

**COMPUTER VISION APPROACH FOR CAULIFLOWER DISEASE
RECOGNITION**

BY

SHAHRUN SIDDIQUE TAKI

ID: 161-15-7195

AND

SYEDA KHADIZATUL MARIA

ID: 161-15-7325

This Report Presented in Partial Fulfillment of the Requirements for the
Degree of Bachelor of Science in Computer Science and Engineering.

Supervised By

Md. Jueal Mia

Senior Lecturer

Department of CSE

Daffodil International University

Co – Supervised By

Shah Md. Tanvir Siddiquee

Assistant Professor

Department of CSE

Daffodil International University



DAFFODIL INTERNATIONAL UNIVERSITY

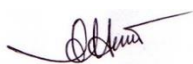
DHAKA, BANGLADESH

SEPTEMBER 2020

APPROVAL

This Project/internship titled “**Computer Vision Approach for Cauliflower Disease Recognition**”, submitted by Shahrin Siddique Taki , ID No:161-15-7195 and Syeda Khadizatul Maria , ID:161-15-7333to the Department of Computer Science and Engineering, Daffodil International University has been accepted as satisfactory for the partial fulfillment of the requirements for the degree of B.Sc. in Computer Science and Engineering and approved as to its style and contents. The presentation has been held on September 2020.

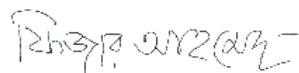
BOARD OF EXAMINERS



Dr. Syed Akhter Hossain
Professor and Head

Department of Computer Science and Engineering
Faculty of Science & Information Technology
Daffodil International University

Chairman



Dr. Fizar Ahmed
Assistant Professor

Department of Computer Science and Engineering
Faculty of Science & Information Technology
Daffodil International University

Internal Examiner



Abdus Sattar
Assistant Professor

Department of Computer Science and Engineering
Faculty of Science & Information Technology
Daffodil International University

Internal Examiner



Dr. Mohammad Shorif Uddin
Professor

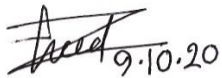
Department of Computer Science and Engineering
Jahangirnagar University

External Examiner

DECLARATION

We hereby declare that, this project has been done by us under the supervision of Md. Jueal Mia, Senior Lecturer, Department of CSE Daffodil International University. We also declare that neither this project nor any part of this project has been submitted elsewhere for award of any degree or diploma.

Supervised by:



9.10.20

Md. Jueal Mia

Senior Lecturer Department of CSE
Daffodil International University

Co-Supervised by:



Shah Md. Tanvir Siddiquee

Assistant Professor
Department of CSE
Daffodil International University

Submitted by:



Shahrin Siddique Taki
ID: -161-15-7195
Department of CSE
Daffodil International University



Syeda Khadizatul Maria
ID: -161-15-7333
Department of CSE
Daffodil International University

ACKNOWLEDGEMENT

First we express our heartiest thanks and gratefulness to almighty God for His divine blessing makes us possible to complete the final year project/internship successfully.

We really grateful and wish our profound our indebtedness to **Md. Jueal Mia, Senior Lecturer**, Department of CSE Daffodil International University, Dhaka. Deep Knowledge & keen interest of our supervisor in the field of “*Machine learning*” to carry out this project. His endless patience ,scholarly guidance ,continual encouragement , constant and energetic supervision, constructive criticism , valuable advice ,reading many inferior draft and correcting them at all stage have made it possible to complete this project.

We would like to express our heartiest gratitude to Professor Md Juel Mia, Shah Md. Tanvir Siddiquee and Dr. Syed Akhter Hossain, Head, Department of CSE, for his kind help to finish our project and also to other faculty member and the staff of CSE department of Daffodil International University.

We would like to thank our entire course mate in Daffodil International University, who took part in this discuss while completing the course work.

Finally, we must acknowledge with due respect the constant support and patients of our parents.

ABSTRACT

Cauliflower dominates a major share in terms of total winter cropping area and production in Bangladesh. It has many health benefits like decrease the risk of obesity, diabetes, heart disease etc. It is a cultivated and winter crop and has huge demand in the country. But if proper care is not taken many serious disease will effects on plants and will reduce productivity, quantity and quality of cauliflower. Manually monitoring of plant disease is very difficult as it requires tremendous amount of work and excessive time. Automatic recognition of disease through computer vision approach is becoming more popular day by day. So in this paper we introduced a modern technique to recognize diseases that occur on plants in cauliflower. The most common disease in cauliflower disease is Bacteria Soft Rot, Black Rot, Buttoning, Downy mildew in Bangladesh. Our proposed solution would support agriculture industry of Bangladesh to grow cauliflower more effectively and will increase its production by taking proper step after automated recognize of these diseases. In our work, for image segmentation, k-means clustering is used after image preprocessing method is applied, ten relevant features are extracted. For classification, we compared various classification technique. Random Forest algorithm achieves overall 81.68% accuracy.

TABLE OF CONTENTS

CONTENTS	PAGE
Board of examiners	i
Declaration	ii
Acknowledgements	iii
Abstract	iv
Table of contents	v-vi
List of figures	vii
List of tables	viii
CHAPTER	
CHAPTER 1: INTRODUCTION	1-3
1.1 Introduction	1-2
1.2 Motivation	2
1.3 Rationale of the Study	3
1.4 Outcome	3
1.5 Report Layout	3
CHAPTER 2: BACKGROUND	4-5
2.1 Introduction	4
2.2 Related Works	4-5
2.3 Comparative Study	5

CHAPTER 3: RESEARCH METODOLOGY	6-15
3.1 Data Collection	6
3.2 Description of Cauliflower Disease	7-9
3.3 Cauliflower Disease Detection Process	10-15
CHAPTER 4: CLASSIFICATION AND MODEL STUDY	16-20
4.1 Classification	16
4.2 Model Study	16-20
CHAPTER 5: EXPERIMENTAL RESULTS AND DISCUSSION	21-24
5.1 Results	21
5.2 Descriptive Analysis1	21-24
CHAPTER 6: CONCLUSION AND FUTURE WORK	25
6.1 Conclusion	25
6.2 Future Work	25
REFERENCES	26-27
PLAGIARISM REPORT	28

