

**Country bean (desi shim) (Lablab purpureus L.) leaf disease detection: Texture features  
extraction based on GLCM and Classification using SVM**

**BY**

**MD. TOUKIR AHMED MEHEDI**

**ID: 171-15-1194**

**AND**

**ARIFUL ISLAM**

**ID: 171-15-1168**

**AND**

**MD. SAFIQUUL ISLAM**

**ID: 171-15-1160**

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

Supervised By

**Zarin Tasnim Shejuti**

Lecturer

Department of CSE

Daffodil International University

Co-Supervised By

**Md. Sabab Zulfiker**

Lecturer

Department of CSE

Daffodil International University



**DAFFODIL INTERNATIONAL UNIVERSITY**

**DHAKA, BANGLADESH**

**MAY 2021**

## **APPROVAL**

This Project titled “**Country bean (desi shim) (Lablab purpureus L.) leaf disease detection: Texture features extraction based on GLCM and Classification using SVM**”, submitted by Md. Toukir Ahmed Mehedi and Ariful Islam and Md. Safiqul Islam 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 May 19, 2021.

## **BOARD OF EXAMINERS**

**(Name)**

**Designation**

Department of CSE

Faculty of Science & Information Technology

Daffodil International University

**Chairman**

**(Name)**

**Designation**

Department of CSE

Faculty of Science & Information Technology

Daffodil International University

**Internal Examiner**

**(Name)**

**Designation**

Department of -----

Jahangirnagar University

**External Examiner**

## **DECLARATION**

We hereby declare that, this project has been done by us under the supervision of **Zarin Tasnim Shejuti, 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:**

**Zarin Tasnim Shejuti**  
Lecturer  
Department of CSE  
Daffodil International University

### **Co-Supervised by:**

**Md. Sabab Zulfiker**  
Lecturer  
Department of CSE  
Daffodil International University

### **Submitted by:**

**(Md. Toukir Ahmed Mehedi)**  
ID: -171-15-1194  
Department of CSE  
Daffodil International University

**(Ariful Islam)**

ID: -171-15-1168  
Department of CSE  
Daffodil International University

**(Md. Safiqul Islam)**  
ID: -171-15-1160  
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 successfully.

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

We would like to express our heartiest gratitude to Zarin Tasnim Shejuti, Md. Sabab Zulfiker, and 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**

Most of the people in Bangladesh are dependent on agriculture. Bangladesh farmers produce crops with great difficulty. If the leaf disease of the crop can be diagnosed then many hardships of the farmers can be removed. Country bean (desi shim) (*Lablab purpureus* L.) leaf disease detection is performed by image processing techniques. In this research paper, which Country bean (desi shim) leaf is healthy or unhealthy is done by analyzing the image and the classification is done by classification techniques. To implement image analysis (analyzing the image); the proposed framework consists of four parts, which has been done at MATLAB. The four parts are (I) Image acquisition; resizing the images (II) Image pre-processing; image enhancement (III) Image segmentation; K - means clustering algorithm to trace the diseased areas (IV) Feature extraction using GLCM; some of the main ones are Variance, Correlation, Energy, Standard Deviation, Mean, Homogeneity and Contrast. Classification of diseases has been done at Weka with the help of classification techniques. Classification is done by conducting Support Vector Machine (SVM) and the accuracy is 78%.

## TABLE OF CONTENTS

<b>CONTENTS</b>	<b>PAGE</b>
Board of examiners	i
Declaration	ii
Acknowledgements	iii
Abstract	iv
<b>CHAPTER</b>	
<b>CHAPTER 1: Introduction</b>	<b>1-3</b>
1.1 Introduction	1
1.2 Motivation	2
1.3 Objectives	2
1.4 Features	3
1.5 Social Impact	3
1.6 Expected Outcome	3
<b>CHAPTER 2: Literature Review</b>	<b>4-5</b>
<b>CHAPTER 3: Methodology</b>	<b>6-20</b>

3.1 Introduction	6
3.2 Data Collection	8
3.3 Image Acquisition	9
3.4 Image Pre-Processing	9
3.5 Image Segmentation	11
3.6 Feature Extraction	13
3.7 Classification	18
3.8 Output of Recognition	20
<b>CHAPTER 4: Result and Discussion</b>	<b>21-25</b>
<b>CHAPTER 5: Conclusions</b>	<b>26-27</b>
5.1 Conclusions	26
5.2 Future Work	26
<b>APPENDIX</b>	<b>28-37</b>
<b>Appendix A</b>	28
<b>Appendix B</b>	36
<b>Appendix C</b>	37
<b>REFERENCES</b>	<b>38-39</b>

## LIST OF FIGURES

FIGURES	PAGE NO
Figure 3.1: The framework design of an expert image processing system, which is used to send captured images through internet for the detection of country bean (desi shim) leaf disease	6
Figure 3.2: Overview of the proposed method of country bean (desi shim) leaf disease detection	7
Figure 3.3: Country bean (desi shim) leaves	9
Figure 3.4: Image pre-processing of diseased leaf using histogram equalization technique	10
Figure 3.5: Country bean (desi shim) leaf clusters using k-means clustering algorithm	12
Figure 3.6: Convert to grayscale image from RGB image	13
Figure 5.1: Future system	27
Figure B.1: Input data in Weka	36
Figure B.2: Random Forest Tree in Weka	36
Figure C.1: Input Table Data	37



## LIST OF TABLES

<b>TABLES</b>	<b>PAGE NO</b>
TABLE 4.1: Confusion matrix for country bean (desi shim) leaf disease detection	22
TABLE 4.2: Derivations of the SVM classifier	22
TABLE 4.3: GLCM based texture features of segmented images	24
TABLE 4.4: Different classifiers accuracy in SVM	24

# CHAPTER 1

## Introduction

### 1.1 Introduction

The agricultural sector of Bangladesh creates employment for about 63% of the total population. For most of countries food security means agriculture, but the only livelihood of the huge population of Bangladesh is agriculture. The agricultural sector in Bangladesh helps in reducing poverty and fostering sustainable economic development. From 2005 to 2010, the agricultural sector reduces 90 percent of poverty in Bangladesh. As of 2016, Bangladesh has 70.63% agricultural land and 59.6% arable land. According to 2016 year, about 70% of the population of Bangladesh lives in villages (rural areas). And, about 87% of the rural households are dependent on agricultural work. Now, Bangladesh has about 16.5 million farmers, according to a National Agricultural Census report (2019). In 2019, Bangladesh's GDP grew by 12.68% due to the agricultural sector. Most of the farmers in Bangladesh work in agriculture without using modern technology. Most of the farmers in Bangladesh work in agriculture without using modern technology. For this reason, just as they have to face a lot of hardships, they also have to produce crops with a lot of hardships. Much of their suffering can be overcome through modern technology image processing. It takes a lot of farmers to check if any crop leaves have been eaten or damaged by insects. This is a waste of time and a lot of money is wasted. Without modern technology, observing crop leaves in such a way as to grow crops is costly and time consuming and no good crop can be obtained accordingly. It saves time and money if the farmer can only take pictures of his crop and identify diseased leaves or insectivorous leaves through image processing. The crop will be several times more than before.

