Abstract:

The food processing industry has been working with different kind of fruits. Most of the industries who are classifying the classes and the damages of fruit by manual. As a result of manual classification, it will become a challenging task for the human. So, in this research work, we have proposed a method with a machine learning technique that classify various kinds of fruits automatically. Also, we will detect damage on fruit using image processing technique like Graph-cut or HSV color segmentation technique.

Introduction:

The huge contribution of modern technology, the world agriculture rate is increasing. Though, a lot of industries/farmers they have been monitoring their fruit to classify the good fruit from bad by manual. The lack of better observation a number of fruit destroyed in a garden or in an industry every year. So in this study, will developed a model to identify the fruit and the damage of the fruit to increase agricultural yield. For this purpose, we can apply numerous machine learning (ML) and image processing techniques.

We will use a high-quality, dataset of images containing fruits. The dataset was named Fruits-360 and can be downloaded from the addresses pointed by references [14] and [15]. This dataset containing 55244 images of 81 fruits.

Then we will perform different machine learning algorithm like Support Vector Machine (SVM for short), K-nearest Neighbors, KNN for short, Randomforest, RF for short, Decision Tree, Naive Bayes Classifier, NB for short, Linear discriminant Analysis, LDA for short and Logistic Regression for LR. We will combine all these algorithms together. For is purpose we will use the most popular python library which is Scikit-learn.

Related Work/ Literature Review:

An enough literature is available for damage detection and also images classification of fruits using various approaches.

Mureşan, Horea, and Mihai Oltean et al. [1] have introduced a new, high-quality, dataset [4] and [5] of images containing fruits that is essential for obtaining a good classifier. A method for identifying fruits. They have trained a deep neural network (DNN) which is an artificial neural network with different layers. A convolutional neural network is used for classifying the images. The training and validation data used in this paper consists of 55244 images including test and training images over 81 fruits. For the purpose of implementing, training and testing by CNN they used TensorFlow library. They have trained dataset with a different scenario in terms of image format like grayscale, RGB, HSV, HSV, and grayscale combination and also HSV, Grayscale, hue/saturation change with flips. According to the trained neural network with the different conversion of the image, they obtained an average 94.66% of accuracy to identify the fruit.

Bhange, Manisha, and H. A. Hingoliwala et al. [2] proposed a web-based tool that helps farmers for identifying fruit disease by uploading a fruit image to the system. The proposed system used already a trained dataset for the pomegranate fruit. They resized the images and then the features are extracted based on the different parameters such as color, morphology, and CCV. According to these features, morphology gave them the best result. Here, they have used k-means algorithm for clustering the dataset. And then, they have used a most popular and more useful machine learning algorithm SVM is for classifying the images in terms of infected or non-infected image. The proposed system gives 82% of accuracy for identifying pomegranate disease. From now on, they will work with increasing the performance of the proposed system for reaching a good result.

Awate et al. [3] introduced a technique which will diagnose and classify external disease of fruit. In this research work, they worked with three fruits such as grape, apple, and pomegranate for the purpose of detecting the disease of this fruit. The proposed system uses two image dataset one for testing and another one for training purpose. The images are classified and mapped to their respective disease categories on the basis of four feature vectors such as color, morphology,

texture, and structure of hole on the fruit. In this proposed system they have used different types of algorithm for different purposes. The image segmentation is done by using K-Means clustering methodology. SURF (Speed up Robust Feature) algorithm is applied for extracting the features. For the purpose of testing and training of the fruit dataset, an Artificial Neural Network is used to identify the disease of fruit. The overall implementation is done by using OpenCV library and the proposed is able to identify the disease with 90% of accuracy.

In the above literature, numerous method has applied to classify fruit and detect fruit damage such as Affective Neural Network, Support Vector Machine, Deep Neural Network, k-means clustering. SURF (Speed up Robust Feature) etc. Though these methods are familiar to detect damage and classify fruit but there are many methods that have not been implemented yet. Machine learning and image processing techniques will be used for the identification and damage detection of fruit and it will contribute to getting better performance, we believe.

A digital color imaging technique is proposed by Saeideh Gorji Kandi et al. [4] for quality control of food and agricultural products, which was traditionally done by manual inspection by using Machine vision. Basically, Saeideh Gorji Kandi et al [4] tried to detect a defect on single-color fruit like banana and plum. He mentioned that growing decay and time-aging made surface color changes in bruised parts of the object. Here HSV (Hue Saturation and Value), RGB and grayscale images are used for color quantization of the object. By using his method the percentage of different degree of defect can be computed and this also can use for grading the fruits.

