Study on Paddy Disease Detection using Color Co-occurrence Features

by

Md. Jahirul Islam

Roll: 1307035

&

Protap Chandra Ghose

Roll: 1307046

Department of Computer Science and Engineering

Khulna University of Engineering & Technology

Khulna 9203, Bangladesh

February, 2018

Study on Paddy Disease Detection using Color Co-occurrence Features

by

Md. Jahirul Islam

Roll: 1307035

&

Protap Chandra Ghose

Roll: 1307046

A thesis submitted in partial fulfillment of the requirements for the degree of

"Bachelor of Science in Computer Science and Engineering"

Supervisor: Dr. Sheikh Mohammad Masudul Ahsan

Professor

Department of Computer Science and Engineering

Khulna University of Engineering and Technology.

Signature

Department of Computer Science and Engineering

Khulna University of Engineering & Technology

Khulna 9203, Bangladesh

February, 2018

Acknowledgment

All the praise to the almighty God, whose blessing and mercy helped us to complete this thesis work well. After that, we humbly acknowledge the valuable advice, guidance and co-operation of Dr. Sk. Mohammad Masudul Ahsan, Professor, Department of Computer Science and Engineering, Khulna University of Engineering & Technology, under whose supervision this work was carried out. His intellectual advice, encouragement and guidance make us feel confident and inspire to go through different research ideas. From him, we have learned that scientific research needs much effort in learning and applying and need to have a broad view at problems from different perspective. We would like to convey our heartiest gratitude to all the faculty members, officials and staffs of the Department of Computer Science and Engineering as they have always extended their co-operation to complete this work. Last but not the least, we wish to thank our friends and family members for their constant support.

Abstract

Being an agricultural country, most of the people of Bangladesh are dependent on agriculture directly or indirectly. It is the fourth largest rice producing country in the world. Main hindrance in rice production is paddy diseases. So in this research the main objective is to develop a prototype system for detecting the paddy diseases, which are Paddy Blast, Brown Spot and Narrow Brown Spot diseases. This concentrate on the image processing techniques used to find pattern in the image and artificial neural network technique to classify the diseases. The methodology involves image collection, image processing, feature extraction and classification. Features are extracted from the images using Haralick's texture feature from color co-occurrence matrix. Then an artificial neural network is trained by these features and a trained model is found. In testing phase, all paddy samples are passed through the leaf color analysis to detect the normal paddy leaf image. If the sample passes leaf color analysis, then it is automatically classified as Normal Paddy leaf image. Otherwise, all the segmented paddy disease samples are converted into the features data and is passed through the artificial neural network. Consequently, by employing the artificial neural network technique, the paddy diseases are recognized. The accuracy to detect diseases of this model is good enough to use in practical life.

Table of Contents

List of Figures

List of Tables

Chapter 1

Introduction

This chapter discuss about an overview of the study conducted. The title is “Study on Paddy Disease Detection using Color Co-occurrence Features”. It consists of background, problem statements, objectives and the scope of study. The background briefly describes the identification of thesis and related issues. Problem statements describes the problems that arise and make the selected projects to be undertaken. The objectives are the goals list for the research to be achieved. Scope of study discuss about the limitations of research. Lastly, thesis organization gives a summary of the sequence for each chapter in the thesis.

1.1 Background

Rice known as Oryza Sativa (specific name), is one of the most utilized food plants and widely grown originated in ASIA. Rice is an important crop worldwide and over half of the world population relies on it for food. Many people in the world including Bangladesh eat rice as staple food. However, there are many factors that make paddy rice production become slow and less productive. One of the main factors is paddy disease.

An abnormal condition that injures the plant or leads it to function improperly is called as a disease. Diseases are readily recognized by their symptoms. There are a lot of paddy disease types which are paddy blast, narrow brown spot, brown spot disease and many more. Image processing and computer vision technology are very beneficial to the agricultural industry. They are more potential and more important to many areas in agricultural technology.

As an agricultural country, Bangladesh gets its one-sixth of national income from rice. About 10.5 million hectares' lands produce 25.0 million tons' rice ever year [2014]. Now govt.'s target is to produce another 30 million over the next 20 years. Main Obstacle for gaining the target is those paddy diseases. If diseases can detect easily with image processing, taking action will be faster. In the fig 1.1 the red line represents annual consumption and the blue line represent annual production from year 1961 to 2006.

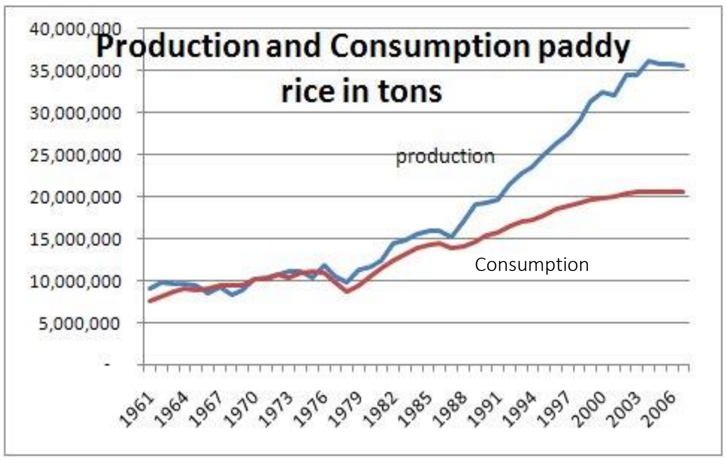


Fig 1.1: Rice Production and Consumption Statistics in Bangladesh from 1961-2006

Paddy Disease Detection System is one of the very beneficial systems. It can help the paddy farmer detect the disease faster. This study aims to develop a prototype system to automatically detect and classify the paddy diseases by using image processing technique as an alternative or supplemental to the traditional manual method.

1.2 Problem Statement

Paddy will be harvest twice in a year. Most of paddy farmer faces many problems to harvest their paddy because they used to attack by snail, worm and fungi. Furthermore, when the paddy had been infected or attacked, the others areas had been exposed to be infected. Thus, it will decrease paddy farmer's income and lead to significance losses to farmer. Currently, the paddy farmer determines the type of disease manually. The errors might occur in order to determine the type of diseases. Paddy farmer also have to spend a lot of time to detect the type of disease. It also takes a time as the paddy farmers manually check the disease since the paddy field is in wide area.

1.3 Objective

There are three goals to accomplish in this thesis:

- to propose a model of Paddy Disease Detection framework
- to distinguish the Paddy Disease by utilizing Color Co-occurrence Features
- to apply Image Processing strategy to investigate the example of Paddy Disease

1.4 Scope of Study

- The users of the system are general farmer
- The prototype is designed in python
- 220 total samples of normal, brown spot disease, narrow brown spot disease and blast disease is used in this

1.5 Thesis Organization

This thesis consists of 5 chapters ranging from Chapter 1 until Chapter 5. Chapter 1 gives an overview of the topic. It additionally comprises of Problem Statements, Objectives and the Scope of Study. In the meantime, Chapter 2 reviews the past research works that was led by others explores. All the significant specialized paper, diaries and books taken from those inquires about will be talked about in detail. Chapter 3 centers about the procedure for framework improvement and process stream in detail of this examination. It uncovers the method and the calculations that will be utilized as a part of playing out this investigation. Chapter 4 that comprise of expected outcome or yield, limitation of task and further research. In conclusion, Chapter 5 finishes up the general examination, research and limitation.