With the help of automatic disease detection, crop leaf infestation or damage can be detected. The main purpose of this research paper is detection and classification of Country bean (desi shim) leaf diseases. This purpose is accomplished with five phases.

Phase 1: Obtaining the images from the data set.

Phase 2: Image enhancement is used to enhance the superiority of the image.

Essential for supplementary processing and analysis.

Phase 3: By using k-means clustering algorithm, segmenting the images into clusters.

Phase 4: From the segmented images pull out the features using texture features GLCM.

Phase 5: Extracted features will determine and classify whether the leaf is healthy or sick using SVM.

To implement all the processes, first step is collecting the images and import them in the database. Then image acquisition process, which is gathering the images from the record. These image file extensions are in the formula of .jpg. There may be other forms as well, such as .png, .bmp, .gif and so on. Resize the images and enhance the images. Before segmentation convert RGB color space image into L\*a\*b color space image. With the assistance of K-means clustering algorithm, clustering the images which is called image segmentation. Then texture features process done by using Gray-Level Co-Occurrence Matrix (GLCM).

## **1.2 Motivation**

- i. To save both time and money of the farmer in diagnosing leaf disease in country bean (desi shim) tree.
- ii. To create a situation where the farmer can produce more crop with less effort.
- iii. Accurately diagnose the leaves of the country bean (desi shim) plant.
- iv. Combining artificial intelligence with the old method of diagnosing country bean (desi shim) leaf disease.

## **1.3 Objectives**

- i. The leaves of the country bean (desi shim) plant will be diagnosed.
- ii. To identify diseased leaves of the country bean (desi shim) plant.
- iii. The output will be sent to the user (farmer) via smartphone.

#### **1.4 Features**

- i. Captured images will be stored in the trained system.
- ii. The stored images will be analyzed with different algorithms.
- iii. The image will be classified according to multi SVM classifier.
- iv. Lastly, the leaves of the country bean (desi shim) plant will diagnose whether it is diseased or not.
- v. Test results will be sent to the user (farmer) via smartphone.

#### **1.5 Social Impact**

There is a lot of demand for food in the society. Society needs more crops to meet this food shortage. Good food has to be produced using the resources that society has at present. So, by using modern technology it is possible to produce more crops by making good use of the resources available to the society. It is possible to meet the food needs of the society by producing more food through image technology or machine learning. In addition, farmers in the society can produce more food using less labor and less money.

#### **1.6 Expected Outcome**

- i. Accurately diagnose leaf disease.
- ii. Calculate good accuracy.
- iii. Automatic systems can work properly.

## **CHAPTER 2**

### **Literature Review**

In paper [1] authors proposed four parts in their framework. To extract features, they used Gray-Level Co-Occurrence Matrix (GLCM) and for classification they used Support Vector Machine (SVM). They try to spot the symptoms of the leaf diseases automatically, in order to help the farmers to produce more crops. They used the K - Means Clustering algorithm for image Segmentation. Their input image is resized to 256x256 pixels. Their minimum classification accuracy is 0.90.

In paper [2] authors had worked on two types of leaves. One is cotton leaf and the other is tomato leaf. For classification they applied Artificial Neural Network (ANN) and Feature Extraction has done by Gray-Level Co-Occurrence Matrix (GLCM). They have used three clusters and these are Cluster 1, Cluster 2 and Cluster 3. They diagnosed four diseases; septoria leaf spot, bacterial leaf spot, leaf mold and target spot. Accuracies for four diseases are 100%, 90%, 100% and 80%.

In paper [3] authors have not only identified the leaves of the tree but also the disease of the fruit of the tree. They have used multi SVM (Support Vector Machine) algorithm in MATLAB. There are five parts in their framework. Such as Image Acquisition, Image Pre-processing, Image Segmentation, Feature Extraction and Training & Classification.

In paper [4] authors proposed approaches are GLCM and shape features using connected regions. They had worked on images with benign tumor, malignant tumor and normal brain (without tumor). The features are separated into two sets; training and testing. They used four special orientations ( $0^{\circ}$ ,  $45^{\circ}$ ,  $90^{\circ}$  and  $135^{\circ}$ ) in co-occurrence matrix. If there is no tumor then the values for perimeter, circularity and area are 0. They found value for contrast feature as malignant tumor 61.7%, benign tumor 112.2% and normal brain 15.9%.

In paper [5] authors detecting stem diseases of plants. They developed a mobile application to send the results to the farmers. There are five phases in their system; Mobile Application Development, Client and server interaction, Shell execution, Parsing the result and Server-Side Operation. Their test images accuracy is 86%.

## CHAPTER 3

### Methodology

#### 3.1 Introduction

The methodology of this research paper can be used to design a system, which can be used in such way that the farmer does not require much technical knowledge to detect the disease leaf. The framework design of this system is highlighted in Figure 3.1. With the help of this system, both experienced or inexperienced farmers will be able to diagnose country bean (desi shim) leaf disease.