A computer vision system was proposed by Mr.S V.Phakade et al. [5] for defect detection of fruit surface in the agricultural field. In this entire work, they detect defected part based on RGB and HSV color space. Among these two color space HSV color space gives much better accuracy that RGB color space. Using this system the percentage of defect on a fruit surface can be detected. The limitation is in this work is that the automatic defect detection is not possible for other kinds of fruit instead of the fruits they have used. The reason behind this is that different fruits have different color and some of the defect color is matched with a good part of the fruit.

The color segmentation technique is used to identify the defected region of fruits and the corresponding percentage of frequency components from its Spectrogram. This technique is

implemented by Md. Imran Hosen et al. [6] using L*a*b color space. They find the defected portion on fruit image from four directions to get the appropriate result of a 3D image. Here, also measured the percentage of the defected area on fruit using scatterplot of the colors of the image.

Mohana S.H., Prabhakar C.J., and Praveen Kumar P.U. et al. [7] proposed a novel method to detect surface defects of an apple using RGB images and apples are graded based on these identified defects. They only use an apple fruit and consider the outer surface to grade the apples. They achieve much higher accuracy in grading the apples using fused features, which are color statistical, color texture and shape features compared to individual features. K-Nearest Neighbor (K-NN), Linear Discriminant Classifier (LDC), AdaBoost and Support Vector Machine (SVM) classification Classification accuracy are 85.00%, 75.00%, 70.00%, and 80.00%. Among them, K-Nearest Neighbor (K-NN) classifier results show that highest classification accuracy 85.00% compared to other classifiers.

A new method for classifying fruits using image processing technique is proposed by PL. Chithra, M.Henila et al. [8]. Sample images of apples and bananas were alone taken for experimentation. They are working only two levels of fruits for their experiment. They are proposed two kinds of classifier algorithms Support Vector Machine (SVM) and K-Nearest Neighbor (K-NN). Among them, SVM gave 100% accuracy when compared to KNN classifier. Sample images should be acquired at 360 degrees in order to obtain 100% accuracy in real-time classification of any fruit or vegetable in the agriculture industry. In future few more fruits or vegetables can be taken as samples for their experimentation. Their research work can be extended to help the agriculturist to classify different varieties of apples and bananas.

Zalak R. Barot, Narendrasinh Limbad et al. [9] presented a survey on an approach for detection and classification of fruit disease. Their literature review consists of nine papers. Those paper conclude different segmentation, feature extraction and classification techniques for plant disease detection using there leaf or fruit. There is a various technique they are using for classification, those are Artificial Neural Network, Backbone Propagation Neural Network, Feedforward Backpropagation Neural Network, Probabilistic Neural Network, Support Vector Machine, Multiclass Support Vector Machine and etc. Each and every technique has some merit and demerits. Several classifiers

are used in different papers. Among all the papers different classifier ANN and SVM give better accuracy than another classifier.

Sachin Khirade and A. B. Patil et al. [10] discussed the main steps of image processing to detect disease in plant and classify it. It involves steps like image acquisition, image pre-processing, image segmentation, feature extraction, and classification. For segmentation, methods like otsu's method, converting an RGB image into HIS model and k-means clustering are there. Among all, the k-means clustering method gives an accurate result. After that, feature extraction is carried out like color, texture, morphology, edges, etc. Among this, morphology feature extraction gives a better result. After feature extraction, classification is done using classification methods like Artificial Neural Network and Back Propagation Neural Network.

Suhaili Kutty et al. [11] proposed the process to classify Anthracnose and Downey Mildew, watermelon leaf diseases. For this Region of Interest need to be identified from infected leaf sample based on RGB color component. Then to reduce noise and for segmentation Median Filter is used. And for classification, Neural Network Pattern Recognition Toolbox is used. Proposed method achieved 75.9% of accuracy based on its RGB mean color component.

Ms. Kiran R. Gavhale et al. [12] presented number of image processing techniques to extract diseased part of leaf. For Pre-processing, Image enhancement is done using DCT domain and color space conversion is done. After that segmentation take place using k-means clustering method. Feature extraction is done using GLCM Matrix. For classification of canker and anthracnose disease of citrus leaf, SVM with radial basis kernel and polynomial kernel is used.