LIST OF FIGURES

FIGURES	PAGE NO
Figure 1.1.1: Provides the area of cauliflower crop	1
Figure 3.1.1: Cauliflower field during harvest	6
Figure 3.2.1: Bacterial Soft rot diseases	7
Figure 3.2.2: Black rot diseases	8
Figure 3.2.3: Downy Mildew diseases	8
Figure 3.2.4: Buttoning diseases	9
Figure 3.2.5: Healthy Cauliflower	9
Figure 3.3.1: Working Process of Cauliflower Disease Recognition	10
Figure 3.3.2: Original Image of diseased cauliflower	11
Figure 3.3.3: Contrast Enhancement	12
Figure 3.3.4: Segmented Image	13
Figure 4.1.1: Selected Machine Algorithm Category	15
Figure 5.2.1: Step by step process	21
Figure 5.2.2: Comparison of the other classifiers	23

LIST OF TABLES

TABLE	PAGE NO
Table 2.3.1: Comparison of our work and other's work	5
Table 4.2.1: Performance of k star	16
Table 4.2.2: Performance of Random Committee	16
Table 4.2.3: Performance of IBK	16
Table 4.2.4: Performance of Multilayer Perceptron	17
Table 4.2.5: Performance of Randomizable Filtered Classifier	17
Table 4.2.6: Performance of PART	18
Table 4.2.7: Performance of Multiclass Classifier	18
Table 4.2.8: Performance of Bagging	18
Table 4.2.9: Performance of random subspace	19
Table 4.2.10: Performance of Random Forest	19
Table 5.2.1: Confusion Matrix	22
Table 5.2.2: Performance of Random Forest Classifier	22

CHAPTER 1 INTRODUCTION

1.1 Introduction

Bangladesh is an agricultural country and approximately 75 % of the population depends on agriculture. It provides a livelihood for a significant population in Bangladesh and according to the World Bank, almost half of all workers in Bangladesh and two-thirds in rural areas are directly engaged in agriculture [5]. The Nature of Bangladesh has adorned itself with six seasons. Every season has come with a variety of fruits and vegetables. Mostly the varieties of vegetable grows in the winter season. The most favorite and popular vegetable in winter is cauliflower. It is one of the significant vegetable crop. It is popular not only for its taste but also its health benefits. Approximately 2500-3000 cauliflower can be produced per Bigha (1 Bigha = 0.16055846 hectare) of a land. In Bangladesh, cauliflower was cultivated in a total area of 10,000 hectares in 1996-97 and the production of cauliflower was approximately 76,000 tons [4].

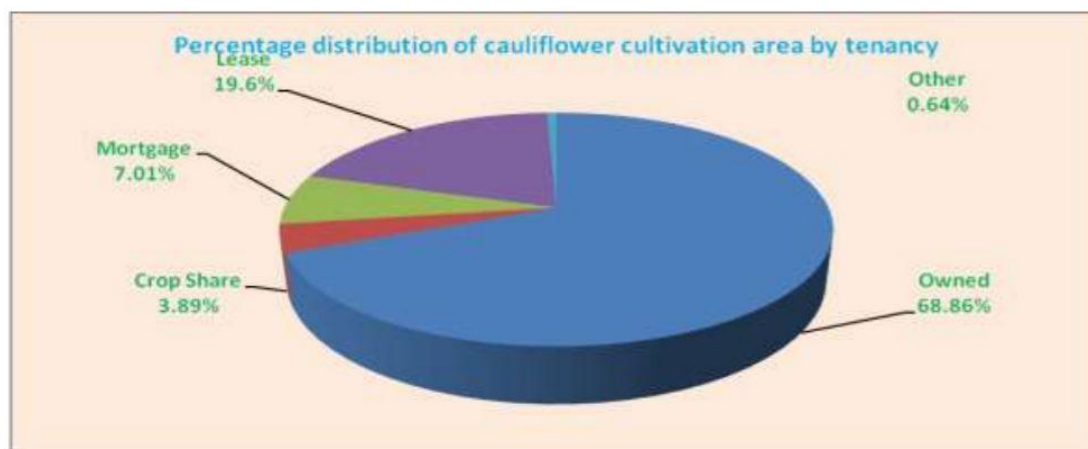


Fig 1.1.1 : Provides the area of cauliflower crop. Adapted from Report on The Productivity Survey of Cauliflower Crop

But the disease in plants reduce its productivity also reduce its quality and quantity. The disease especially on its leaves and flower has hindered its quality. Around 10 percent of

crops are lost every year due to plant diseases before they reach in dining table, resulting in substantial economic discrepancy to the farmers in developing countries. If the proper solution is not taken the production will be reduced and then it will cause poverty, unhealthiness and food insecurity. Various diseases of cauliflower that can damage or even destroy the plant. Mainly cauliflower disease type classified into bacterial, viral, fungus etc. Common disease of cauliflower is black rot, club rot, downy mildew, bacterial soft rot, buttoning, stem rot etc. In this work, we mainly focused on bacterial soft rot, downy mildew, black rot and buttoning.

Manual monitoring is very difficult and does not give satisfactory result. Besides manual monitoring is old techniques and can cause false detection. It is also time consuming. Automatic recognition of plant disease technique offers solution to maximize the productivity and also ensure the food security. Various modern technique are used in these fields. One of the popular application is computer vision. It is used for the recognition of plant disease.

1.2 Motivation

Bangladeshi economy depends on agriculture because it contributes almost 20 percent to national GDP. Besides it is the largest employment sector. But plant disease are important cause for reduction of productivity in Bangladesh. Reducing productivity in agriculture means it will cause poverty and food insecurity. So detecting plant disease is important to increase productivity. In our research based project we are proposing a system where user can automatic recognize cauliflower disease.