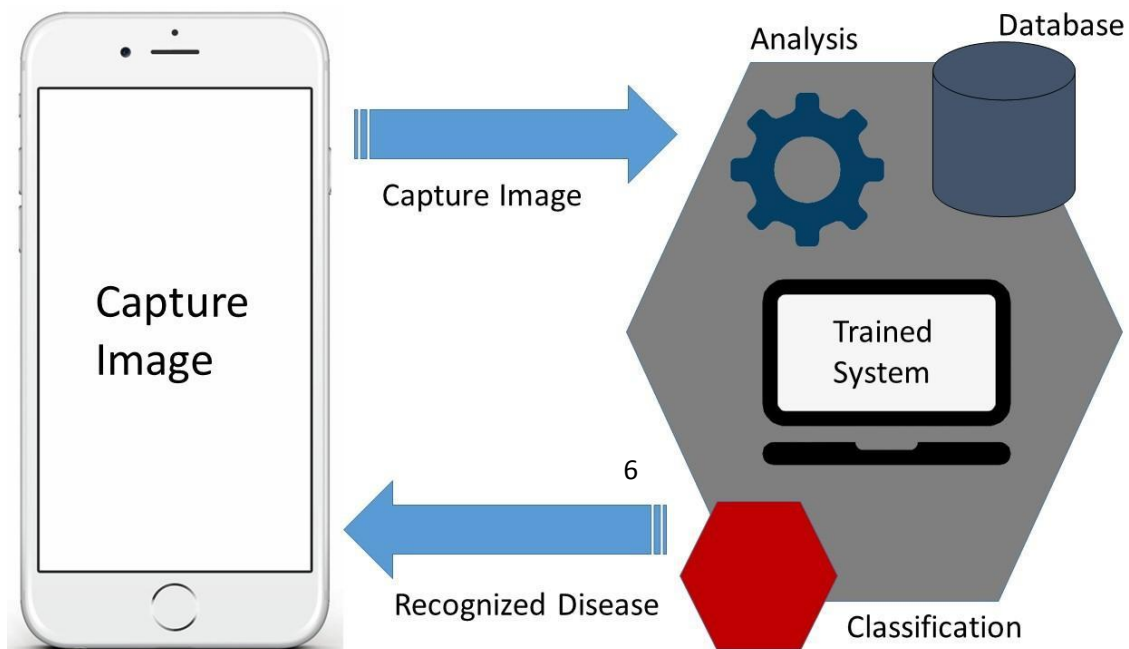
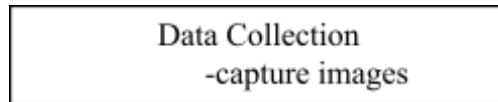
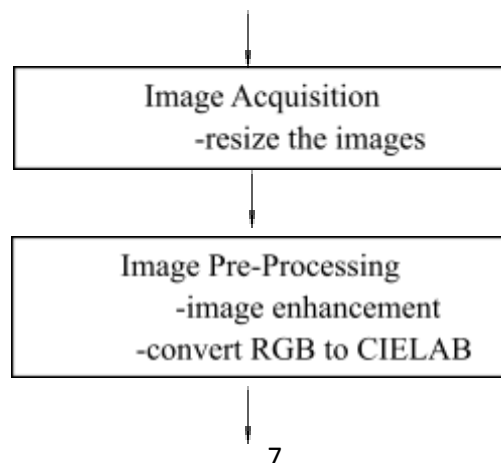


Figure 3.1: The framework design of an expert image processing system, which is used to send captured images through internet for the detection of country bean (desi shim) leaf disease



Farmers or users will capture the image of the bean (country bean) leaves using smartphone or iPhone. Now, the capture image will be sent to the database through android application or iOS application. In database, the trained system will analyse the image and then classify the image according to multi SVM classifier. According to classification, the system will give the final result that, the country bean (desi shim) leaf is healthy or unhealthy. Then the result will be sent to farmers android application or iOS application.

The block diagram of the anticipated methodology which is discussed in this research paper is shown in the Figure 3.2 below.





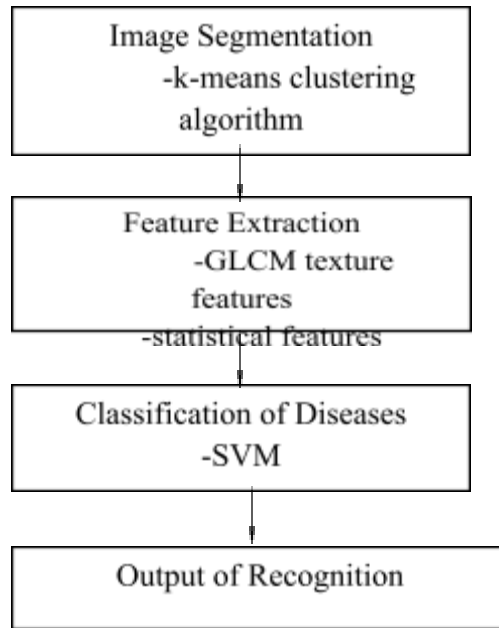


Figure 3.2: Overview of the proposed method of country bean (desi shim) leaf disease detection

### 3.2 Data Collection

The first thing to do is, carefully take images of the healthy and unhealthy leaves of the country bean (desi shim) plant with required resolution for better quality. For this research paper various kind of country bean (desi shim) leaf images are gathered from numerous bean fields with the help of the farmers. The images must be carefully stored in the database. In Figure 3.3 shows that, healthy and unhealthy leaves of country bean (desi shim).





(a) Healthy Leaves

(b) Unhealthy Leaves

Figure 3.3: Country bean (desi shim) leaves

### 3.3 Image Acquisition

The images will be collected from the database to classified properly. 154 healthy and 152 unhealthy leaves have been used for this research paper. Then collect the captured images from the database to resize them. The leaves need to be resized 256x256x3 pixels.

### 3.4 Image Pre-Processing

For further processing and analysis, the quality of the images needs to be enhanced, which is known as image preprocessing. At this stage two thing need to be done, one is image enhancement and the other is color space conversion. In this research paper histogram equalization technique is used to enhance the images. To apply histogram equalization technique in image enhancement, the images need to be transformed RGB color space to HSV color space. Now histogram equalization technique can be applied. After applying histogram equalization technique, the images need to be transformed again HSV color space to RGB color space.



(a) Input image



(b) Histogram equalization image

Figure 3.4: Image pre-processing of diseased leaf using histogram equalization technique

After improvement in the complexity of the RGB images of leaves are transformed into L\*a\*b\* colour space for image segmentation.

