Automated grading of mangoes based on surface defect detection using a combined approach of image segmentation

Kumari N.,✉[1] **Bhatt K. A.,**[2] **Dwivedi K. R.,**[3] **Belwal R.**[4]

Abstract

Image segmentation is an essential and critical step in huge number of applications of image processing. Accuracy of image segmentation influence retrieved information for further processing in classification and other task. In image segmentation algorithms, a single segmentation technique is not sufficient in providing accurate segmentation results in many cases. In this paper we are proposing a combining approach of image segmentation techniques for improving segmentation accuracy. As a case study fruit mango is selected for classification based on surface defect. This classification method consists of three steps: (a) image pre-processing, (b) feature extraction and feature selection and (c) classification of mango. Feature extraction phase is performed on an enhanced input image. In feature selection PCA methodology is used. In classification three classifiers BPNN, Naïve bayes and LDA are used. Proposed image segmentation technique is tested on online dataset and our own collected images database. Proposed segmentation technique performance is compared with existing segmentation techniques. Classification results of BPNN in training and testing phase are acceptable for proposed segmentation technique.

Key Words: *BPNN , Classification, Computer vision, Machine learning, Naïve bayes*

Introduction

Mango (*Mangifera indica* L.) called king of fruits in India is one of the most important fruits of the world, found mostly in tropical and subtropical regions (Nandi *et al.,* 2014). Image analysis success in digital image processing depends on segmentation algorithm accuracy used to divide input image in various segments for retrieving information according to requirement, (Gonzalez and Woods 2002). A huge number of researchers have worked on various segmentation algorithms for increasing segmentation accuracy of existing segmentation algorithms. Blasco *et al.* (2007), worked on peel defect detection of citrus fruit using computer vision. In segmentation process region oriented segmentation algorithm was used. Accuracy of defect detection and classification was 95%. Lucchese and Mitra (1999), proposed unsupervised low frequency driven segmentation for colored image segmentation using K-means clustering algorithm for segmentation. Arivazhagan *et al.* (2013), worked on

lesion spot segmentation and used K-means clustering algorithm for image segmentation. Qin *et al.* (2008), used Otsu's thresholding algorithm in segmentation phase of image. Leemans *et al.* (1998), worked on classification of golden delicious apples. In segmentation phase for surface defect detection K-means clustering algorithm is used. Kumari *et al.* (2019), worked on classification of mangoes based on external quality features .In segmentation of this study k-means clustering and FCM algorithms was used. Pham and Lee (2015), worked on surface defect detection of fruit using image segmentation technique. In image segmentation a combined approach of k-means clustering and graph based algorithm was proposed. In its first step k-means clustering was used for splitting original image in various segments based on pixel intensity and after that graph methodology was used for merging similar regions for increasing homogeneity of segmented area. Momin *et al.* (2017), worked on mangoes classification based on estimated mass using image processing. The above research discussed results of various shortcomings. Like, in case of low contrast image, contrast enhancement is essential as it affects the defected

Author's Address

[1] U.K.T.U. Dehradun, Uttrakhand, India.
[2] Birla Institute, Bhimtal, India
[3] CCSIT, TMU, Moradabad
[4] Amrapali Institute Haldwani, India
E-mail.: *neelu.rajpoot80@gmail.com*

part segmentation accuracy and segmentation accuracy will affect classification accuracy decay in classification phase of machine learning. In case of feature extraction and selection, affective feature extraction and selection approach is also crucial step for such type of applications. In this article considered problems are (a) problem of noise due to capturing devices and poor environment (b) image segmentation accuracy (c) feature extraction and selection. The above-listed problems are resolved by implementation of a new hybrid technique which is comprised of five major steps: (a) implementation of background removing and contrast enhancement technique for highlighting defected region on mango surface. (b) defected part segmentation by a combination of existing segmentation techniques (c) construction of codebook by simple concatenation based fusion of color, texture and geometric features, (d) selection of most prominent features by feature selection method, PCA (e) BPNN, Naïve bayes, LDA are used for classification.

Our prime contributions in this research are as follow:

(a). In preprocessing step median filter is used to remove noise and image morphological functions are used for enhancing image contrast which helps in effectively segmenting the defected part of mango. Enhanced defected part of mango is extracted by a combination of various segmentation methods like: binarization of image, K-mean clustering, FCM technique and Otsu's image segmentation. After that best result is selected by using fuzzy rules.