Chapter 2

Literature Review

2.1 Introduction

This chapter quickly surveys, clarifies and talks about on existing writing audit related with our research topic which is "Study on Paddy Disease Detection using Color Co-occurrence Features". This part includes three areas. The primary segment portrays the outlines of paddy. The subsections are the definition, kind of paddy infection, paddy manifestation and paddy administration. The second area is the survey of some current framework that utilized same strategies and techniques. The third area talks about the survey on strategy and technique utilized by the framework. The subsections are picture securing, picture division and manufactured neural system.

2.2 Paddy Overviews

In this area, firstly introduces a meaning of paddy. From that point onward, this subsection quickly examines on symptoms of paddy diseases.

2.2.1 Definition of Paddy

Paddy otherwise called rice is the dull seeds of a yearly south-east Asian grain grass (Oryza sativa) that are cooked and utilized for sustenance. This grain grass that is broadly developed in warm atmospheres for its seeds and results. Rice is a standout amongst the most used sustenance plants and generally developed began in ASIA. Rice is a critical product worldwide and over portion of the total populace depends on it for sustenance. Numerous individuals on the planet including Bangladesh eat rice as staple nourishment.

2.2.2 Paddy Diseases Symptoms

There are numerous elements that influence paddy to rice generation turn out to be moderate and less profitable. One of the fundamental elements is paddy disease. The statements beneath will indicate three kind of paddy disease, the symptom of paddy disease and the management of paddy disease. This inquires about spotlight on three sorts of diseases, which are paddy blast, brown spot disease and narrow brown spot disease.

2.2.2.1 Paddy Blast Symptoms

- Disease infect paddy at growth stages and aerial parts of plant (leaf, neck and node)
- Among the three leaves and neck infections are more severe
- Small specks originate on leaves
- Several spots coalesce to big irregular patches

2.2.2.2 Brown Spot Symptoms

- Initial lesions are water-soaked to greenish gray and later become grayish white with brown margin
- Lesions on leaf sheaths near waterline
- Presence of sclerotic
- Lesions may coalesce death of the whole leaf
- Partially filled or empty grains

2.2.2.3 Narrow Brown Spot Symptoms

- Short, narrow, elliptical to linear brown lesions usually on leaf blades but may also occur on leaf sheaths, pedicels, and glumes or rice hulls
- Lesions about 2-10 mm long and 1 mm wide
- Lesions narrower, shorter, and darker brown on resistant varieties
- Lesions wider and lighter brown with gray necrotic centers on susceptible varieties
- Leaf necrosis may also occur on susceptible varieties
- Lesions occur in large numbers during the later growth stages

Why Occurs?

- The disease is observed on rice crops grown on soil deficient in potassium.
- Temperature ranging from 25-28° C is favorable for the optimum growth of the disease
- Susceptibility of the variety to the fungus and the growth stage of the rice crop are other factors that affect the development of the disease
- Although rice plants are susceptible to the fungus at all stages of growth, they are more susceptible from panicle emergence to maturity
- Thus, becoming more severe as rice approaches maturity.

2.3 Case Study on Existing System

An examination led by Kurniawati et al. from University Kebangsaan Malaysia means to build up a model framework to consequently and effectively recognize and characterize the paddy diseases with Blast Disease (BD), Brown Spot Disease (BSD), and Narrow Brown Spot Disease (NBSD) utilizing picture preparing system as an option or supplemental to the conventional manual technique.

In the paper, Batule et al. from Trinity College of Engineering and Research, Pune, Trinity College of Engineering and Research, Pune, Maharashtra, India gave a method to detect the disease caused to the leaf calculating the RGB and HSV values. Primarily the image is blurred in order reduce noise. Then the image is converted from RGB to HSV form, after this color thresholding is done. After thresholding foreground or background detection is performed. Background detection leads to feature extractions of the leaf. Then k-means algorithm is applied which can help to clot the clusters. The following system is a software based solution for detecting the disease with which the leaf is infected.

In the paper, R.Preethi et al. from Panimalar Engineering College,Chennai, Tamilnadu-123, (India) proposed a system which will automatically detect the symptoms of diseases as soon as they appear on plant leaves. These images are made to undergo a set of pre-processing methods for image enhancement. Later, a satisfying set of visual features from the region of interest are extracted by applying histogram for detecting diseases accurately. The advisory helps farming community in effective decision making to protect their crop from diseases and increase its productivity.

In the paper, Phadikar and Sil from Dept. of CSE, West Bengal University of Technology, Kolkata700064, India, described a software prototype system for rice disease detection based on the infected images of various rice plants. Images of the infected rice plants are captured by digital camera and processed using image growing, image segmentation techniques to detect infected parts of the plants. Then the infected part of the leaf has been used for the classification purpose using artificial neural network. The methods evolved in this system are both image processing and soft computing technique applied on number of diseased rice plants.

In the paper, Paul and Sharma, Department of Electronics and Telecommunication, Bhilai Institute of Technology, Durg, India, evaluated a software solution for fast, accurate and automatic detection of plant diseases through Image Processing. Identification of the plant diseases is the key to preventing losses in the quality and quantity of the agricultural product. Health monitoring and disease detection of plant is critical for sustainable agriculture. The typical method of studying plant disease is to rely on visually observable patterns on the plant leaves. Visually identifying plant diseases is inefficient, difficult, time consuming, requires expertise in plant diseases and continuous monitoring which might be expensive in large farms. Therefore; a fast, automatic and accurate method to detect plant disease is of great importance. Hence, image processing technique is employed for the detection of plant diseases. The implementation of these technologies will lead to improved productivity.

2.4 Conclusion

In this chapter, this thesis presented the common method for detecting paddy disease. By previewing various method, we learned advantage and disadvantage of various method and their characteristic for simple and clean paddy disease detection.

Chapter 3

Proposed Methodology

3.1 Introduction

The purpose of this chapter is to discuss the approach and system for the thesis. Method, technique or approach that has been used while designing and implementing the thesis are included in this chapter. This chapter also explains about the justification of method or approach used and hardware and software necessity.

