

**COMPUTER VISION APPROACH FOR CUCUMBER DISEASE
RECOGNITION**

BY

Salowa Binte Sohel
ID: 162-15-7820

Joyanta Basak
ID: 162-15-7740

AND

Md. Abu Ishak Mahy
ID: 162-15-7893

This Report is presented in Part to Meet the Requirements for the
Bachelor Degree 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
2020

APPROVAL

This Project titled “**Computer Vision Approach for Cucumber Disease Recognition**”, submitted by Salowa Binte Sohel, ID: 162-15-7820, Joyanta Basak, ID: 162-15-7740 and Md. Abu Ishak Mahy, ID: 162-15-7893 to 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

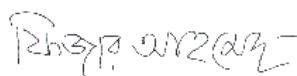
Chairman

Professor and Head

Department of Computer Science and Engineering

Faculty of Science & Information Technology

Daffodil International University



Dr. Fizar Ahmed

Internal Examiner

Assistant Professor

Department of Computer Science and Engineering

Faculty of Science & Information Technology

Daffodil International University



Abdus Sattar

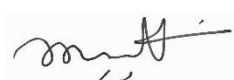
Internal Examiner

Assistant Professor

Department of Computer Science and Engineering

Faculty of Science & Information Technology

Daffodil International University



Dr. Mohammad Shorif Uddin

External Examiner

Professor

Department of Computer Science and Engineering

Jahangirnagar University

DECLARATION

We are hereby announcing that this project has been supervised by **Md. Jueal Mia, Senior Lecturer, Department of CSE** Daffodil International University. We are also hereby announcing that no any part of this project has been submitted anywhere.

Supervised by:



9.10.20

Md. Jueal Mia

Senior Lecturer

Department of CSE

Daffodil International University

Co-Supervised by:



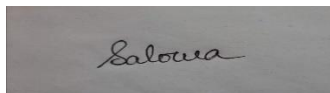
Shah Md. Tanvir Siddiquee

Assistant Professor

Department of CSE

Daffodil International University

Submitted by:

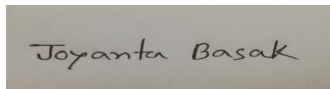


Name: **Salowa Binte Sohel**

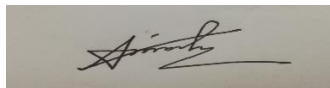
Id: 162-15-7820

Department of CSE

Daffodil International University



Name: **Joyanta Basak**
Id: 162-15-7740
Department of CSE
Daffodil International University



Name: **Md. Abu Ishak Mahy**
Id: 162-15-7893
Department of CSE
Daffodil International University

ACKNOWLEDGEMENT

At first, we evolve our gratefulness and gratitude to all-powerful Allah for His blessing makes us feasible to complete the final year project effectively.

We are obliged and grateful admit of our indebtedness to **Md. Jueal Mia, Senior Lecturer**, Department of CSE, Daffodil International University, Dhaka. Deep consciousness & sharp sensibility of our supervisor in the field of Machine Learning bring to pass this project. His infinite charity, studious conduct, inspiration, diligent supervision, formative detraction, worthy consultation at all state have formed it feasible to finish this project.

We express our gratefulness to Md Jueal Mia, Shah Md. Tanvir Siddiquee, and Dr. Syed Akhter Hossain, Head, Department of CSE and other faculty members and the staffs of CSE department of Daffodil International University.

We also thankful to our friends in Daffodil International University, who accepted in this consult while ending the course.

Finally, we must remember our parents to whom department needs no mention. Without their support and patients, we could not do anything.

ABSTRACT

Agriculture is the greatest labor agency in Bangladesh. As our country is an agricultural country; Most of the people of Bangladesh depends on agricultural products for their livelihood. Lack of opportunities and facilities, natural disasters, diseases of crops are main obstacles in growth of agricultural sector in our country. Diseases in plants are a significant yield and quality constraint for growers of broad acre crops. As important agricultural crops are under threat due to various plant diseases and pests and the quality of the crop can be affected by various diseases. So, it is very necessary to detect the disease at the proper period to the cultivator. When crop disease could be easily track out, then it can be effective for monitoring and restrain disease for agriculture and food safety. We have worked with cucumber disease detection in our project because cucumber is much needed grain in our country. If the cucumber is affected, it will cause huge damage to the economy of the country. Since cucumber has many qualities, if it is transited, it will have a huge detrimental effect on nutrition. A system called computer vision and machine learning is used to detect crop disease promptly and accurately which we used to detect cucumber diseases. Diagnosis of plant disease through this technology is very profitable and easy. This is because it reduces huge workload of crop monitoring and it can detect the symptoms of all diseases at very early stage. This system includes pattern recognition and creating database and classification, approximation, optimization, and data clustering. We are going to implement this goal by thinking of the people and helping the farmers to cultivate properly.

TABLE OF CONTENTS

CONTENTS	PAGE
Title	I
Approval	II
Board of Examiners	II-III
Declaration	IV-V
Acknowledgement	VI
Abstract	VII
Table of Contents	VIII-X
List of Figures	XI
List of Tables	XII

CHAPTER

Chapter 1: INTRODUCTION 1-5

1.1 Introduction	1-2
1.2 Motivation	3
1.3 Objectives	3
1.4 Research Question	4
1.5 Expected Output	4
1.6 Report Layout	5

Chapter 2: BACKGROUND 6-9

2.1 Introduction	6
2.2 Related Works	7-8
2.3 Scope of the Problem	9

2.4 Challenges	9
Chapter 3: RESEARCH METHODOLOGY	10-22
3.1 Introduction	10
3.2 Research Topic and Equipment	10
3.3 Diseases Description	11-12
3.4 Work Flow	13
3.5 Collecting Images	14
3.6 Image Preprocessing	14
3.7 Data Training	14
3.8 Clustering	15
3.9 Extraction of Features	16
3.10 Features	17-20
3.11 Classification	21-22
Chapter 4: OUTPUT VALUES AND RESULT	23-26
4.1 Confusion Matrix values	23-24
4.2 Project Overview and Result	25-26

Chapter 5: CONCLUSION AND FUTURE WORK	27-28
5.1 Summary and discussion of the study	27
5.2 Conclusion	27
5.3 Solicitation	28
5.4 Future work and development	28
REFERENCES	29
PLAGIARISM REPORT	30-33