Monika Jhuria et al. [13] investigated an approach of image processing for the detection of disease and fruit grading. They have used an artificial neural network for the classification of disease. They consider three feature vectors, namely, color, textures, and morphology. Among all, the morphological feature gives a better result. It can detect two diseases of grape which is Black Rot and Powdery Mildew and two of apple which are Apple Scab and Rot. Two methods are used for fruit grading which is the spread of disease and automated calculation of mango weight.

Proposed Method:

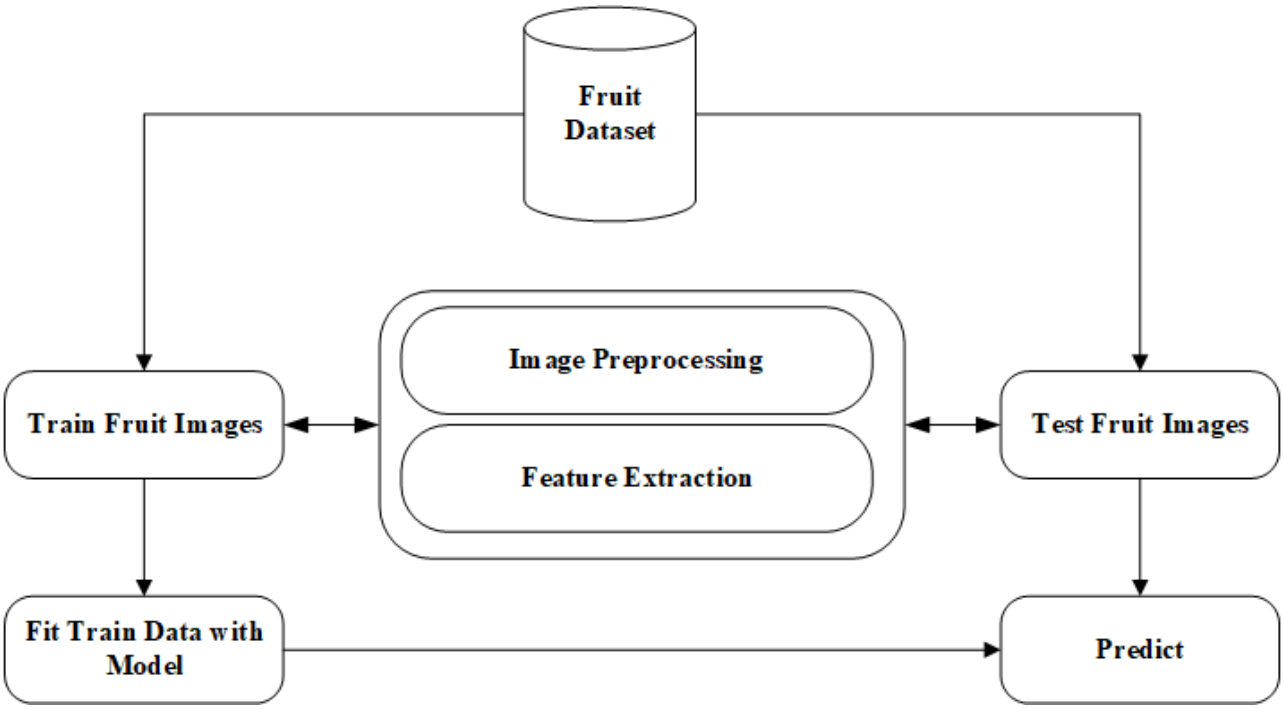


Figure 1: Proposed Method for Fruit Identification.

