

CHAPTER 1

INTRODUCTION

CHAPTER 1

INTRODUCTION

This chapter is going to give an overview of this thesis that includes the significant of the problem and the problem statement in details. Besides, we will discuss the different research challenges that we are going to face in the whole research work. After that, we will present the thesis objectives and contributions. This chapter will ends with a short description of the organization of this thesis.

1.1 Motivation

First thing first is how we classify anything It means how a doctor can classify good blood cell from a bad one, how photographers can classify their latest shot was beautiful or not, how a musician can classify which sound is good, and what doesn't, in a piece of music. Whenever we think to classify something, the first question arises how we classify them. The same question arises when we talking about classifying different kind of fruits. Recently, classification of fruits are done by manually. We are motivated to doing something in automatic way for the classification problem of fruit for helping people to give them repose from the manual classification of fruits. Here is another problem is to detect damage and find the percentage of damage on fruit. We are also motivated for creating an approach to detect and measure the percentage of damage on fruit which would be a great job for helping people who are doing this stuff.

From this motivation, we are going to develop an approach that will be used for classifying fruit and will be capable of detecting damage and the percentage of damage of fruit that makes the people repose from manual identification and detection of damage of fruits.

1.2 Objective and Research Challenges

The principal objectives of our research area are to reduce the classification problem means the identification of different kinds of fruits and to make an approach to detect damage of fruit and find the percentage of damage on fruits.

If we are able to identify the fruits and to detect the damage of fruits it will really be a helpful job that will help the people or industry who are identifying and detecting damage by manual. The proposed system will save a lot of time and help to get a good profit. For conducting our research, we will face some challenges like:

- To identify fruits, those have same color.
- To detect damage on the fruit surface.
- To detect the percentage of damage to fruit.

1.3 Contribution to Knowledge and Statement of Significance

In this research work, we have proposed a system to identify what type of fruit it is and detect damage on the fruit surface by using a machine learning technique. The overall contribution of this research work are summarized as follows:

- We have applied a huge amount of dataset containing different types of fruits.
- We were able to troubleshoot the classification problem for the same color fruit.
- We classified fruit by using Random Forest Classifier.
- We validated the training and testing dataset by using K-fold cross-validation to compare the validation result of the different machine learning algorithms.
- We have extracted three different features (color, shape and histogram) to make the classifier to give better accuracy.
- We able to detect damage and the percentage of damage by using K-means unsupervised machine learning algorithm.

1.4 Thesis Outline

We have divided our thesis into five chapters. Chapter 2, is the discussion of related work that we have collected from different research work which is related with this thesis work. Chapter 3, will discussed our proposed method. Chapter 4, presents an analysis of the result of the experiments applied to the fruit dataset. Chapter 5, represents discussion, chapter 6, conclusion & future work.

CHAPTER 2

LITERATURE REVIEW

CHAPTER 2

LITERATURE REVIEW

This chapter will explain about damage detection and classification of fruit and summarize several related works. Already there are exists various type of methods of classification and damage detection of fruits.

2.1 Damage Detection and Classification of Fruit

2.1.1 Damage Detection:

Damage detection is actually related to damage diagnosis. As we are working on fruit damage detection, then a question come in mind is that what exactly the damage detection on fruit. The damage detection on the fruit surface is to identify the bad portion from the good. Here we have summarized several research work that is related to damage or defect detection on fruit surfaces are mentioned bellow.

Saeideh Gorji Kandi in 2010, proposed a digital color imaging technique [4] for quality control of food and agricultural products, which was traditionally done by manual inspection by using Machine vision. Basically, he 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 degrees of defect can be computed and this also can be used for grading the fruits.

In 2014 Mr. S V. Phakade et. al. [5] proposed a computer vision system for defect detection of fruit surface in the agricultural field. In this entire work, he detect defected parts based on RGB and HSV color space. Among these two color space, HSV color space gives much better accuracy from RGB color space. Using this system the percentage of defects 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 colors could be matched with color of the 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] by using L*a*b color space. They find the defected portion of the 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. 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 in 2013. He 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.

Monika Jhuria et. al. [13] in 2013, investigated an approach of image processing for the detection of disease and fruit grading. They have used artificial neural network for the classification of disease. They consider three feature vectors, 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.

Ms. Kiran R. Gavhale et. al. [12] in 2014, presented a number of image processing techniques to extract disease part of leaf. For Pre-processing, Image enhancement is done by using DCT domain and color space conversion. After that, the segmentation take place by 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.

2.1.2 Fruit Classification:

Classification of fruit is to classify or identify one fruit from another, for example, a doctor can classify good blood cells from bad ones, photographers can classify their latest shot was beautiful or not, etc. Some of the classification problem those are solved by using different approach some of them are mentioned below.