LIST OF FIGURES

FIGURES	PAGE NO
Fig 3.3.1: Cucumber crop disease (a) Mosaic virus (b) Belly rot (c) Scab Disease (d) Pythium fruit rot (Cottony leak) (e) Fresh cucumber	12
Fig 3.4.1: The architecture of Computer vision based system for Cucumber Disease Recognition	13
Fig 3.4.2: Approached Methodology for Cucumber Disease recognition system	13
Fig 3.8.1: Changes of the image in stepwise (a) Names (b) Captured Image (c) Resized image (d) Contrast-enhanced image (e) Segmented Image	15
Fig 3.9.1: Features extracted values of some images (a) Name, (b) Captured Image, (c) Extracted feature values [$x = (\mu, \sigma, S, \sigma^2, \kappa, \gamma, C, \rho, E, H)$]	16
Fig 4.2.1: Experimental Outcome (a) Name, (b) Image, (c) Cluster, (d) Extracted Feature vector, (e) Recognized Diseases	26

LIST OF TABLES

TABLES	PAGE NO
Table 4.1.1: Confusion Matrix for RandomCommittee Classifier	23
Table 4.1.2: Class wise Confusion Matrix values of RandomCommittee	24
Table 4.1.3: Average values (accuracy, specificity, sensitivity, Precision, FNR, FPR) for every Classifier	24

CHAPTER 1

Introduction

1.1 Introduction

Everyone directly or indirectly depends on agriculture. According to FAO (Food & Agriculture Organization) about 60% people around the world depends on agricultural sector for their existence. In Bangladesh economy, agriculture is the most significant department. [1] In Bangladesh economy, agriculture is the most significant department. This department provides employment more than 63% population of this country. Agriculture contributes 19.6% to the nationwide GDP. [2] Cucumber, Cucumis sativus is also a part of this department. It is herbaceous warm season crop in the family Cucurbitaceae. It is widely cultivated in our country. In this country about Twenty-two thousand two hundred and eighty-six acres of land is used for cucumber farming. Every year farmers produce approximately 54854 metric tons cucumbers all over the country. [2] This cucumber is a great source of vitamin B and good for skin and hair. It also helps to fight and prevent cancer. Cucumber also plays a significant role in weight loss which reduces cholesterol and blood pressure. But disease infected cucumber causes harmful effects. Cucumber disease have features of severe disease casualty, numerous and fast transit. Agricultural productivity is fully dominated by Bangladeshi economy. Hence, disease detection of plants can play a significant role in the sector of agriculture. This is a special and favorable procedure to spy on an herbaceous disease in the early stages and as an innate disease detection strategy. According to our project we can find if there are any mentioned diseases on cucumber alongside, we can also find a fresh cucumber. [3] Physically approach is slow and difficult for identification of these kind of diseases. But automated system method is quite easy and accurate. So, we work with automated system to help the farmers by identifying the diseases before their cultivation and it also saves money.

There has been some related works on disease detection. Pawar proposed a system for the detection of two cucumber diseases using Artificial Neural Network. In this system, the input will be to capture the image of the infected crop and trained by neural network. Then find out the features that would be appropriate to diagnosis crop diseases. [4] Another approach is Machine Vision Based Papaya Disease Recognition by Habib, they proposed an agro-medical system which can recognize diseases from the image of papaya. This proposed system is almost in real time system. The disease classification has been sourced with support vector machine. Their project attained 90.15% accuracy. [5] Computer Vision Based Local Fruit Recognition by Mia, which is an expert system based on computer vision that is working on captured pictures and recognize what fruits this is. The accuracy of their system is 94.61% which is quite well. [6] The other approach was cucumber disease recognition which is based on (SVM) Support Vector Machine and image processing system by Youwen. The system of cucumber disease detection based on color feature is more accurate and faster. [7]

In our research paper, the computer vision system is raised. It works on capturing images through mobile device and can be able to identify which disease the cucumber is affected with. We worked on four types of cucumber diseases as well as fresh cucumber, those diseases - Belly rot, Pythium fruit rot, Scab cucumber, Mosaic virus. We collected images then trained them and then extracted the features from images by clustering. We have extracted 337 data features through clustering.

From, the previous research papers, it can be concluded that our project will be very up to date. And, it has an excellent accuracy level which is the most needed in case of making a successful research paper.

The extent of this research paper is embodied as follows: In chapter one, introduction, motivation, objectives, research question, expected output, report layout have been discussed. The second chapter is organized with introduction, some previous related works, scope of the problems and some challenges. In the third chapter, research methodology is described. Output value and outcome is discussed in the fourth chapter. Summary, conclusion, future work is described in the fifth chapter.

1.2 Motivation

Our economy highly depends on agricultural productivity. Diseases of plants are very natural. If proper care is not taken, it will cause serious effects on the plants. So, identification and disease detection of plants can play a significant role in agriculture field. Hence, classification of plant disease is necessary in agricultural producing, increasing market value and meeting quality standards. If the identification and sorting is done through physical techniques, the process will be slow. The labor's classify based on size, color etc. If these methods are recorded into artificial automated system by using appropriate programming language then it will be error free and faster. Plant disease detection with machine-learning approach is speedy and accuracy. And the other approach is image processing techniques. This method will prove useful technique for farmers and conscious them at the proper time before contagious the disease.

1.3 Objectives

Diseases in cucumber cost a farmer a great loss and also effects in economy of the country. We decided to work on this project with a main objective to reduce the losses for both Farmer and country. Our objectives for the project are given below:

- ❖ Identifying the disease using computer vision.
- ❖ It also identifies the insects and pests responsible for epidemic.
- ❖ It completes each of the process sequentially achieving each of the output.
- ❖ Reduce frail rate.
- ❖ Save time and provide efficient service.

1.4 Research Questions

- How can we recognize whether it is fresh cucumber or not?
- How can we recognize which disease the cucumber is affected by?
- How to select appropriate cluster of images?
- What kind of training data should be used for better disease detection?
- How efficient is the WEKA and MATLAB for training and predicting?
- How much of training data is needed for this research?

1.5 Expected Outcome

One can easily identify the disease of cucumber as well as fresh cucumber. So farmers will be able to take the right action within the appropriate time. Shopkeepers will be able to choose the right one and this will save them from buying rotten grain. They will be able to check the percentage of fresh cucumber in their delivery product and in affected cucumber, they can find out what diseases it has. Farmers can take grain sample from their land to see which diseases have spread in their land. Having an accurate idea of all of the diseases of cucumber, it will make easier for them to produce more grain. As such they can give medicine to their land. The agriculture officer can calculate what kind of cucumber disease has spread in any of the projects under his control.

1.6 Report Layout

We have decorated this research paper with five chapters. Each chapter has been nicely decorated and very nicely described as well as explained appropriately and well merged. The surface of the body for the first chapter of this report contains the information of motivation of the study, objectives and to illustrate the expected outcomes. In the second chapter. We explained about the study background of our research paper and described in several related works. We also tried to explain the challenges we faced in solving the previous problems and also give a research summery in this second chapter. In chapter three, we have discussed the details of our work. Here we have explained about the collection of images, preprocessing, work flow diagram, methodology diagram data training, extraction of features, data clustering, statistical process of the study, classification with confusion matrix. For chapter four, we have explained outcome value of the project with average values and the result. In chapter five, we have focused on the summarized of our research and marked as conclusion. The results of the research involved for future studies and the use of strategies for further use and also some drawbacks are described with accurate and precise explanations.

