Proceedings of the Second International Conference on Electronics and Sustainable Communication Systems (ICESC-2021)
IEEE Xplore Part Number: CFP21V66-ART; ISBN: 978-1-6654-2867-5

Identification of Citrus Fruit Defect using Computer Vision System

Akshatha Prabhu [1], Likhitha S [2], Sangeetha K.V [3]
Department of Computer Science,
Amrita School of Arts and Sciences,Mysuru
Amrita Vishwa Vidyapeetham, India

akshathaprabhu06@gmail.com
likhithasudhir13@gmail.com
sangeethakv22@gmail.com

Abstract— Agronomy production accounts for a significant portion of the Indian economy. Since there is a high demand for fruits in the market, fruit disease detection is a critical component, the incorporation of computer vision systems in agriculture also deal with the identification and classification of fruit diseases. There is an increasing need for detecting flaws in various fruits, which is laborious and ineffective. The procedure identifies citrus fruit illnesses using several classifiers and assesses the accuracy and effectiveness of each algorithm. This is an involuntary recognition system and the procedure is both labor-intensive and cost-effective. This research work has performed a comparative analysis of several classifiers such as KNN, SVM, and DT by providing the image of a sick fruit as an input. By using the proposed classicification model, upto 97% accuracy can be obtained. The proposed system has major advantages such as user friendliness, good reliability etc. The primary goal of the proposed strategy is to increase the consumer knowledge about fruit illness and whether or not they should be consumed.

Keywords— K-Means, KNN, SVM, Decision Tree, Adaptive Histogram Equilisation, GLCM.

I. INTRODUCTION

Citrus fruit is considered as the most general fruit by means of ironic nutritious cost and sole flavor. It is majorly developed in more than 100 states and districts. Citrus production in India is accounted as the third largest fruit trade of the country[1]. India positions at nineth amongst highest orange producing countries underwriting 3% to the world's entire orange production and only 1.72% of the country's production is distributed[2]. It is crowded with Vitamin C and used in agro industries as raw materials.. According to NABARD (National Bank for Agriculture and Rural Development), India misplaces around 30% of its yields owing to pests and diseases each year. The citrus diseases have extensively affected the citrus production. The exterior apperance of fruit is usually connected with the superiority of fruit and thus the presence of skin defects plays a vital factor that influences the consumption of fresh fruit.. Thus the computer vision systems play an significant part in packaging lines to detect the citrus defects.There is a constant strive to design a novel approach for defect detection in citrus fruits. [3] have used texture feature based on discriminant analysis to identify 5 peel diseases using K-Means cluster, training part is majorly classified on SVM the accuracy obtained by this system is nearly 93%. [4]have employed Principal Component Analysis and Linear Discriminant Analysis. Defected zones have been identified by using shade and surface features,further neural networks have been applied with accuracy of 84.65%. [5] presents the design of the framework, review of different proposed strategies to recognize and classify natural product illness [6], the defects are tracked by using shaded graph and the features are required to classify the diseases. The chromatic aberration map, a circularity threshold and global Otsu segmentation method is used in preprocessing stage and Convolution Neural Networks have been used to detect citrus defects on 1200 images. In [7], authors have detected the defects on fruits. In particular, authors have utilized SVM algorithm to identify the passion fruit diseases with an accuracy of 79%. In [8], authors have used color features, K means clustering and SVM to detect bacterial blight disease in pomogranate. In [9], authors have used color covariance vector to detect apple scab. This research work has reported various shortcomings. In [10], authors have analyzed changes in shape and color of the disease and various orientations. Henceforth, to overcome these short comings, Contrast stretching is employed. To locate the defected areas, K means clustering is used. [11] have used features such as Mean, Entropy, Skewness, Energy, Standard Deviation, and Homogeneity have been used. Numerous classifiers like SVM, KNN and Decision Tree have been applied upon these features.The correctly classified diseases and the misclassified ones are determined . Accuracy in each classifiers are measured. [12] The utilization of picture handling for distinguishing the quality can be applied not exclusively to a specific organic product. This technique can be applied to distinguish nature of vegetables with more exactness. Consequently, this will empower the innovation to be applied in numerous items. Here, the k-means can be used rather than c-mean. In [13], authors have used preprocessing, edge detection, calculating the area and detrmining the quality ratio. Further if the ratio >0 is identified as fresh fruit else it is defected one. B. [14] presents the equipment for organic product evaluation, where all the mangoes were reviewed by utilizing PC vision procedures. Depending on the yield of the image handling segment, the equipment moves the mangoes to various receptacles after selecting surrendered/non-abandoned and unripe. The precision obtained by characterization is 96%. In [15], authors have classified the various fruits by converting to HIS for applying otsu thresholding, extracting 36 statistical and textural features by using wavelet transformation using Haar filter. In [16], authors have developed a productive organic product discovery for utilizing various element based calculation is created and proposed in this paper. Different highlights like force, shading, edge and direction are investigated. In [17], authors have usedbi-directional wavelet transform (BWT) for feature extraction and a rule based linear classifier model for detection and classification of the defects. Totally, 74 mandarin images are used. Fuzzy image thresholding segments the defective regions, splitting and pitting defects are more than 90% and overall performance is 80%. Many works have used neural networks and classifiers [18][19] [20] for the detection of fruits and leaves.