(b) In feature selection PCA technique is used for best features selection. A total of 40 features are used in classification.

(c) In experimental results, two databases are used one, online dataset and second own created dataset. These two datasets are further divided in one local dataset and one mixed dataset. Each local and mixed dataset has three categories of mango: 1. unripe, 2. ripe (mature), 3. Over ripe. Based on extracted features mango is classified in healthy and defected classes. The chronological order of the article is as follows: Section 2 presents the methodology and data collection. Image segmentation and feature extraction is presented in section 3. Experimental results and conclusion are explained in section 4 and 5 respectively.

Material and Methods

Methodology of this study is displayed in figure 1. In this complete study experimentation is performed on MATLAB 2017b that includes image pre-processing, image feature extraction and selection and classification. Rest part of this section includes data collection, image pre-processing and color feature extraction phase for maturity prediction.

Data collection

In this experiment total 1256 mango images are collected from a mango orchid in the same environment between May 2018 and June 2018. Mango data was collected in three stages of mango that was unripe, ripe and over ripe. Camera was placed at a distance of 30 cm from the sample. An individual sample image is captured two times in two directions for accurate maturity prediction. Collected data set samples are displayed in figure 3. Collected data set is divided in 80:20. 80% of dataset is used in training phase of machine and 20% is used in the testing phase of machine. In this study we also compared classification accuracy by varying number of samples used in training phase of machine.

Image pre-processing

Preprocessing is an essential step in image processing and used to enhance the visual quality of input image. This process is used for removing various problems brightness effects, illumination etc and main source of these problems are poor contrast image. In this study, image pre-processing phase includes two processes like: a) background removing b) contrast enhancement.

Color Feature extraction and maturity prediction

In maturity prediction of mango RGB, HSV, HSI, L*a*b color space models are used after image preprocessing. From each color space model mean, SD, median and variance color features are extracted from each color band. A total of 1x48 vector is created for color features. RGB is one of the most popular color model used by huge number of researchers in various applications of image processing like agriculture applications, medical applications, remote sensing applications etc. In
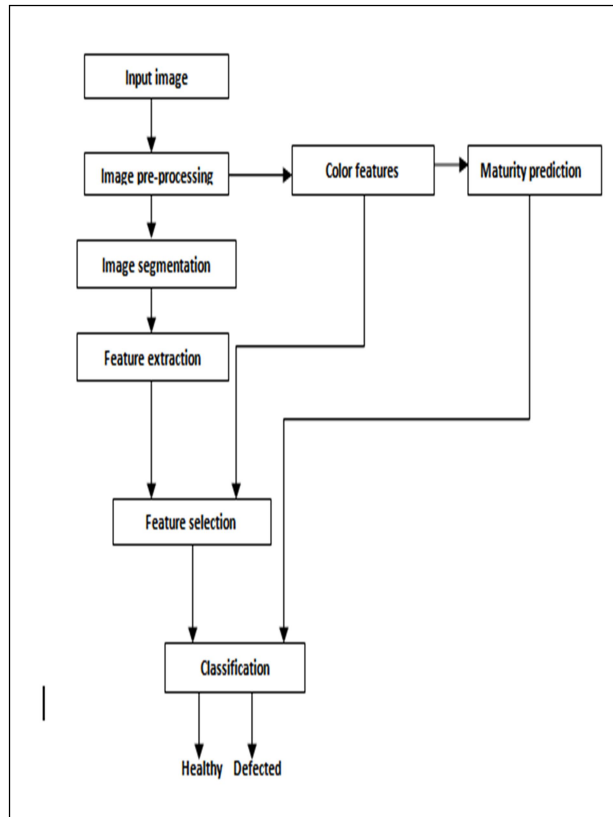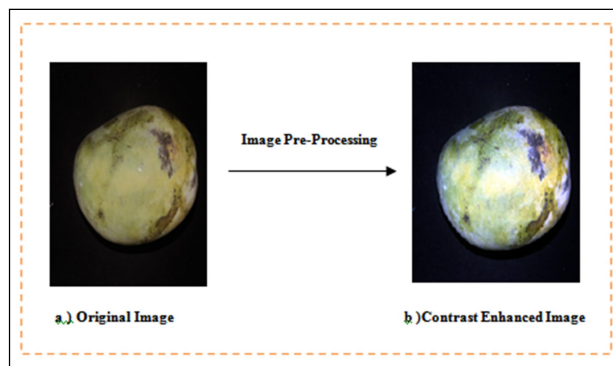
Fig 1. Proposed methodology.