CHAPTER 2

Background

2.1 Introduction

Cucumber disease is the normal part of nature. Ours is an agricultural country, where most of the population fully depends on agriculture sector. So, agriculture has been playing an important conduct in overall economic development of Bangladesh. Agricultural sector provides large amount of employment to around 47% of people in our country. Cucumber is a seasonal crop which is one of the main sources of the economy of Bangladesh and this crop has various types of quality. Cucumber disease is a normal part of nature. But it is a result of hunger and starvation. Also, it is a great issue for our economy sector. The identification of cucumber disease with the help of computer vision can solve many problems. So, we think this project will play a significant induction in our agricultural sector broadside on it saves time in this sector.

2.2 Related Works

A system is proposed for the identification of two cucumber diseases using Artificial Neural Network by Pawar et al. [4] Those diseases are: Downy mildew and Powdery mildew. They will give an input to this system and this input will be to capture the image of the infected crop. Those images need to be trained through neural network. Then find out the features that would be appropriate to diagnosis crop diseases. Image features usually include color, size and texture in image processing method. They are used ANN which will give symptoms of disease for Diagnosis and Treatment of cucumber.

Another approach is Machine Vision Based Papaya Disease Recognition by Habib et al. [5] Where they proposed an agro-medical system which can recognize diseases from the image of papaya. They set a feature for diseases recognition. To substance the characteristics of images, here Image processing method is used. They used support vector machine (SVM) for classification of diseases. This proposed system is almost in real time system. The disease classification has been sourced with support vector machine. Their project attained 90.15% accuracy, which is excellent as well as pledge. Where there remains a dynamic future conduct with huge data set of pictures with wide range of this papaya disease.

An approach is Computer Vision Based Local Fruit Recognition by Mia et al. [6] In this project, an expert system is approached, which is based on computer vision that is working on captured pictures and recognize what fruits this is. In the midst of endemic fruits, they choose six fruits called amla, sugar-apple, bilombo, elephant apple, orboroi and sapota for experiment. They perform a preprocessing method on those images and they use segmentation for extracting the features of images. After that, support vector machines (SVMs) used for classification. An expert system is called machine vision for recognition of fruit. They used two feature set and that consists ten features. To extract the features Image processing system is used. Their system accuracy level is 94.61% which is good enough.

Another is An In-Depth Exploration of Automated Jackfruit Disease Recognition abstract by Mia et al. [8] In this project, they used an agro-medical expert system, which results with digital images that recognizes diseases. At first, they select a feature. A segmentation is k-means clustering is used to detect disease affected area of the images of jackfruit and extracting nine classifiers separately in the features and Image processing technique is used to do this. For classification of the diseases, the classifiers of the index are seven eminent metrics. Random forest is got of all other classifiers which attains 90% accuracy. They attempt to detect the jackfruit disease with strong features in to classify jackfruit diseases. In order to simplify future research, an in-depth of classification of jackfruit disease recognition is used. Their approached project is very competitive.

The other approach was cucumber disease recognition which is based on (SVM) Support Vector Machine and image processing system by Tian et al. [7] At first, they used vector median filter to remove cucumber disease leaf color images. Then the pattern of statistics and the morphology of mathematics were introduced in the images of segment. Finally, cucumber disease spot on leaves and the color properties were extracted. The system of detection cucumber disease is based on color feature is more accurate and faster.

Another related work is a local fish recognition. It is a Machine vision based project by Sharmin et al. [9] Then they converted images in gray-scale and they formed gray-scale histogram. Then the image segmentation took place where they used histogram-based method. Then after calculation SVM gives them the accuracy of 94.2%.

Another one is Registration Status Prediction using Machine Learning for students of private university Mia et al. [10] They have done this for saving private university from direct in marketing strategy, terms of budget and sustainability. They applied total seven classifiers. They have used a data set of a thousand private university students in Bangladesh. They have got a top accuracy of 85.76%.

2.3 Scope of the Problem

We have already seen that some current application already works with this topic which method is expensive and time consuming. Further, in some developing countries, farmers may need to go long distances to contact experts. This approach is speedy and accuracy. So, this will prove useful for farmers and conscious them at the proper time as it detects symptoms of diseases at very early stage and this reduces the large workload of monitoring system of large farms. Accurate identification of crop diseases keeps significant role in technical sector besides economy sector in agriculture.

2.4 Challenges

We face many problems when we are going to collect data from a different area. It is a seasonal crop and that is why cucumber is not found when and then. This project deals with hundreds of images of cucumber. Those images require pre-processing before training. So computational system has to be very strong to require in this research. Our computer was not so enough strong to handle it. And, that was quite a big problem for us. We were completely new in machine learning and so it was the biggest challenge for us. We did not have any knowledge about it. Basically, we focused on studying more than implementing. That is why it proved to be extremely difficult.

CHAPTER 3

Research Methodology

3.1 Introduction

The main intention of our project is to diagnosis the cucumber disease very fast and accurate. Machine learning has been required for its successful implementation. First of all, we have collected raw data and hundreds of images with the help of field and website and trained them by machine learning using MATLAB.

3.2 Research Topic and Equipment

Through our project we have created a model that can detect and identify four types of cucumber disease and also fresh cucumber and extract features from the pictures through clustering. To implement this project, some steps to be taken such as image collection, preprocessing, clustering data, extraction of features which is in the data train part and then the other part is data mining, classification, confusion matrix and diagnosis.

Equipment

- Machine Learning
- Matlab
- Microsoft Excel
- Weka

3.3 Diseases Description

(a) Mosaic virus:

Due to the Cucumber mosaic virus (CMV), cucumber mosaic is happened. CMV is the most destructive viruses which is widely contagious in temperate as well as tropical areas. Cucumber can be transited with CMV at any term, and in infected cucumber, the virus expands systemically. Symptoms come into sight 7 to 14 days after taint at the temperatures of 79 to 89 °F. Light- to dark-green spots appear first. Infected cucumbers are smaller and formless, shows mosaic patterns on aril with light as well as dark green color. [11]

(b) Belly rot:

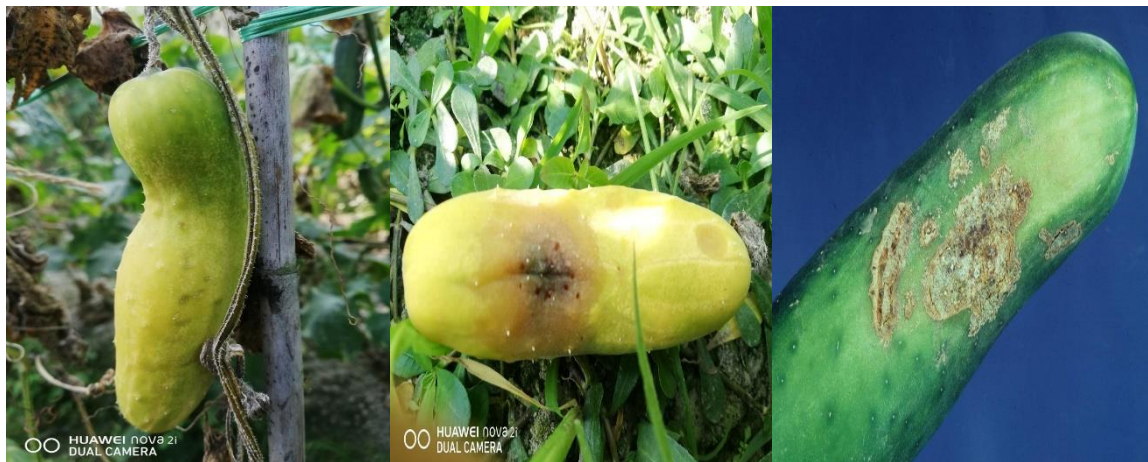
Due to the soil borne fungus *Rhizoctonia solani* Belly rot is produced. The disease is deadly to cucumbers, the germs can attack all the cucurbits. *R. solani* is a very common soil inhabitant which invades cucumbers in contact with soil. This disease can ensue over (46 to 95°F) temperatures, but greater damaging portion is near 80°F. For this disease uplift, moisture is less significant than temperature. But intensity of dampness and exceeding moisture is avail for this infection. Belly rot symptoms ensue on underside as well as blossom end which produces water-drabble, tan to brown bruise. Bruise become submerged when the disease grows, made irregular shape, gradually dried. Infected cucumbers are rigid and a tender rot ever occurs. It reduces the virtue and saleable output of cucumbers. [12]

(c) Scab:

Cladosporium cucumerinum fungus is reason for scab disease of cucumber. Cucumber spots maybe white to gray and may have a yellowish tinge. As the age of infection, leaf spot falls out the center. The Scab travels as a seed transmitted over infected seed or moist air motion. Disease spreads in cold (around 70° F), wet climate. [13]

(d) **Pythium fruit rot:**

Pythium fruit rot is result from a few fungal organisms. Pythium spp. are found in most, if not all soils, of the world. When cucumber contacts with wet ground, those infection occurs. The Pythium fruit rot pathogens are rapidly propagated among fields in watering. The symptoms of the disease begin as tawny, water-soaked blighting that speedily become massive, watery and limp and putrefied. This is usually starts infected by contacting the soil. Cucumbers are evident on a brown to dark green. This Pythium fruit rot is most cute in ailing-drained fields since wet climate. This disease can pay the fruit as unsaleable. [14]



(a)

(b)

(c)



(d)

Fig 3.3.1: Cucumber crop disease (a) Mosaic virus (b) Belly rot (c) Scab Disease (d) Pythium fruit rot (Cottony leak)

3.4 Work Flow

System Model:

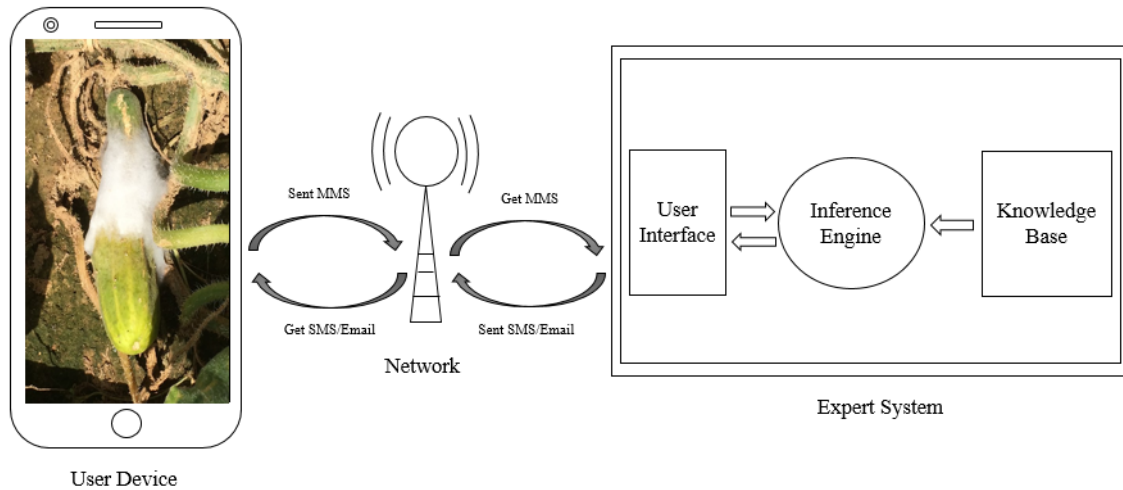


Fig 3.4.1: The architecture of Computer vision based system for Cucumber Disease Recognition

Methodology Diagram:

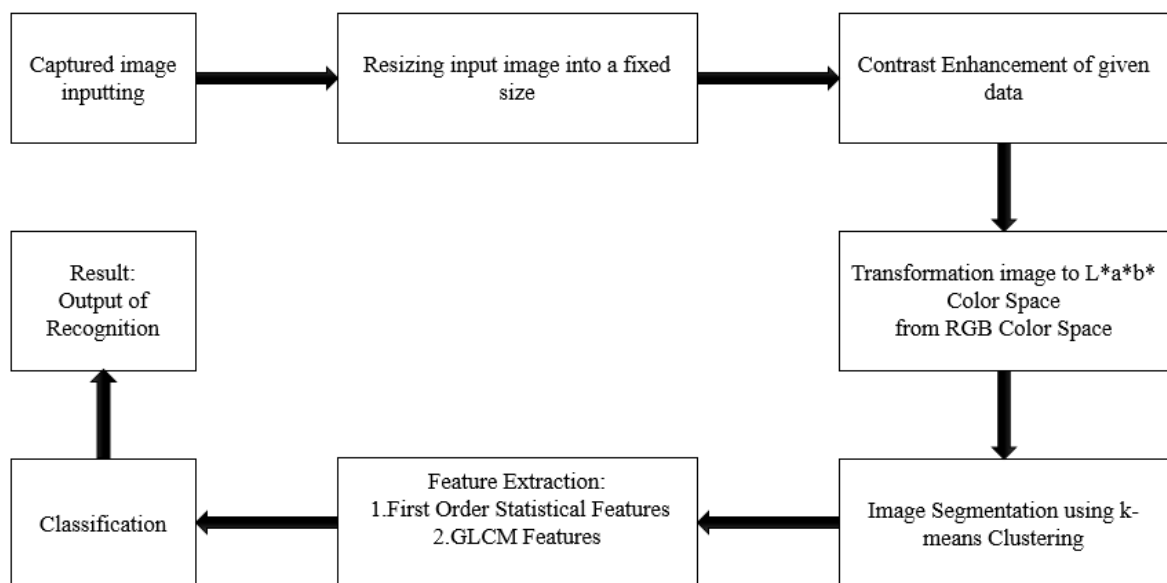


Fig 3.4.2: Approached Methodology for Cucumber Disease recognition system

3.5 Collecting Images

The first step of this project is image collection. We have gone to different fields and collected infected cucumber images during the cucumber production time. We captured these images at different angles using our cell phone. As it is a seasonal crop so we have need to download several pictures from Internet. Then we use machine learning to train these collected images. We work with four diseases of cucumber and fresh cucumber pictures data.

