **Advances in Computer Science: an International Journal,** Vol.4(4), 119-125.

Huang, X. et al. (2018). Integration of computer vision and colorimetric sensor array for nondestructive detection of mango quality. **Journal of Food Process Engineering,** Vol.41(8), e12873.

Joshi, C., Ruggeri, F., & Wilson, S. P. (2018). Prior Robustness for Bayesian Implementation of the Fault Tree Analysis. **IEEE Transactions on Reliability,** Vol.67(1), 170-183.

Kotsiantis, S. B. (2014). Integrating global and local application of naive bayes classifier. **Int. Arab J. Inf. Technol.,** Vol.11(3), 300-307.

Liao, C. H., & Wen, C. H. P. (2018). SVM-Based Dynamic Voltage Prediction for Online Thermally Constrained Task Scheduling in 3-D Multicore Processors. **IEEE Embedded Systems Letters,** Vol.10(2), 49-52.

Mim, F. S., Galib, S. M., Hasan, M. F., & Jerin, S. A. (2018). Automatic detection of mango ripening stages–An application of information technology to botany. **Scientia Horticulturae,** Vol.237, 156-163.

Nagle, M. et al. (2016). Determination of surface color of 'all yellow'mango cultivars using computer vision. **International Journal of Agricultural and Biological Engineering,** Vol.9(1), 42-50.

Nandi, C. S., Tudu, B., & Koley, C. (2014). Computer vision based mango fruit grading system. **In Proceedings of International Conference on Innovative Engineering Technologies (ICIET),** 1-5.

Othman, M., Bakar, M. N. A., Ahmad, K. A., & Razak, T. R. (2014). Fuzzy ripening mango index using RGB colour sensor model. **Researchers World,** Vol.5(2), 1.

________. (2016). Mango Size Classification Using RGB Color Sensor and Fuzzy Logic Technique. **In Regional Conference on Science, Technology and Social Sciences (RCSTSS 2014)** (pp. 287-296). Singapore, Springer

Rungpichayapichet, P., Mahayothee, B., Nagle, M., Khuwijitjaru, P., & Müller, J. (2016). Robust NIRS models for non-destructive prediction of postharvest fruit ripeness and quality in mango. **Postharvest Biology and Technology,** Vol.111, 31-40.

Sa'ad, F. S. A., Ibrahim, M. F., Shakaff, A. M., Zakaria, A., & Abdullah, M. Z. (2015). Shape and weight grading of mangoes using visible imaging. **Computers and Electronics in Agriculture,** Vol.115, 51-56.

Thairat news. (2016). **Thai mango export market,** [online]. Available https://www.thairath.co.th/clip/45706. access on 08/04/2016. (in Thai)

Vyas, A. M., Talati, B., & Naik, S. (2014). Quality Inspection and Classification of Mangoes using Color and Size Features. **International Journal of Computer Applications,** Vol.98(1).

Wang, J. C., Lian, L. X., Lin, Y. Y., & Zhao, J. H. (2015). VLSI design for SVM-based speaker verification system. **IEEE Transactions on Very Large Scale Integration (VLSI) Systems,** Vol.23(7), 1355-1359.

Wei, Q., Wu, B., Xu, D., & Zargari, N. R. (2017). Natural sampling SVM-based common-mode voltage reduction in medium-voltage current source rectifier. **IEEE Transactions on Power Electronics,** Vol.32(10), 7553-7560.

Zhao, N., Basarab, A., Kouamé, D., & Tourneret, J. Y. (2016). Joint segmentation and deconvolution of ultrasound images using a hierarchical Bayesian model based on generalized Gaussian priors. **IEEE transactions on Image Processing,** Vol.25(8), 3736-3750.