# Hybrid Approach for Apple Fruit Diseases Detection and Classification Using Random Forest Classifier

Bhavini J. Samajpati and Sheshang D. Degadwala

***Abstract*—Nowadays, abroad trade has expanded definitely in numerous nations. Plenty fruit products are foreign from alternate countries, for example, oranges, apples and so forth. Manual distinguishing proof of infected fruit is extremely tedious. The utilization of image processing procedures is of outstanding implication for the analysis of agro based applications. In any case, detection of infections in the fruit products utilizing images is still risky because of the regular changes of skin color in distinctive sorts of fruit products. In this paper three normal infections of apple fruit are considered i.e. Apple scab, apple rot and apple blotch. The image processing based proposed methodology is made out of the accompanying some state of the art color and texture features are extracted from the test image, then color and texture features are fused together and random forest classifier is used for diseases classification and if the fruit is infected by any of the one disease then the infected part is segmented using k-means clustering technique. The accuracy of the diseases classification will improve by feature level fusion.**

***Index Terms*—K-Means clustering; Color features; Texture features; Random forest classifier**

## I. INTRODUCTION

In the production of fruits india has a second rank[2].Sixty percent population is having employment by agriculture domain[3].The established methodology for recognition of fruit product illnesses depends on the naked eye perception by the specialists. In some creating nations, counselling specialists are costly and tedious because of the inaccessible area of accessibility.

The other areas like twigs, leaves, and branches of the tree is also infected by this infections. Apple rot, apple blotch and apple rot are some most mutual diseases of apple fruits[1], as shown in Fig.1.If apple fruit is having gray or brown corky spots then it is apple scab diseases[1]. If the surface of the apple fruit is slightly sunken, circular brown or black spots that may be covered by a red halo then it is referred as apple rot type diseases[1]. The surface of the apple fruit is having dark, irregular or lobed edges then it is apple blotch type diseases[1]. Which is the fungal diseases of the apple fruit.

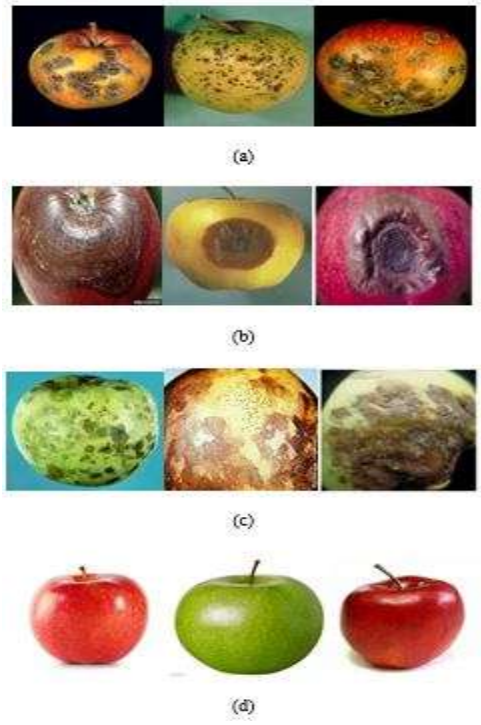


Fig. 1. Three apple fruit diseases: (a) apple scab (b) apple rot (c) apple blotch (d) normal apple.

## II. RELATED WORK

Shivram Dubey et al(and other)(2012)[1].In this approach three apple diseases have been taken apple scab, apple rot and apple blotch[1].Image segmentation is carried out using K-means clustering. In the next step features are extracted like global color histogram(GCH), color coherence vector(CCV), local binary pattern(LBP) and complete local binary pattern(CLBP).

Shivaram Dubey et al(and others)(2013)[4]. Defect segmentation is done into two stages. Utilizing this two stage system, the computational efficiency can be improved. Likewise also gives a brief overview of the various segmentation methods.

Bhavini J. Samajpati is M.E. Student of Computer Department at Sigma Institute of Engineering, Vadodara, Gujarat, India.(e-mail: bhavini10x11@gmail.com)

Sheshang D. Degadwala is Assistant Professor of Computer Department at Sigma Institute of Engineering, Vadodara, Gujarat, India.(e-mail: sheshang13@gmail.com)

978-1-5090-0396-9/16/$31.00 ©2016 IEEE

R.Sivamoorthi et al (and other)(2015)[5].The features exraction is carried out from the images using Color Invariant, Color histogram and Local Binary pattern algorithm. The neural network classifier is used for classification.