3.2 Work Flow

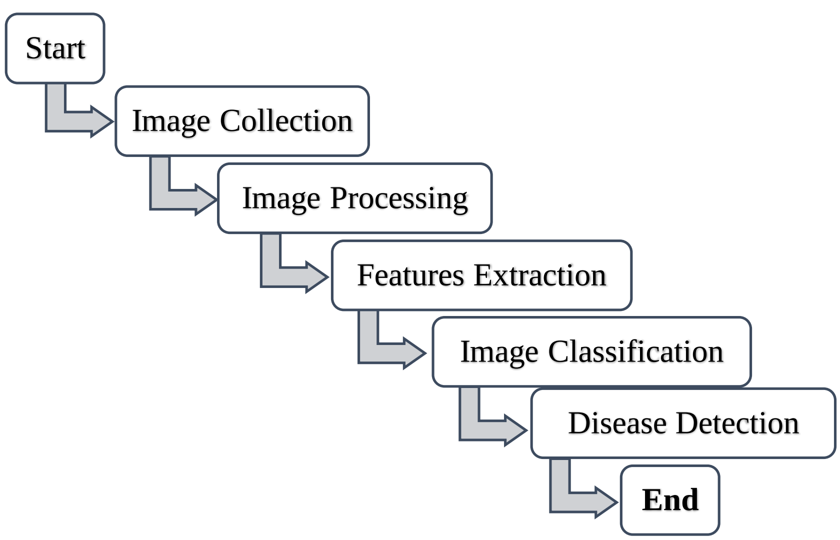
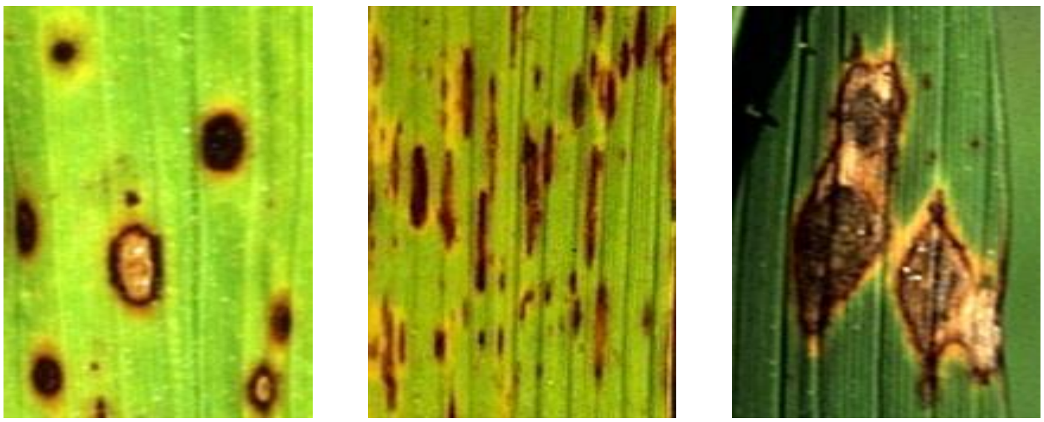


Fig 3.1: Work Flow of proposed methodology

3.2.1 Image Collection

The RGB images of paddy leaf are collected from Internet. Those image cropped into a smaller image with dimension of 64 x 64 pixels as training data. We have collected about 180 data samples with the four rotation from each images. It consists of three types of paddy diseases (Paddy Blast, Brown Spot, Narrow Brown Spot) as shown in Fig. 3.2. Images are stored in jpg format.



(a) (b) (c)

Fig 3.2: Sample of collected images (a) Brown Spot Disease; (b) Narrow Brown Spot Disease; (c) Blast Disease

3.2.2 Image Processing

The main objective of this process is to obtain an image with an approximation of human color perception. The RGB image (Fig. 3.4(a)) is converted into Lab as abbreviation for CIEL*a*b* 1976 color space (also CIELAB), as shown in Fig. 3.4(b).

Lab Color Space

The LAB color model is a three axis color system and LAB colors are absolute, meaning that the color is exact. It's what's known as device independent; meaning that the LAB color space is the only way to communicate different colors across different devices. An object's color is measured in LAB color with a spectrophotometer.

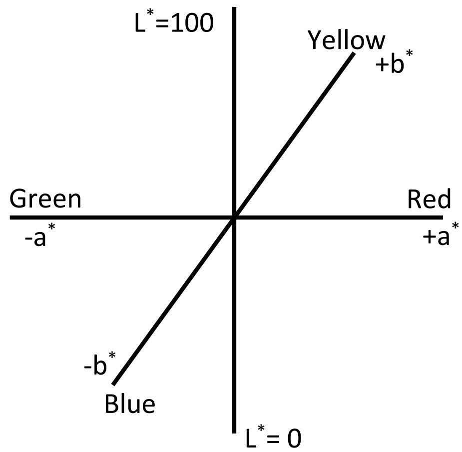
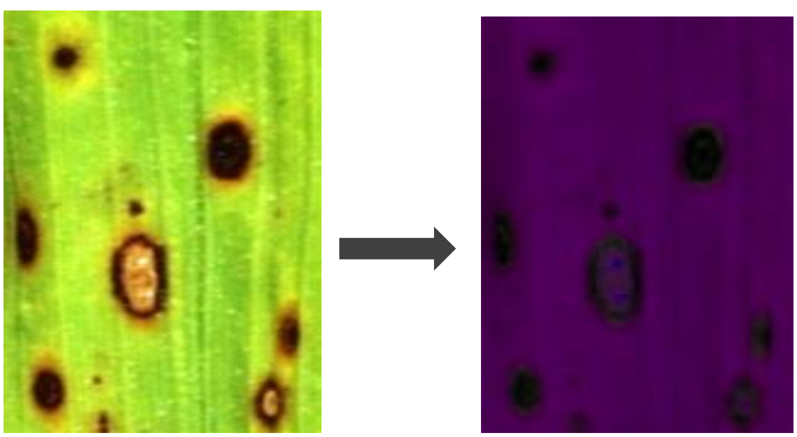


Fig 3.3: L*a*b* Color Space in Dimensional Graph

These three coordinates of CIELAB (Fig 3.3) represent-

- The lightness of the color L* = 0 yields black and L* = 100 indicates diffuse white; specular white may be higher
- Its position between red/magenta and green (a*, negative values indicate green while positive values indicate magenta) and
- Its position between yellow and blue (b*, negative values indicate blue and positive values indicate yellow)



(a) (b)

Fig 3.4: Processing image from RGB to Lab (a) RGB image; (b) Lab image

3.2.3 Feature Extraction

Feature Extraction a sort of dimensionality lessening that productively speaks to intriguing parts of a picture as a smaller component vector. Features are extracted from the color co-occurrence matrix which is calculated previously.