Mureşan et. al. [1] in 2018, introduced a new, high-quality, dataset [4, 5] of images containing fruits that is essential for obtaining a good classifier. A method for identifying fruits he 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. He had 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 et. al. [2] in 2015, 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 a k-means algorithm for clustering the dataset. And then, they have used the most popular and more useful machine learning algorithm SVM is for classifying the images in terms of infected or non-infected images. By using his proposed system he achieve 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] in 2015, introduced a technique that diagnose and classify the external diagnoses 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 purposes. 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 an algorithm for different purposes. The image segmentation is done by using the 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 the OpenCV library and the proposed system is able to identify the disease with 90% accuracy.

A new method for classifying fruits using image processing technique is proposed by PL. Chithra and M.Henila [8] in 2019. 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.

Sachin Khirade and A. B. Patil [10] in 2015, 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] in 2013, 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.

Zalak R. Barot and Narendrasinh Limbad [9] in 2015, 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.

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.

CHAPTER 3

PROPOSED METHOD

CHAPTER 3

PROPOSED METHOD

This chapter will explain the proposed system in detail. We divide this section into several parts. First of all, we discuss the dataset we have used in our system. Then we explain all the procedures, how our system identify fruit and detect damage on fruit through machine learning.

3.1 Fruit Identification and Damage Detection System

Fruit identification and damage detection are some of the very challenging tasks for helping people to detect and identify fruit and damage automatically. Many research works have been done in the last couple of years to identify the fruit and detect damage on the fruit surface. Most of the research has done with Matlab. The most important thing is that most of the previous method is done for a few fruits whereas we have worked with fifteen different fruits.

To build up our framework, we utilize four design phases such as image preprocessing, feature extraction, training model, fruit identification and damage detection. There we have applied Machine Learning techniques (illustrated in Figure 3.1). This approach delivers an optimal result for identifying the fruit and detecting damage.

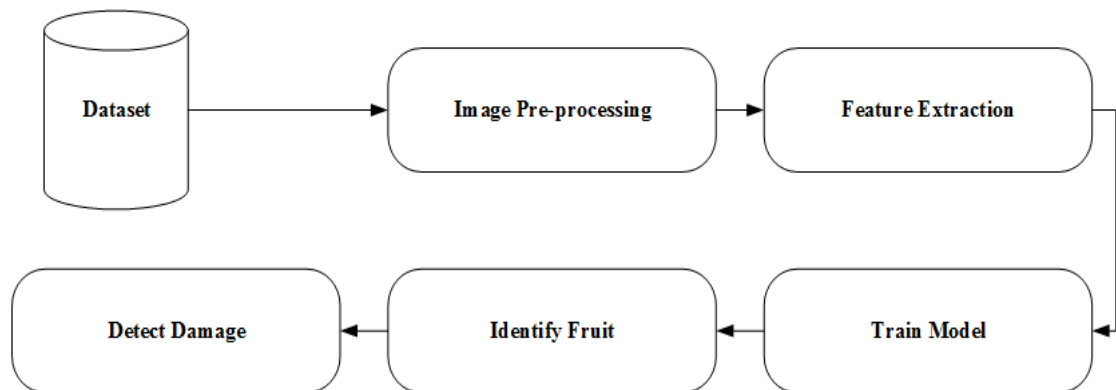


Figure 3.1: Fruit Identification and Damage Detection System

3.1.1 Image Preprocessing

Before extracting feature of fruit image, we resized our images those are included with the dataset fruits-360 [14]. We have resized our images for getting the large image that really helps to take features in detail. We have resized every training and testing image into 500×500 to achieve more data points those will help to achieve best accuracy of fitting ML model. Color enhancement is also a part of preprocessing that we used while detecting damage in our proposed method. The enhancement of image have done for making different the damage color from the good portion of fruit.

3.1.2 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 is a wider range of feature extraction algorithms in Computer Vision. In terms of feature extraction basically we used three terms like color, shape, and texture. The reason for taking these three features is that the “Tomato Cherry Red” is almost similar to “Tomato Maroon” according to color but the shape is different. We measure color, shape, and texture with Color Histogram, Hu Moment and Harlinck Texture. Features are extracted from different classes that we mentioned before in the 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 save it along with its label in an HDF5 file format. Instead of using HDF5 file-format, we could use “.csv” file-format to store the features. But, as we are working with large amounts of data for becoming familiar with HDF5 format is worth it. In the following sub-section, we are going to explain three features descriptors such as Color Histogram, Hu Moment and Harlinck Texture.

a) Color Histogram

A color histogram is one of the way we have used to extract features. The color histogram represents the arrangement of color in an image. Actually, a color histogram is the arrangement of the pixel that contains color with a range of (0-255). Another purpose of the color histogram is to represent the color position in an image.

b) Hu Moment

The weighted average of image pixel intensities is called image moments. We find moment by using the following equation.

$$M = \sum_x \sum_y I(x, y) \quad (1)$$

In the above equation x , and y represents the position of the image pixel. Here, $I(x, y)$ is representing the intensity of the location of the pixels of (x, y) . For calculating Hu Moment we need to find out the central moments of the image moment. We can calculate the central moments by using the following equation.