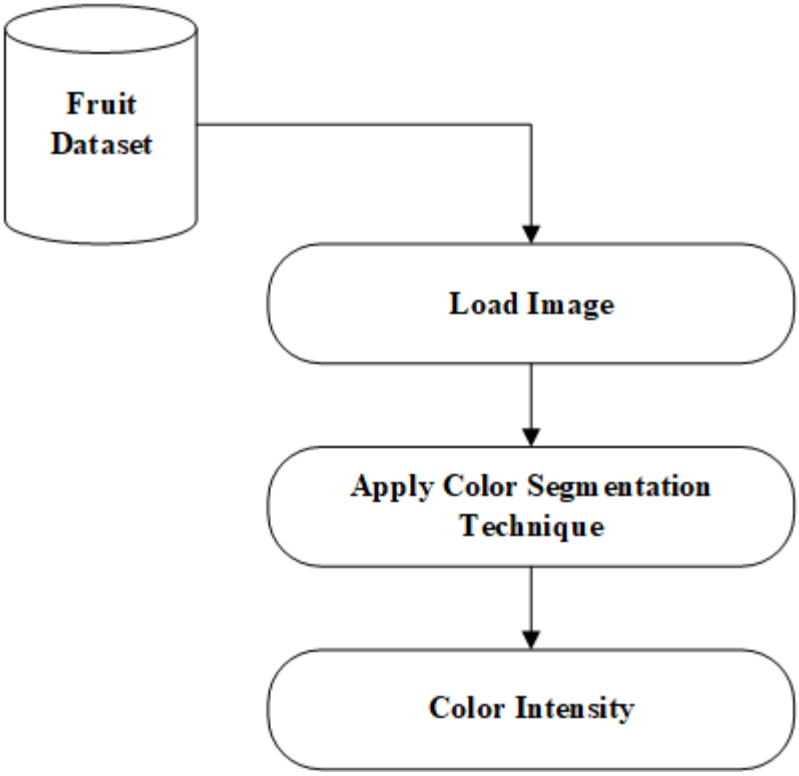


Figure 2: Proposed Method for Fruit Damage Detection.

Dataset Information:

We use the fruits-360 dataset, the addresses pointed by references [14] and [15]. This dataset is a highly challenging dataset with 81 classes of fruit image, each having almost 492 images. So, totally we have 55244 images to train our model. But, among of them we take only 15 class containing 15 different fruits. Our dataset will be as like as given table 1.

No	Data Set	No. of Features	Types of features	Working Instances	Training Instances	Testing Instances	Classes	Category
01	Fruits -360	3	Color, Shape, Texture	8685	7185	1500	15	Semi Supervised

Table 1: Training & Testing Dataset.

Methodology:

In this section, we will discuss the whole procedure that really expose our proposed method in details.

Image Preprocessing: Before extracting feature of fruit image, we resized our images those are included with the dataset fruits-360, the addresses pointed by references [14] and [15]. We resized our images for getting large image that really help to take features in detail.

Feature Extraction: Features are the information or list of numbers that are extracted from an image. These are real-valued numbers (integers, float or binary). There are a wider range of feature extraction algorithms in Computer Vision. In terms of feature extraction basically we are going to use three terms like color, shape and texture. The reason of taking these three features is that the "Tomato Cherry Red" is almost similar with "Tomato Maroon" according to color but the shape is different. We measure color, shape and texture with Color Histogram, Hue Mument and Harlinck Texture. Features are extracted from different classes that we mentioned before in section of dataset information. We take one image at a time, extract three global features, concatenates the three global features into a single global feature and saves it along with its label in a HDF5 file format. Instead of using HDF5 file-format, we could use ".csv" file-format to store the features. But, as we

will be working with large amounts of data in future, becoming familiar with HDF5 format is worth it.