3.2.3.1 Color Co-occurrence Matrix

A co-occurrence matrix or co-occurrence distribution is a matrix that is defined over an image to be the distribution of co-occurring pixel values (grayscale values, or colors) at a given offset:

- The offset, $(\Delta x, \Delta y)$, is a position operator that can be applied to any pixel in the image (ignoring edge effects): for instance, $(1, 2)$ could indicate "one down, two right".
- An image with p different pixel values will produce a $p \ x \ p$ co-occurrence matrix, for the given offset.
- The $(i, j)^{th}$ value of the co-occurrence matrix gives the number of times in the image that the i^{th} and j^{th} pixel values occur in the relation given by the offset.

$$C_{\Delta x, \Delta y}(i,j) = \sum_{x=1}^{n} \sum_{y=1}^{m} \left\{ \begin{matrix} 1, if \ I(x,y) = i \ and \ I(x+\Delta x, y+\Delta y) = j \\ 0, otherwise \end{matrix} \right\}$$

Where: i and j are the pixel values; x and y are the spatial positions in the image I; the offsets $(\Delta x, \Delta y)$ define the spatial relation for which this matrix is calculated; and $I(\Delta x, \Delta y)$ indicates the pixel value at pixel (x, y).

The offset value $(\Delta x, \Delta y)$ is calculated by the spatial direction. If the direction is

$$0°, \text{then } \Delta x = \ 0, \ \Delta y = 1$$

$$45°, \text{then } \Delta x = \ 1, \ \Delta y = 1$$

$$90°, \text{then } \Delta x = \ 1, \ \Delta y = 0$$

$$135°, \text{then } \Delta x = -1, \ \Delta y = 1$$

Let's take an example of an $5x5$ image and the GLCM matrix is calculated by following procedure:

$$\begin{bmatrix} 22 & 220 & 180 & 111 & 75 \\ 25 & 240 & 103 & 180 & 118 \\ 65 & 110 & 210 & 230 & 191 \\ 123 & 40 & 150 & 5 & 199 \\ 180 & 95 & 70 & 15 & 255 \end{bmatrix}$$

The matrix size is $5x5$. So, the segment size will be ${}^{255}/_{5} = 51$ and all values of the above matrix will be replaced by below:

$$0 - \quad 51 \rightarrow 0$$

$$52 - 103 \rightarrow 1$$

$$104 - 154 \rightarrow 2$$

$$155 - 205 \rightarrow 3$$

$$206 - 255 \rightarrow 4$$

By applying above mapping the resultant matrix is given below:

$$\begin{bmatrix} 0 & 4 & 3 & 2 & 1 \\ 0 & 4 & 1 & 3 & 2 \\ 1 & 2 & 4 & 4 & 3 \\ 2 & 0 & 2 & 0 & 3 \\ 3 & 1 & 1 & 0 & 4 \end{bmatrix}$$

When the direction is 0° and the offset value is 1, then the matrix which will be found is given below:

$$\begin{bmatrix} 0 & 0 & 1 & 1 & 3 \\ 1 & 1 & 1 & 1 & 0 \\ 2 & 1 & 0 & 0 & 1 \\ 0 & 1 & 2 & 0 & 0 \\ 0 & 1 & 0 & 2 & 1 \end{bmatrix}$$

The transpose matrix of the above matrix is given below:

$$\begin{bmatrix} 0 & 1 & 2 & 0 & 0 \\ 0 & 1 & 1 & 1 & 1 \\ 1 & 1 & 0 & 2 & 0 \\ 1 & 1 & 0 & 0 & 2 \\ 3 & 0 & 1 & 0 & 1 \end{bmatrix}$$

After Adding the above two matrices, we get the following matrix:

$$\begin{bmatrix} 0 & 1 & 3 & 1 & 3 \\ 1 & 2 & 2 & 2 & 1 \\ 3 & 2 & 0 & 2 & 1 \\ 1 & 2 & 2 & 0 & 2 \\ 3 & 1 & 1 & 2 & 2 \end{bmatrix}$$

Now, the determinant of this matrix = 40. By normalizing the matrix, we get resultant matrix:

$$\begin{bmatrix} 0 & 0.025 & 0.075 & 0.025 & 0.075 \\ 0.025 & 0.050 & 0.050 & 0.050 & 0.025 \\ 0.075 & 0.050 & 0 & 0.050 & 0.025 \\ 0.025 & 0.050 & 0.050 & 0 & 0.050 \\ 0.075 & 0.025 & 0.025 & 0.050 & 0.050 \end{bmatrix}$$

This is the resultant GLCM matrix when the offset value is 1 and rotation is 0°.

If we apply CCM in the Fig 3.5 image of size $64\ x\ 64$, a $20\ x\ 20$ CCM matrix will be found.

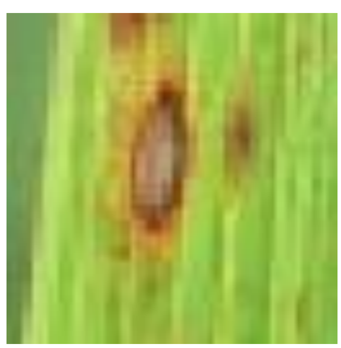


Fig 3.5: Brown Spot affected paddy image

Table 3.1: A $20x20$ matrix of CCM values

105	52	25	8	4	4	4	3	4	3	3	3	2	2	2	1	0	1	0	0
42	50	38	20	13	9	8	4	2	2	3	4	2	1	3	1	1	1	0	0
26	31	39	31	17	14	7	7	5	2	2	2	2	3	1	2	2	2	0	0
9	16	27	42	26	15	12	12	4	6	4	4	4	4	2	4	4	2	0	0
8	12	16	22	30	23	13	11	10	4	5	4	2	5	3	3	1	2	0	0
5	11	13	15	30	35	25	17	14	6	7	7	4	4	2	3	4	1	1	0
7	7	8	11	14	27	47	37	19	11	8	6	5	5	5	3	2	3	0	0
5	3	7	10	9	22	39	55	28	18	15	9	6	5	10	4	1	2	5	1
3	3	7	6	15	21	37	42	23	19	13	9	6	4	5	2	4	1	2	3
2	2	3	4	5	7	10	17	25	25	23	17	10	6	4	5	4	3	3	0
1	6	2	5	5	8	7	14	21	22	29	28	18	11	7	4	4	4	2	1
5	4	3	3	3	10	9	9	14	15	26	31	26	18	15	8	9	7	3	2
3	3	4	3	3	3	5	5	10	7	19	26	23	21	22	12	8	5	4	1
2	2	2	4	4	6	7	6	6	6	12	19	21	26	30	18	10	8	4	2
2	2	1	3	2	3	4	8	7	6	7	17	20	30	29	23	18	14	8	6
1	1	1	3	2	3	4	5	5	5	5	9	11	18	28	34	24	17	12	3
0	2	1	2	0	3	3	4	3	5	4	11	10	11	18	24	28	27	13	5
1	1	1	1	1	1	2	1	2	3	4	4	5	8	9	13	20	30	49	35
0	0	0	0	0	0	1	3	2	3	3	4	4	6	9	10	14	34	43	14
0	0	0	0	0	0	0	0	1	2	1	1	2	2	3	3	6	12	14	4

3.2.3.2 Textural Features

Notations:

$p(i, j)$ — (i, j)th entry in the normalized gray level co-occurrence matrix

$p_x(i)$ — ith entry in the marginal probability matrix obtained by summing the rows of $p(i, j)$

N_g — number of distinct gray levels in the quantized image.