$$\mu_{ij} = \sum_x \sum_y (x - \bar{x})^i (y - \bar{y})^j I(x, y) \quad (2)$$

Hu Moments (or rather Hu moment invariants) are a set of 7 numbers calculated using central moments that are invariant to image transformations and the Hu Moment is actually used here to find the shape of the fruit. The central moment is calculated by using equation (2). Now we are ready to find the Hu Moment by using the following equation.

$$\begin{aligned} h1 &= \mu_{20} + \mu_{02}, \\ h2 &= (\mu_{20} - \mu_{02})^2 + 4\mu_{11}^2, \\ h3 &= (\mu_{30} - 3\mu_{12})^2 + (3\mu_{21} - \mu_{03})^2, \\ h4 &= (\mu_{30} + \mu_{12})^2 + (\mu_{21} + \mu_{03})^2, \\ h5 &= (\mu_{30} - 3\mu_{12})(\mu_{30} - \mu_{12}) [(\mu_{30} + \mu_{12})^2 - 3(\mu_{21} + \mu_{03})^2] + \\ &\quad (3\mu_{21} - \mu_{03})(\mu_{21} + \mu_{03}) [3(\mu_{30} + \mu_{12})^2 - (\mu_{21} + \mu_{03})^2], \\ h6 &= (\mu_{20} - \mu_{02}) [(\mu_{30} + \mu_{12})^2 - (\mu_{21} + \mu_{03})^2] + 4\mu_{11}(\mu_{30} + \mu_{12})(\mu_{21} + \mu_{03}), \\ h7 &= (3\mu_{21} - \mu_{03})(\mu_{30} - \mu_{12}) [(\mu_{30} + \mu_{12})^2 - 3(\mu_{21} + \mu_{03})^2] - \\ &\quad (\mu_{30} - 3\mu_{12})(\mu_{21} + \mu_{03}) [3(\mu_{30} + \mu_{12})^2 - (\mu_{21} + \mu_{03})^2] \end{aligned} \quad (3)$$

The equation (3) is used to find the Hu Moment of fruit image for finding the shape of the fruit image. In this equation μ represents the central moment as mentioned earlier in equation (2).

c) Harlinck Texture: As we have worked with fruit image we need to find the image texture. Image texture is basically a set of metrics calculated for extracting features and to partition images into regions of interest and classify those regions. In another word, texture provides us some information about the arrangement of color or intensities in an image or selected region of an image.

3.1.3 Train Model

After extracting feature we apply a machine learning classifier to train our extracted data as saved into an HDF5 file. In this part, we create our machine learning model with the help of one of the popular python library which is Scikit-learn. Some of the models we mentioned here and from the following algorithms, we used RF classifier in our system to train our dataset and others are used to validation including RF.

- a) K-NN (k Nearest Neighbor):** The k-nearest neighbor is a supervised machine learning algorithm is used to solve the classification problems. This algorithm is one of the simplest algorithms from others. It takes training features vector as an input and classifies new data based on majority class for the k neighbors. Where k represents the number of training samples closest to the point entry.
- b) Support Vector Machine (SVM):** This algorithm is possibly more famous and more used supervised machine learning algorithm and it is used to categorize data in a high-dimensional space. It constructs hyper plane or a set of a hyper plane in a high-dimensional space that has the largest distance to the nearest training data point of any class. Training dataset D is the set of n couples of the element (x_i, y_i) where the y_i are 1 or -1 value each indicating the class to which the point x_i belongs. Any hyper plane can be written as the set of point's x_i satisfying $W \cdot x_i + b = 0$.
- c) Decision tree:** Decision Trees are a non-parametric supervised learning method used for classification and regression problems. The overall mechanism of the decision tree is used in the RF classifier.
- d) Logistic Regression:** This is another supervised machine learning algorithm dedicated to the classification problems. Logistic Regression is a Machine Learning classification algorithm that is used to predict the probability of a categorical dependent variable.

In logistic regression, the dependent variable is a binary variable that contains data coded as 1 (yes, success, etc.) or 0 (no, failure, etc.). In other words, the logistic regression model predicts $P(Y=1)$ as a function of X .

- e) **Random Forest (RF):** The random forest is a supervised machine learning algorithm used for both classification and regression problems. In RF it creates a forest with a number of trees. Each individual tree in the random forest spits out a class prediction and the class with the most votes becomes our model's prediction.

Different algorithms are used in a wide variety of applications. Below are some the application where the random forest algorithm is widely used.

1. Banking
2. Medicine
3. Stock Market
4. E-commerce

We used the RF algorithm in our proposed method to train and test our dataset.

- f) **K-fold Cross Validation:** Cross-validation is a statistical method used to estimate the skill of machine learning models. It is commonly used in applied machine learning to compare and select a model for a given predictive modeling problem because it is easy to understand, easy to implement, and results in skill estimates that generally have a lower bias than other methods.

In K-fold cross-validation the dataset is split in k number of groups. When a specific value for k is chosen, it may be used in place of k in the reference to the model, such as $k=10$ becoming 10-fold cross-validation.