There are also some specific motives in our research based project:

- To reduce amount of work and processing time to recognize disease.
- To increase the production of cauliflower.

1.3 Rationale of the Study

In research area, there is so much research that were prosecuted to automatic disease detection. But for Cauliflower disease detection is so scarce. Very few work has been done in this topic. Also, Cauliflower disease detection carries much challenges because of collecting data as it is seasonal vegetable.

1.4 Outcome

Our work aims that to automatic recognize cauliflower disease (Black rot, Bacterial soft rot, Buttoning, Downy Mildew)

It will analysis disease part of the plant and show the accuracy of that disease.

1.5 Report Layout

In this chapter we have discussed about the introduction of the Cauliflower disease detection, motivation, rational of the study and the outcome of the thesis. Later followed by the report layout.

In chapter 2, we will discuss about the background of our research topic.

In chapter 3, we will discuss about the research methodologies employed in our study.

In chapter 4, we will discuss about classification and the model study.

In chapter 5, we will discuss about the acquired experimental results and discussion.

In chapter 6, we will discuss about the conclusion and future work.

CHAPTER 2 BACKGROUND

2.1 Introduction

Now the world is digital platform where everything is relying on the internet. So manually and naked eye disease recognition is really old technique. Besides it requires much time to identify diseases also the new comer in this agricultural field can predict false which leads complexity. Cauliflower disease recognition through computer vision can help farmer to increase productivity of cauliflower production. Previously several researches have been conducted to find plant disease recognition. But rarely can find any research which can recognize Cauliflower Disease.

2.2 Related Works

Computer vision is the new buzz word in the world because of its potential growth in a short time. A day will come when it will successfully mimic human sight. Though a lot of research work has been done based on disease recognition using computer vision but very few work done on cauliflower disease detection.

Automated plant disease recognition using computer vision approach plays an important role in agricultural field thus it is now important research topic. Some existing method of computer vision for plant disease detection is discussed in this section:

In [1] authors have proposed detect rice leaf diseases using texture and color Descriptors. Their proposed algorithm in classification is SVM which gives 92.5% accuracy and ANN gives 87.5% accuracy. Their database consists of 120 image dataset. They extracted 13 features from the rice leaf. Image segmentation was done by thresholding method.

In [2] authors used support vector machine to recognize and detect tea leaf's diseases. They introduced grayscale image segmentation using color space. Then they extracted 10 features during classification. In their work, 300 tea leaves images was tested

by their system. Their proposed algorithm SVM achieved 93% accuracy and neural network achieved 91% accuracy.

In [6] authors proposed a system using computer vision approach to recognize local fruit. They proposed a system application where users can send the image of a fruit and their expert system can analyze that image and sent a feedback to the user. In their system they used resizing, contrast enhancement, conversion of RGB color space to L*a*b color space for image preprocessing, and for image segmentation they used k-means clustering. They divided their dataset into 6 class of different local fruits. After segmentation, feature extraction is done by 2 feature set. SVM is implemented for classification. Their system achieved 94.61% accuracy.

2.3 Comparative Study

Research from different species' disease recognition found in this field but as per our knowledge, there has no such work on cauliflower disease recognition so we compare this work to other existing work on various plant diseases with an average accuracy of classification method and other attributes. The table is shown the comparison of our work to other existing work.

Reference	Plant Disease	Dataset	Class	Feature Extracted	Classification	Average Accuracy
This work	Cauliflower	500	5	10	Random Forest	81.68%
Ghyar et al[1]	Rice leaf	120	3	13	SVM,ANN	92.5%, 87.5%
Hossain et al [2]	Tea leaf	300	3	10	SVM, Neural Network	93%, 91%
Mia et al [6]	Local Fruit	240	6	10	SVM	94.61%
Habib et al [8]	Jackfruit	480	5	10	Random Forest	90%

Table 2.3.1: Comparison of our work and other's work

CHAPTER 3

RESEARCH METHODOLOGY

3.1 Data Collection:

The importance of data collection is essential in the field of research area. It is important to obtain reliable and sufficient data in order to achieve higher accuracy. We collected real time data for cauliflower disease such as bacterial soft rot, buttoning, black rot and downy mildew. These data were acquired on November 2019 to February 2020. Very few data are collected from internet. Real time data was collected mainly the area of Ashulia in Dhaka.



Fig 3.1.1: Cauliflower field during harvest

3.2. Description of Cauliflower Disease:

Cauliflower has many nutrition. It is mostly used as vegetables. It is popular vegetable in Bangladesh because of its very low calories with high in almost every vitamins. It reduces several disease like heart disease, defend against cancer etc. It grows well in cold weather therefor only available in winter season. But some disease are disrupted its production. Bacterial soft rot, black rot, downy mildew and buttoning are some of those disease. A short description these diseases is provided in this section:

1. Bacterial Soft rot:

It is caused by several bacteria mostly common is *Erwinia caratovora*. Symptom: water soaked areas from its leaves and flower head.



Fig 3.2.1: Bacterial Soft rot diseases

2. Black rot:

It is caused by bacterium known as *Xanthomonas campestris*. Symptom: affected leaves turn dull yellow to brownish. Yellow 'V' shaped appear on the leaf margin.



Fig 3.2.2: Black rot diseases

3. Downy Mildew.

It is caused by fungus like *Peronospora parasitica* and attacked head of cauliflower, stem and leaves. Small yellow spot which turns into brown can be seen in underside of the leaves.



Fig 3.2.3: Downy Mildew diseases

4. Buttoning:

It is physiological disorder which is caused because of insufficient of nitrogen. Due to this disorder, head of the cauliflower remain small and buttoned.