Fig 2. Contrast Enhancement.

this study RGB model is used for maturity prediction and rest of color models are used in validation phase of maturity.

Image segmentation & feature extraction
This proposed method of classification passes through the primary steps used in image classification using computer vision. Those are in sequence (a) image enhancement, (b) color feature

extraction for maturity prediction (c) image segmentation in infected surface detection of mango, (d) feature extraction, selection and construction of codebook and (e) classification displayed in figure 1. In rest part of this section we discussed positive and negative points of used segmentation techniques

Image segmentation
In classification image segmentation plays a vital role. In past researchers has worked on various image segmentation algorithms like K_means clustering, Otsu's ,FCM, threshold, region growing etc, segmentation accuracy varies from one algorithm to another algorithm . A single algorithm does not provide satisfactory results. Variation in segmented result accuracy will influence classification result also. In agriculture science fruit quality testing and grading is also based on defected part of fruit. Each used segmentation algorithm has some positive and negative points listed below:

Image binarization: Image binarization is one of the pixels based technique in segmenting process. This is simplest and widely used method in grey level image. Advantage of this method is to reduce data complexity in retrieving required information. A threshold value is selected and based on that value pixel intensity is converted into white and black color. Disadvantage of this technique is to provide information in two classes only, black pixels class and white pixels class. This technique does not provide satisfactory segmentation of multichannel image.

K-means: K-means clustering is simple, flexible, and suitable in large dataset in image segmentation. It improves clustering accuracy. By using K-means clustering results can be easily interpreted. Compared to other segmentation algorithms k-means produce tighter clusters especially with spherical clusters. Due to some negative points like no-optimal set of clusters, lack consistency, uniform effect, sensitivity of scale, specification of k values initially, prediction issues and working only on numerical data in many cases satisfactory results of image segmentation are not produced by this algorithm.

Otsu's: Otsu's method working is similar to other binarization algorithms used for segmentation. In its advantages as compared to other threshold segmentation technique, are seedy and easy in

(a) Un-ripe mango

(b) Ripe mango

(c) Overripe mango

Fig 3. Collected dataset samples (a), (b), (c).