- g) **Linear Discriminant Analysis (LDA):** Linear Discriminant Analysis or Normal Discriminant Analysis or Discriminant Function Analysis is a dimensionality reduction technique which is commonly used for the supervised classification problems. It is used for modeling differences in groups i.e. separating two or more classes. It is used to project the features in higher dimension space into a lower dimension space.

3.1.4 Identify Fruit

By using training data we train a model. When our model is ready, then we load our test dataset and extract features of them in the same way that we had done before in the feature extraction part for preparing our training dataset. Then we predict the result with the testing image data. From fifteen different classes, we successfully identified each of the fruit belongs to each class with our proposed method.

3.1.5 Damage detection

The idea of damage detection on fruit is come from color segmentation, because, the color of the damage of fruit is different from the actual good fruit. So, if we can segment the color of the fruit we can detect damage and can measure the percentage of the portion of the damaged color on the fruit image.

From this idea, we implemented our detection system in two different ways, one of them is k-means color segmentation and another one is graph cut color segmentation technique.

Before segmenting the color of the fruit image by using the k-means clustering algorithm first we enhanced image color for increasing the brightness of the fruit image so that we can easily segment the damaged color of the fruit. After that, we segment the image color using the k-means clustering algorithm. We segmented the fruit image into three different clusters that mean we used $k = 3$. Where k is the number of cluster we measured. We segment the fruit image into 3 different cluster is because to keep the background of the image is in one cluster, the damage is another cluster and the good portion of the fruit image is another cluster. After that, we calculated the percentage each of the segmented cluster to measure the percentage of the damage portion of the fruit image. One thing is to remember that is we can segment the color of the damaged portion of the fruit image with good accuracy only for single color fruit.

In graph cut segmentation we have followed the same procedure as we have done for k-means clustering. First of all, we enhance the image color an then we apply graph cut segmentation technique to segment the color of fruit image. We used Scikit-learn and OpenCV python library for implementing this part.

3.2 Proposed Method

The following Figure 3.2 that represents the proposed method of this thesis. Here we first load the fruit image dataset. The dataset made by two-part one is for test data another one is for train data.