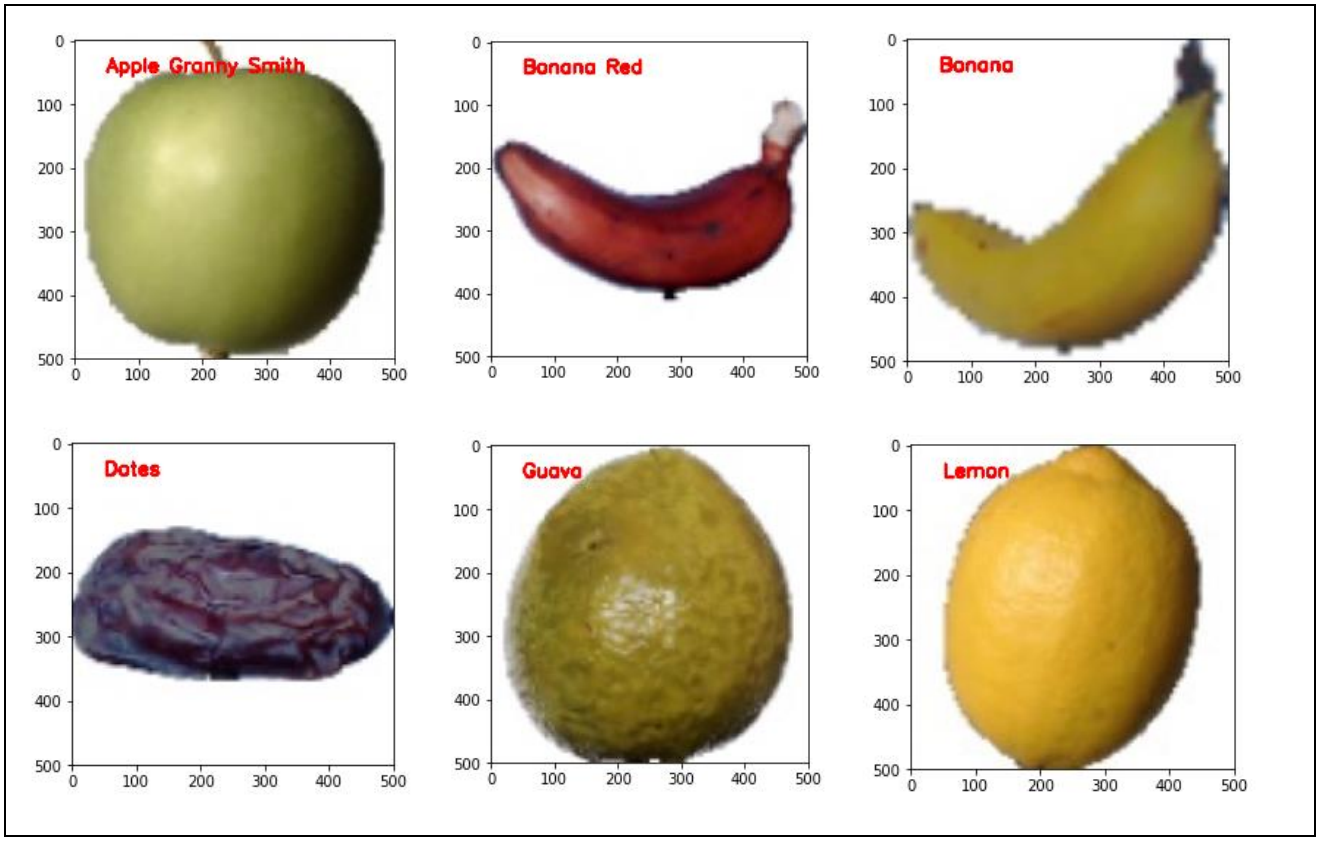
Train and Test Fruit Image: After extracting feature we apply machine learning classifier to train our extracted data as saved into HDF5 file. In this part we create our machine learning model with the help of one of the popular python library Scikit-learn.

We will choose Logistic Regression, Linear Discriminant Analysis, K-Nearest Neighbors, Decision Trees, Random Forests, Gaussian Naive Bayes and Support Vector Machine as our machine learning models. Among of them we apply "Random Forest" model to train our dataset. After that, we fit the model with train dataset that we save to HDF5 file. After that, we load our test dataset and extract feature of them with the same way that we had done before in feature extraction part. After that, we successfully classify fruit from testing dataset.

Damage Detection: In this section we are trying to detect damage on fruit surface. We were able to segment color of fruit. Here, we use Graph-cut segmentation technique to separate different color on fruit image. After segmentation we represent the color intensity by using histogram plot to measure the intensity of damage part of the fruit surface. But we are not able to detect exact damage part on fruit. We have some other options like using of HSV color space is also can be a solution for damage detection. We already convert our RGB image into HSV color space and separate H channel, S channel and V channel as. The result of this part is mention in the discussion section.

Experimental Analysis:

We had a dream to make the automatic fruit classification to help farmer and fruit processing industries to make them smarter than before. Now, it's time to make our dream true because we are successfully able to classify fruits. Here, we apply a machine learning technique which is Random Forest and it gives us a better accurate result in terms of fruit classification. The predicted results are given in figure 3.