Jagadeesh. D. Pujari et al (and other)(2013)[6]. The RGB color features are reduced from 18 to 2 and GLCM texture features are reduced from 30 to 2. The reduced features set of comprises of green mean, saturation mean, red GLCM sum mean and green GLCM sum mean. A BPNN classifier is suitable for this work.

Monika Jhuria et al (and other)(2013)[7]. In this paper three defects of grapes fruit and two of apple fruit have been picked. From these features 90% accuracy is achieved.

Greg Pass et al(and others)[8].In this paper color histograms are used to look at images in various applications. A color coherence vector (CCV) contains the several coherent pixels with every color.

Wen-Hung Liao et al(2010)[11]. In this approach they attempted to upgrade the original LBP by a novel expansion that is extended local ternary pattern (ELTP).

Basvaraj .S. Anami et al (and other)(2011)[12].In this paper an ANN based classifier is received which utilizes the combination of color and texture features to perceive and categorized different agriculture produce.

Madasamy Raja et al(and other)(2013)[13]. In this approach Local Ternary Pattern(LTP) is described which is dependable on the LBP feature. This system is strong to image rotation, gray scale transformation, histogram exploit and noise.

## III. METHODOLOGY

In this paper we have proposed some color and texture features are extracted from the test images. After the feature extraction some color and texture features are fused together for better accuracy. Finally the apple fruit diseases are classified using random forest classifier and then if the fruit is infected by any of the diseases then the diseased part is segmented using k-means clustering technique.

### *A. Image Segmentation*

Segmentation means finding region of interest from the image[3].

K-means is a clustering algorithm discovered by MacQueen in 1967 [4]. K-means is generally carried to find the infected regions in an image.

### *B. Feature Extraction*

Features are extracted from the fruit. Based on the features of the fruit the classification of fruit diseases is done. Extract various features from fruit like color of fruit, texture and shape [3].

In the proposed approach some color and texture features are used.

*1) Color Features*

The color features utilized as a part of the fruit and vegetable classification/fruit diseases identification are Global Color Histogram, Color Coherence Vector.

*a) Global Color Histogram:* It is the simple methodology to translate the information available in an image [1]. For every different color GCH is a group of values which shows the probability. For decreases the different color and scaling preference constant normalization and quantization are used[1]. It the Simplest approach for separate color.

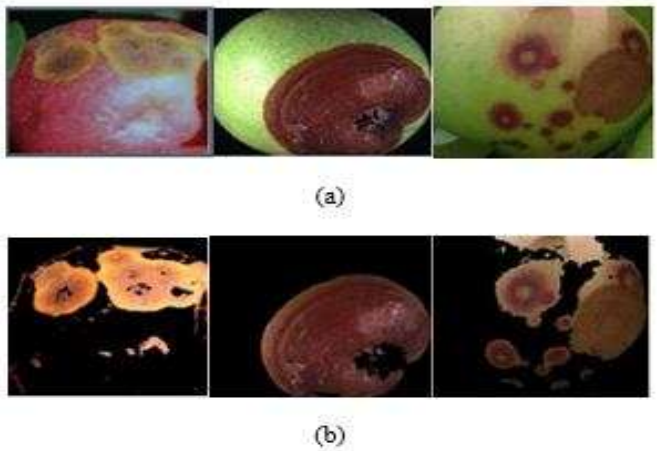


Fig. 2. Some defect segmentation of diseased fruit[4].

*b) Color Coherence Vector:* Some sizable contiguous region is having coherent pixels, while incoherent pixels are not[1]. The color space and the variations between the neighboring pixels are eliminate to calculate ccvs[1].After that the classification of that pixels either coherent or incoherent by the connected parts in the image[1]. It is efficient and invariant to minor changes.

*2) Texture Features*

The texture features utilized as a part of the fruit and vegetable classification/fruit diseases identification are local binary pattern, complete local binary pattern, and local ternary pattern.