Fig. 3.2.4: Buttoning diseases

5. Healthy Cauliflower:

The edible head should be white and compact and the leaf should be fresh and green.



Fig. 3.2.5: Healthy Cauliflower

3.3 Cauliflower Disease Detection Process:

The basic step for disease recognize consist of following approaches:

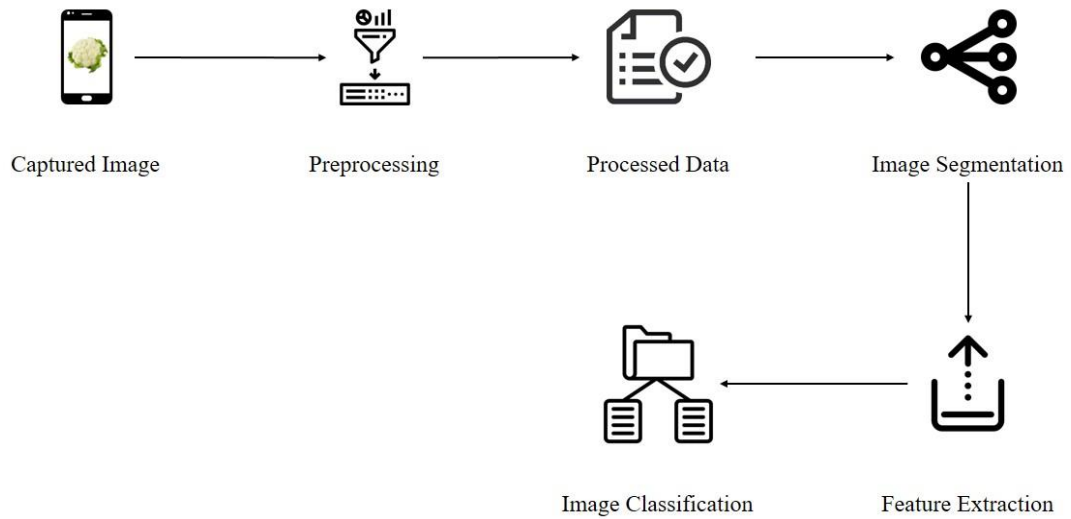


Fig. 3.3.1: Working Process of Cauliflower Disease Recognition

Image Acquisition:

Image acquisition is the first step of our method where sample images are collected for training and testing purpose. Images are captured through common method-photography with the help of digital camera. In this work, images are captured by digital camera and few images are collected from internet. Sample images include both healthy and affected plants. The cauliflower image is captured through digital camera and stores into standard digital format in digital media. These images are in RGB form. Sample image of bacterial soft rot as shown in fig 3.3.2.



Fig. 3.3.2: Original Image of diseased cauliflower

Image preprocessing:

Image collected from various source is called raw picture which cannot directly use in next step. So after collecting sample image, the next step is image preprocessing which is used for converting image in suitable form for improving the quality of the image for further process. Image processing requires removing noise or unwanted object removal, cropping image to size and shape, enhancing and filtering image etc. The following step we followed in this work:

- **Image resizing:** It is important to resize the input image for classification. Images are resized for further processing.
- **Image Filtering:** Image filtering is done by removing noise, smoothing filter etc.
- **Contrast Enhancement:** The technique is used to enhance contrast is histogram equalization. Images are enhanced for better perform.
- **Color conversion:** We increase in value: green color (59%) because green color mostly represent the healthy part of the plant and also have to decrease in value: red color (30%) and blue color (11%). So the equation of the conversion is:

$$\text{New Grayscale Image} = (0.3 * R) + (0.59 * G) + (0.11 * B)$$

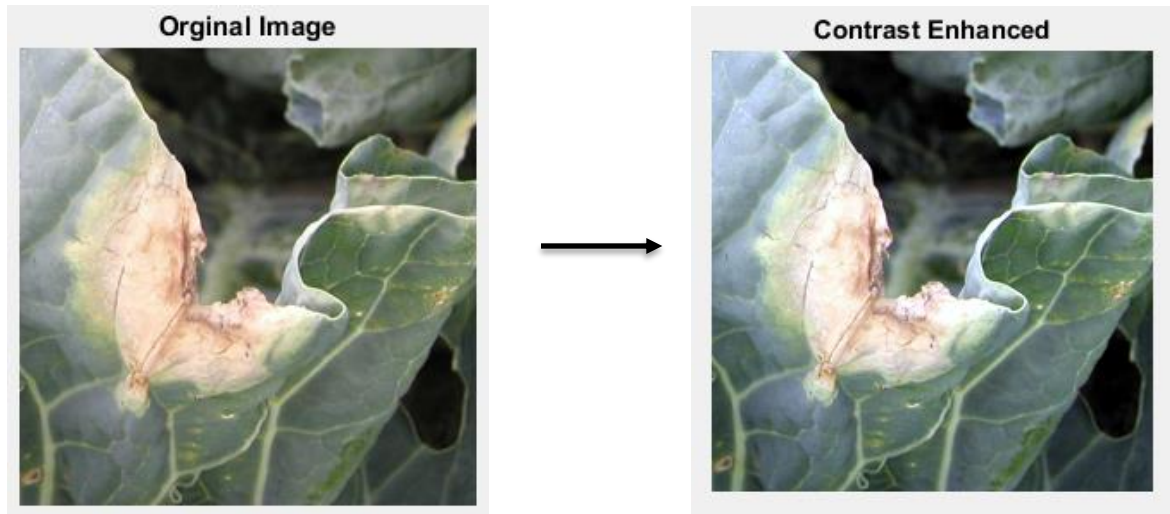


Fig 3.3.3: Contrast Enhancement

Image Segmentation:

In our work, we applied K Means clustering algorithm for image segmentation into three clusters as shown in fig 3.3.4. This algorithm is used for grouping the similar pixel of images according to some set of features into k number of classes. It is used to minimize the total squared of distances between the training images and the corresponding cluster. The first step we followed to convert image from RGB to L*a*b* color space where L is luminosity layer 'L*' and a*b* is chromaticity-layer.