I. DATASETS

The datasets have been collected from online source. Over 110 images have been collected to represent the various defects in citrus fruit. Fig 1 shows the various

Fig 1: Sample Dataset of Diseases

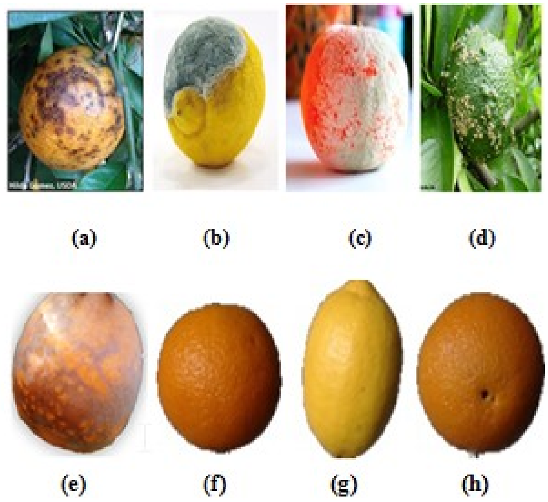

Fig 2: Samples of Various Defects a) Black spot , (b-c) Powdery Mildew d) Citrus Cancker e) Wind Scar, (f-h) Healthy

The commonly occuring defects in citrus fruits are black spots, powdery mildew , cancker and wind scar.

A. Black Spots

Pre-harvest disease causes severe lesions on the rind. The symptoms occur on leaf and organic product of citrus fruit and are described as slight, curved, lesions, dark brown with gray squares for ensuring spot diameter of 0.12 to 0.4

B. Powdery Mildew

It is a fungal disease and specially develops in young leaves and when infected on fruit drops prematurely. The deposited whitish powdery mass distorts the leaves and twigs.

C. Scab

In citrus leaf and fruits, the scab acne is a combination of fungal and organism tissue. It looks like an raised acne that varies its color from pink to light brown in color. Sometimes its color varies from yellow-brown and dirty grey.

D. Wind Scar

It is caused by the wind on the young fruit. It occurs when the leaves, thorns, and twigs rub against the rind.The scars expand and enlarge during its growth

II. PROPOSED METHODOLOGY

The method is a combination of four primary steps that consist of 1) Preprocessing 2) Identification of defect using K means clustering 3)Feature selection and 4) Classification. The entire process is set into training phase and testing phase as represented in Fig.1 and Fig.2 correspondingly. In the training phase, different imagesundergo pre-processing. A training model is built to run through the input data samples using algorithm and correlates the processed output against the sample output to create a trained model. The trained model would then be applied over the test images, which would undergo preprocessing followed by mid level process. The mid level process takes input as image and extracts the features. Each image in the training set have been labelled and stored in label vector. The features extracted over the test data are compared with the features of the trained model and the corresponding defect is allocated from the labelled vector. This label determines the category of test image, which would be classified according to its disease class and the end result would be displayed along with all the feature values (Fig 11) as well as the name and image of the identified fruit defect name and its image respectively (Fig 12).