3.6 Image Preprocessing and Resizing

We were able to take pictures of diseased cucumber very well and pictures were enough sharp and looked very clear but that was the end time of cucumber season. So we couldn't find necessary samples for picturing. So we downloaded some picture of cucumber form Internet. Then we organize all images as class. Then we resize the images into image size (350 X 300) pixels. We have done Contrast enhancement and color intensity of those images. Then we have used k-mean clustering for data training.

3.7 Data training

In this stage, pictures have to be organized separately due to different diseases. We need to make it clear to our program that if there is such a characteristic it will be called by this name. For example, if there are yellow spots, ring spots on the cucumber, then the program would identify it as a mosaic disease. If there is tan to brown color lesions and water soaked produced then it is called belly rot. If the infected leaves looked green or gray and oozing or sunken spots are shown then it is identified as scab disease. And pythium rot can be identified by its brownish color, water soaked and soft symptoms. The full process of this step is called train data. We use Matlab to do this step.

3.8 Clustering

To extract features of cucumber diseases, we first need to cluster the images. In this case, the images are in the form of three cluster. Then we select the first best and the other best cluster from the three cluster of images. And through this cluster, there are thirteen features of cucumber diseases came out. We have extracted 337 data features through clustering which has been collecting in Excel. The process of clustering picture is shown in Fig.2.

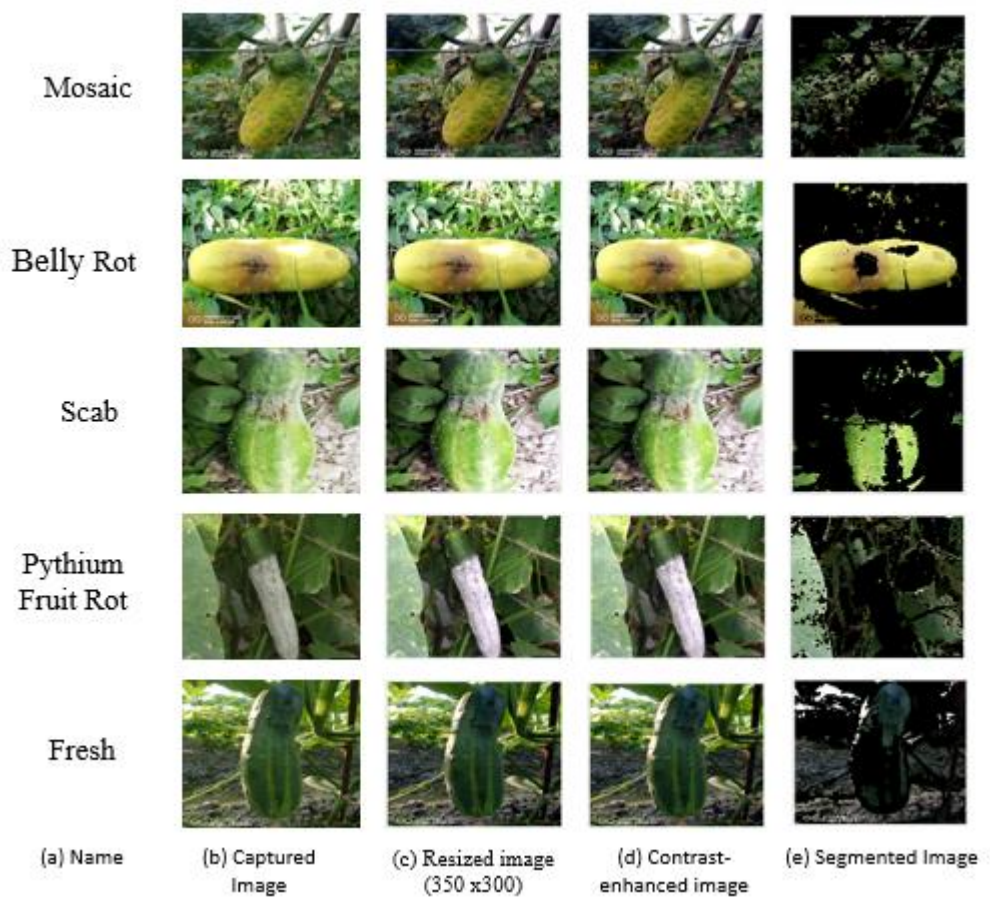


Fig 3.8.1: Changes of the image in stepwise (a) Names (b) Captured Image (c) Resized image (350 x 300) (d) Contrast-enhanced image (e) Segmented Image

3.9 Extraction of Features

The purpose of these step is to find out and extract the features and this is used to detect the crop disease. Through image processing where color, shapes and textures are included.

Mosaic		0.62, 0.76, 0.44, 0.91, 23.06, 40.31, 3.19, 1575.50, 4.93, 1.70
Belly Rot		0.48, 0.95, 0.59, 0.97, 37.05, 73.26, 2.45, 3107.64, 4.20, 1.68
Scab		2.09, 0.79, 0.31, 0.88, 52.20, 74.67, 4.16, 5105.54, 2.81, 1.13
Pythium Fruit Rot		0.93, 0.85, 0.28, 0.89, 41.80, 60.02, 4.41, 2076.95, 3.66, 1.36
Fresh		1.13, 0.66, 0.38, 0.89, 24.97, 44.35, 3.65, 1833.04, 8.69, 2.30
(a) Name	(b) Captured image	(c) Extracted feature values [$x = (\mu, \sigma, S, \sigma^2, \kappa, \gamma, C, \rho, E, H)$]

Fig 3.9.1: Features extracted values of some images (a) Name, (b) Captured Image, (c) Extracted feature values [$x = (\mu, \sigma, S, \sigma^2, \kappa, \gamma, C, \rho, E, H)$]