The CIELAB colour space stand for,

- i. L\*: Lightness
- ii. a\*: Red/Green Value
- iii. b\*: Blue/Yellow Value

From CIELAB to CIEXYZ color space conversion equations are,

$$X = X_n f^1\left(\frac{L^*+16}{116} + \frac{a^*}{500}\right)$$

$$Y = Y_n f^1\left(\frac{L^*+16}{116}\right)$$

$$Z = Z_n f^{-1}\left(\frac{L^*+16}{116} - \frac{b^*}{200}\right)$$

Where,

$$f^1(t) = \{t^3 \text{ if } t > \delta \text{ } 3\delta^2\left(t - \frac{4}{29}\right) \text{ otherwise and where } \delta = 6/29.$$

### 3.5 Image Segmentation

The procedure of image segmentation is, breakdown an image into various segments and the target of this process is, revolution the delegation of an image hooked on roughly that is more significant and calmer to analyze. Such as to locate lines, objects, curves, boundaries, etc. In this research paper, the technique of image segmentation is used k-means clustering technique. Sometimes k-means clustering algorithm referred to as the Lloyd–Forgy algorithm. Before image

segmentation, RGB images need to convert into CIELAB images because the CIELAB enables k-means clustering algorithm to quantify colors distinguish easily from one color to another.

The steps of k-means clustering algorithm are given below,

Step one: Initialize cluster centers.

Step two: Assign observations to the closest cluster center.

Step three: Revise cluster centers as mean of assigned observations.

Step four: Repeat steps 2 and 3 until convergence.

After clustering, four clusters were found in each image. 4th cluster is Nuclei Segment. Cluster 3 image holds the green substances. There are shady and light green substances in the cluster 3 image. By using the 'L\*' layer in the L\*a\*b\* colour space, separate dark green from light green. The cell nuclei are shady green. So, Segment the Nuclei. The illumination values of each colour are in 'L\*' layer. So, recall 'L\*' layer. In this cluster after extracting the illumination values of the pixels, threshold them. From cluster 3 duplicate the mask of green substances, then eliminate the light green pixels from the mask. Final result will be found by applying the new mask to the original image.



(a) Cluster 1

(b) Cluster 2



(c) Cluster 3

(d) Cluster Nuclei

Figure 3.5: Country bean (desi shim) leaf clusters using k-means clustering algorithm

### 3.6 Feature Extraction

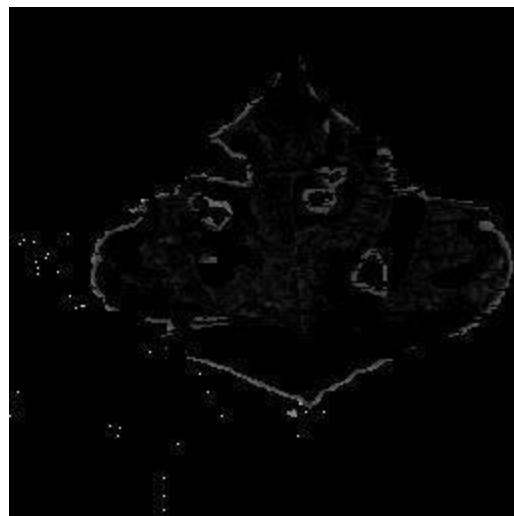
In this research paper the technique used to extract the features from the segmented images is texture features extraction technique based on Gray-Level Co-Occurrence Matrix (GLCM). A matrix that is demarcated over an image (grayscale image) to be the allocation of grayscale pixel values or cooccurring values or colours at a given balance is called gray-level co-occurrence matrices (GLCMs) or co-occurrence distribution or co-occurrence matrix. Give a specific gray level image in GLCM and it will create a formulation of how oftentimes various mixtures of gray levels co-occur in an image or image dataset. Now, GLCM's contents are used to work out texture feature calculations. These calculations are used to stretch a statistical quantity of the

mutation in strength at the pixel of awareness. The texture of an image is characterized by the GLCM functions.

Before extracting the GLCM texture features, the segmented images have to transformed into grayscale color space if the images in RGB color space.



(a) Segmented image



(b) Grayscale image

Figure 3.6: Convert to grayscale image from RGB image

Two types of features were found after feature extraction. One is texture and the other is statistical. That 13 features were extracted through the proposed method mentioned in this research paper are: Contrast (C), Correlation (Corr), Energy (E), Homogeneity (H), Entropy (h), Inverse Difference Moment (IDM) (H), Mean ( $\mu$ ), Standard deviation ( $\sigma$ ), Variance (Var) ( $\sigma^2$ ), Root Mean Square (RMS) ( $y$ ), Smoothness (R), Kurtosis (K), Skewness (S). The brief description with equation of all the features is given below,

## 1. Texture Features

### i. Contrast

In the gray-level co-occurrence matrix (GLCM), contrast procedures the local disparities. Over the whole image, it measures strength contrast of a pixel and its fellow citizen pixel. For a constant image contrast is 0. The equation of the contrast is given below,

$$Contrast = \sum_{i,j=0}^{N-1} (P_{ij})(i - j)^2$$

### ii. Correlation

Correlation procedures the combined probability incidence of the specified pixel couples. Over the entire image, correlation procedures how correlated a pixel is to its neighbor. Its range is [-1 1]. For a constant image correlation is NaN (Not a Number). The equation of the correlation is,

$$Correlation = \sum_{i,j=0}^{N-1} P_{ij} \left( \frac{(i-\mu)(j-\mu)}{\sigma^2} \right)$$

### iii. Energy

In the grey level co-occurrence matrix (GLCM), energy or the pointed second moment or uniformity revenues the sum of squared essentials. Its range is [0 1]. For an endless image energy is 1. The equation of the energy is,

$$Energy = \sum_{i,j=0}^{N-1} (P_{ij})^2$$

### iv. Homogeneity

In the GLCM to the GLCM slanting, homogeneity procedures the closeness of the distribution of elements. Its range is [0 1]. For a slanting GLCM homogeneity is 1. The equivalence of the homogeneity is,



$$Homogeneity = \sum_{i,j=0}^{N-1} \frac{(p_{ij})^2}{[1+(i-j)^2]}$$

v. Entropy

To illustrate the texture of the input image, entropy statistically measures the randomness. Entropy is expressed by  $h$ . The equation of the entropy is given below,

$$Entropy = \sum_{i,j=0}^{N-1} - \ln(P_{ij}) P_{ij}$$

vi. Inverse difference moment (IDM)

The local homogeneity of an image is measured by IDM. It is a quantity of image texture. Its range is [0 1]. IDM is expressed by  $H$ . The equation of the IDM is given below,

$$IDM = \sum_{i,j} \frac{P(i,j)}{1+|i-j|}$$

## 2. Statistical Features

i. Mean

If central clustering occurs, then the mean evaluations the worth in the image and in the defined window  $m$  of the pixel values. Mean denoted by  $\mu$ . The equation of the mean is,

$$Mean = \frac{1}{MN} \sum_{i=1}^M \sum_{j=1}^N p(i,j)$$

ii. Standard Deviation (SD)