*a) Local Binary Pattern:* It is calculated by comparing the image pixels with its neighbors[1]:

$$LBP_{N,R} = \sum_{n=0}^{n-1} s(v_n - v_c)2^n, s(x) = \begin{cases} 1, x \geq 0 \\ 0, x < 0 \end{cases} \quad (1)$$

Where, vc = central pixel value,

vn = value of neighbours,

R = radius of the neighbourhood

N = total number of neighbours.

Assume the coordinate of vc is (0,0), then the coordinates of vn are (Rcos(2π n / N), Rsin(2π n /N)) .Image dimensions is I*J.Then calculated the texture image formed a histogram:

$$H(k) = \sum_{i=1}^{I} \sum_{j=1}^{J} f(LBP_{N,R}(i,j),k), k \in [0.K],$$

$$f(x,y) = \begin{cases} 1, x = y \\ 0, otherwise \end{cases} \quad (2)$$

Where, K is the maximal LBP code value.

*b) Complete Local Binary Pattern*: It is a feature which considers both sign (S) and magnitude (M) and center gray

level (C)[1]. It is the mixture of three features, that is CLBP_S, CLBP_M, and CLBP_C[1]. CLBP_S is that simillar and used to code the sign details of local differences. The code magnitude information of local information uses by CLBP_M [1]:

$$CLBP_{N,R} = t(g_c, c_1), t(x,c) = \begin{cases} 1, x \geq c \\ 0, x < c \end{cases} \quad (3)$$

Where, c is threshold and fixed to the mean value of the input image[1].

$$CLBP_{N,R} = \sum_{n=0}^{n-1} t(m_n, c)2^n, t(x,c) = \begin{cases} 1, x \geq c \\ 0, x < c \end{cases} \quad (4)$$

Where, threshold c is set to the average gray level of the input image. It achieves much better rotation invariant texture classification results than conventional LBP-based schemes.

*c) Local Ternary Pattern:* When a 3´3 neighborhood of a centre pixel in an image is found, the LTP operator takes the form Equation[13]:

$$LTP_8 = \sum_{n=1}^{8} 3^{n_s}(i_n - i_c) and s(u) = \begin{cases} -1 if u \leq i_c - t, \\ 0 if i_c - t < u < i_c + t \\ 1 if u \geq i_c + t \end{cases} \quad (5)$$

Where, t = user-defined threshold, n = the number of neighboring pixels close the center pixel c, ic = pixel value of c and in are the pixel values of n.

*d) Gabor Features:* Gabor features is used for many purpose like texture analysis and segmentation etc. After finding the parameters of every filter, compute the convolution of each filter and image, the mean and standard deviation of the original image and each filtered image[18].

For original image mean and standard deviation:

$$\mu_{or} = \frac{\sum_x \sum_y |f(x,y)|}{P \times Q}$$

$$\sigma_{or} = \sqrt{\frac{\sum_x \sum_y (|f(x,y)| - \mu_{or})^2}{P \times Q}} \quad (6)$$

where $f(x, y)$ is original image, $P \times Q$ is original image size.

For filtered image mean and standard deviation:

$$\mu_{sn} = \frac{\sum_x \sum_y |I_{sn}(x,y)|}{P \times Q}$$

$$\sigma_{sn} = \sqrt{\frac{\sum_x \sum_y (|I_{sn}(x,y)| - \mu_{sn})^2}{P \times Q}} \quad (7)$$

where $s$ is number of frequency, $n$ is number of directions, $I(x, y)$ $sn$ is filtered image, $P \times Q$ is size of filtered image.

*C. Random Forest Classifier*

Random Forests (RF) is one of the finest machine learning classification and regression method. It is suitable for classification of large number of dataset[17]. It is having a groups of the tree-structured classifiers[17]. The tree is depends on the random values sampled and the forest. The input is given at top of the tree then down the tree[17]. The data is sampled is randomly, but it is having reduced sets. The sample class is found by random forests trees, which are of a random number [17]. The randomizing variable found how the cuts are found regularly. At the time of construction of the tree by selecting the node and the coordinate to divide and the position of the divided [17].

## IV. PROPOSED WORK

In this section proposed work is deal with feature level fusion. Image fusion is that the technique of joining related information from more than one images into a single image.

*A. Diseases Detection and Classification Procedure*

Input: Apple fruit image.