3.10 Features

First Order Statistical Features

In our project we select the first feature as a 1st order statistical process to detect diseases or freshness in cucumber. It predicts features separate pixel values by waving the interaction between picture pixels. The main characteristics of 1st order statistical process are Mean, Standard Deviation, Kurtosis, Skewness and Variance. Their equations are given below: [8]

a) Mean

Mean is the average equilibrium of redundancy of the image or texture. Mean as defined in Equation (1) as-

Mean:

$$\mu = \frac{1}{n} \sum_{i=1}^n x_i \quad (1)$$

b) Standard Deviation

This statistic is used to determine variability. It shows how much variation or scattering exists from the average. It indicates a low deviation that points to a high standard deviation tends to close to the mean and those data is scattered over the points. Standard deviation is defined as in Equation (2) as-

Standard Deviation:

$$\sigma = \sqrt{\frac{1}{n} \sum_{i=1}^n (x_i - \mu)^2} \quad (2)$$

c) Kurtosis

Kurtosis is the rule of “tailed ness” of the chance of ordering random value from a real value. Kurtosis is defined as in Equation (3) as-

Kurtosis:

$$K = E \left[\left(\frac{x - \mu}{\sigma} \right)^4 \right] \quad (3)$$

d) Skewness

Skewness is used to measure a dataset's symmetry. Here, normal distribution is 0, and symmetric which have skewness is zero. Skewness is defined as in Equation (4) as-

Skewness:

$$\gamma = E \left[\left(\frac{x-\mu}{\sigma} \right)^3 \right] \quad (4)$$

e) Variance

In statistical process Variance measures how far the set values are spread out And their average values. Variance is defined as in Equation (5) as-

Variance:

$$\sigma^2 = \frac{\sum_{i=1}^N (GS_i - \mu)^2}{N} \quad (5)$$

GLCM Features

The GLCM is used for extracting second order texture data from given images. It is a well set device. It is a kind of matrix in which the number of columns and rows of the matrix are equal to the pixel values of given images of the texture. This matrix narrates frequency of relationship between two gray levels within the area of supervision. [6] In GLCM, there are fourteen characteristics among them the most effective characteristics are Contrast, Correlation, Energy, Homogeneity and Entropy. Their equations are given below:

a) Contrast

Contrast measures the intensity of a pixel of the image. It is fixed with the difference in the color and luster of an object. Contrast is defined as in Equation (6) as-

Contrast:

$$C = \sum_{i=0}^{G-1} \sum_{j=0}^{G-1} (i - j)^2 [p(i, j)] \quad (6)$$

b) Correlation

It is the method of measuring the gray levels linear dependency from its neighboring pixels. Correlation is defined as in Equation (7) as-

Correlation:

$$\rho = \sum_{i=0}^{G-1} \sum_{j=0}^{G-1} \frac{ijP(j, j) - \mu_{XY}}{\sigma_x \sigma_y} \quad (7)$$

c) Energy

It measures the limitation of repetitions of pixel pairs. It also measures the uniformness of an image. Hence, pixels are similar and energy value would be large. Energy is defined as in Equation (8) as-

Energy:

$$E = \sum_{i=0}^{G-1} \sum_{j=0}^{G-1} [P(i, j)]^2 \quad (8)$$

d) Homogeneity

IDM is homogeneity. IDM feature is obtained to measure the local homogeneity and measures the closeness distribution of GLCM elements to GLCM diagonal. It has a large range of values to determine the picture is textured or not textured. Homogeneity is defined as in Equation (9) as-

Homogeneity:

$$H = \sum_{i=0}^{G-1} \sum_{j=0}^{G-1} \frac{P(i,j)}{1+(i-j)^2} \quad (9)$$

e) Entropy

Entropy characterizes the texture of an input image. When the elements of the matrix are same entropy's value will be maximum. Entropy is defined as in Equation (10) as-

Entropy:

$$S = - \sum_{i=0}^{G-1} \sum_{j=0}^{G-1} P(i,j) \log[p(i,j)] \quad (10)$$

3.11 Classification

Classifiers and Confusion matrix

Confusion matrix is a method of calculating functionality for classification of Machine learning. It helps to know the functionality of the given classification model on a set of given test data. As, we have 5 type of data set for our project so we need to take dimension of (n*n) where n=5, which means we take 5 class problem. So, for a confusion matrix of 5-class problem summarizes the number true positives (TP), true negatives (TN), false positives (FP), and false negatives (FN). [6] Now these values can be calculated as shown in equation no (11-14).

$$TP = a_{ii} \quad (11)$$

$$FP_i = \sum_{j=1, j \neq i}^n a_{ji} \quad (12)$$

$$FN_i = \sum_{j=1, j \neq i}^n a_{ij} \quad (13)$$

$$TN_i = \sum_{j=1, j \neq i}^n \sum_{k=1, k \neq i}^n a_{jk} \quad (14)$$

After calculating these values, we need to calculate the accuracy, precision, specificity, sensitivity, FNR (false negative rate) and FPR (false positive rate) of the system by using those values of confusion matrix. We have been divided our data into two individual data sets. One of them is training data set and another one is testing data set. We have kept the highest number of data in our training data set. We can measure the performance of the system in phrase of these given metrics with the help of those test data set. [5] The rules of calculating the percentage values of accuracy, specificity, sensitivity, precision, FNR, FPR are given in equation no (15-20).

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \times 100 \% \quad (15)$$

$$Precision = \frac{TP}{TP+FP} \times 100 \% \quad (16)$$

$$Specificity = \frac{TN}{FP+TN} \times 100 \% \quad (17)$$

$$Sensitivity = \frac{TP}{TP+FN} \times 100 \% \quad (18)$$

$$FNR = \frac{FN}{FN+TP} \times 100 \% \quad (19)$$

$$FPR = \frac{FP}{FP+TN} \times 100 \% \quad (20)$$

We calculate the total performance of classifiers with the help of given equation of evaluation matrix are shown in equation no (15-20). We have taken eight classifiers and calculated the average values of those equations for our all 5-class problems.

CHAPTER 4

Output values and Result

4.1 Confusion Matrix values:

We have discussed about Confusion Matrix and its equations above. Now we have calculated all Confusion Matrix values (TP, FN, FP and TN) and equations (accuracy, specificity, sensitivity, precision, FNR, FPR) for every classifier. Here is an example for Confusion Matrix of a classifier given below:

Table 4.1.1: Confusion Matrix for RandomCommittee Classifier.