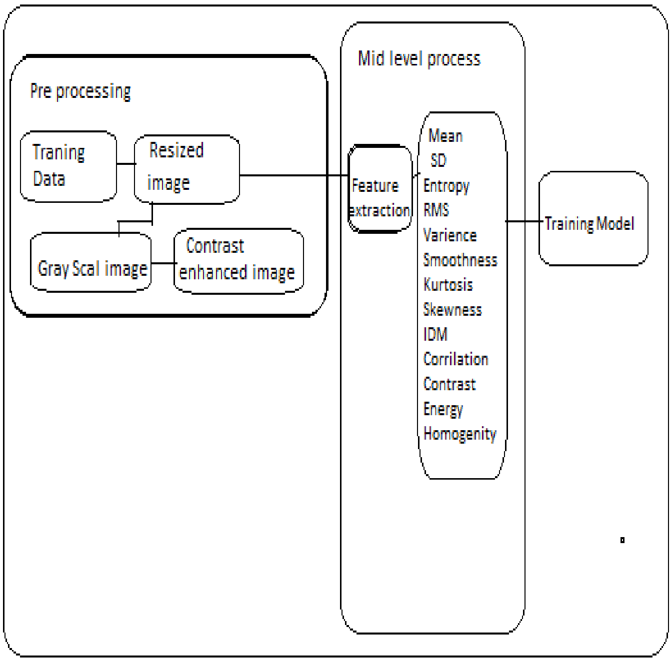

FIG. 3: THE ARCHITECTURE OF TRAINING MODULE

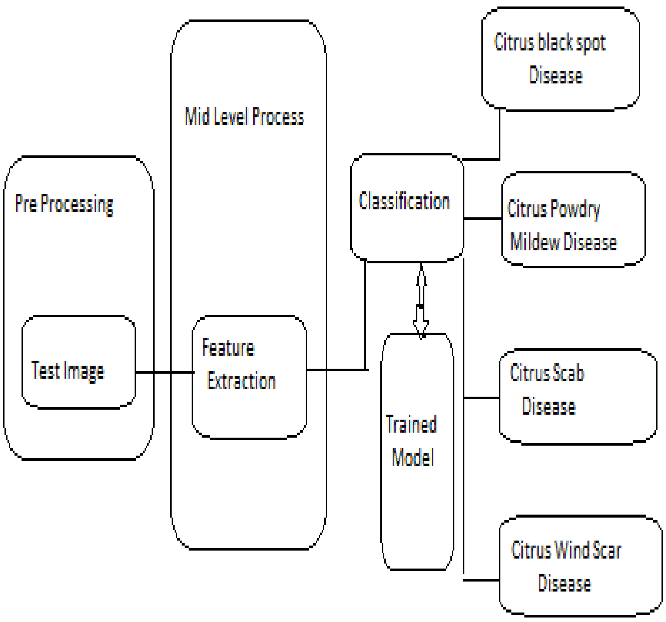

Fig. 4. The Architecture of Testing Module

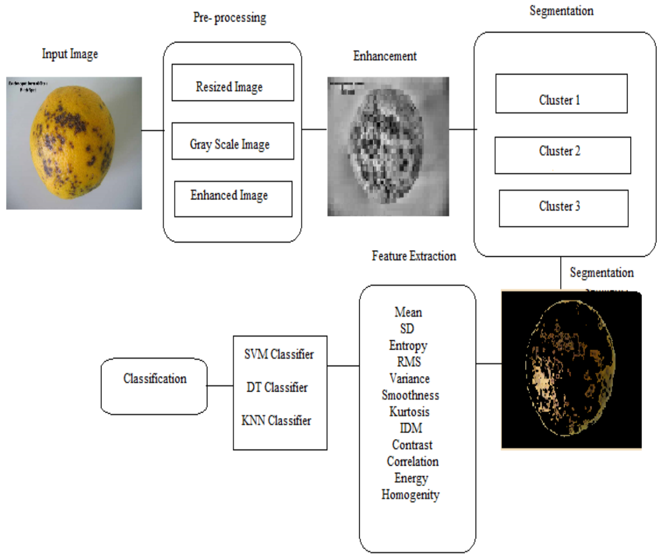

Fig 5: Overview of Proposed System

A. Pre-Processing

The images are obtained during the acquirement stage could not be appropriate for the identification and classification. Several issues, such as low contrast, brightness effects, and illumination must be addressed in order to enhance segmentation accuracy. As a result, pre-processing is used to improve the visual value of the image. This phase includes 3 steps

a) Image Resize
b) Gray Scale Conversion
c) Contrast Enhancement.