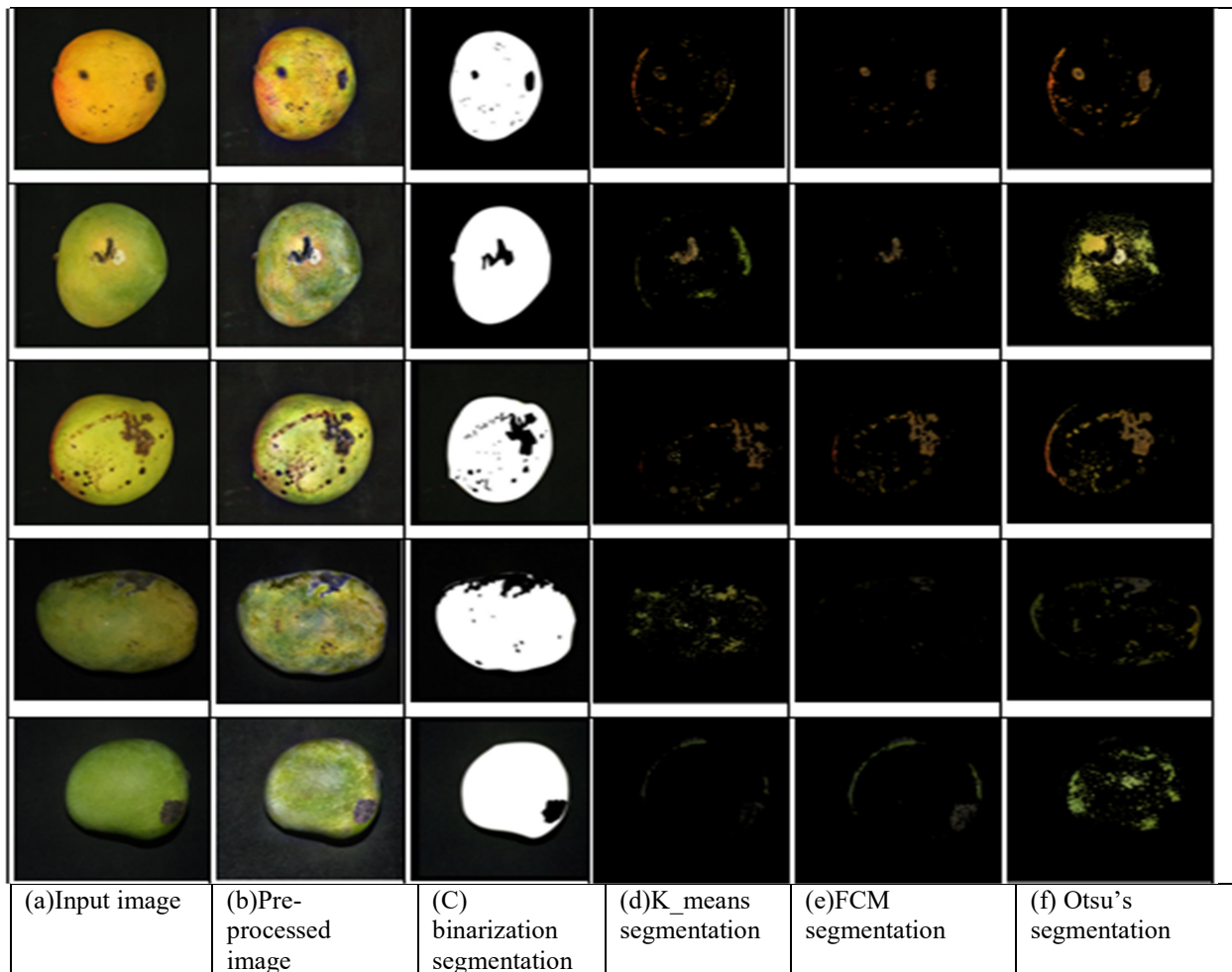| (a)Input image	(b)Pre-processed image	(C) binarization segmentation	(d)K_means segmentation	(e)FCM segmentation	(f) Otsu's segmentation
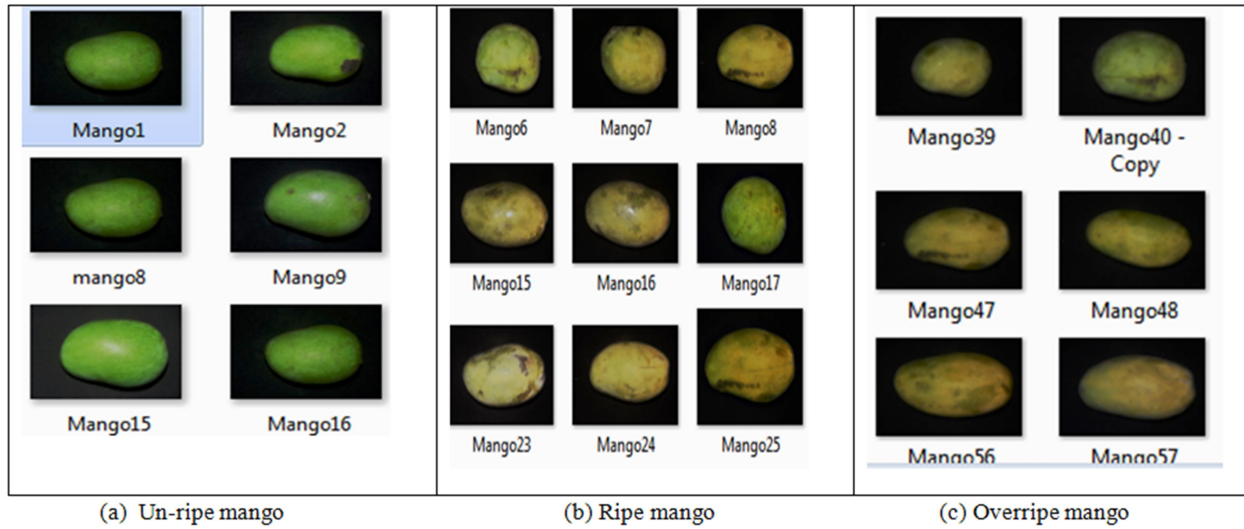
Fig 4. Segmentation results. (a) Input image (b) Pre-processed image (c) binarized segmentation (d) K_means segmentation (e) FCM segmentation (f) Otsu's segmentation.

implementation. In disadvantages it includes assumption on uniform illumination, bimodal histogram, assumption on uniform statistics in non-local version and not using any specific structure of object.