Output: Classified apple fruit diseases.

Algorithm:

Steps:

1. Load training images.
2. Read testing image.
3. Perform feature extraction using color features i.e. Global color histogram (GCH), Color coherence vector (CCV) and texture features i.e. Local binary pattern (LBP), complete local binary pattern (CLBP), Local ternary pattern (LTP), Gabor Features.
4. Perform feature level fusion using color and texture features.
   - GCH + Gabor
   - GCH + LTP
   - Gabor + GCH + LBP
   - Gabor + LBP
   - Gabor + CLBP
   - Gabor + LTP
   - Gabor+CLBP+LTP

5. Apply Random forest classifier on the segmented image.
   a) Draw Mtree bootstrap samples from the original data.
   b) Draw an un-pruned classification tree for each bootstrap samples.
   c) For every internal node,for N predictors randomly select ntry then find the best split using only those predictor.
   d) Save tree as is, nearby those built thus far (Do not perform cost complexity pruning).
   e) Predict new data by aggregating the Predictions of the Mtree trees.

6. If fruit is infected by any of the disease then goto step 7 otherwise goto step 8.

7. Apply K-means clustering for image segmentation.
   Use Euclidean distance for K-Means clustering.
   a) Load input image.
   b) Convert image from RGB color space to L*a*b* color space.

c) Categories colors utilizing K-Means clustering in 'a*b*' space.

d) Tag every pixel in the image from the outcomes of K-Means.

e) Make images that segment the image by color.

f) Selection of the segmentation area containing disease.

8. Fruit is normal or ordinary fruit.
9. Results the classified apple fruit diseases.

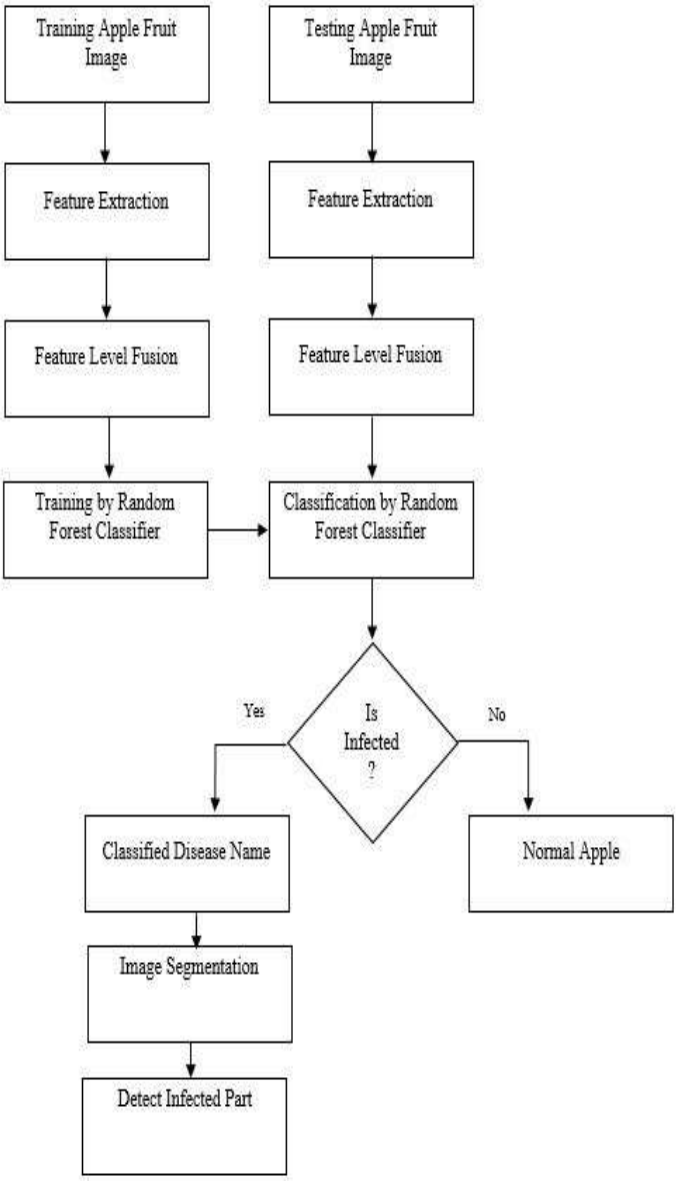


Fig. 3. Framework of the proposed approach.