The standard deviation is the estimate of the mean square deviation of grey pixel value from its mean value. Standard Deviation denoted by  $\sigma$ . The equation of the standard deviation is,

$$SD = \sqrt{\frac{1}{MN} \sum_{i=1}^M \sum_{j=1}^N (p(i,j) - \mu)^2}$$

iii) Variance

The square root of standard deviation is the variance. Variance can be denoted by Var or  $\sigma^2$ . The equation of the variance is,

$$\text{Variance} = \sqrt{SD}$$

iv) Root Mean Square (RMS)

The rms value of each column or row of the input, along with vectors of an earmarked dimension of the entire input is computed by Root Mean Square (RMS). The equation of the RMS is,

$$RMS = \sqrt{\frac{\sum_{i=1}^M |u_{ij}|^2}{M}}$$

v) Smoothness

Relative smoothness, R can be used to launch descriptors of comparative smoothness for the measure of grey level contrast. The equation of the smoothness is,

$$\text{Smoothness} = 1 - \frac{1}{1+\sigma^2}$$

vi) Kurtosis

Kurtosis procedures the flatness or peakedness of a circulation comparative to a normal circulation. Kurtosis denoted by K. The equation of the kurtosis is,

$$\text{Kurtosis} = \left\{ \frac{1}{MN} \sum_{i=1}^M \sum_{j=1}^N \left[ \frac{p(i,j)-\mu}{\sigma} \right]^4 \right\} - 3$$

vii) Skewness

In the earmarked window around its mean, skewness characterizes the degree of asymmetry of a pixel distribution. Skewness only characterizes the shape of the distribution. The equation of the skewness is,

$$Skewness = \frac{1}{MN} \sum_{i=1}^M \sum_{j=1}^N \left( \frac{p(i,j) - \mu}{\sigma} \right)^3$$

Where:

$P_{ij}$  = the pixel value at point (i,j) of an image

N = Number of gray levels in the image

$\mu$  = the GLCM mean

SD =  $\sigma$  = the GLCM Standard Deviation

$\sigma^2$  = the GLCM variance

MxN = size of image

u = M-by-N input matrix

### 3.7 Classification

The method used for classification in this research paper is Support Vector Machine (SVM). To examine data for classification and regression analysis, need supervised learning models with associated learning algorithms. These supervised learning models are support-vector machines (SVMs). SVMs are founded on statistical learning frameworks and these methods are one of the strongest prediction techniques. To discoveries an optimal edge between the conceivable outputs SVM have to transform input data. For these transformations, it practices a method called the

kernel trick. By comparing the predicted and actual values, the presentation of the classifier can be measured. For the calculation of statistical classification, confusion matrix can be used.

Equations of all derivations can be determined from the confusion matrix.

i) Accuracy (ACC)

$$ACC = \frac{TP+TN}{P+N} = \frac{TP+TN}{TP+TN+FP+FN}$$

ii) Precision or Position Predictive Value (PPV)

$$PPV = \frac{TP}{TP+FP} = 1 - FDR$$

iii) False Discovery Rate (FDR)

$$FDR = \frac{FP}{FP+TP} = 1 - PPV$$

iv) False Omission Rate (FOR)

$$FOR = \frac{FN}{FN+TN} = 1 - NPV$$

v) Negative Predictive Value (NPV)

$$NPV = \frac{TN}{TN+FN} = 1 - FOR$$

vi) Prevalence

$$Prevalence = \frac{Actual\ Yes}{P+N}$$

vii) Sensitivity, Recall, Hit Rate, or True Positive Rate (TPR)

$$TPR = \frac{TP}{P} = \frac{TP}{TP+FN} = 1 - FNR$$

viii) Fall-Out or False Positive Rate (FPR)

$$FPR = \frac{FP}{N} = \frac{FP}{FP+TN} = 1 - TNR$$

ix) Miss Rate or False Negative Rate (FNR)

$$FNR = \frac{FN}{P} = \frac{FN}{FN+TP} = 1 - TPR$$

x) Specificity, Selectivity, or True Negative Rate (TNR)

$$TNR = \frac{TN}{N} = \frac{TN}{TN+FP} = 1 - FPR$$

xi) Informedness or Bookmaker Informedness (BM)

$$BM = TPR + TNR - 1$$

xii) Markedness (MK) or deltaP

$$MK = PPV + NPV - 1$$

xiii) Matthews Correlation Coefficient (MCC)

$$MCC = \frac{TP \times TN - FP \times FN}{\sqrt{(TP+FP)(TP+FN)(TN+FP)(TN+FN)}}$$

xiv) Prevalence Threshold (PT)

$$PT = \frac{\sqrt{TPR(-TNR+1)} + TNR - 1}{(TPR+TNR-1)}$$

xv) Misclassification Rate (MR)

$$MR = \frac{(FP+FN)}{P+N}$$

Where:

TP = True Positive

TN = True Negative

FP = False Positive

FN = False Negative

### **3.8 Output of Recognition**

Now, the last part is recognition. After classification, SVM classify the country bean (desi shim) leaves as healthy or unhealthy then send the recognition output to the user.

## **CHAPTER 4**

### **Result and Discussion**

In this research paper the planned technique of image processing is done in MATLAB R2020a software and the classification is done in Weka 3.8.4 software. For the proposed method of this research paper 154 healthy leaves and 152 unhealthy leaves have been used. Total 306 country bean (desi shim) (Lablab purpureus L.) leaves have been used.

First of all, the image enhancement has to be done by resizing the images of the country bean (desi shim) leaves in the MATLAB. Image enhancement is done by histogram equalization technique. Then transform RGB images into CIELAB images for image segmentation, which is done by k-means clustering algorithm. The last process in MATLAB is texture features extraction, which is done by Gray-Level Co-Occurrence Matrix (GLCM).

Now in Weka, the texture features are given as input to the classifier, which is done by Support Vector Machine (SVM). The features are separated haphazardly into training vectors and testing vectors. In smoothness feature almost every value is same and in IDM feature every value is same. So, these two features have been removed during the final input in Weka.

The confusion matrix for country bean (desi shim) (Lablab purpureus L.) leaf disease detection is given in table 4.1. This confusion matrix generated by trees.RandomForest classifier in SVM.