1) Angular Second Moment:

$$f_1 = \sum_i \sum_j (p(i, j)^2)$$

Angular Second Moment measure the smoothness of the image. There are two cases, If all pixels have same gray level I=k,

$p(k,k) = 1$ if $(i = j)$ and $p(i,j) = 0$ if otherwise.

$ASM = 1$

If all pixels have different gray level,

$$p(i,j) = {}^{1}/_{R} \qquad \& \qquad ASM = {}^{1}/_{R}$$

ASM value of Fig 3.5 image is: 0.00528

2) Contrast:

$$f_2 = \sum_{n=0}^{N_g-1} n^2 \left\{ \sum_{i=1}^{N_g} \sum_{j=1}^{N_g} p(i,j) \right\}$$

Contrast measures the image contrast (locally gray level variations). The term n^2 is used to take of the largest contrast value.

Contrast value of Fig 3.5 image is: 26.67

3) Correlation:

$$f_3 = \frac{\sum_i \sum_j (ij)\, p(i,j) - \mu_x \mu_y}{\sigma_x \sigma_y}$$

Correlation measures how the pixels are correlated with each other. Where μ_x, μ_y are the standard deviation and σ_x, σ_y are means of p_x, p_y

Correlation value of Fig 3.5 image is: 0.557

4) Sum of squares: Variance

$$f_4 = \sum_i \sum_j (i - \mu)^2\, p(i,j)$$

Sum of squares value of Fig 3.5 image is: 128.231

5) Inverse Difference Moment(Homogeneity)

$$f_5 = \sum_i \sum_j \frac{1}{1+(i-j)^2} p(i,j)$$

Inverse Difference Moment takes care of low contrast images. It takes care of low contrast images because of the inverse $(i-j)^2$.

Homogeneity value of Fig 3.5 image is: 0.30597

6) Sum Average

$$f_6 = \sum_{i=2}^{2N_g} i p_{x+y}(i)$$

Sum Average value of Fig 3.5 image is: 19.807

7) Sum Variance

$$f_6 = \sum_{i=2}^{2N_g} (i - f_8)^2 p_{x+y}(i)$$

Sum Variance value of Fig 3.5 image is: 426.844

8) Sum Entropy

$$f_8 = -\sum_{i=2}^{2N_g} p_{x+y}(i) log\{p_{x+y}(i)\}$$

Sum Entropy value of Fig 3.5 image is: 1.554

9) Entropy

$$f_9 = -\sum_i \sum_j p(i,j) log\{p(i,j)\}$$

Entropy takes low values for smooth images. It measures the randomness.

Entropy value of Fig 3.5 image is: 2.4052

10) Difference Variance

$$f_{10} = variance\ of\ p_{x-y}$$

Difference Variance value of Fig 3.5 image is: 0.00436

11) Difference Entropy

$$f_{11} = -\sum_{i=0}^{N_g-1} p_{x-y}(i)\ log\{p_{x-y}(i)\}$$

Difference Entropy value of Fig 3.5 image is: 0.99

12) &

13) Information Measure of Correction

$$f_{12} = \frac{HXY - HXY1}{max\{HX, HY\}}$$

$$f_{13} = (1 - exp[-2.0(HXY2 - HXY)])^{1/2}$$

$$HXY = -\sum_{i} \sum_{j} p(i) log\{p(i)\}$$

Since some of the probabilities becomes zero and $log(0)$ is very high so arbitrary small positive constant is added to avoid the infinite number.

Where, HX and HY are entropies of p_x and p_y and

$$HXY1 = -\sum_{i} \sum_{j} p(i,j) log\{p_x(i)p_y(j)\}$$

$$HXY2 = -\sum_{i} \sum_{j} p_x(i)p_y(j) log\{p_x(i)p_y(j)\}$$

Information Measure of Correlation 1 value of Fig 3.5 image is: -0.1358

Information Measure of Correlation 2 value of Fig 3.5 image is: 0.4915

14) Maximal Correction Coefficient(Energy)

$$f_{14} = (second\ largest\ eigenvalue\ of\ Q)^{1/2}$$

Where,

$$Q(i,j) = \sum_{k} \frac{p(i,k)p(j,k)}{p_x(i)p_y(k)}$$

Energy value of Fig 3.5 image is: 0.0711

3.2.3.3 Feature Selection

Selecting correct features is an important issue for the system. Not all the features are suitable for classifying the different classes. In our study around all 14 features we found only 5 of them are appropriate. For selecting those features a method called Subset Choosing method is used.

The methodology is:

```
ans = 0.0
subset = 1
for i in range (1, (1<<15)):
    ret = fun(i)
    if (ret>ans):
        ans = ret
        subset = i
```

There are 2^{15} possible subsets among them the subset which is the possible best subset. So, the selected features are:

1) Homogeneity
2) Angular Second Moment(ASM)
3) Energy
4) Information Measure of Correlation 1
5) Information Measure of Correlation 2

3.2.4 Classification

After extracting features from the images, now a classifier is needed to classify the images. Here an artificial neural network with three hidden layer is used as a classifier. In the classifier two steps are followed. If an image does not pass leaf color analysis, classifier algorithm will be used to detect the diseases.

3.2.4.1 Leaf Color Analysis

First the whole image is scanned through and calculate the minimum and maximum value for each channel. The RGB calculation will be passed.

$$93 \leq R_{min} \leq 211 \;\; \& \;\; 93 \leq R_{max} \leq 211$$

$$142 \leq G_{min} \leq 222 \;\; \& \;\; 142 \leq G_{max} \leq 222$$

$$64 \leq B_{min} \leq 155 \;\; \& \;\; 64 \leq B_{max} \leq 155$$

If an image passes all the above conditions, then the image is normal leaf image. Otherwise, it is an affected image.

To calculate those values first we took some normal paddy leaf images and calculate the histogram values. From those histogram values the minimum and maximum values were calculated.

3.2.4.2 Artificial Neural Network

All the selected features are in the input of an Artificial Neural Network and the output is used in classification. The equation which is minimized is

$$WX + B$$

Where,

$$W = weights$$

$$B = biases$$

$$X = Input\ Features$$

The above equation is minimized such that the cost function produces as small error as possible.

In an artificial neural network, there are some nodes and the nodes are connected by some edges. The edges represent some arbitrary values and the nodes represent some activation functions. In our proposed methodology, Tensor Flow is used as the deep learning framework. TensorFlow is a typical computational graph that can be executed much more efficiently then if the same calculations were to be performed directly in python. TensorFlow can be more efficient then NumPy because TensorFlow knows the entire computation graph that must be executed, while NumPy knows only the computation of a single mathematical operation at a time. TensorFlow can also automatically calculate the gradients that are needed to optimize the variables of the graph so as to make the model perform better. This is because the graph is a combination of simple mathematical equations. So, the gradient of the entire graph can be calculated through iterative chain rule method.