After generating the image into three different clusters we measure the distance using Euclidean Distance Metric is given as follows:

$$\text{Euclidean Distance (d)} = \sqrt{(X2 - X1)^2 + (Y2 - Y1)^2}$$

Where (X1, Y1) and (X2, Y2) is two pixel points.



Fig 3.3.4: Segmented Image

Feature Extraction:

The feature extraction is a useful part of image analysis because it extracts useful information from an image. We know that every disease has its unique characteristics

feature which interprets the disease. These characteristics include color, shape, and texture. In our work, we used 5 texture analyses using the Gray-Level Co-Occurrence Matrix (GLCM) which is contrast, energy, homogeneity, correlation, and entropy. Other 5 statistical features are also extracted from the input image like mean, skewness, standard deviation, variance, and kurtosis. Mathematical formula is shown in below:

1. Contrast: $\sum_{i,j} |i - j|^2 p(i, j)$

2. Correlation: $\sum_{i,j} \frac{(i - \mu_i)(j - \mu_j)p(i, j)}{\sigma_i \sigma_j}$

3. Energy: $\sum_{i,j} p(i, j)^2$

4. Homogeneity: $\sum_{i,j} \frac{p(i, j)}{1 + |i - j|}$

5. Mean: $m = \sum_{i=0}^{L-1} x_i p(x_i)$

6. Standard Deviation: $s = \sqrt{\{\sum_{i=0}^{L-1} [(x_i - m)^2 p(x_i)]\}}$

7. Entropy: $\sum_{i=0}^{L-1} (p(x_i) \log_2 p(x_i))$

8. Variance: $\sum_{i=0}^{L-1} [(x_i - m)^2 p(x_i)]$

9. Kurtosis: $\frac{1}{s^k} \sum_{i=0}^{L-1} (x_i - m)^k p(x_i)$

10. Skewness: $\frac{1}{s^3} \sum_{i=0}^{L-1} (x_i - m)^3 p(x_i)$

Classification:

The image classification technique is used to classify images into predefined classes according to their extracted features. Our proposed system consists of 5 classes: Bacterial Soft Rot, Black Rot, Downy Mildew, Buttoning, and Healthy. So we used multiclass classification method for our system. There are various classifiers that exist in this field. We compare various classifiers like Random Forest, IBK, K-star, Multilayer Perceptron, PART, Random Committee, etc. But Random Forest provides higher disease recognition accuracy (81.68%).

CHAPTER 4 CLASSIFICATION AND MODEL STUDY

4.1 Classification

For recognize disease of cauliflower we used 10 models to compare the average accuracy rate.

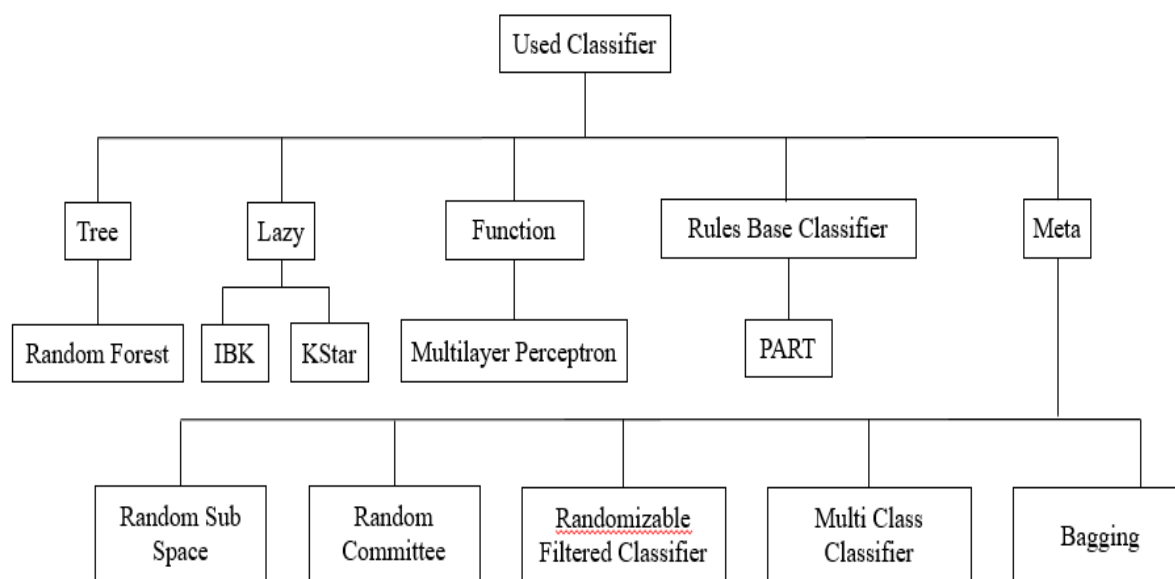


Fig. 4.1.1: Selected Machine Algorithm Category

Here we discussed 10 models of the classification.

4.2 Model Study

K-star:

k star model is a lazy classifier which is simple and an instance-based classifier. k star algorithm mainly used for entropy measure. Through k-star classification we get good accuracy but random forest gives the best accuracy of all the classification.

	Accuracy	Precision	Specificity	Sensitivity	FNR	FPR
Bacterial Soft Rot	78.24%	42.47%	89.76%	31.00%	69.00%	10.24%
Black Rot	79.22%	46.88%	87.56%	45.00%	55.00%	12.44%
Buttoning	85.10%	58.45%	85.61%	83.00%	17.00%	14.39%
Downymildew	81.57%	52.94%	88.29%	54.00%	46.00%	11.71%
Healthy	77.45%	47.42%	87.25%	41.82%	58.18%	12.75%

Table 4.2.1: Performance of k star

RandomCommittee:

It is a supervised algorithm under the meta group. It is a supervised learning algorithm which is even simpler that gives us 80.40% average accuracy.