TABLE 4.1: Confusion matrix for country bean (desi shim) leaf disease detection

====Confusion Matrix====

n = 306		Predicted		
		Unhealthy	Healthy	
Actual	Unhealthy	117	35	152
	Healthy	31	123	154
		148	158	

Derivations can be determined from the confusion matrix. Some of them are accuracy 78.43%, precision 77.85%, sensitivity or recall 79.87%, specificity 76.97%, Fall-out 23.03%, Miss rate 20.13%, and prevalence 50.33% so on. Derivations values are given in table 4.2.

TABLE 4.2: Derivations of the SVM classifier

Derivations	Output
ACC	78.43%
PPV	77.85%
FDR	22.15%
FOR	20.95%
NPV	79.05%
Prevalence	50.33%



TPR	79.87%
FPR	23.03%
FNR	20.13%
TNR	76.97%
BM	56.7%
MK	56.9%
MCC	56.87%
PT	34.92%
MR	21.57%

Another feature has been added which attribute name is "Disease Class". It is also known as a decision class. So, total features are 14. The results are shown in Table 4.3.

TABLE 4.3: GLCM based texture features of segmented images

Contrast	Correlation	Energy	Homo-geneity	Mean	Standard-Deviation	Variance	Entropy	RMS	Smoothness	Kurtosis	Skewness	IDM	Disease Class
0.17	0.55	.96	0.99	2.93	15.77	228.08	1.59	3.25	0.99	116.93	9.97	255	Healthy
0.05	0.71	.84	0.98	4.67	12.18	140.88	1.99	5.51	1	12.64	3.06	255	Healthy
0.03	0.59	.92	0.99	2.59	8.62	61.16	1.41	3.37	0.99	23.05	4.19	255	Healthy
0.03	0.63	.89	0.99	2.62	8.44	48.51	1.37	2.81	0.99	17.77	3.76	255	Healthy
0.04	0.55	.88	0.98	2.25	8.34	57.47	1.26	2.52	1	21.61	4.31	255	Healthy
.....	.....	...	.....	.....	.....	.....	.....	....	....	.....	.....	....	.....
0.02	0.49	.95	0.99	1.11	6.45	39.69	0.81	1.99	0.99	65.84	7.65	255	Healthy

In this research paper, trees.RandomForest classifier algorithm gives highest accuracy, which is 78.4314% And also has different types of classifier in SVM.Such as trees.DecisionStump, trees.HoeffdingTree,trees.J48, trees.LMT, trees.RandomTree, trees.REPTree,bayes.BayesNet, bayes.NaiveBayes, and functions.Logistic,etc. The comparative outcomes of this classifiers are shown in table 4.4.

TABLE 4.4: Different classifiers accuracy in SVM

Classifier	Accuracy
trees.RandomForest	78.4314%
trees.DecisionStump	61.1111%
trees.J48	73.5294%

trees.LMT	74.8366%
trees.RandomTree	72.2222%
trees.REPTree	70.915%
bayes.BayesNet	59.8039%
bayes.NaiveBayes	59.1503%
functions.Logistic	70.915%

## **CHAPTER 5**

### **Conclusions**

#### **5.1 Conclusions**

In this research paper, GLCM texture feature method and SVM classification method are implemented. In practice, the farmers can benefit a lot by implementing these methods. Here are the techniques used in this research paper, histogram equalization, RGB to CIELAB color space, k-means clustering algorithm, Grey Level Co-occurrence Matrix (GLCM), Support Vector Machine (SVM). These methods have been tested on various diseased leaves. Work will be done to further improve these approaches in the future and try to increase accuracy.

#### **5.2 Future Work**

- i. Create a system capable of diagnosing all types of bean plants.
- ii. Make the user interface more accessible to the user (farmer).
- iii. Creating a system that will monitor the country bean (desi shim) plant through the camera all the time.
- iv. The system will automatically capture the image, analyze it and send the results to the user (farmer).

This is what the system may look like in the future (Figure 5.1);



Figure 5.1: Future system [6]

## APPENDIX

### Appendix A

#### Source Code

```
%% image acquisition
%% clearing the windows

clc;
clear all;
close all;

%% loading the image dataset

source_file1 = 'C:\Users\Toukir
Ahmed\Desktop\source_file';
filenames1 = dir(fullfile(source_file1,
 '*.jpg'));
total_images1 = numel(filenames1);

%% setting the location of resized image
dataset

resized_file = 'C:\Users\Toukir
Ahmed\Desktop\resized_file';
```

```

%% resizing the images

for n = 1:total_images1
    im = imread(fullfile(source_file1,
        filenames1(n).name));
    resized = imresize3(im, [256 256 3]);
    imshow(resized);
    imwrite(resized, fullfile(resized_file,
        strcat(num2str(n), '.jpg')));
end

%% image pre-processing
%% loading the resized image dataset

source_file2 = 'C:\Users\Toukir
Ahmed\Desktop\resized_file';
filenames2 = dir(fullfile(source_file2,
    '*.jpg'));
total_images2 = numel(filenames2);

%% setting the location of histogram
equalization image dataset

histogram_equalization_file =
'C:\Users\Toukir
Ahmed\Desktop\histogram_equalization_file';

%% histogram equalization process

for n = 1:total_images2

```

```

        im = imread(fullfile(source_file2,
filenames2(n).name));
        con_hsv = rgb2hsv(im); %% convert RGB
image to histogram
        con_hsv(:,:,3) =
histeq(con_hsv(:,:,3)); %% 3rd value
process through histeq function
        con_rgb1 = hsv2rgb(con_hsv); %% convert
histogram to RGB
        imshow(con_rgb1);
        imwrite(con_rgb1,
fullfile(histogram_equalization_file,
strcat(num2str(n), '.jpg')));
end

%% convert RGB to L*a*b & image
segmentation
%% loading the image dataset

source_file3 = 'C:\Users\Toukir
Ahmed\Desktop\histogram_equalization_file';
filenames3 = dir(fullfile(source_file3,
 '*.jpg'));
total_images3 = numel(filenames3);

%% setting the location of k means cluster
image dataset

cluster1_file = 'C:\Users\Toukir
Ahmed\Desktop\cluster1_file';

