# Study and Development of Automated System to Identify Leaf Spot Rice Disease Using Image Processing Techniques

**Shashank Chaudhary[1*], Dr.Upendra kumar[2], Dr. Shwetank Parihar[3]**

[1] Department of Computer Science, Research Scholar, Dr. Abdul Kalam Technical University Lucknow 226031 UP India, s_sha_123@yahoo.co.in

[2]Assistant Professor, Department of Computer Science Institute of Engineering Technology Lucknow UP India 226027 UP India, ukumar@ietlucknow.ac.in

[3] Assistant Registrar MNNIT Allahabad UP India, shwetankp@gmail.com

## Abstract

Smart farming is a concept that involves the use of high-quality algorithms to make agriculture more productive. The process of machine learning is a subfield of Artificial Intelligence (AI) that enables machines to learn and improve their skills without being programmed. In this work a methodology was developed using Gray-Level co-occurrence matrix (GLCM) for feature extraction and Support Vector Machine (SVM) for classification, to detect the disease with desired level of accuracy. This tool was developed in MATLAB for the precision agriculture to detect the leaf spot disease. Approximately 1000 images data sets were used to accurately (more than 90%) classify the rice disease.

**Keywords-** Image processing, AI, GLCM, SVM.

## Introduction

In India, agriculture is a vital aspect of economy. Unfortunately due to natural disasters & diseases crops are damaged and for the increasing demand for high-quality and safety-assured agricultural products, various methods have been devised to evaluate the quality of these products. The use of manual techniques and expensive equipment for identifying plant diseases has been criticized as inefficient and prolonged. Due to the increasing importance of identifying plant diseases, the use of precise and automatic methods has been widely considered [1]. This is why it is important to develop an automated system that can detect diseases in plants. Image processing is a technology that is helping in the identification of various diseases that can damage a plant. This process can detect diseases that are already on the plant.

Rice crop is one of the most susceptible crops to various diseases. This plant is mainly prone to bacterial leaf blight, brown spot, and leaf smut. It is estimated that around 30% to 40% of crops are lost annually due to diseases. The economic impacts of these losses are significant. Various studies have been performed to identify the diseases that cause these losses. This type of disease management practice requires a team of experts to thoroughly diagnose and control the spread of the disease. Through a computer vision system, it can detect the presence of the disease in the plant. The performance of different image processing techniques is often used to improve the classification of images of rice plants [2]. The main challenge in controlling diseases is to prevent their spread and minimize the loss of crop. The diagnostic process involves both visual and scientific methods.

Automatic techniques using image processing were used in the field for plant protection to improve the quality of the product. This paper aims to develop a tool that can detect the presence of a leaf spot disease in the rice plants. It can then inform the user about the steps needed to prevent or cure the disease.

Image processing techniques such as image acquisition, pre-processing, feature extraction and classification are utilized in order to improve the accuracy of the system's decisions.

Due to the significant variability of global crop and vegetation conditions, monitoring these conditions is highly relevant for food insecure regions. Data collected by remote sensing tools can help in this monitoring. The SPIRITS (Software for the Processing and Interpretation of Remotely sensed Image Time Series) software is a set of tools that help environmental monitoring professionals collect and interpret data related to crop production. It provides a variety of tools to extract vegetation indicators from images. SPIRITS is a flexible and integrated analysis environment that can be used for various tasks. [3].

Diseases can be found in different parts of a plant, such as the root, leaf, and kernel. The classification and identification of these diseases can be performed through various tools. Camargo et al. (2009) presented an image processing technique to identify the visual symptoms of plant diseases. They used an RGB color space of a leaf and converted it into H, I3a, and I3b conversions. By analyzing the scattering of intensities in a histogram, they were able to identify a disease segment even when its level of intensities varies widely [4]. The method of disease identification for cotton leaf was proposed. The process involves capturing and filtering the images, and then highlighting the edges with a histogram process. The enhanced images are then converted into three different color models: the RGB, the HIS, and the YCbCr. The results of the tests revealed that the YCbCr color format was the best fit for identifying the injured leaf image [5]. A low-resolution image enhancement algorithm is used to improve the quality of the data. Then, a method for image segmentation is presented, which helps identify the symptoms of the disease. The leaf image is also segmented to classify the diseases. This method works by using the statistical model GLCM and the different kernel method [6]. Image segmentation is an important step in image processing. It involves identifying and processing images based on their color. The use of genetic algorithms can help minimize the complexity of the problem and improve its quality [7]. An image processing technique was implemented to identify the leaf spot disease. It was based on the concept of Otsu thresholding. The affected regions were then segmented according to the H element. The Sobel operator was then used to identify the areas with the most leaf spots. The leaf areas and the diseased region were then assessed to get an accurate result [8]. When you have a large data set, feature extraction is a technique that helps reduce the amount of data that needs to be collected. It also helps in building a robust machine learning model. Author studied the effectiveness of the spatial gray level dependence method (SGDM) in detecting diseases in grape leaves. They used different leaf subclasses for their study. The S and I components were discarded since they did not provide useful information. The green and blue pixel values were masked using Otsu method. Then, the features based on the texture and color was extracted. The authors noted that the proposed method has minimal computational efforts and can be improved with the use of hybrid algorithms [9]. Zhenghong et al. (2014) propose a method for extracting specularity-invariant crop data using the MRF random field. This method is based on a rule that gradually changes the intensity of the highlight areas as it moves between the non-highlight areas. The probability of each node in a label field is computed by the Belief Propagation algorithm. This method is more accurate than the other extraction methods [10].