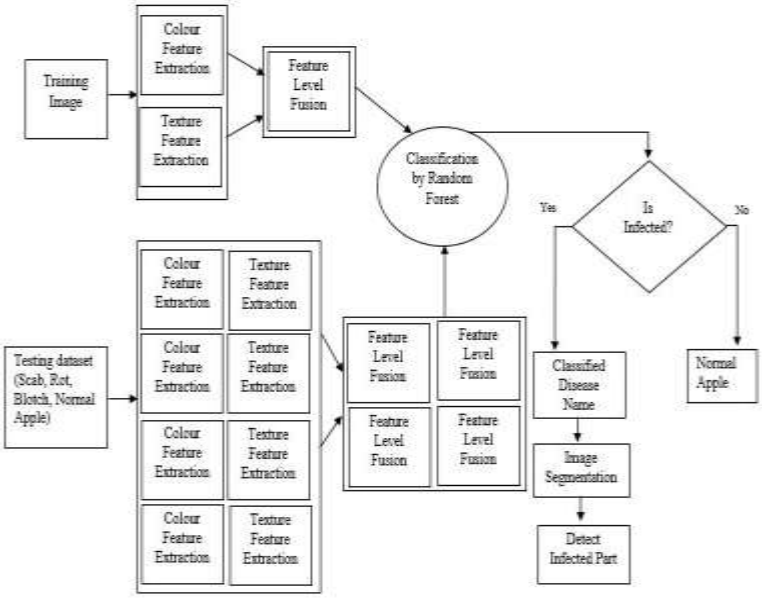


Fig. 4. Block Diagram of the proposed approach.

## V. EXPERIMENTS RESULT

In the phase for finding the best order of defects. If N images is used per class for training then other images are used for testing. The accuracy of the proposed approach is defined as[1],

$$\text{Total Accuracy (\%)} = \frac{\text{number of images correctly classified}}{\text{Total number of images used for testing}} * 100 \quad (8)$$

Compare Accuracy of Random Forest Classifier to Recognize the Apple Diseases based on Features by taking 70 images in training dataset and 10 images in testing dataset per class.

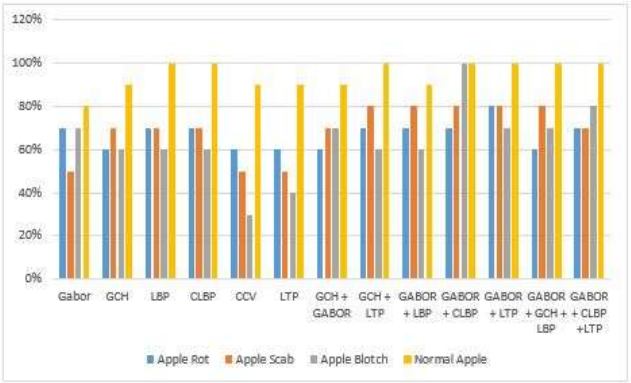


Fig. 5. Comparison of features accuracy based on random forest classifier

TABLE I

APPLE FRUIT DISEASES CLASSIFICATION ACCURACY

| Features | Apple rot | Apple Scab | Apple Blotch | Normal Apple |
|---|---|---|---|---|
| Gabor | 70% | 50% | 70% | 80% |
| GCH | 60% | 70% | 60% | 90% |
| LBP | 70% | 70% | 60% | 100% |
| CLBP | 70% | 70% | 60% | 100% |
| CCV | 60% | 50% | 30% | 90% |
| LTP | 60% | 50% | 40% | 90% |
| GCH + Gabor | 60% | 70% | 70% | 90% |
| GCH + LTP | 70% | 80% | 60% | 100% |
| Gabor + LBP | 70% | 80% | 60% | 90% |
| Gabor + CLBP | 70% | 80% | 80% | 100% |
| Gabor + LTP | 70% | 80% | 60% | 100% |
| Gabor + GCH + LBP | 60% | 80% | 70% | 100% |
| Gabor + CLBP + LTP | 80% | 80% | 80% | 100% |

## VI. CONCLUSION

Here by it is concluded that in apple fruit diseases detection system image segmentation is performed using K-Means clustering techniques. Color and texture features are extracted. Feature level fusion is carried out. By fusing two or more than two color and texture features like GCH, LBP, CCV, LTP, CLBP,gabor the apple fruit diseases classification accuracy may improve and the performance of the diseases classification will improve. Gabor with CLBP and Gabor with LTP gives better result than other single features. That means features fusion give better accuracy than single features. Diseases is classified by random forest classifier.