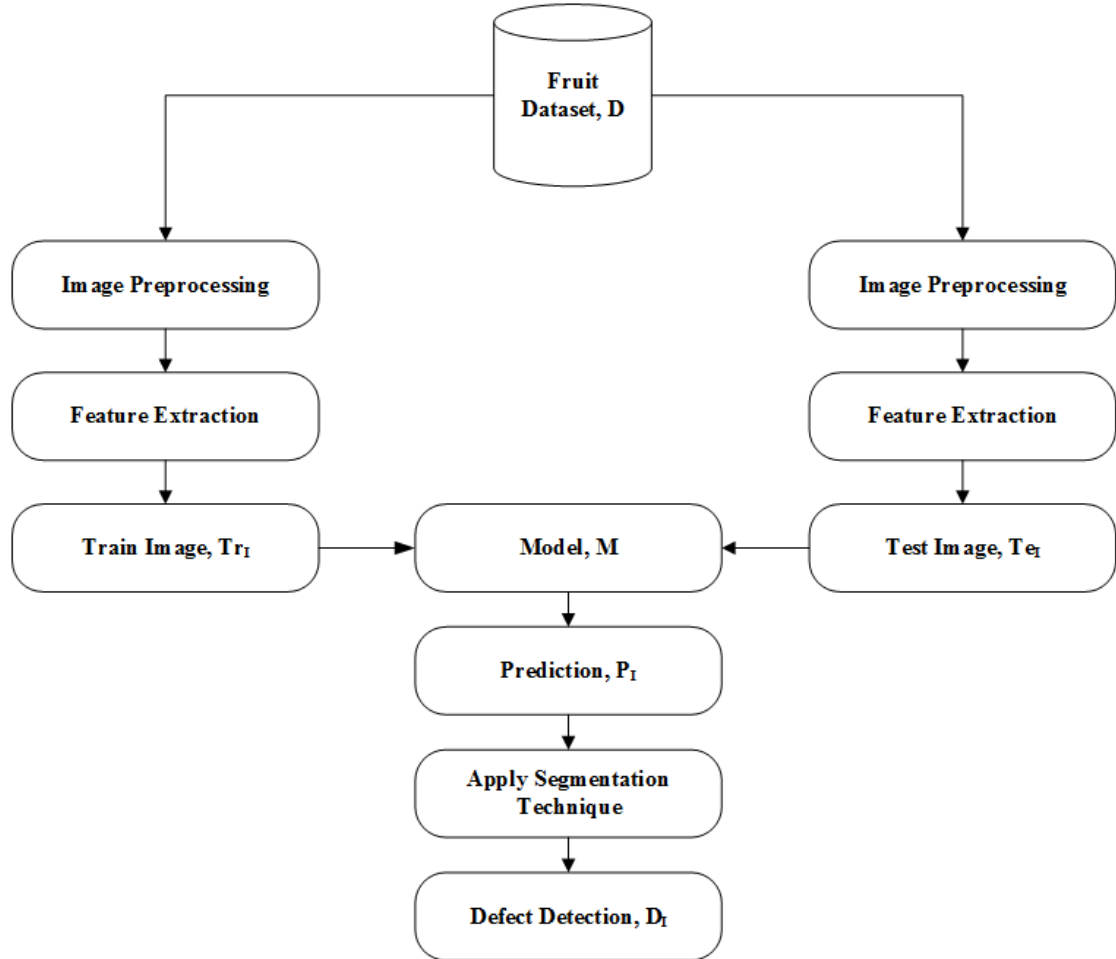


Figure 3.2: Proposed Method

After that, it loads every single image from the training dataset and then it does some processing on those images. After that our method goes to the next section to extract features from the processed images. After feature extraction, this method goes to the next section for the train ML model with extracted features. After that, the system loads fruit image from test dataset and do the same as done before. With the extracted feature of the test fruit image the model predicts which class it belongs to and the return label of the class. Finally, the fruit is predicted.

After predicting fruit this model applies segmentation technique for segmenting color of the predicted fruit image and the damage and good portion of the predicted fruit will be segmented.

The following algorithm is to identify the fruit and detect damage on fruit. Here are some steps that are followed by our proposed method.

Algorithm 1: Identify the Fruit and Detect Damage on Fruit

Input: Semi-supervised fruit image data, D ;

Output: Fruit identification & damage detection;

Method:

- 1: Load images for training data, Tr_I ;
 - 2: Image preprocessing;
 - 3: Feature extraction;
 - 4: Train image, Tr_I ;
 - 5: Create model, M ;
 - 6: Load image for test data, Te_I ;
 - 7: Repeat 2 and 3;
 - 8: $M \leftarrow Te_I$;
 - 9: $M \leftarrow Tr_I$;
 - 10: $P_I \leftarrow M$;
 - 11: Apply segmentation technique;
 - 12: Defect detection, D_I ;
-

CHAPTER 4

EXPERIMENTAL ANALYSIS

CHAPTER 4

EXPERIMENTAL ANALYSIS

In this entire chapter, first we will discuss about how and where from we have collected data. After that we explain our dataset we used. After that we explain how we setup environment for implementing proposed system. At the end of this section, we explain experimental analysis and then the accuracy we gained is to be discussed.

4.1 Data Collection

Actually we have used an existing fruit image dataset that we collect from kaggle. The name of the dataset that we used in our system is fruits-360.

4.2 Dataset Information

We use the fruits-360 dataset [14]. 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 4.2.1.

Table 4.2.1: Dataset Information.

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

Here we have used label data as training ML model and for test ML model we have used unlabeled data.

4.3 Environmental Setup

To evaluate performance and effectiveness of our experiment, we applied several software and open source tools. The experiments were carried out in a computer with Windows 10 Operating System, Intel Core I5 Processor, and RAM 4 GB.

Here, the table 4.1.1 shows the name of the Package, Descriptions and URL link used in this thesis. The column file says that total number of files are considered from the thesis sources.

Table 4.3.1: Open Source Software/Project found from different Enterprise Repository online.

Package Name	Description	URL Link
Python v3.73	Python is an interpreted, general-purpose, dynamic programming language. Python provides object oriented feature.	https://www.python.org/downloads/release/python-373/
Anaconda3 v2019.03	Anaconda is a free and open-source distribution of the Python and R programming languages for scientific computing, that aims to simplify package management and deployment.	https://www.anaconda.com/distribution/
Scikit Learn v0.20.3	Scikit-learn (also known as sklearn) is a machine learning library for python programming language. It provides a lots of machine learning related packages.	https://scikit-learn.org/stable/install.html
OpenCV v4.1.0	OpenCV (Open Source Computer Vision Library) is an open source computer vision and machine learning software library.	https://opencv.org/

4.4 Experimental Analysis

This section describes how the total experiment analysis has been carried out. Here we have shown the validation result that we achieve when our system perform the task for identifying fruit. We used k-fold cross validation is a statistical method used to estimate performance of different machine learning models. In this experiment we used seven different machine learning models (LR, RF, SVM, CART, KNN, LDA, NB) to measure performance of them using k-fold cross validation. The following Figure 4.5.1 shows the validation result achieved by the k-fold cross validation.