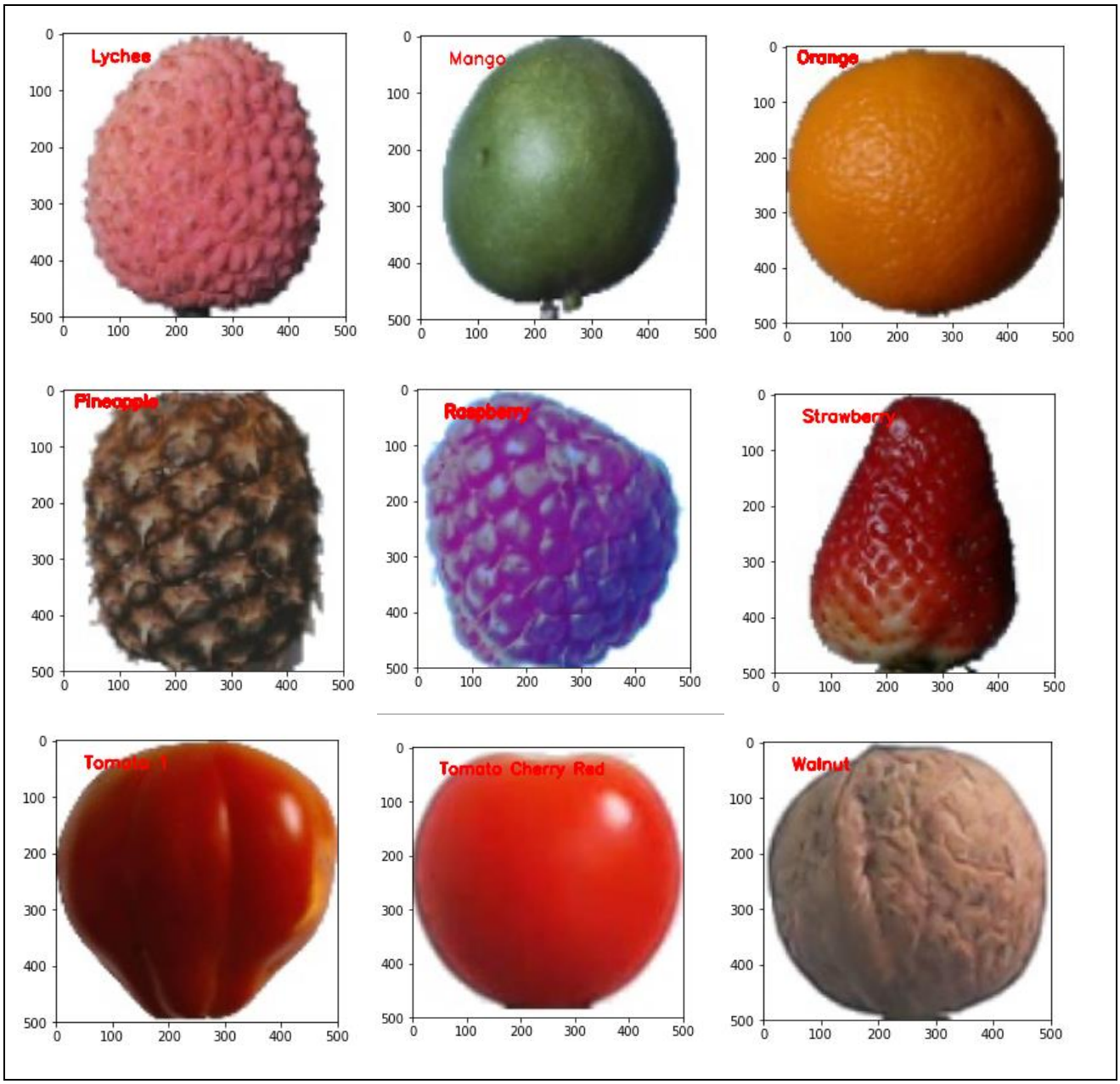


Figure 3: Fruit Identification Result.

Here is the result of damage detection part that we initially have done. In the rest of the day, we will work through it. The following figures are showing Graph-cut color segmentation and HSV color space result.