B. Resize and Gray Scale Conversion

Image resizing is used to minimize training time since the total of pixels in an image is greater than processing time. All images are 1000 × 1000 pixels in size and have been compressed to 256 x 256 pixels. The image is then transformed to gray scale.

Contrast in the areas is considerably brighter or dimmer and it can be increased using Adaptive histogram equalization (AHE) by altering each pixel with a alteration purpose resulting from a locality area. Consider, a distinct grayscale image X and let ni be the digit of incidences in gray level i.The likelihood of incidence of pixel of level i is ;

$$p(i) = ni / n \quad 0 < i < L \qquad - \quad (1)$$

Where: L is the amount of gray levels
n is the overall integer of pixels
p(i) is the image's histogram .

Conversion of the method y = T(x) to create a original image Y and a increasing supply of task for this given by

$$cfy(y) = cfy(T(k))$$

Where: k is in the range [0 L]

C. Feature Extraction

It is used to identify and extract a specific pattern or a feature from the sample image data. It functions as a dimensionality reduction, displaying just the most interesting image characteristics.The use of color features is done in order to maximize the defect spot information.

Texture based features have been used. Gray equal co-occurrence medium is created by characterizing the texture features that defines the pairs of pixel with specific values and in a specified spatial relationship, statsitical measures are extracted from this matrix. Hence, the fusion of shade and quality structures generates a feature vector which consists of much information as compared to selection of each category of feature. The system utilizes GLCM function in order calculate various simulation or feature measurement parameters that is represented in tables- 1.1 and 1.2; while also extracting the different sets of various line types that together form a single particular image region, while also concentrating on the positional features of the same. Zoning, a fruit sample or image segment is usually divided into predefined zones, thus by reduction of complexity, the feature-specific grid or zone would have opted. The pixel or the texture features of the defective fruit segment is one among the main criteria in determining the diesease, these identified feature values(Fig 11) will then be forwarded onto the next module of classification.

Table 1.1: Simulation Parameter Formulas

Feature Name	Description	Equation
Mean	Average colour in the image	$\sum_{j=1}^{M} 1/M(Pij)$
Standard Deviation	Measure of variables	$1/M \sum_{j=1}^{M} (Pij - Ei)^2$
Energy	Measures closeness of distribution of elements	$\sum_{i,j=0}^{M-1} (Pij)^2$
Contrast	Variations in grey level matrix	$\sum_{i,j=0}^{M-1} Pij(i-j)^2$
Correlation	Probability occurrence of the specified pixel pairs	$\sum_{i,j=0}^{M-1} Pij(i-\mu)(j-\mu)/\sigma^2$
Entropy	Measures the non-uniformity of pixels in image	$\sum_{i,j=0}^{M-1} -\ln(Pij)\,Pij$
Homogeneity	Image region depends on pixel	$\sum_{i,j=0}^{M-1} \frac{Pij}{1} + (i-j)^2$

Table 1.2: Simulation Parameter Formulas

Variance	It is the square of SD	Sum((x-mean(x))^2)/(length(x)-1)
Root Mean Square	It represents the average power of a signal	$1/M \sum_{j=1}^{M} yj^{2}\text{^}1/2$
Smoothness	Reduce the amount of intensity variation between pixels.	$1 - 1/1 + \sigma^2$
Kurtosis	It describes the peakedness of frequency distribution	$\sum_{i,j=0}^{M-1} ((Pij - Pijmean)/N))\text{^}4/\,s\text{^}4$
Skewness	It is a measure of symmetry	$\sqrt[3]{1/M \sum_{i=1}^{M} (Pij - Mj)^3}$
IDM	Reduce the number of operations	$IDM = \sum_{i=0}^{N_g-1}\sum_{j=0}^{N_g-1} \frac{1}{1+(i-j)^2} p(i,j)$

D. Classification of the fruit

Once the features are extracted, the detection of fruit diseases are performed based on their surface defects such as scab, scars, spot, powdery and so on. The defected fruits is identified by generating a boundary of outlines on the defective part of fruit images and the outlines are filled with white pixels foradjustment. Later the circumstances will be applied. Then, if the ratios are associated and are more than the specified onset value, the fruit is defective with a disease; otherwise, the fruit is fresh. Decision Tree, SVM, and KNN are the three primary classifiers utilized here.