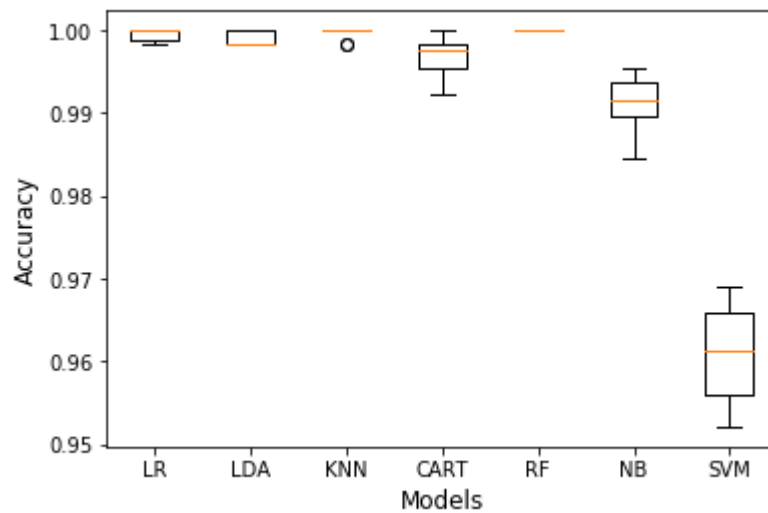


Figure 4.4.1: Cross Validation Result

Figure 4.4.1 shows that the RF classifier gives more accuracy than others algorithms.

Our system successfully identify different fruit with best accuracy. We apply our system to identify fruit from fifteen classes, the result is mentioned bellow.

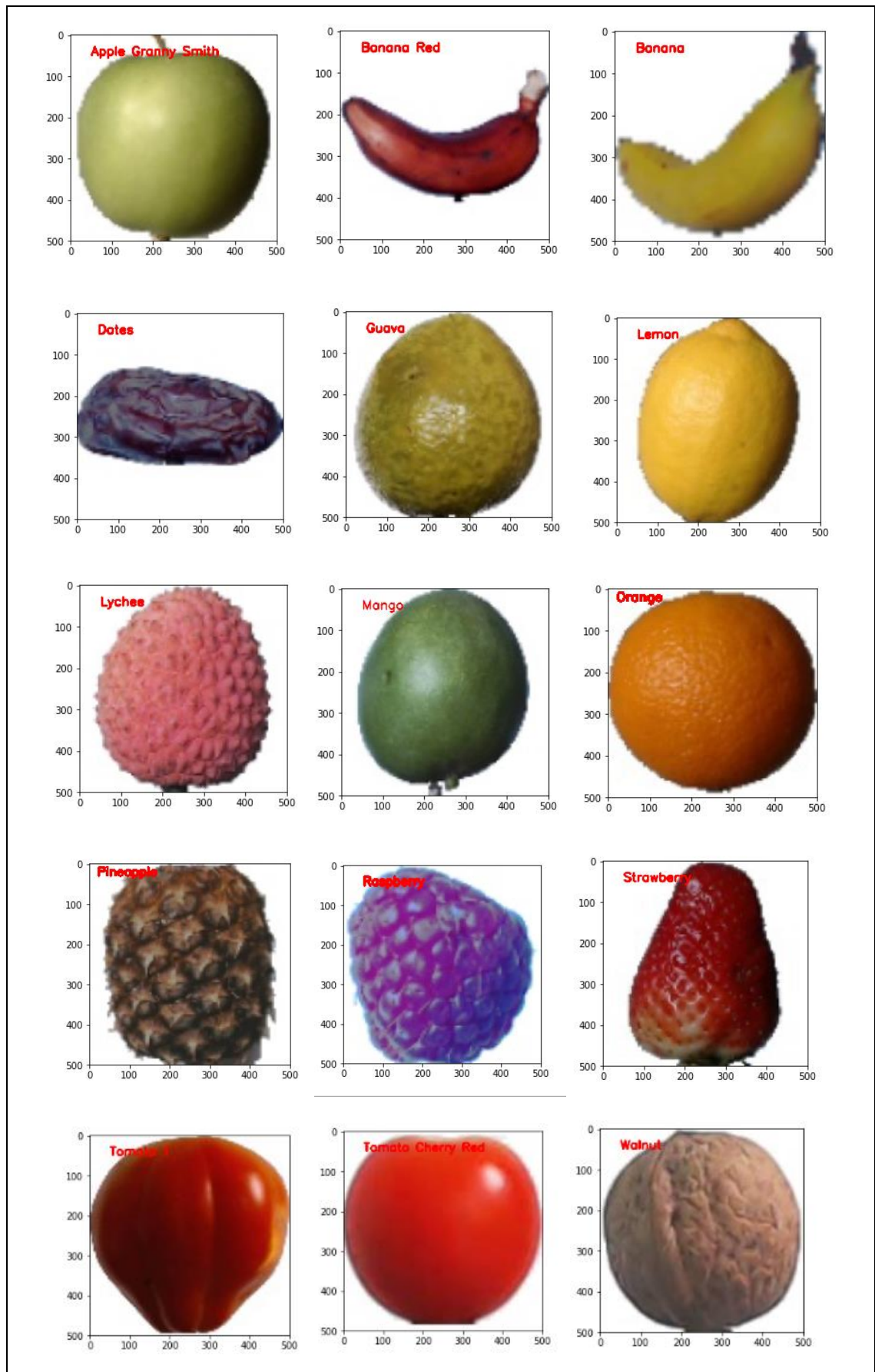


Figure 4.4.2: Fruit Identification Result

Fruit damage detection is actually a big problem. In our proposed system damage detection problem is actually remove a bit. The following Figure 4.4.3 shows the damage detection result of a fruit that has a single color with damage.