Fig 3.6 shows our used system, where the network architecture 15 – 50 – 50 – 50 – 4 nodes in 1 input, 3 hidden and 1 output layer respectively. It is a fully connected artificial neural network. Each hidden layer has 50 nodes which gives better result in our research. The hidden nodes are chosen by comparative study.

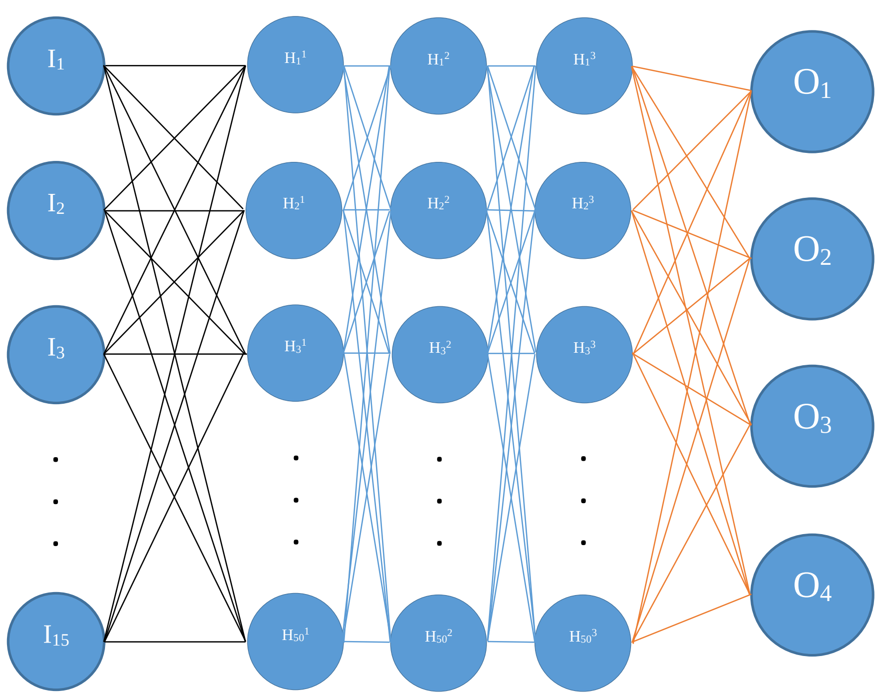


Fig 3.6: Proposed Artificial Neural Network

Chapter 4

Experimental Analysis

4.1 Introduction

There are 220 samples of paddy image used as sample data in the testing phase of this development. The paddy images samples had gone through the phases as discuss in the chapter 3. This chapter will briefly describe about the output result of each phases.

4.2 Experimental Setup

The experiments and related analysis are done in this chapter. The experiments and analysis processes are done on a computer with Core-I5 processor having 4 cores with each core having 2.5GHz Speed. Also the system had 4GB of RAM, and 1GB of internal intel HD video memory. For software, PyCharm Community Edition 2017.1 is used and program is done with python language with OpenCV and deep learning framework, TensorFlow. OpenCV (Open Source Computer Vision) is a library of programming functions mainly aimed at real-time computer vision. The reason for using OpenCV because it gives easy functionality to do different processes without going into implementations. Moreover, it gives the benefit to use GPU by which processes can be made faster than using CPU only for computation works. TensorFlow™ is an open source software library for numerical computation using data flow graphs. TensorFlow was originally developed by researchers and engineers working on the Google Brain Team within Google's Machine Intelligence research organization for the purposes of conducting machine learning and deep artificial neural networks research, but the system is general enough to be applicable in a wide variety of other domains as well.

4.3 Performance Measures – Definition

Confusion Matrix – Confusion matrix is to summarize the performance of a classification technique. Classification accuracy can be misleading for an unequal number of observations in each class or more than two classes in dataset. Calculating confusion matrix gives better idea of what classification model is getting right and what types of errors it is making.

True Positives (TP) - These are the correctly predicted positive values which means that the value of actual class is yes and the value of predicted class is also yes. E.g. if actual class value indicates that this leaf is Paddy Blast and predicted class tells the same thing.

True Negatives (TN) - These are the correctly predicted negative values which means that the value of actual class is no and value of predicted class is also no. E.g. if actual class says this leaf isn't Paddy Blast and predicted class tells you the same thing.

False positives and false negatives, these values occur when your actual class contradicts with the predicted class.

False Positives (FP) – When actual class is no and predicted class is yes. E.g. if actual class says this isn't Paddy Blast but predicted class tells that this leaf is Paddy Blast.

False Negatives (FN) – When actual class is yes but predicted class in no. E.g. if actual class value indicates that this is Paddy Blast and predicted class tells that leaf is Other.

Accuracy – Accuracy is the most intuitive performance measure and it is simply a ratio of correctly predicted observation to the total observations. One may think that, if we have high accuracy then our model is best. Yes, accuracy is a great measure but only when you have symmetric datasets where values of false positive and false negatives are almost same. Therefore, one has to look at other parameters to evaluate the performance of the model.

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN}$$

Precision – Precision is the ratio of correctly predicted positive observations to the total predicted positive observations. The question that this metric answer is of all passengers that labeled as survived, how many actually survived? High precision relates to the low false positive rate.

$$Precision = \frac{TP}{TP + FP}$$

Recall – Recall is the ratio of correctly predicted positive observations to the all observations in actual class - yes. The question recall answers, is: Of all the leaves that truly are in this particular class, how many did we label?

$$Recall = \frac{TP}{TP + FN}$$

4.4 Result Analysis

Let's indicate Paddy Blast, Brown Spot, Narrow Brown Spot and Other, 4 classes as PB, BS, NBS and O respectively. And predicted class and actual class are on column and row. So, the table of confusion matrix is as below:

Table 4.1: Confusion matrix of proposed system

		Predicted Class			
		PB	BS	NBS	O
Actual Class	PB	17	1	0	1
	BS	0	13	0	1
	NBS	0	1	7	1
	O	1	0	1	3

4.4.1 Random Test(k-fold)

Random test is used to determine the robustness of a system. Here all the data are taken and from the data randomly some data are selected for testing and some data are selected for training. Here k-fold means, the whole data are divided into k-fold, then each fold is used as testing data and rest of the data are used as training data. And there are n-number of runs.

5-fold is used within 3 distinct run with our collected data and the resulted values are showed in the table 4.1 with the accuracy, precision and recall individually, as average and total.