III. RESULTS & DISCUSSION

The total images collected for the experiment were 240.The citrus black spot, powdery mildew, citrus scab and wind scar consists of 70, 50, 60 and 60 images respectivley. Each ariety of citrus fruit is shared among the training and testing set equally.

Table 2: DataSet Count

Diseases Name	Total Dataset	Training Dataset	Testing Dataset
Citrus Black Spot	70	35	35
Powdery Midew	50	25	25
Citrus Scab	60	40	20
Citrus Wind Scar	60	30	30

Fig. 6 is the input image, and with that the preprocessing has been performed. Further, the K means clustering is performed and the cluster number has been selected.

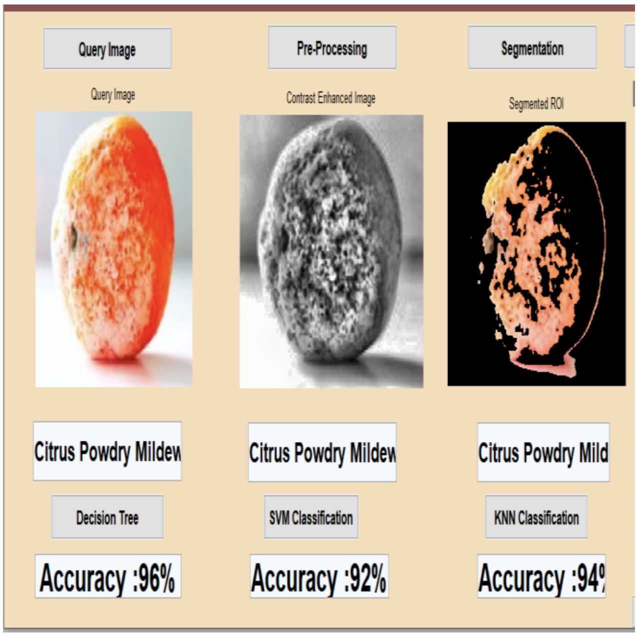

Fig 6: Output of the low level process

The output of the K means clustering with K=3 is shown. In the Fig 7, the defects are highlighted in Cluster 3.

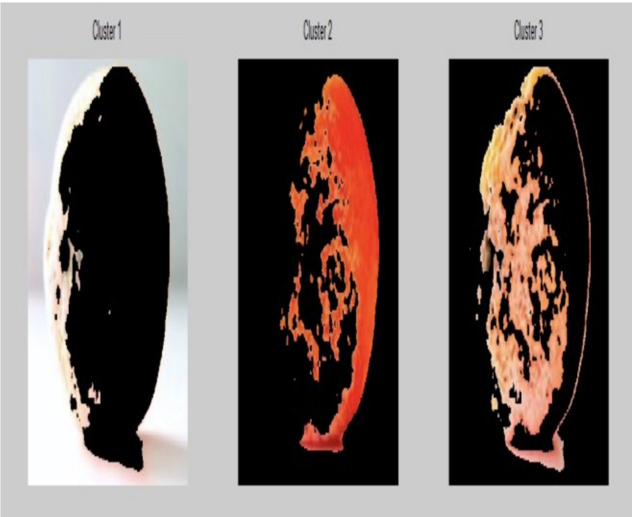

Fig 7: The output of K means displaying Cluster 1 , Cluster 2 and Cluster 3.

The features extracted are shown in Fig 8. 13 features have been extracted. Mean , SD, Entropy ,RMS ,Variance ,Smoothness, Kurtosis , Skewness , IDM ,Contrast , Correlation, Energy , Homogenity are extracted.