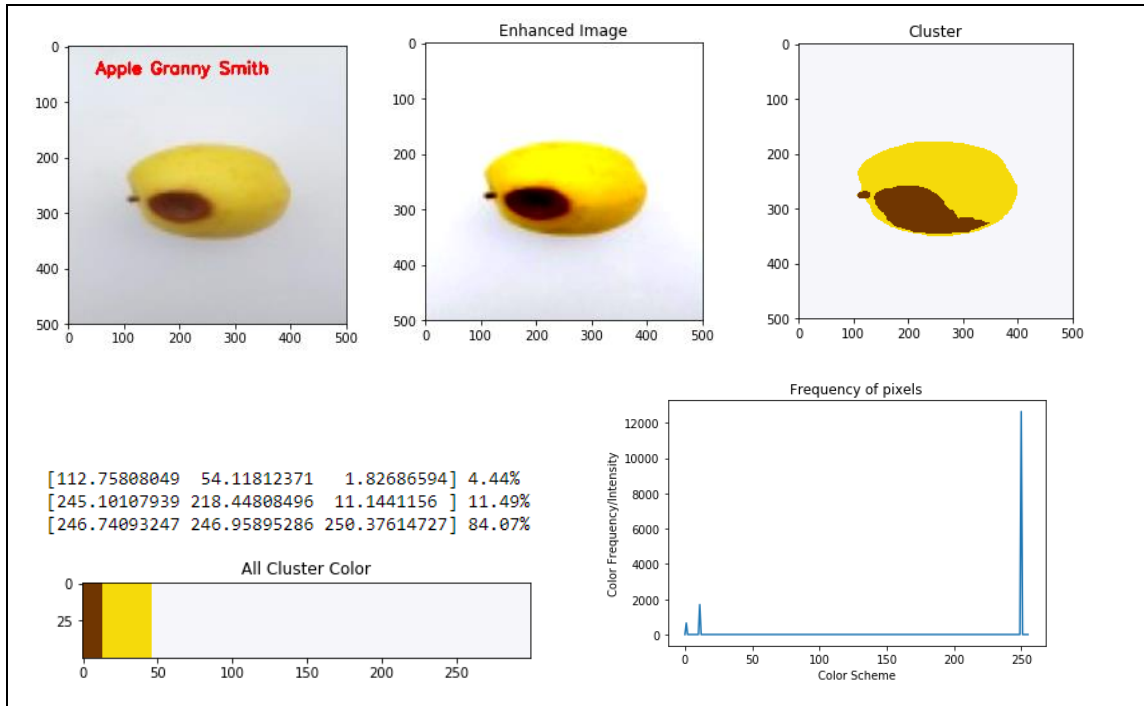


Figure 4.4.3: Damage Detection Using K-Means Algorithm

According to Figure 4.4.3 here we segmented the fruit image into three cluster by using k-means algorithm and measured the percentage each of these cluster. The reason of taking three clusters, so that we can separate background from the actual fruit part that belongs to one cluster. Another reason is to differentiate damage part from the actual good portion of the fruit part that belongs to another cluster and rest of the portion for good part which is in another cluster. These are the reason for taking three clusters.

We segmented the fruit image using another algorithm which is graph-cut that we mentioned at the last part of the section 3.2.5. The following Figure 4.4.4 and 4.4.5 show the detection result.