FCM: FCM algorithm of image segmentation is more powerful due to its flexibility and producing satisfactory results in real life applications data. FCM is quite straightforward due to its membership function closeness to human perception. However, there are some shortcomings in FCM method like taking large computational time, objective function is not so good for clusters having unequal size, working only on same type of clusters, requirement of parameters in initialization phase of FCM.

Image segmentation using proposed method: This method passes through two phases. In first phase defected area of mango surface is detected using existing algorithms discussed above. Defected area is calculated in form of defected number of pixels. In second phase mean of defected pixels from all segmentation algorithm is calculated and classification is done based on mean value of defected area.

Results and Discussion

Performance of the proposed algorithm of classification is evaluated on mango dataset (www.cofilab.com) and own collected mangoes dataset. Testing of proposed method is performed in three cases: 1) on online dataset 2) on own dataset 3) on mixed dataset. In the third experiment the recognition results are obtained on own collected database and online dataset having three types of mangoes (unripe, ripe and overripe). In classification three classifiers BPNN, LDA, and Naïve bayes are used. Mango is classified in two categories healthy and defected. Measurement units used in this study are True Positive (TP), True Negative (TN), False Positive (FP), False Negative (FN), False Positive Volume (FPV), False Negative Volume (FNV) and Accuracy.

Classification based on proposed segmentation method is tested through discussed classifiers. Classifiers results of training dataset are displayed in table 1, table 2, table 3, table 4 and table5. A total of 1256 mango images are used in training phase of machine. In classification results displayed accuracy represents healthy mangoes classification accuracy Confusion matrix of training and testing phase for healthy and defected mangoes classification using proposed method of segmentation is displayed in table 6, table 7, table8, table 9, table10 and table 11. In confusion matrix displayed results, BPNN is used because results of BPNN are satisfactory in all cases (own dataset, online dataset and mixed dataset) as compared to Naïve bayes and LDA classifiers.

Table 1. Training dataset performance measures of classification based on proposed segmentation method.

Classifiers	TP	TN	FP	FN	FPV	FNV	Accuracy (%)
Healthy/ defected mangoes							
BPNN	328	924	4	0	0.0043	0	99.6815
Naïve bays	382	765	58	51	0.0705	0.1178	91.3217
LDA	433	710	51	62	0.067	0.1253	91.0032

Table 2. Training dataset performance measures of classification based on K-means clustering method

Classifiers	TP	TN	FP	FN	FPV	FNV	Accuracy (%)
Healthy/defected mangoes							
BPNN	208	872	124	52	0.1245	0.2	85.9873
Naïve bays	312	522	172	250	0.2478	0.444	66.4013
LDA	282	664	158	152	0.1922	0.3502	75.3185

Table 3. Training dataset performance measures of classification based on FCM method.

Classifiers	TP	TN	FP	FN	FPV	FNV	Accuracy (%)
Healthy/defected mangoes							
BPNN	132	924	200	0	0.1779	0	84.0764
Naïve bays	484	316	0	456	0	0.4851	63.6943
LDA	282	664	158	152	0.1922	0.3502	75.3185

Table 4. Training dataset performance measures of classification based on Otsu's method.

Classifiers	TP	TN	FP	FN	FPV	FNV	Accuracy (%)
Healthy/defected mangoes							
BPNN	355	654	124	23	0.1032	0.2367	80.0253
Naïve bays	309	519	0	428	0.0230	0.4031	65.3662
LDA	277	645	157	167	0.1922	0.3502	73.4076

Table 5. Training dataset performance measures of classification based on binarization method.