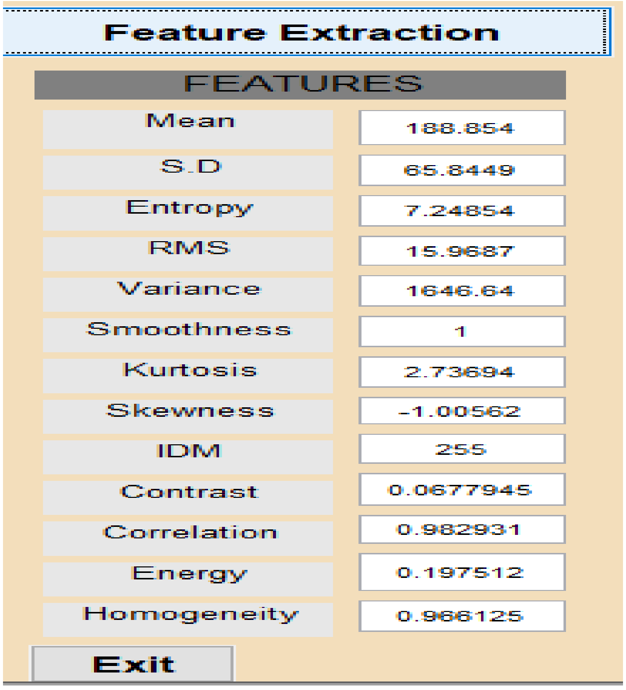

Fig 8: The Output of Feature Extraction

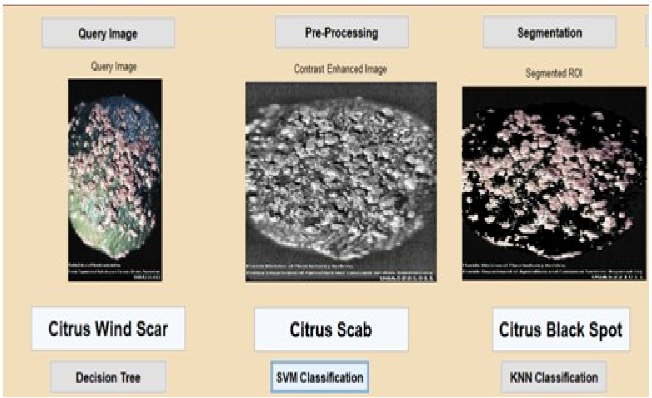

Fig 9: Citrus Scab Detection by Various Classifiers

Fig 10 : Citrus Wind Scar Detection using Decision Tree, SVM and Citrus Black Spot.

Fig 11 : Citrus Black Spot Detection using Classifiers.

Fig. 9 , Fig. 10 & Fig. 11 display the detection of different defects by using various classifiers. Fig 9 shows the Citrus scab, which is correctly identified by SVM but KNN and DT will only display black spot and wind scar respectively. Fig 10 detects citrus wind scar correctly by SVM and Decision Tree but is misclassified as black spot by KNN. SVM and Decision Tree misclassify the blackspot as wind scar and powdry mildew respectively.However KNN has classified the black spot correctly.

IV. CONFUSION MATRIX & PERFORMANCE MEASURES

The confusion matrix for defects in citrus has been tabulated in Table 3 , Table 4 and Table 5. SVM classifier is able to identify 30 citrus black spts , 23 of 25 powdery mildew and 18 of 20 citrus scab and 27 of 30 wind scar correctly.

Table 3: Confusion Matrix for defects using SVM

	Citrus Black Spot	Powdery Midew	Citrus Scab	Citrus Wind Scar
Citrus Black Spot	30	0	0	5
Powdery Midew	0	23	0	2
Citrus Scab	0	0	18	2
Citrus Wind Scar	2	0	1	27

The Decision Tree can identify 26 Citrus black spot , 18 powdery mildew ,14 citrus scab and 24 of 30 wind scar defects correctly.

Table 4: Confusion Matrix for Decision Tree

	Citrus Black Spot	Powdery Midew	Citrus Scab	Citrus Wind Scar
Citrus Black Spot	26	2	1	1
Powdery Midew	1	18	2	1
Citrus Scab	1	2	14	2
Citrus Wind Scar	2	2	2	24

The KNN can identify 24 Citrus black spot , 18 powdery mildew ,11 citrus scab and 22 of 30 wind scar defects correctly.