```



```

cluster2_file = 'C:\Users\Toukir
Ahmed\Desktop\cluster2_file';
cluster3_file = 'C:\Users\Toukir
Ahmed\Desktop\cluster3_file';
blue_nuclei_file = 'C:\Users\Toukir
Ahmed\Desktop\blue_nuclei_file';

%% convert the rgb images to L*a*b images &
image segmentation

for n = 1:total_images3
    im = imread(fullfile(source_file3,
filenames3(n).name));
    lab_transform = makecform('srgb2lab',
'AdaptedWhitePoint', whitepoint('D65'));

%% RGB to L*a*b with function
makecform('rgb2lab')

    con_lab1 = applycform(im,
lab_transform);

    ab = con_lab1(:, :, 2:3); %% Classify the
Colours in 'a*b*' Space Via K-Means
Clustering
    ab = im2single(ab);
    nColors = 3;
    pixel_labels =
imsegkmeans(ab, nColors, 'NumAttempts', 3);

```

```
%% Produce Images that Segment the Image by  
Colour
```

```
    mask1 = pixel_labels==1; %% Cluster 1  
    cluster1 = im .* uint8(mask1);  
    imshow(cluster1);  
    imwrite(cluster1,  
fullfile(cluster1_file, strcat(num2str(n),  
' .jpg')));
```

```
    mask2 = pixel_labels==2; %% Cluster 2  
    cluster2 = im .* uint8(mask2);  
    imshow(cluster2);  
    imwrite(cluster2,  
fullfile(cluster2_file, strcat(num2str(n),  
' .jpg')));
```

```
    mask3 = pixel_labels==3; %% Cluster 3  
    cluster3 = im .* uint8(mask3);  
    imshow(cluster3);  
    imwrite(cluster3,  
fullfile(cluster3_file, strcat(num2str(n),  
' .jpg')));
```

```
%% Segment the Nuclei
```

```
L = con_lab1(:, :, 1);  
L = double(L);
```

```

    L_blue = L .* double(mask3);
    L_blue = rescale(L_blue);
    idx_light_blue =
imbinarize(nonzeros(L_blue));

    blue_idx = find(mask3);
    mask_dark_blue = mask3;

mask_dark_blue(blue_idx(idx_light_blue)) =
0;

    blue_nuclei = im .*
uint8(mask_dark_blue);
    imshow(blue_nuclei);
    imwrite(blue_nuclei,
fullfile(blue_nuclei_file,
strcat(num2str(n), '.jpg')));
end

%% GLCM texture features
%% loading the image dataset

source_file4 = 'C:\Users\Toukir
Ahmed\Desktop\blue_nuclei_file';
filenames4 = dir(fullfile(source_file4,
'*jpg'));
total_images4 = numel(filenames4);

%% location of grayscale image

```

```

grayscale_image_file = 'C:\Users\Toukir
Ahmed\Desktop\grayscale_image_file';

%% GLCM texture features process

for n = 1:total_images4
    im = imread(fullfile(source_file4,
filenames4(n).name));
    Mean = mean2(im);

    %% Translate to grayscale if image is
RGB

    if ndims(im) == 3
        gray_img = rgb2gray(im);
    end
    imshow(gray_img);
    imwrite(gray_img,
fullfile(grayscale_image_file,
strcat(num2str(n), '.jpg')));
    %% Generate the Gray Level Cooccurrence
Matrices (GLCMs)

    glcms = graycomatrix(gray_img);

    %% Derive Statistics from GLCM

    stats = graycoprops(glcms, 'Contrast
Correlation Energy Homogeneity');
    Contrast = stats.Contrast;
    Correlation = stats.Correlation;

```

```

Energy = stats.Energy;
Homogeneity = stats.Homogeneity;

%% Derive features from GLCM
segmentation image

Mean = mean2(im);
Standard_Deviation = std2(im);
Variance = mean2(var(double(im)));
Entropy = entropy(im);
RMS = mean2(rms(im));
a = sum(double(im(:)));
Smoothness = 1-(1/(1+a));
Kurtosis = kurtosis(double(im(:)));
Skewness = skewness(double(im(:)));
IDM = idm_fun(im);
features(n,:) = [Contrast, Correlation,
Energy, Homogeneity, Mean,
Standard_Deviation, Variance, Entropy, RMS,
Smoothness, Kurtosis, Skewness, IDM];
end
%% end of code

```

IDM (Inverse Difference Moment) function;

```
%% Inverse Difference Moment
```

```
function IDM = idm_fun(seg_img)
    m = size(seg_img,1);
    n = size(seg_img,2);

```

```

    in_diff = 0;
    for i = 1:m
        for j = 1:n
            temp =
seg_img(i,j) ./ (1+(i-j).^2);
            in_diff = in_diff+temp;
        end
    end
    IDM = double(in_diff);
end

```

## Appendix B

### Screenshots of Weka Software

The screenshot shows the Weka Explorer interface. The 'Selected attribute' panel displays the following statistics for the 'Contrast' attribute:

Statistic	Value
Minimum	0
Maximum	1.604
Mean	0.109
StdDev	0.17

The 'Attributes' panel shows a list of attributes with 'Contrast' selected as the first attribute. The 'Current relation' panel shows 'Relation: f1anl\_features3' with 306 instances and 12 attributes.