Table 4.2: Accuracy, Precision and Recall with k-fold(5-fold) method

Run	Fold	Fold Accuracy(%)	Fold Precision(%)	Fold Recall(%)	Run Accuracy(%)	Run Precision(%)	Run Recall(%)	Total Accuracy(%)	Total Precision(%)	Total Recall(%)
1	1	86.4	90.3	73.8	86.38	87.52	80.02	85.69	83.36	76.17
	2	84.1	85.6	78.3						
	3	81.8	78.9	83.3						
	4	93.2	94.2	88.0						
	5	86.4	88.6	76.7						
2	1	84.1	73.4	75.0	84.38	78.72	73.28			
	2	86.4	66.9	70.3						
	3	90.0	92.4	80.5						
	4	77.3	78.3	63.6						
	5	84.1	82.0	77.0						
3	1	79.5	76.7	68.9	86.32	84.74	75.2			
	2	86.3	89.6	74.4						
	3	88.6	89.8	73.9						
	4	88.6	91.3	81.2						
	5	88.6	76.3	77.6						

4.4.2 Classification Result

4.4.2.1 Paddy Blast

Total 87 leaves which are affected by Paddy Blast disease are used for experiment. In those leaves, 10 images were misclassified and the other 77 images are correctly classified.

Examples:

Correctly classified

Fig 4.1: Correctly classified 4 Paddy Blast images

Incorrectly classified

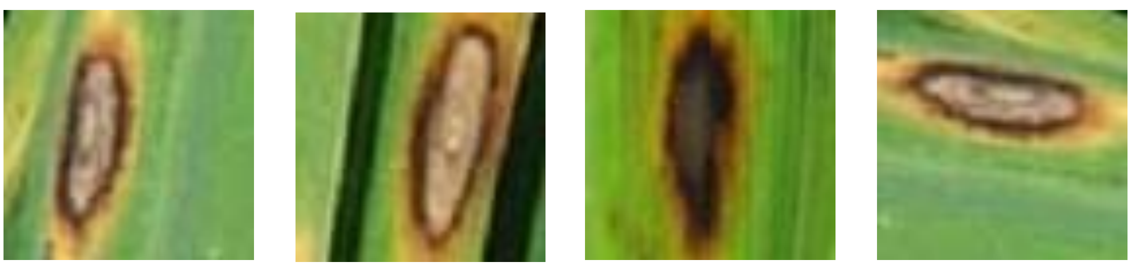


Fig 4.2: Incorrectly classified 4 Paddy Blast images

Accuracy of Paddy Blast = 88.51%

4.4.2.2 Brown Spot

Total 56 leaves which are affected by Paddy Blast disease are used for experiment. In those leaves, 14 images were misclassified and the other 42 images are correctly classified.

Examples:

Correctly classified

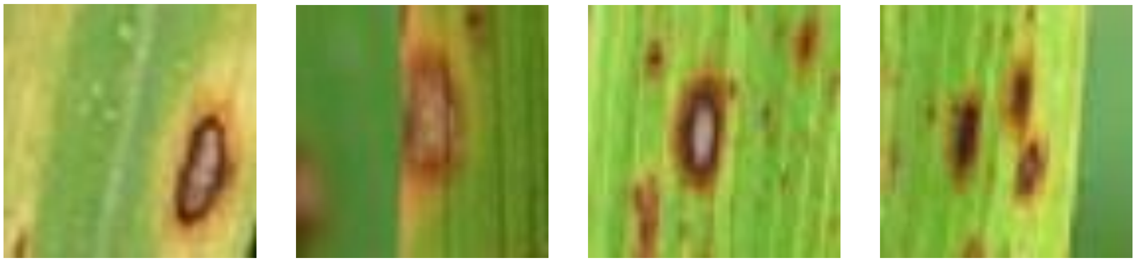


Fig 4.3: Correctly classified 4 Brown Spot images

Incorrectly classified

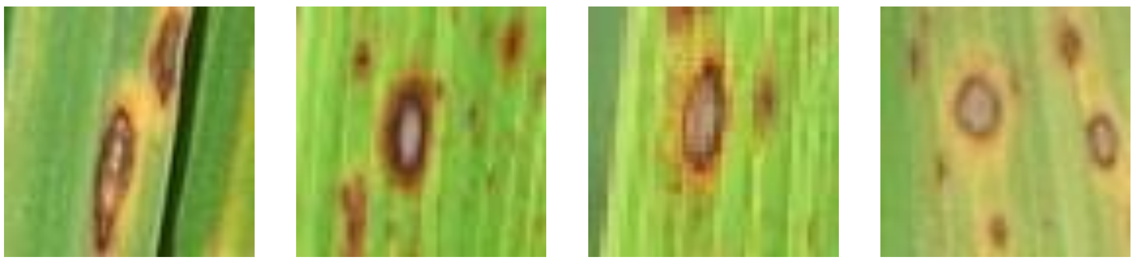


Fig 4.4: Incorrectly classified 4 Brown Spot images

Accuracy of Brown Spot = 75.00%

4.4.2.3 Narrow Brown Spot

Total 33 leaves which are affected by Paddy Blast disease are used for experiment. In those leaves, 4 images were misclassified and the other 29 images are correctly classified.

Examples:

Correctly classified

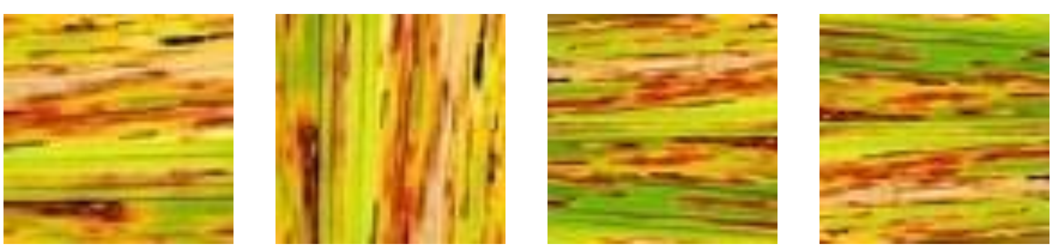


Fig 4.5: Correctly classified 4 Narrow Brown Spot images

Incorrectly classified

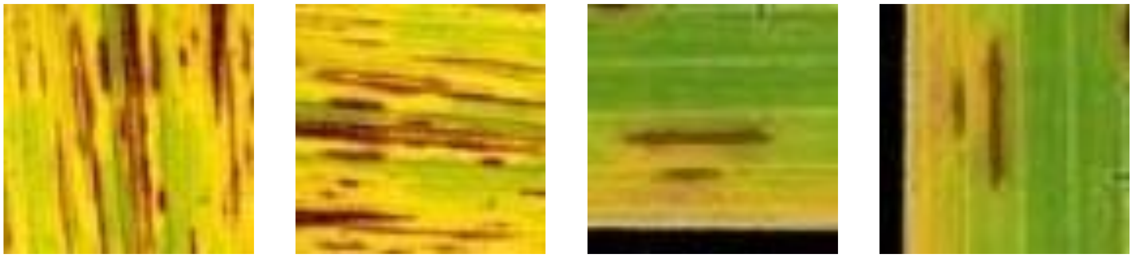


Fig 4.6: Incorrectly classified 4 Narrow Brown Spot images

Accuracy of Narrow Brown Spot = 87.87%

4.5 Conclusion

In this chapter, a system for diagnosis the paddy disease has been developed using the PyCharm application. The image processing techniques are applied to find the features and neural network is used to build a trained model. In testing phase, the model is tested with the test images. Before testing each image is passed through leaf color analysis to find out the normal leaf images. The accuracy of testing around 85 percent which is good for practical use.