Table 5: Confusion Matrix for KNN

	Citrus Black Spot	Powdery Midew	Citrus Scab	Citrus Wind Scar
Citrus Black Spot	24	3	3	4
Powdery Midew	2	18	2	3
Citrus Scab	5	3	11	1
Citrus Wind Scar	5	4	1	20

The performance measures -Accuracy , Precision ,Recall and F-Measure of 3 classifiers for various defects are mentioned in Table 4 , 5 , 6 respectively. It can be observed that the

Table 6: SVM Performance Measures

	Accuracy	Precision	Recall	F-Measure
Citrus Black Spot	93.64	0.86	0.94	0.9
Powdery Midew	98.18	0.92	1	0.96
Citrus Scab	97.27	0.9	0.95	0.92
Citrus Wind Scar	89.09	0.9	0.75	0.82

Table 7: DT Performance Measures

	Accuracy	Precission	Recall	F-Measure
Citrus Black Spot	92.08	0.87	0.87	0.87
Powdery Midew	90.01	0.82	0.75	0.78
Citrus Scab	90.01	0.74	0.74	0.74
Citrus Wind Scar	90.01	0.8	0.86	0.83

Table 8: KNN Performance Measures

	Accuracy	Precission	Recall	F-Measure
Citrus Black Spot	79.82	0.71	0.67	0.69
Powdery Midew	84.4	0.72	0.64	0.68
Citrus Scab	86.24	0.55	0.65	0.59
Citrus Wind Scar	83.49	0.67	0.71	0.69

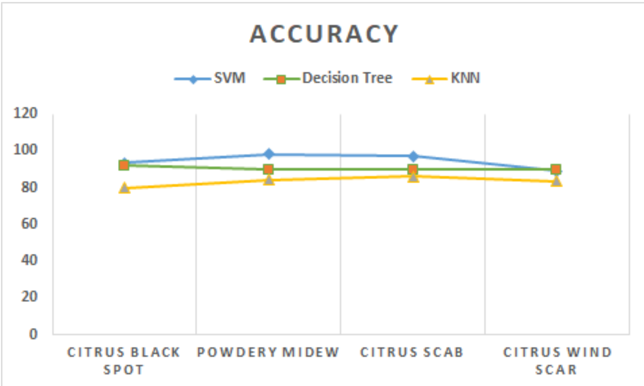

Fig 9: Comparitive Accuracy Chart

From table 6 , Table 7 and Table 8 it is observed that SVM provides an accuracy of 95%, while accuracy of decision tree is 90% and 83.43 % for KNN.Knn displays poor performance since it misclassifies 13 fruits out of 35 of citrus black spot , 7 misclassifications for powdery mildew , 9 misclassifications for citrus scab and 10 for wind scar.
The above graph represents the overall efficiency of the proposed model. The performance evaluation measuring parameters being the F-measure, Recall, Precision and Accuracy as observed from table- 6, 7, 8. By utilizing the confusion matrix (tables- 3, 4, 5) the values are based out of the experimental results of the outcome of the model. The values are based on the defective fruit samples classified accordingly; the defects have been identified and classified on which the proposed methodology has been implemented. The highest and the lowest values of the accuracy based out the 3 techniques can be observed in the Fig 9.

V. CONCLUSION

Computer vision systems are in constant need for fruit defect inspection. In this work, four defects of citrus fruits have been identified- powdery mildew , black spot , cancker and scab.The preprocessing stage includes resize , gray scale conversion and adaptive histogram equialisation.13 features have been extracted that is fed to the classifers.Among 3 classifeirs the accuracy of SVM is better than Decision Tree and KNN. SVM detects 98.18% accuracy for powdery mildew followed by 97.27% accuracy for citrus scab , 93.64% accuracy for black spot and 89.09% for wind scar. However, Decision Tree has achieved average of 90% for all the types of defects.The accuracy of KNN is ~ 83 %. The precsion, recall and F measure of SVM is better than the other 2 classifiers.

The work utilizes 240 images. The cirus black spot, powdery mildew, citrus scab and wind scar consists of 70 , 50 , 60 and 60 images respectivley.The work can be further validated by increasing the data set and using other classifiers.

REFERENCES

[1] Zhang, H., Zhang, S., Dong, W., Luo, W., Huang, Y., Zhan, B., & Liu, X. (2020). Detection of common defects on mandarins by using visible and near infraredhyperspectralimaging.Infrared Physics & Technology, 108, 103341.

[2] http://nhb.gov.in/

[3] Dubey, S. R., & Jalal, A. S. (2012, November). Detection and classification of apple fruit diseases using complete local binary patterns. In Proceedings of the 3rd international conference on computer and communication technology (pp. 346-351).