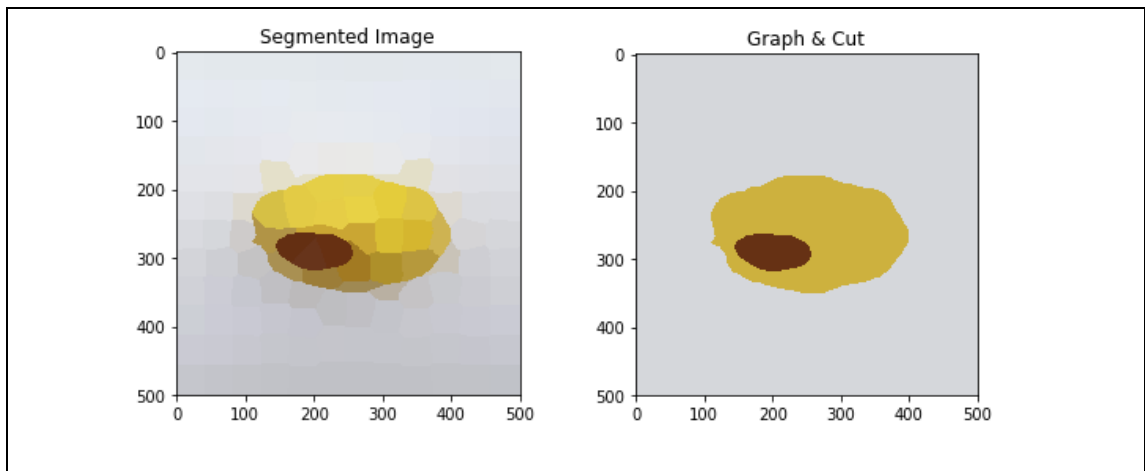


Figure 4.4.4: Damage Detection Using Graph-Cut

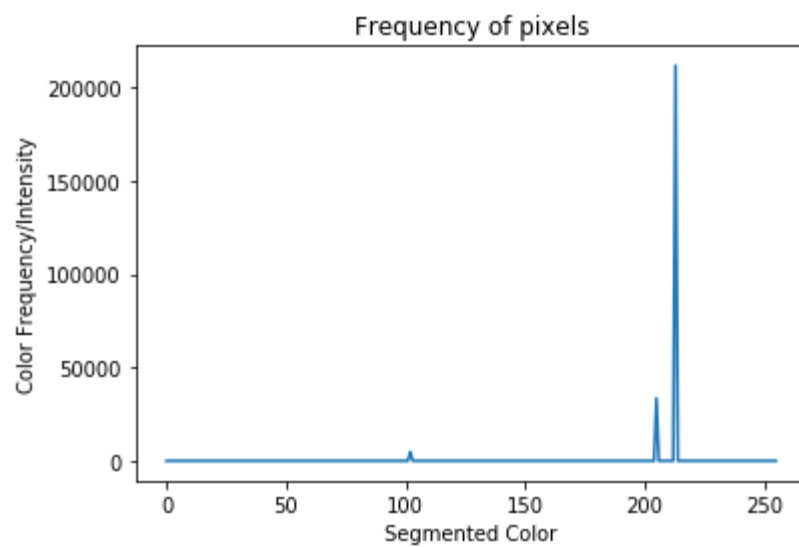


Figure 4.4.5: Damage Detection Using Graph-Cut (Histogram Plot)

Figure 4.4.5 represents the color intensity of segmented portion of the fruit image and this is a histogram plot. Here the y-axis represents the frequency of image color and x-axis represents the color scheme.

CHAPTER 5

DISCUSSION

CHAPTER 5

DISCUSSION

This thesis presented a construct design for informing the importance of fruit classification and damage detection and possible solution's provisions so a fully-functional decision support application in future can be developed. We have discussed the performance of different ML classifiers in identifying fruit in a shorter time. The experiment was conducted using Anaconda3 v2019.03. We have applied Random Forest classifier for identifying fruit. We showed validation result in Figure 4.4.1 that shows a comparison with different ML classifier. Among seven classifier (LR, RF, SVM, CART, KNN, LDA, NB), RF classifier gives us maximum accuracy. Damage detection part is done by using two different algorithm. Using k-means clustering algorithm we divide a fruit image into three clusters and the percentage of each cluster color is measured. Here is a limitation of doing clustering is that our system divide a fruit into three clusters if there exist damage or not. If a fruit is carrying only one color it is so much easy to segment damage portion from that fruit image. Using graph-cut segmentation technique is also segments the fruit image according to color. In graph-cut algorithm we cut our fruit image into 150 segments. Graph-cut algorithm is much more efficient for segmenting fruit image according to color of the image than k-means algorithm. We believe that the current study has laid the ground for future research on inferences and discovery of classification and detecting damage problem based on cause-event relation.

5.1 Theoretical Implication for Classifying Fruit and Detecting Fruit Damage

Fruit classification is actually referred to classify or separate one fruit from another fruit. From these ideas, it should be easier to identify the fruit. Our proposed method is theoretically stable enough for solving many kinds of classification problems like fruits, vegetables, and flower classification.

5.2 Practical Implication for Classifying Fruit and Detecting Fruit Damage

Our system highlights the use of machine learning technology in the field of fruit industry. This system are presently extensively used for quality evaluation of fruits.

CHAPTER 6

CONCLUSION

CHAPTER 6

CONCLUSION

In this research work, we tried to classify different kinds of fruits and also detecting the damaged part on the fruit surface. Here actually, we just apply the color segmentation technique for segmenting damage portion from the good part. We implemented the entire work with the help of various supervised and unsupervised machine learning algorithm and, we achieve best accuracy by using RF classifiers. We got an average 99.9% of accuracy for the different classifier which is done by using K-fold cross validation technique. We worked with a large number of fruit images for getting the best accuracy and we did it. Our method will be very effective for fruit processing industries who have been classifying fruit by manual. Our system comes with a great opportunity for the farmer and for the fruit processing industrialist for helping them to reduce more use of man power and cost.

6.1 Future Work

From our point of view, one of the main objectives for the future is to improve accuracy. Another objective is to expand the data set to include more fruits. Machine learning techniques will be very effective to classify and detect of fruit damage. The more using of a large dataset will give better accuracy. In the future we also try to identify fruit with real-time data that will be captured by the camera. The damage will be detected in a more specific way.

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 Kaggle. <https://www.kaggle.com/moltean/fruits>. last visited on 23.04.2019 =>1, 10