	Predicted					
		Mosaic	Belly Rot	Scab	Pythium Fruit Rot	Fresh
Actual	Mosaic	62	19	2	7	11
	Belly Rot	21	46	7	8	7
	Scab	12	10	18	0	4
	Pythium Fruit Rot	12	9	3	10	2
	Fresh	20	10	4	2	30

Table 4.1.2: Class wise Confusion Matrix values of RandomCommittee

Class	Accuracy	Precision	Specificity	Sensitivity	FNR	FPR
Mosaic	69.05%	48.82%	72.34%	61.39%	38.61%	27.66%
Belly Rot	72.92%	48.94%	80.57%	51.69%	48.32%	19.43%
Scab	87.50%	52.94%	94.52%	40.91%	59.09%	5.48%
Pythium Fruit Rot	87.20%	37.04%	94.33%	27.78%	72.22%	5.67%
Fresh	82.14%	55.56%	91.11%	45.46%	54.55%	8.89%

Now we have calculated the average values of every equations (accuracy, specificity, sensitivity, precision, FNR, FPR) for all those classifiers. The average values for every classifier given below:

Table 4.1.3: Average values (accuracy, specificity, sensitivity, precision, FNR, FPR) for every Classifier.

Classifier name	Accuracy	Precision	Specificity	Sensitivity	FNR	FPR
RandomCommittee	79.76%	48.66%	86.57%	45.44%	54.56%	13.43%
RandomForest	79.64%	51.09%	86.37%	45.12%	54.88%	13.63%
IBK	78.69%	43.66%	86.24%	43.58%	56.42%	13.76%
Kstar	78.45%	43.33%	86.12%	43.71%	56.29%	13.88%
RandomSubSpace	77.26%	46.92%	84.38%	34.66%	65.34%	15.62%
Bagging	76.91%	41.71%	84.55%	37.50%	62.49%	15.45%
RandomTree	76.91%	39.32%	85.07%	39.30%	60.70%	14.93%

4.2 Project Overview and Result

Recognition of cucumber diseases using computer vision system is shown in table 4.1.2. Firstly, we have grabbed 337 color pictures of four diseased cucumber which are captured with different view of points and size. These pictures have been sent to a proposed system. These datasets are made to train with machine learning, MATLAB 7.10.1 is used for machine learning toolbox. The size of the images has been changed into 350×300 pixels and completed with contrast enhancement. Then k-means clustering is used to extract all the features from these images. In this case, the images are in the form of three cluster. Then we select the first best and the other best cluster from the three cluster of images. And through this cluster, there are thirteen features of cucumber diseases came out. We have taken 337 data features through clustering. We extract the features to detect the diseases of cucumber.

In our project we extract the feature from two statistical feature process: 1st order statistical feature and 2nd order GLCM feature. The main characteristics of 1st order statistical features are Mean, Standard Deviation, Kurtosis, Skewness and Variance. The GLCM is used for extracting second order texture data from given images. In GLCM, there are fourteen characteristics among them the most effective characteristics are Contrast, Correlation, Energy, Homogeneity and Entropy. To calculate functionality for classification of Machine learning, confusion matrix is used.

Then we calculate the total performance of classifiers such as accuracy, precision, specificity, sensitivity, FNR (false negative rate) and FPR (false positive rate) of the system. Our experiment. Our experimental output is shown in Fig 4.2.1. finally we can say that, our expert system gives better outcome.

To achieve the desire result we need this Classifier values. After all calculation, we have got our desire result. We got 79.76% Accuracy which is good persentage. As we took just 337 data set we got this accuracy. We will increse our data set. Then our accuracy level will be increased. Then we have done some expreimental steps to find some output. The experimental output in given fig 4.2.1.

(a) Name	(b) Captured image	(c) Cluster	(d) Extracted feature vector	(e) Recognized Diseases
Mosaic			0.62, 0.76, 0.44, 0.91, 23.06, 40.31, 3.19, 1575.50, 4.93, 1.70	Mosaic
Belly Rot			0.48, 0.95, 0.59, 0.97, 37.05, 73.26, 2.45, 3107.64, 4.20, 1.68	Belly Rot
Belly Rot			4.09, 0.79, 0.31, 1.88, 56.29, 73.66, 4.76, 5110.55, 2.81, 1.77	Mosaic
Pythium Fruit Rot			0.93, 0.85, 0.28, 0.89, 41.80, 60.02, 4.41, 2076.95, 3.66, 1.36	Pythium Fruit Rot
Fresh			1.56, 1.62, 0.58, 0.83, 25.67, 42.35, 3.78, 1837.24, 8.22, 2.35	Pythium Fruit Rot

Fig 4.2.1: Experimental Outcome (a)Name, (b) Image, (c) Cluster, (d) Extracted Feature vector, (e) Recognized Diseases

CHAPTER 5

Conclusion and Future Work

5.1 Summary and discussion of the study

Computer vision is expanding exponentially and it is flourishing vastly. The progress of computer vision sector has made in recent years remarkable and also played a significant role. A positive result has been produced through our research although it is not that kind of unerring model but we are all hopeful that this model will be very useful in the future and will be helpful for everyone especially in economy sector as well as agriculture sector in Bangladesh, besides it will be a significant model which can be applied in real life. This model needs a lot of data, it should be required more data training to prepare this system appropriate and usable. A lot of development should be required in Neural Network architecture and also required feasible training which can be a suitable solution to make the system best and achieve the higher accuracy.

5.2 Conclusion

In our project, there are thirteen features and, image processing system should be applied to extract all the features in this regard. We have got our outcome based on these features and attain a desired output. But Our departure to complete our project was not so easy and simple as we have had to face many obstacles and even many challenges to come to this positive result. None of our group had any such knowledge because we were completely new to the computer vision and machine learning sector. So, we had to practice a lot with these two sectors. Another challenge was cucumber is a seasonal crop so it is not always available and not found at any time. Again, going to find out all the diseases of cucumber in Bangladesh and to collect all these pictures was quite difficult for us. We spend a lot of time completing this research paper. By analyzing the work of this project, we can say that our system is up to date and work well enough. The accuracy level of our project is 79.76 % and which is quite well.

5.3 Solicitation

We mentioned earlier that we are new to these sectors. Especially at the beginning of the project, we had lot a question awoke in our mind about the system, program and methodology and also confused how to write a research paper appropriately. For getting a positive result we have had to face many challenges and had to work hard and practice more and more. Needless to say, Md Jueal Mia sir, our supervisor has given a lot of positive support, references and proper guidelines that have helped us to complete our research paper. Our Co-supervisor Shah Md. Tanvir Siddiquee sir is also helpful and supportive.

5.4 Future work and development

The main objective of Computer Vision is to make machines know the world around them make them look and learn like humans. A picture contains a lot of information that helps to classify the object. The classification of cucumber diseases will make a huge contribution in science as well as the agricultural sector. We have seen development of computer vision nowadays as well as seen many problems too. And which cannot be solved yet to achieve higher accuracy for existing predictions. What we have achieved as the final outcome can be further improved through a lot of data collection and increasing data training. And it can be said that this cucumber disease detection has become a much more promising future with a huge data set. And in future we will be able to work with huge amount of data set about diseases of cucumber for performing the best experiment.