	Accuracy	Precision	Specificity	Sensitivity	FNR	FPR
Bacterial Soft Rot	74.40%	35.71%	84.25%	35.00%	65.00%	15.75%
Black Rot	78.80%	47.00%	86.75%	47.00%	53.00%	13.25%
Buttoning	84.00%	58.47%	87.75%	69.00%	31.00%	12.25%
Downymildew	81.60%	54.35%	89.50%	50.00%	50.00%	10.50%
Healthy	83.20%	58.70%	90.50%	54.00%	46.00%	9.50%

Table 4.2.2: Performance of Random Committee

IBK:

KNN (k-nearest algorithm) which is alternatively known as IBK (Instance-Based Algorithm). It used Euclidean Distance Matric. Here K value predicts the number of nearest neighbors according to their distance.

	Accuracy	Precision	Specificity	Sensitivity	FNR	FPR
Bacterial Soft Rot	75.00%	35.29%	86.25%	30.00%	70.00%	13.75%
Black Rot	78.00%	45.19%	85.75%	47.00%	53.00%	14.25%
Buttoning	83.00%	55.91%	86.00%	71.00%	29.00%	14.00%
Downymildew	82.80%	57.78%	90.50%	52.00%	48.00%	9.50%
Healthy	81.60%	54.26%	89.25%	51.00%	49.00%	10.75%

Table 4.2.3: Performance of IBK

Multilayer Perceptron:

MLP, the Multi-layer Perceptron classifiers, which uses the back propagation technique along with hidden layers. Many researchers applied MLP because it works better with numerical value.

	Accuracy	Precision	Specificity	Sensitivity	FNR	FPR
Bacterial Soft Rot	78.76%	45.98%	88.25%	40.40%	59.60%	11.75%
Black Rot	83.37%	59.34%	90.73%	54.00%	46.00%	9.27%
Buttoning	76.15%	40.21%	85.46%	39.00%	61.00%	14.54%
Downymildew	83.97%	60.42%	90.48%	58.00%	42.00%	9.52%
Healthy	78.36%	46.88%	82.96%	60.00%	40.00%	17.04%

Table 4.2.4: Performance of Multilayer Perceptron

Randomizable Filtered Classifier:

A simple Filtered Classifier variant that demonstrates the model with a randomizable filter, more precisely, as the base classifier. Each base classifier is constructed using a different random number seed by using a randomizable filter classifier as it is an ensemble base classifier. Averaging the predictions provided by the individual base classifiers calculates the final output.

	Accuracy	Precision	Specificity	Sensitivity	FNR	FPR
Bacterial Soft Rot	73.80%	34.02%	84.00%	33.00%	67.00%	16.00%
Black Rot	76.40%	41.35%	84.75%	43.00%	57.00%	15.25%
Buttoning	84.60%	59.06%	87.00%	75.00%	25.00%	13.00%
Downymildew	81.00%	52.75%	89.25%	48.00%	52.00%	10.75%
Healthy	83.00%	59.26%	91.75%	48.00%	52.00%	8.25%

Table 4.2.5: Performance of Randomizable Filtered Classifier

PART:

PART is a Rules-based classifier. Association rules are used for class prediction among all the properties in this rules-based classifier. Those accurate predictions are considered as coverage. [3]. More than one conclusion can be predicted by them.

	Accuracy	Precision	Specificity	Sensitivity	FNR	FPR
Bacterial Soft Rot	75.20%	35.37%	86.75%	29.00%	71.00%	13.25%
Black Rot	79.00%	46.91%	89.25%	38.00%	62.00%	10.75%
Buttoning	83.20%	56.06%	85.50%	74.00%	26.00%	14.50%
Downymildew	79.60%	49.04%	86.75%	51.00%	49.00%	13.25%
Healthy	81.80%	54.46%	88.50%	55.00%	45.00%	11.50%

Table 4.2.6: Performance of PART

Multiclass Classifier:

Multiclass classifier is a Meta approach which is used for processing multiclass datasets with 2-class classifiers. This classifier is also able to apply error-correcting output codes.

	Accuracy	Precision	Specificity	Sensitivity	FNR	FPR
Bacterial Soft Rot	77.80%	42.86%	89.00%	33.00%	67.00%	11.00%
Black Rot	83.00%	57.28%	89.00%	59.00%	41.00%	11.00%
Buttoning	75.00%	37.62%	84.25%	38.00%	62.00%	15.75%
Downymildew	80.40%	50.89%	86.25%	57.00%	43.00%	13.75%
Healthy	81.80%	54.21%	87.75%	58.00%	42.00%	12.25%

Table 4.2.7: Performance of Multiclass Classifier

Bagging:

The bagging classifier is a meta-algorithm approach of the machine learning ensemble that produces different samples of the training dataset and produces a classifier for each one. It is implemented to evaluate the stability, increase accuracy, and also lowers variation and assists in preventing the over fitting of machine learning.

	Accuracy	Precision	Specificity	Sensitivity	FNR	FPR
Bacterial Soft Rot	75.80%	36.00%	88.00%	27.00%	73.00%	12.00%
Black Rot	78.80%	46.25%	89.25%	37.00%	63.00%	10.75%
Buttoning	79.40%	48.92%	82.25%	68.00%	32.00%	17.75%
Downymildew	82.20%	55.34%	88.50%	57.00%	43.00%	11.50%
Healthy	81.40%	53.40%	88.00%	55.00%	45.00%	12.00%

Table 4.2.8: Performance of Bagging

Random Subspace: The random subspace method, is known as attribute bagging, is an ensemble learning algorithm. It aims to improve the highest accuracy as the constructed decision trees expand in complexity.