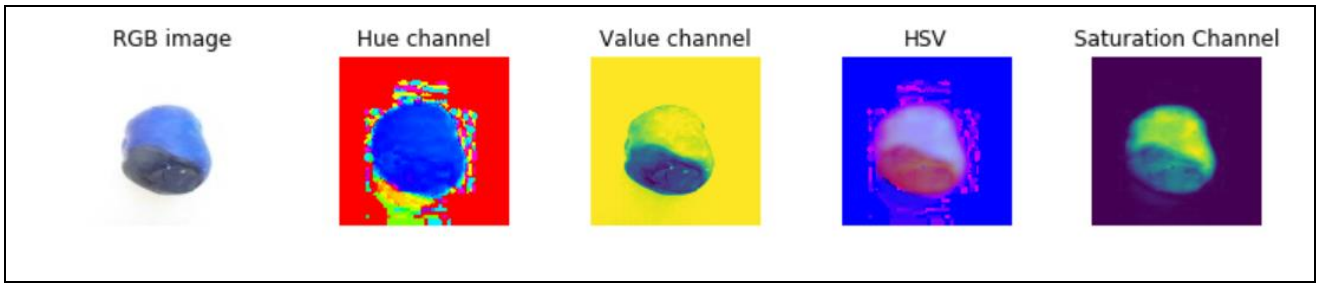


Figure 3: HSV Representation

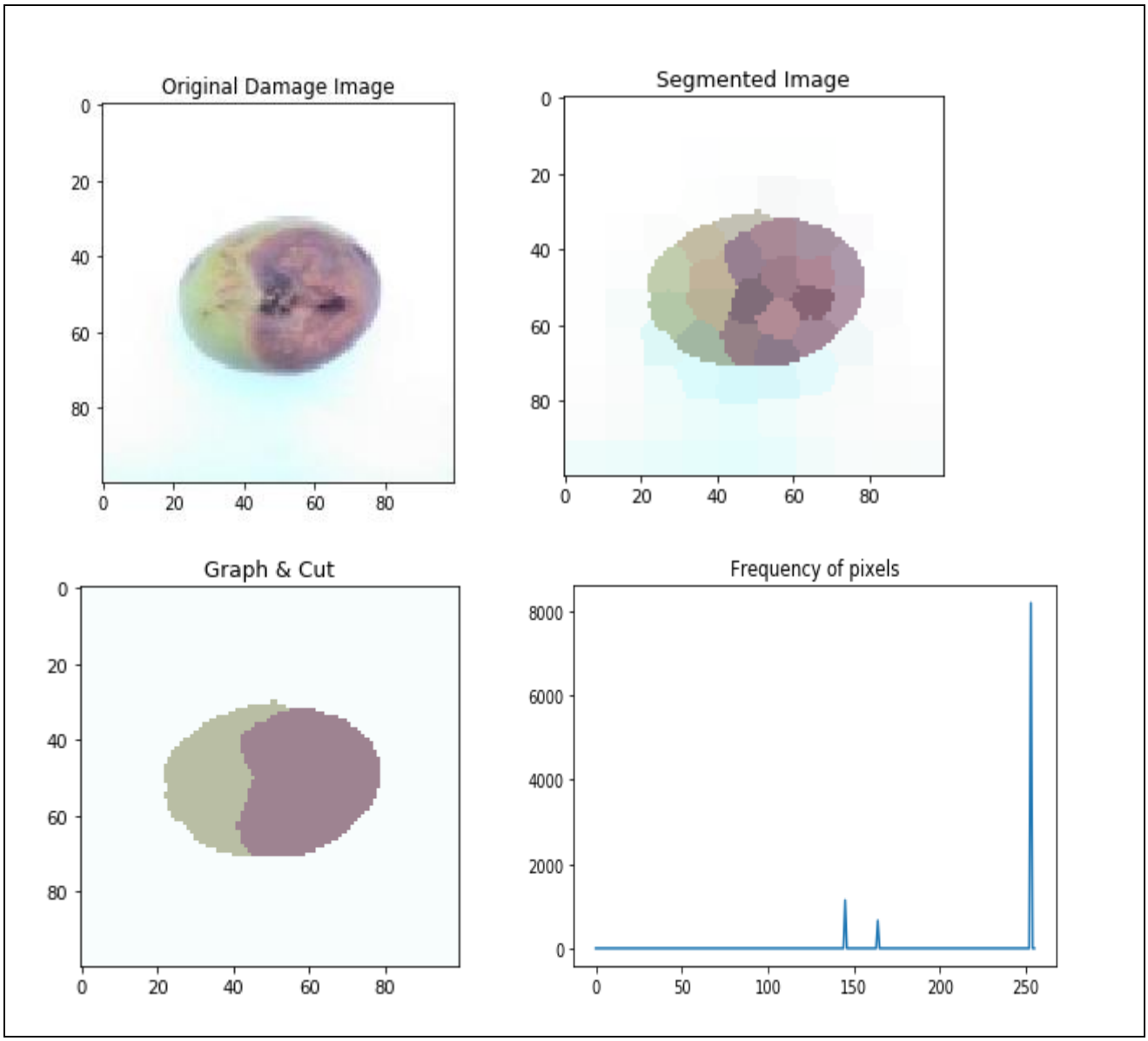


Figure 4: Graph-Cut Color Segmentation

Conclusion & Future Work:

In this research work, we tried to classify a different kind of fruits and also the damaged part on the fruit surface. We implemented the entire work with the help of various machine learning algorithm and image processing techniques and, we got 98.02% for the different classifier. We worked with a large number of fruit images for getting the best accuracy.