References

- [1] EXPO net, "Agriculture Remains Central to the World Economy. 60% of the Population Depends on Agriculture for Survival," *expo2015*, 2014.
- [2] Nations Eyclopedia, "nationsencyclopedia," 2020. [Online]. Available: <https://www.nationsencyclopedia.com/economies/Asia-and-the-Pacific/Bangladesh-AGRICULTURE.html>. [Accessed 6 August 2020].
- [3] Plant Village, "plantvillage," 2020. [Online]. Available: <https://plantvillage.psu.edu/topics/cucumber/infos>. [Accessed 6 August 2020].
- [4] V. T. a. P. P. Pooja Pawar, "Cucumber disease detection using artificial neural network," *IEEE*, p. 5, 2016.
- [5] A. M. A. Z. M. Z. Md. Tarek Habib, "Machine Vision Based Papaya Disease Recognition," *researchgate*, p. 10, June 2018.
- [6] M. R. M. M. T. H. A. M. a. S. S. Md. Jueal Mia, "Computer Vision Based Local Fruit Recognition," *International Journal of Engineering and Advanced Technology*, vol. 9, no. 1, p. 11, October 2019.
- [7] T. Y. L. T. N. Yan, "The Recognition of Cucumber Disease Based on Image Processing and Support Vector Machine," *researchgate*, vol. 2, no. 1, June 2008.
- [8] M. T. H. M. S. U. F. A. Md. Jueal Mia, "An in-depth exploration of automated jackfruit disease recognition," *Journal of King Saud University*, 5 May 2020.
- [9] M. T. H. a. m. Israt Sharmin, "Machine vision based local fish recognition," *researchgate*, 2019.
- [10] A. S. M. T. h. Md Juel Mia, "Registration Status Prediction of Students Using Machine Learning in the Context of Private University of Bangladesh," *researchgate*, 2019.
- [11] last updated 2020. [Online]. Available: <https://www.seminis-us.com/resources/agronomic-spotlights/mosaic-virus-diseases-of-cucumber>. [Accessed 12 August 2020].
- [12] Center for Invasive Species and Ecosystem Health at the University of Georgia., "Bugwood," 27 November 2012. [Online]. Available: [https://wiki.bugwood.org/Rhizoctonia_solani_\(belly_rot\)](https://wiki.bugwood.org/Rhizoctonia_solani_(belly_rot)). [Accessed 12 August 2020].
- [13] M. Grabowski, "extension," 2020. [Online]. Available: <https://extension.umn.edu/diseases/scab-cucurbits>. [Accessed 12 August 2020].
- [14] H. F. S. a. D. H. Gent, "bugwood," 04 April 2016. [Online]. Available: https://wiki.bugwood.org/HPIPM:Pythium_Fruit_Rot. [Accessed 16 August 2020].
- [15] D. V. T. P. P. P. Ms. Pooja Pawar, "Cucumber Disease Detection Using Artificial Neural Network," *scribd*, p. 5, 2020.

Plagiarism Report

Computer Vision Approach for Cucumber Disease Recognition

ORIGINALITY REPORT

15%

SIMILARITY INDEX

%

INTERNET SOURCES

%

PUBLICATIONS

%

STUDENT PAPERS

PRIMARY SOURCES

- | | | |
|---|---|----|
| 1 | "Computer Vision Based Local Fruit Recognition", International Journal of Engineering and Advanced Technology, 2019
Publication | 3% |
| 2 | Pooja Pawar, Varsha Turkar, Pravin Patil. "Cucumber disease detection using artificial neural network", 2016 International Conference on Inventive Computation Technologies (ICICT), 2016
Publication | 3% |
| 3 | Submitted to Daffodil International University
Student Paper | 2% |
| 4 | dspace.daffodilvarsity.edu.bd:8080
Internet Source | 2% |
| 5 | Md. Tarek Habib, Anup Majumder, A.Z.M. Jakaria, Morium Akter, Mohammad Shorif Uddin, Farruk Ahmed. "Machine vision based papaya disease recognition", Journal of King Saud University - Computer and Information Sciences, 2020 | 1% |

6	Md. Tarek Habib, Md. Jueal Mia, Mohammad Shorif Uddin, Farruk Ahmed. "An in-depth exploration of automated jackfruit disease recognition", Journal of King Saud University - Computer and Information Sciences, 2020 Publication	1%
7	"Registration Status Prediction of Students using Machine Learning in the Context of Private University of Bangladesh", International Journal of Innovative Technology and Exploring Engineering, 2019 Publication	<1%
8	bugwoodcloud.org Internet Source	<1%
9	Saeidi, Omid, Jamal Rostami, Mohammad Ataei, and Seyed Rahman Torabi. "Use of digital image processing techniques for evaluating wear of cemented carbide bits in rotary drilling", Automation in Construction, 2014. Publication	<1%
10	Submitted to Volunteer State Community College Student Paper	<1%
11	faculty.daffodilvarsity.edu.bd Internet Source	<1%

12	"Table of Contents - Volume 2", 2008 Congress on Image and Signal Processing, 05/2008 Publication	<1 %
13	Submitted to uu Student Paper	<1 %
14	Mahvish Jan, Hazik Ahmad. "Image Features Based Intelligent Apple Disease Prediction System", International Journal of Agricultural and Environmental Information Systems, 2020 Publication	<1 %
15	www.re3data.org Internet Source	<1 %
16	www.seminis-us.com Internet Source	<1 %
17	repository.up.ac.za Internet Source	<1 %
18	link.springer.com Internet Source	<1 %
19	repub.eur.nl Internet Source	<1 %
20	Saiqa Khan, Meera Narvekar. "Novel fusion of color balancing and superpixel based approach for detection of tomato plant diseases in natural complex environment", Journal of King Saud University - Computer and Information	<1 %

Sciences, 2020

Publication

-
- | | | |
|-----------|---|---------------|
| 21 | Israt Sharmin, Nuzhat Farzana Islam, Israt Jahan, Tasnem Ahmed Joye, Md. Riazur Rahman, Md. Tarek Habib. "Machine vision based local fish recognition", SN Applied Sciences, 2019
<small>Publication</small> | <1% |
| <hr/> | | |
| 22 | digitalcommons.mtu.edu
<small>Internet Source</small> | <1% |
| <hr/> | | |
| 23 | bch.cbd.int
<small>Internet Source</small> | <1% |
| <hr/> | | |
| 24 | oda.hio.no
<small>Internet Source</small> | <1% |
| <hr/> | | |
| 25 | Yanwei Xu, Menghui Chen, Tancheng Xie. "State Recognition of Egg Embryo in Vaccines Production Based on SVM", 2017 5th International Conference on Mechanical, Automotive and Materials Engineering (CMAME), 2017
<small>Publication</small> | <1% |
-

Exclude quotes Off

Exclude matches Off

Exclude bibliography Off