	Accuracy	Precision	Specificity	Sensitivity	FNR	FPR
Bacterial Soft Rot	77.60%	40.91%	90.25%	27.00%	73.00%	9.75%
Black Rot	77.80%	43.04%	88.75%	34.00%	66.00%	11.25%
Buttoning	80.80%	51.45%	83.25%	71.00%	29.00%	16.75%
Downymildew	81.20%	52.88%	87.75%	55.00%	45.00%	12.25%
Healthy	79.00%	47.79%	85.25%	54.00%	46.00%	14.75%

Table 4.2.9: Performance of random subspace

Random Forest:

Random Forest algorithm is a supervised algorithm in traditional machine learning. It is decision tree based learning. It is powerful classifier because it can handle numeric and categorical data. Random forest algorithm can easily understandable to human. . The maximum number of votes from each of the nodes of the decision tree provide the final output for disease identification. This classifier we get higher accuracy 81.68% compare to other classifier.

	Accuracy	Precision	Specificity	Sensitivity	FNR	FPR
Bacterial Soft Rot	78.60%	45.07%	90.25%	32.00%	68.00%	9.75%
Black Rot	79.40%	48.51%	87.00%	49.00%	51.00%	13.00%
Buttoning	84.60%	58.27%	85.50%	81.00%	19.00%	14.50%
Downymildew	82.80%	58.33%	91.25%	49.00%	51.00%	8.75%
Healthy	83.00%	57.14%	88.75%	60.00%	40.00%	11.25%

Table 4.2.10: Performance of Random Forest

CHAPTER 5

EXPERIMENTAL RESULTS AND DISCUSSION

5.1 Results

In our experimental study, 500 images of different cauliflower diseases are acquired for our approach from Savar, Bangladesh. The image database consists of hundred images per class for Bacterial Soft Rot, Black Rot, Downy Mildew, Buttoning, and Healthy images. Acquired images are pre-processed through resizing, filtering, contrast-enhanced, noise reduction, etc. The segmentation of images is achieved by k-means clustering and Euclidean distance metrics. Then 10 features are extracted from the image and stored for further evaluation. The non-necessary features have been removed. Using this, the diseases are recognized using Random Forest to classify them.

5.2 Descriptive Analysis

In our work, Random Forest algorithm gives highest accuracy 81.68%. Performance measure of the Random Forest such as accuracy, precision, specificity, sensitivity, FNR, FPR were calculated through confusion matrix and the table shown in Table 5.2.2. 10 fold cross validation generates an average accuracy of the classifier. So we applied 10 fold cross validation. Other classification algorithm are compared in fig. 5.2.2 Step by step of our process is shown in table 5.2.1.




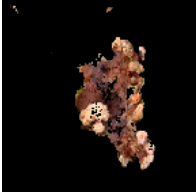







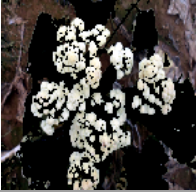



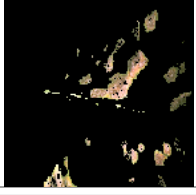




Original Image	Resized Image	Contrast-enhanced Image	Segmented Image	Type
				Bacterial soft rot
				Black rot,
				Buttoning
				Downy Mildew
				Healthy Image

Fig. 5.2.1: Step by step process

Confusion matrix is a visualized way to represent the performance of a classifier. Confusion matrix is made up column and rows where scenario represent as actual class and predicted class. Each row the represents actual class where each column represents the predicted class. In Table 5.2.1 multi class confusion matrix of Random Forest is shown.

Actual	Bacterial Soft Rot	Black Rot	Buttoning	Downy Mildew	Healthy	
	32	10	29	7	22	Bacterial Soft Rot
	7	49	8	22	14	Black Rot
	13	4	81	2	0	Buttoning
	6	26	10	49	9	Downy Mildew
	13	12	11	4	60	Healthy
	Predicted					

Table 5.2.1: Confusion Matrix

Class	Accuracy	Average Accuracy
Bacterial Soft Rot	78.60%	81.68%
Black Rot	79.40%	
Buttoning	84.60%	
Downy Mildew	82.80%	
Healthy	83.00%	