Classifiers	TP	TN	FP	FN	FPV	FNV	Accuracy (%)
Healthy/defected mangoes							
BPNN	215	883	111	47	0.1052	0.19	87.4203
Naïve bays	480	312	8	456	0.1407	0.3921	63.0573
LDA	273	657	163	158	0.1922	0.3502	74.0422

Table 6. Confusion matrix of training using proposed method for online dataset.

Mango image	No. of images	Unhealthy	Healthy
Defected	100	98.7%	1.3%
Healthy	100	1%	99%

Table 7. Confusion matrix of testing using proposed method for online dataset.

Mango image	No. of images	Unhealthy	Healthy
Defected	50	96%	4%
Healthy	50	2.3%	97.7%

Table 8. Confusion matrix of training using proposed method for own dataset.

Mango image	No. of images	Unhealthy	Healthy
Defected	628	97.23%	2.77%
Healthy	628	2%	98%

Table 9. Confusion matrix of testing using proposed method for own dataset.

Mango image	No. of images	Unhealthy	Healthy
Defected	100	100%	0%
Healthy	100	.32%	99.68%

Table 10. Confusion matrix of training using proposed method for mixed dataset.

Mango image	No. of images	Unhealthy	Healthy
Defected	722	100%	0%
Healthy	722	3%	97%

Table 11. Confusion matrix of testing using proposed method for mixed dataset.

Mango image	No. of images	Unhealthy	Healthy
Defected	150	98.02%	1.98%
Healthy	150	5.3%	94.97%

Conclusion

In this study a proposed approach of image segmentation algorithm's produced results are satisfactory as compared to traditional image segmentation algorithms like K-means clustering, FCM, Otsu's and image binarization. Used image pre-processing step highlighted defected portion that helped in affective segmentation of required defected part of mango that in future improved classification accuracy. Proposed method segmentation results comparison is acceptable and satisfactory as compared to existing methods segmentation results. Classification is done for online, own and mixed dataset. BPNN, LDA, and Naïve bayes classifiers classified mixed dataset. BPNN classification results are good as compared to Naïve bayes and LDA. In classification binary classification is performed. In future we will work on deep learning for increasing classification

References

Arivazhagan, S., Shebiah, R. N., Ananthi, S. and Vishnu V. S. 2013. Detection of unhealthy region of plant leaves and classification of plant leaf diseases using texture features. *Agricultural Engineering International: CIGR Journal,* 15(1): 211–217.

Blasco, J., Aleixos, N. and Molto, E. 2007. Computer vision detection of peel defects in citrus by means of a region oriented segmentation algorithm. *Journal of Food Engineering*, *Elsevier.* 81(3): 535-543.

Gonzalez, R. C. and Woods, R. E. 2002. *Digital image processing*. NJ, USA: Prentice Hall.

Kumari, N., Bhatt, A. K., Dwivedi, R. K. and Belwal, R. 2019. Performance analysis of support vecor machine in

defective and non defective mangoes classification," *,IJEAT,* 8(4): 1563-1572.

Leemans, V., Magein, H. and Destain, M. F. 1998. Defect segmentation on 'golden delicious' apples by using colour machine vision. ***Computers and Electronics in Agriculture,*** 20(2): 117-130

Lucchese, L. and Mitra, S. K. 1999. *Unsupervised low-frequency driven segmentation of color images*. In Proceedings of the 1999 IEEE International conference on image processing (pp. 240–244). Kobe, Japan.

Momin, M. A, Rahman, M. T., Sultana, M. S., Igathinathana, C. and Ziauddin, A. T. M. 2017. Geometry based mass grading of mango fruits using image processing", ***Information processing in agriculture,*** 4(2): 150-160.

Nandi, C. S, Tudu, B. and Koley, C. 2014. Computer vision based mango fruit grading system, ***ICIET***. 1-5.

Pham, V. H. and Lee B. R. 2015. An image segmentation approach for fruit defect detection using k-means clustering and graph-based algorithm. ***Vietnam Journal of computer science,*** 2: 25-33.

Qin, J., Burks, T. F., Kim, D. G. and Bulanon, D. M. 2008. Classification of citrus peel diseases using color texture feature analysis. In *Food Processing Automation Conference Proceedings, 28-29 June 2008, Providence, Rhode Island* (p. 31). ***American Society of Agricultural and Biological Engineers.***