Our system will help farmers to increase their productivity and yield with the help of automating tasks in garden/farm. From our point of view, one of the main objectives for the future is to improve the accuracy of the Image processing technique and Machine learning. Another objective is to expand the data set to include more fruits. Machine learning and image processing techniques will be very effective to classify and detect of fruit damage. The more using of a large dataset will give better accuracy.

References:

[1] Mureşan, Horea, and Mihai Oltean. "Fruit recognition from images using deep learning." Acta Universitatis Sapientiae, Informatica 10, no. 1 (2018): 26-42.

[2] Bhange, Manisha, and H. A. Hingoliwala. "Smart farming: Pomegranate disease detection using image processing." Procedia Computer Science 58 (2015): 280-288.

[3] Awate, Ashwini, Damini Deshmankar, Gayatri Amrutkar, Utkarsha Bagul, and Samadhan Sonavane. "Fruit disease detection using color, texture analysis and ANN." In 2015 International Conference on Green Computing and Internet of Things (ICGCIoT), pp. 970-975. IEEE, 2015.

[4] Saeideh Gorji Kandi, "Automatic Defect Detection and Grading of Single-Color Fruits Using HSV (Hue, Saturation, Value) Color Space", Dec. 2010, Volume 4, No.7 (Serial No.32) Journal of Life Sciences, ISSN 1934-7391, USA.

[5] Mr.S V.Phakade, Miss. D'souza Flora, Miss. Halade Malashree, Miss. Joshi Rashmi, "Automatic Fruit Defect Detection Using HSV and RGB Color Space Model", International Journal of Innovative Research in Computer Science & Technology (IJIRCST), ISSN: 2347-5552, Volume-2, Issue-3, May-2014.

[6] Md. Imran Hosen, Tahsina Tabassum, Jesmin Akhter, Md. Imdadul Islam, "Detection of Fruits Defects Using Colour Segmentation Technique", International Journal of Computer Science and Information Security (IJCSIS), Vol. 16, No. 6, June 2018.

[7] Mohana S.H., Prabhakar C.J., and Praveen Kumar P.U, "Surface Defect Detection and Grading of Apples", DOI: 03.AETS.2013.4.33, Association of Computer Electronics and Electrical Engineers, 2013.

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

[9] Zalak R. Barot, Narendrasinh Limbad, “An Approach for Detection and Classification of Fruit Disease: A Survey”, International Journal of Science and Research (IJSR), ISSN (Online): 2319-7064.

[10] Sachin D. Khirade, A. B. Patil, “Plant Disease Detection Using Image Processing,” IEEE, International Conference on Computing Communication Control and Automation, Pune, pp 768-771, 2015

[11] Suhaili Beeran Kutty, Noor Ezan Abdullah, Dr. Hadzli Hashim, A’zraa Afhzan Ab Rahim, Aida Sulinda Kusim,Tuan Norjihan Tuan Yaakub, Puteri Nor Ashikin Megat Yunus, Mohd Fauzi Abd Rahman, “Classification of Watermelon Leaf Diseases Using Neural Network Analysis”, IEEE, Business Engineering and Industrial Applications Colloquium (BEIAC), Langkawi, pp 459 – 464, 2013

[12] Ms. Kiran R. Gavhale, Prof. Ujwalla Gawande, Mr. Kamal O. Hajari, “Unhealthy Region of Citrus Leaf Detection Using Image Processing Techniques”, IEEE, International Conference for Convergence of Technology, Pune, pp 1-6, 2014

[13] Monika Jhuria, Ashwani Kumar, Rushikesh Borse, “Image processing for smart farming: Detection of disease and fruitgrading”, IEEE, Second International Conference on Image Processing, Shimla, pp 521 – 526, 2013

[14] Fruits 360 Dataset on GitHub. https://github.com/Horea94/Fruit-Images-Dataset. last visited on 23.04.2019 =>1, 10

[15] Fruits 360 Dataset on Kaggle. https://www.kaggle.com/moltean/fruits. last visited on 23.04.2019 =>1, 10