## REFERENCES


[1] Shiv Ram Dubey, Anand Singh Jalal," Detection and classification of Apple Fruit Diseases Using Complete Local Binary Patterns", 978-0-7695-4872-2/12 IEEE 2012.

[2] India Rank in Agriculture http://www.apeda.gov.in / apedawebsite/six_head_product/FFV.htm.

[3] Bhavini J. Samajpati, Sheshang D. Degadwala,"A Survey on apple Fruit Diseases Detection and Classification" International Journal of Computer Applications (0975 –8887) Volume 130 – No.13, November 2015.

[4] Shiv Ram Dubey, Pushkar Dixit, Nishant Singh, Jay Prakash Gupta,"Infected Fruit Part Detection using K-Means Clustering Segmentation Technique", International Journal of Artificial Intelligence and Interactive Multimedia, Vol. 2, No 2,2013.

[5] R. Sivamoorthi, 2Dr. N. Sujatha," A Novel Approach of Detection and Classification of Apple Fruit Based on Complete Local Binary Patterns", International Journal of Advanced Research in Computer Science and Software Engineering, Volume 5, Issue 4, April 2015.

[6] Jagadeesh. D. Pujari, Rajesh. Yakkundimath, A. S. Byadgi," Reduced Color and Texture features based Identification and Classification of Affected and Normal fruits' images", International Journal of Agricultural and Food Science 2013.

[7] Monika Jhuria, Ashwani kumar,Rushikesh Borse,"Image processing for smart farming: detection of diseases and fruit grading ", 978-1-4673-6101-9/13/2013 *IEEE* .

[8] Greg Pass, Ramin Zabih,, Justin Miller," Comparing Images Using Color Coherence Vectors", Computer Science Department Cornell University O. Kleynen , V. Leemans, M.-F. Destain," Development of a multi-spectral vision system for the detection of defects on apples",Elsevier 2004.

[9] Nikita Rishi, Jagbir Singh Gill," An Overview on Detection and Classification of Plant Diseases in Image Processing", International Journal of Scientific Engineering and Research (IJSER), Volume 3 Issue 5, May 2015.

[10] Wen-Hung Liao," Region Description Using Extended Local Ternary Patterns", 1051-4651/10 2010 IEEE.

[11] Basvaraj .S. Anami1, J.D. Pujari2, Rajesh. Yakkundimath," Identification and Classification of Normal and Affected Agriculture/horticulture ProducemBased on Combined Color and Texture Feature Extraction", International Journal of Computer Applications in Engineering Sciences, vol i, issue iii, september 2011.

[12] Madasamy Raja, G. and V. Sadasivam," optimized local ternary patterns: a new texture model with set of optimal patterns for texture analysis", Journal of Computer Science 9 (1): 1-15, 2013.

[13] Guang-Hai Liu a,b,n, Jing-YuYang," Content based image retrieval using color difference histogram" Elsevier 2012.

[14] Wen-Hung Liao," Region Description Using Extended Local Ternary Patterns", 1051-4651/10 2010 *IEEE* .

[15] Uravashi Solanki, Udesang K. Jaliya and Darshak G. Thakore ," A Survey on Detection of Disease and Fruit Grading", International Journal of Innovative and Emerging Research in Engineering Volume 2, Issue 2, 2015.

[16] Hossam M. Zawbaa,Maryam Hazman, Mona Abbass, Aboul Ella Hassanien,"Automatic fruit classification using random forest algorithm",978-1-4799-7633-1/14/2014 I*EEE*.

[17] zhi-chun huang, patrick p. k. chan, wing w. y. ng, daniel s. yeung," content-based image retrieval using color moment and gabor texture feature", Proceedings of the Ninth International Conference on Machine Learning and Cybernetics, Qingdao, 11-14 July 2010 *IEEE*.