[4] Pujari, J. D., Yakkundimath, R., & Byadgi, A. S. (2013). Grading and classification of anthracnose fungal disease of fruits based on statistical texture features. International Journal of Advanced Science and Technology, 52(1), 121-132.

[5] Dakshayini Patil. (2019, March). Fruit Disease Detection using Image Processing Techniques. International Journal for Research in Engineering Application & Management (IJREAM) ISSN : 2454-9150

[6] Rupam Thakur, Priyanka Mehta (2017). Automatic detection of fruit diseases: A Review. International Journal of Engineering Development and Research. Volume 5, Issue 3 | ISSN: 2321-9939.

[7] Dharmasiri, S. B. D. H., & Jayalal, S. (2019, March). Passion Fruit Disease Detection using Image Processing. In 2019 International Research Conference on Smart Computing and Systems Engineering (SCSE) (pp. 126-133). IEEE.

[8] Ananthi N, Akshaya S, Aarthi B, Aishvarya J, Kumaran K. (2019, November). An Image Processing Based Fungus Detection System For Mangoes. International Journal of Innovative Technology and Exploring Engineering (IJITEE) ISSN: 2278-3075, Volume-9 Issue-1.

[9] Dameshwari Sahu, Chitesh Dewangan.(2016). Identification and Classification of Mango Fruits Using Image Processing. International Journal of Scientific Research in Computer Science, Engineering and Information Technology Volume 2 | Issue 2 | ISSN : 2456-3307.

[10] Ketki Tarale, Anil Bavaskar. (2017, February). Fruit Detection Using Image Processing Technique. National Conference on Advances in Engineering and Applied Science (NCAEAS). ISSN: 2395-602X.

[11] Pradeepkumar Choudhary, Rahul Khandekar, Aakash Borkar , Punit Chotaliya.(2017, March). Image Processing Algorithm For Fruit Identification. International Research Journal of Engineering and Technology (IRJET) ISSN: 2395 -0056 Volume: 04 Issue: 03.

[12] S.V.Phakade, Edake Vaishali, Chikode Pooja, Ovhal Rupali.(2017, April). The Fruit Quality Identification System in Image Processing Using Matlab. International Journal of Innovative Studies in Sciences and Engineering Technology (IJISSET) ISSN 2455-4863.

[13] Siddhika Arunachalam, Harsh H. Kshatriya , Mamta Meena.(2018, October). Identification of Defects in Fruits Using Digital Image Processing. International Journal of Computer Sciences and Engineering Vol.-6, Issue-10, ISSN: 2347-2693.

[14] M.Pushpavalli.(2019, August). Image Processing Technique for Fruit Grading. International Journal of Engineering and Advanced Technology (IJEAT) ISSN: 2249 – 8958, Volume-8 Issue-6.

[15] PL.Chithra , M.Henila. (2019, March). Fruits Classification Using Image Processing Techniques. International Journal of Computer Sciences and Engineering Vol.-7, Issue,5 ISSN: 2347-2693.

[16] Hetal N. Patel, Dr. R.K.Jain, Dr. M.V.Joshi. (2011, January). Fruit Detection using Improved Multiple Features based Algorithm. International Journal of Computer Applications (0975 – 8887) Volume 13– No.2.

[17] Kamalakannan, A., & Rajamanickam, G. (2012, December). Surface defect detection and classification in mandarin fruits using fuzzy image thresholding, binary wavelet transform and linear classifier model. In 2012 Fourth International Conference on Advanced Computing (ICoAC) (pp. 1-6). IEEE.

[18] Kamel, Khaled, S. Smys, and Abul Bashar. "Tenancy Status Identification of ParkingSlots Using Mobile Net Binary Classifier." Journal of Artificial Intelligence 2, no. 03(2020): 146- 154.

[19] Shree, B. D., Brunda, R., & Rani, N. S. (2019). Fruit detection from images and displaying its nutrition value using deep Alex network. In Soft Computing and Signal Processing (pp. 599-608). Springer, Singapore.

[20] Pushpa, B. R., Megha, N., & Amaljith, K. B. (2020, June). Comparision and Classification of Medicinal Plant Leaf Based on Texture Feature. In 2020 International Conference for Emerging Technology (INCET) (pp. 1-5). IEEE.