Figure B.1: Input data in Weka

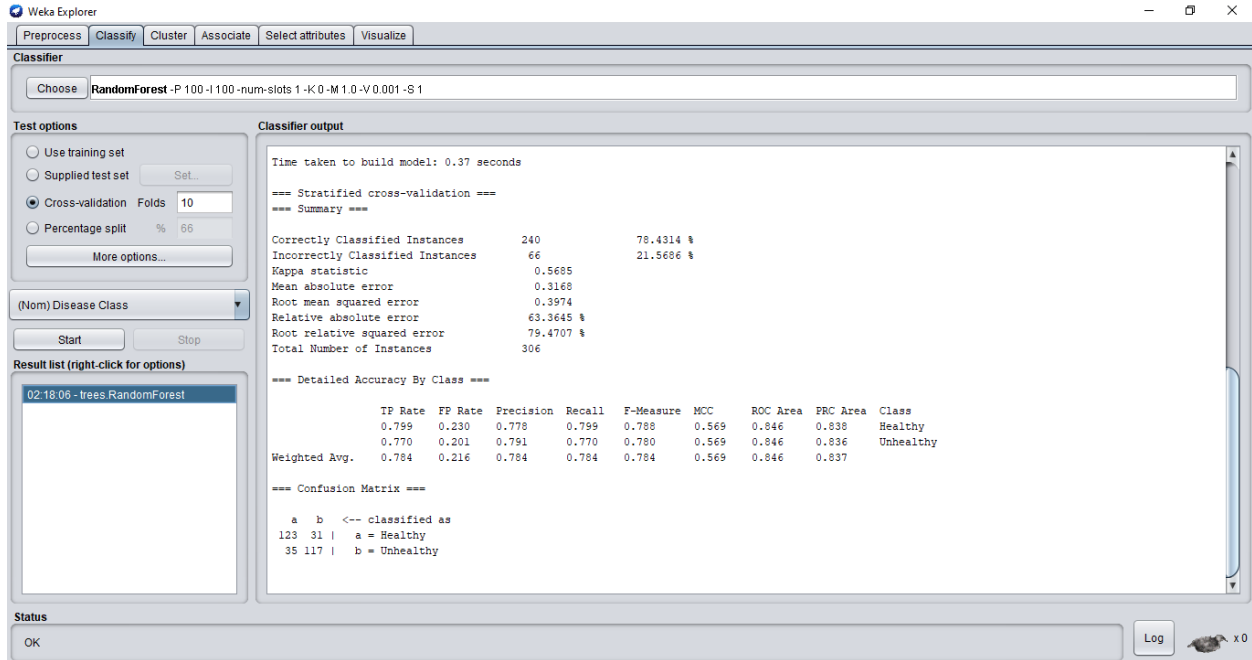


Figure B.2: Random Forest Tree in Weka

## Appendix C

### Screenshot of Input Table

	A	B	C	D	E	F	G	H	I	J	K	L	
1	Contrast	Correlatio	Energy	Homogen	Mean	Standard	Variance	Entropy	RMS	Kurtosis	Skewness	Disease Class	
2	0.170803	0.54947	0.957264	0.987912	2.930954	15.77078	228.08	1.556916	3.24884	116.9344	9.970813	Healthy	
3	0.046507	0.714321	0.838972	0.980404	4.670929	12.18192	140.8889	1.995731	5.510532	12.64044	3.057167	Healthy	
4	0.029657	0.593203	0.919011	0.987377	2.591573	8.615471	61.16362	1.408116	3.367947	23.05357	4.185803	Healthy	
5	0.029779	0.629337	0.893363	0.985539	2.624873	8.444456	48.50705	1.372534	2.814843	17.76782	3.760051	Healthy	
6	0.038664	0.552042	0.876951	0.980729	2.253911	8.340058	57.46952	1.264786	2.520377	21.61459	4.311168	Healthy	
7	0.304948	0.244513	0.952608	0.986141	2.885376	17.56927	195.1872	1.439274	2.491376	102.8007	9.553619	Healthy	
8	0.238572	0.875973	0.457361	0.97209	26.52036	39.20988	493.6011	3.879809	7.56081	2.981529	1.166024	Healthy	
9	0.004504	0.067363	0.990688	0.997748	1.384786	5.059544	18.9995	1.087609	2.073034	23.08338	4.36778	Healthy	
10	0.21633	0.918061	0.416195	0.975269	32.34899	44.56909	1155.226	4.094344	8.214	2.500829	0.982424	Healthy	

Figure C.1: Input Table Data

**Reference:**



- [1] R. M. Prakash, G. Saraswathy, G. Ramalakshmi, K. Mangaleswari, and T. Kaviya, "Detection of leaf diseases and classification using digital image processing," in 2017 international conference on innovations in information, embedded and communication systems (ICIIECS). IEEE, 2017, pp. 1–4.
- [2] C. U. Kumari, S. J. Prasad, and G. Mounika, "Leaf disease detection: Feature extraction with k-means clustering and classification with ann," in 2019 3rd International Conference on Computing Methodologies and Communication (ICCMC). IEEE, 2019, pp. 1095–1098.
- [3] S. Raut and A. Fulsunge, "Plant disease detection in image processing using matlab," International journal of innovative Research in science, Engineering and Technology, vol. 6, no. 6, pp. 10 373–10 381, 2017.
- [4] S. K. PS and V. Dharun, "Extraction of texture features using glcm and shape features using connected regions," International Journal of Engineering and Technology, 2016.
- [5] Z. N. Reza, F. Nuzhat, N. A. Mahsa, and M. H. Ali, "Detecting jute plant disease using image processing and machine learning," in 2016 3rd International Conference on Electrical Engineering and Information Communication Technology (ICEEICT). IEEE, 2016, pp. 1–6.
- [6] DOLPHIN PG COLLEGE, available at <<<https://dolphinlifesciences.com/top-agriculture-colleges-in-himachal-pradesh/>>>, last accessed on March 16, 2021 at 10:00 PM.
- [7] M. U. Mojumdar and N. R. Chakraborty, "A computer vision technique to detect scab on malabar nightshade," in 2020 11th International Conference on Computing, Communication and Networking Technologies (ICCCNT). IEEE, 2020, pp. 1–4.
- [8] N. Pradeep, H. Girisha, and K. Karibasappa, "Segmentation and feature extraction of tumors from digital mammograms," Computer Engineering and Intelligent Systems, vol. 3, no. 4, pp. 37–46, 2012.
- [9] S. D. Khirade and A. Patil, "Plant disease detection using image processing," in 2015 International conference on computing communication control and automation. IEEE, 2015, pp. 768–771.
- [10] W. K. Mutlag, S. K. Ali, Z. M. Aydam, and B. H. Taher, "Feature extraction methods: A review," in Journal of Physics: Conference Series, vol. 1591, no. 1. IOP Publishing, 2020, p. 012028.
- [11] J.-S. Wang and X.-D. Ren, "Glcmm based extraction of flame image texture features and kpca-glvq recognition method for rotary kiln combustion working conditions," International Journal of Automation and Computing, vol. 11, no. 1, pp. 72–77, 2014.

- [12] T. Youwen, L. Tianlai, and N. Yan, "The recognition of cucumber disease based on image processing and support vector machine," in 2008 congress on image and signal processing, vol. 2. IEEE, 2008, pp. 262–267.
- [13] C. Fernandez-Maloigne, *Advanced color image processing and analysis*. Springer Science & Business Media, 2012.
- [14] X. Yang, S. Tridandapani, J. J. Beitler, D. S. Yu, E. J. Yoshida, W. J. Curran, and T. Liu, "Ultrasound glcm texture analysis of radiationinduced parotid-gland injury in head-and-neck cancer radiotherapy: An in vivo study of late toxicity," *Medical physics*, vol. 39, no. 9, pp. 5732–5739, 2012.