Chapter 5

Conclusion

5.1 Summary

A system for diagnosis the paddy disease has been developed using the PyCharm application. The image processing techniques is applied to improve and enhance the image to a better quality. Besides, the artificial neural network and leaf color analysis is used to classify the paddy diseases which are paddy blast, brown spot disease, narrow brown spot disease and normal paddy leaf. The methodology involves image collection, image processing, analysis and classification of the paddy disease. All the paddy sample will be passing through the leaf color analysis before it proceeds to the artificial neural network. If the sample is in the range of normal paddy RGB, then it is automatically classifying as Normal paddy leaf and otherwise all the disease affected sample will be passed through image processing to get the features which will be forwarded to the artificial neural network for training and testing. Consequently, by employing the artificial neural network technique, the paddy diseases are recognized about 86 percent accuracy rates. This study has a very great potential to be further improved in the future.

5.2 Limitation

There are not many impediments in our proposed strategy. One of them is accuracy percentage. The accuracy is at least 86 percent that is not high enough. For better result this number should be risen. There may be also some implementation bug to fix. Best features selection is computationally hard and lengthy process, so, we could not check all possible subset of the features. We only checked some random features based on previous best features. There is a possibility of losing some best features. And also it takes a lot of time, so, our implementation is little slower. Thus it can't be applied on real time for now.

5.3 Future Scope of Work

Even though, we've tried our best to get the desired output and the accuracy value is quite good for this method, yet there is still room for improvement as long as it is not close to 100 percent.

- By using more efficient way to choose best features, this method can be made to work in real time
- Can be implemented in LBP which could rise the accuracy
- Can be implement in mobile application thus farmers can easily detect the disease before it is too late.

References

[1] M. R. Tejonidhi, B. R. Nanjesh, J. G. Math and A. G. D'sa, "Plant disease analysis using histogram matching based on Bhattacharya's distance calculation," *2016 International Conference on Electrical, Electronics, and Optimization Techniques (ICEEOT)*, Chennai, 2016, pp. 1546-1549.

[2] S. Mutalib, M. H. Abdullah, S. Abdul-Rahman and Z. A. Aziz, "A brief study on paddy applications with image processing and proposed architecture," *2016 IEEE Conference on Systems, Process and Control (ICSPC)*, Bandar Hilir, 2016, pp. 124-129.

[3] R. P. Narmadha and G. Arulvadivu, "Detection and measurement of paddy leaf disease symptoms using image processing," *2017 International Conference on Computer Communication and Informatics (ICCCI)*, Coimbatore, 2017, pp. 1-4.

[4] N. N. Kurniawati, S. N. H. S. Abdullah, S. Abdullah and S. Abdullah, "Investigation on Image Processing Techniques for Diagnosing Paddy Diseases," *2009 International Conference of Soft Computing and Pattern Recognition*, Malacca, 2009, pp. 272-277.

[5] R. M. Prakash, G. P. Saraswathy, G. Ramalakshmi, K. H. Mangaleswari and T. Kaviya, "Detection of leaf diseases and classification using digital image processing," *2017 International Conference on Innovations in Information, Embedded and Communication Systems (ICIIECS)*, Coimbatore, 2017, pp. 1-4.

[6] P. Revathi and M. Hemalatha, "Classification of cotton leaf spot diseases using image processing edge detection techniques," *2012 International Conference on Emerging Trends in Science, Engineering and Technology (INCOSET)*, Tiruchirappalli, Tamilnadu, India, 2012, pp. 169-173.

[7] D. Al Bashish, M. Braik and S. Bani-Ahmad, "A framework for detection and classification of plant leaf and stem diseases," *2010 International Conference on Signal and Image Processing*, Chennai, 2010, pp. 113-118.

[8] A. N. I. Masazhar and M. M. Kamal, "Digital image processing technique for palm oil leaf disease detection using multiclass SVM classifier," *2017 IEEE 4th International*

Conference on Smart Instrumentation, Measurement and Application (ICSIMA), Putrajaya, 2017, pp. 1-6.

[9] J. Shijie, J. Peiyi, H. Siping and s. Haibo, "Automatic detection of tomato diseases and pests based on leaf images," *2017 Chinese Automation Congress (CAC)*, Jinan, 2017, pp. 2537-2510.

[10] A. A. Sarangdhar and V. R. Pawar, "Machine learning regression technique for cotton leaf disease detection and controlling using IoT," *2017 International conference of Electronics, Communication and Aerospace Technology (ICECA)*, Coimbatore, 2017, pp. 449-454.

[11] C. G. Dhaware and K. H. Wanjale, "A modern approach for plant leaf disease classification which depends on leaf image processing," *2017 International Conference on Computer Communication and Informatics (ICCCI)*, Coimbatore, 2017, pp. 1-4.

[12] B. Sandika, S. Avil, S. Sanat and P. Srinivasu, "Random forest based classification of diseases in grapes from images captured in uncontrolled environments," *2016 IEEE 13th International Conference on Signal Processing (ICSP)*, Chengdu, 2016, pp. 1775-1780.

[13] N. Satya Priya, E. Nivetha and Rashmita Khilar., "Efficient knowledge based system for leaf disease detection and classification", *International Journal of Advance Research in Science and Engineering (IJARSE),* Vol. No.4, Special Issue (01), March 2015, ISSN-2319-8354(E).

[14] Santanu Phadikar and Jaya Sil, "Rice Disease Identification using Pattern Recognition Techniques", *Proceedings of 11th International Conference on Computer and Information Technology (ICCIT 2008)*, 25-27 December, 2008, Khulna, Bangladesh.

[15] Shoumi Paul, Reema D. Sharma, "Plant Disease Detection Using Image Processing Technique", *International Journal of Innovative Research in Electrical, Electronics, Instrumentation and Control Engineering(IJIREEICE),* Vol. 4, Issue 9, September 2016, ISSN 2321 – 5526.

[16] R. M. Haralick, K. Shanmugam and et., "Textural Features for Image Classification," in *IEEE Transactions on Systems, Man, and Cybernetics*, vol. SMC-3, no. 6, pp. 610-621, Nov. 1973.

[17] J. B. Cunha, "Application of image processing techniques in the characterization of plant leafs," *2003 IEEE International Symposium on Industrial Electronics (Cat. No. 03TH8692)*, 2003, pp. 612-616 vol. 1.

[18] S. Dave and K. Runtz, "Image processing methods for identifying species of plants," *IEEE WESCANEX 95. Communications, Power, and Computing. Conference Proceedings*, Winnipeg, Man., 1995, pp. 403-408 vol.2.