Table 5.2.2: Performance of Random Forest Classifier

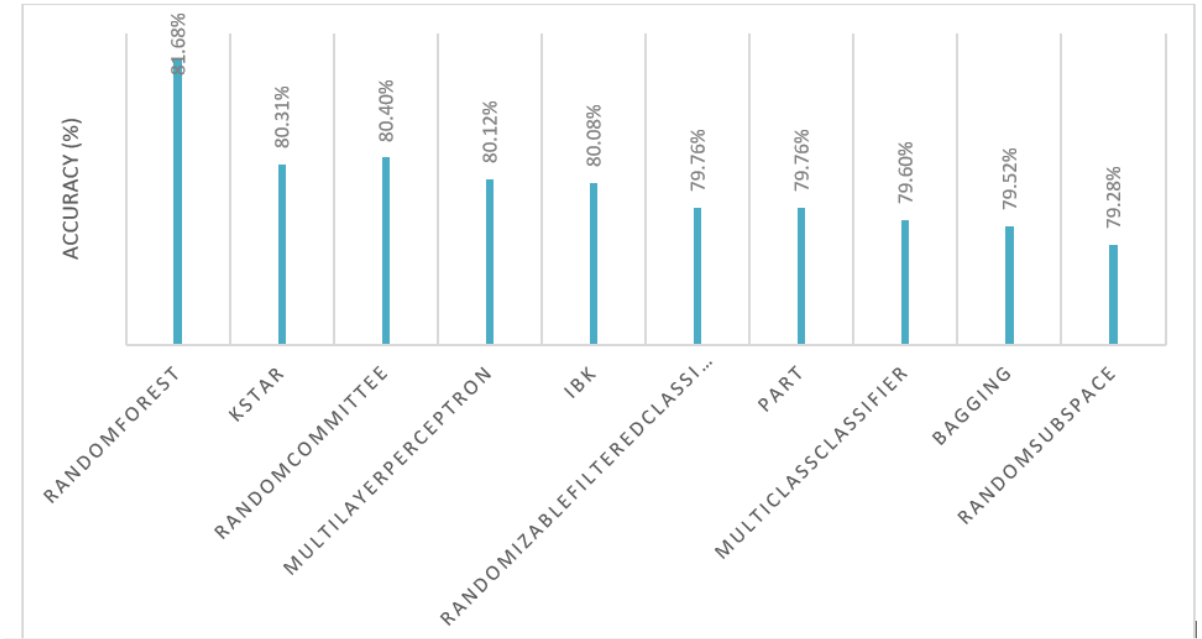


Fig. 5.2.2: Comparison of the other classifiers

CHAPTER 6

CONCLUSION AND FUTURE WORK

6.1 Conclusion

As far as we know, automatic cauliflower disease recognition is first initiative in Bangladesh. Recognize the diseases of cauliflower is the main purpose of our proposed system. Mainly our system focus on significant four disease of cauliflower which is bacterial soft rot, black rot, downy mildew, buttoning. After capturing the images our methodology begins. Captured images are preprocessed by resizing images with fixed size, contrast enhancement and conversion of RGB color space .For image segmentation, we used k means clustering and Euclidean distance technique in our proposed system. After segmentation, features are extracted using two feature set of GLCM consists of 10 features in total. Extracted features are used for classify data. The average accuracy for disease recognition of cauliflower disease is obtained for our proposed system is 81.68%.

6.2 Future plan

Our future plan is to develop a mobile application where farmers will take picture of the disease part of the plant and send it to our expert system and our system will analyze the image and a feedback will send to the farmer. Thus farmers will get benefit through our system. Through our system, accuracy is 81.68% but this result has no satisfied as our expectation. Our future goal is to acquire more accuracy using different various model with large number of data set. Besides cauliflower, a lot of other disease of plant is yet to work. So in near future we will work diseases of other plants.

REFERENCES

- [1] Bhagyashri S. Ghyar and Gajanan K. Birajdar, "Computer Vision Based Approach to Detect Rice Leaf Diseases using Texture and Color Descriptors". Proceedings of the International Conference on Inventive Computing and Informatics (ICICI 2017) IEEE Xplore Compliant - Part Number: CFP17L34-ART, ISBN: 978-1-5386-4031-9
- [2] Md. Selim Hossain, Rokeya Mumtahana Mou, Mohammed Mahedi Hasan, Sajib Chakraborty, M. Abdur Razzak, "Recognition and Detection of Tea Leaf's Diseases Using Support Vector Machine". 2018 IEEE 14th International Colloquium on Signal Processing & its Applications (CSPA 2018), 9 -10 March 2018, Penang, Malaysia
- [3] Tanya Garg ,Surinder Singh Khurana, "Comparison of Classification Techniques for Intrusion Detection Dataset Using WEKA". IEEE International Conference on Recent Advances and Innovations in Engineering (ICRAIE-2014), May 09-11, 2014, Jaipur, India
- [4] Cauliflower Cultivation Method. Available online: <https://www.ajkerkrishi.com/cauliflower-cultivation-method/> [Last accessed on September 23, 2020]
- [5] Bangladesh growing economy through advances in agriculture. Available online: <https://www.worldbank.org/bangladesh-growing-economy-through-advances-in-agriculture/> [Last accessed on September 23, 2020]
- [6] Md. Robel Mia, Md. Jueal Mia, Anup Majumder, Soummo Supriya and Md. Tarek Habib, "Computer Vision Based Local Fruit Recognition". International Journal of Engineering and Advanced Technology (IJEAT) ISSN: 2249 – 8958, Volume-9 Issue-1, October 2019
- [7] The Productivity Survey of Cauliflower Crop. Available online: <http://203.112.218.65:8008/cauliflower/> [Last accessed on September 15,2020]
- [8] Md Tarek Habib, Md Jueal Mia, Mohammad Shorif Uddin, and Farruk Ahmed. "An in-depth exploration of automated jackfruit disease recognition". Journal of King Saud University-Computer and Information Sciences (2020).

- [9] Sachin D. Khirade and A. B. Patil, "Plant disease detection using image processing." In 2015 International conference on computing communication control and automation, pp. 768-771. IEEE, 2015.
- [10] Md. Tarek Habib, Anup Majumder, A. Z. M. Jakaria, Morium Akter, Mohammad Shorif Uddin, and Farruk Ahmed. "Machine vision based papaya disease recognition." Journal of King Saud University-Computer and Information Sciences 32, no. 3 (2020): 300-309.
- [11] Aakanksha Rastogi, Ritika Arora, and Shanu Sharma. "Leaf disease detection and grading using computer vision technology & fuzzy logic." In 2015 2nd international conference on signal processing and integrated networks (SPIN), pp. 500-505. IEEE, 2015.
- [12] Farhana Tazmim Pinki, Nipa Khatun, and SM Mohidul Islam. "Content based paddy leaf disease recognition and remedy prediction using support vector machine." In 2017 20th International Conference of Computer and Information Technology (ICCIT), pp. 1-5. IEEE, 2017.

PLAGIARISM REPORT

Cauliflower Disease Recognition

ORIGINALITY REPORT

18%	13%	8%	12%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	Submitted to Daffodil International University Student Paper	5%
2	www.ijeat.org Internet Source	2%
3	Choudhury, Sumouli, and Anirban Bhowal. "Comparative analysis of machine learning algorithms along with classifiers for network intrusion detection", 2015 International Conference on Smart Technologies and Management for Computing Communication Controls Energy and Materials (ICSTM), 2015. Publication	1%
4	Submitted to Yeshwant Rao Chavan College of Engineering Student Paper	1%
5	Submitted to University of Moratuwa Student Paper	1%
6	Submitted to whu Student Paper	1%