An artificial neural network was developed to identify Phalaenopsis seedling diseases using texture and color analysis. It was tested against a set of lesions with varying backgrounds and textures. Gray Level Co-occurrence matrix (GLCM) was used to evaluate texture features of the lesion area. The system then learns how to classify the lesions into different categories. It could also detect the presence of infected areas in the lesions [11]. A method were developed a method to extract tobacco leaf lesions using neural networks. The method involves the use of contrast stretching and morphological operations to segment the lesions. The

Probabilistic Neural Network is used to classify tobacco leaf diseases into anthracnose and frog-eye spots [12]. A method was proposed to identify two diseases known as powdery mildew and downy mildew. The diseases are commonly found in the cucumber leaf. The proposed algorithm is performed by SVM through a series of kernel functions, such as the linear kernel function and the radial basis kernel function [13]. The authors noted that the combination of the shape feature and the texture feature can provide good performance. The objective of this paper was to develop an algorithm for identifying cotton leaf diseases using an existing Neuro-fuzzy inference system. The diseases that were analyzed were Myrothecium, Alternaria, and Bacterial Blight [14].

## Methodology

To design an automatic rice plant disease detection tool, there should be a database to store disease information and find out the affected diseases. So need to diagnose different type's rice plant diseases. Detection of rice plant diseases involves two phases: image processing and machine learning.

The first phase collects images from the farm and then categorizes the disease based on its features. Image processing includes various steps such as background removal, noise removal, and image segmentation. Machine learning focuses on feature selection and classification of the rice diseases. This framework shown in the figure 1 describes the various steps and procedures that are used in order to collect and pre-process the data related to the problem. It also includes a set of tools and techniques that are used for analyzing the data.

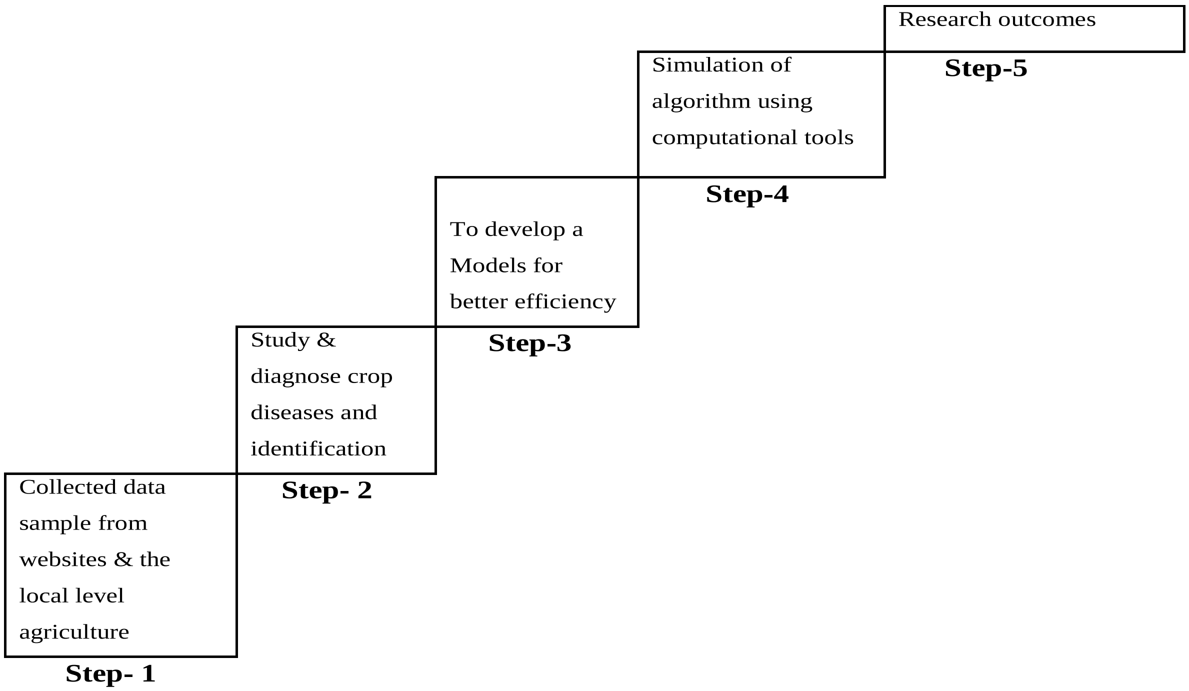



**Figure 1 Procedural framework to identify rice leaf disease**

Then segmentations can be performing on the processed rice leave data to identify the latent characteristics of the plant. We then extract the region of interest that only has visible diseases symptoms. For a particular disease, the regions have specific texture and color properties. The method used to extract the feature sets is called the GLCM Method. This procedure takes into account the various features of an image, such as its color and texture. The GLCM method is a statistical procedure that describes the shape of a region. It allows us to detect and distinguish the presence of diseases. By training a SVM classifier with features that allow it to detect and distinguish diseases.

## Image processing Approach

Image processing is a technology that is helping in the identification of various diseases that can damage a plant. This process can detect diseases that are already on the plant. The process of detecting the plant disease based on the images is shown in Figure 2. The process of getting images of infected leaves is referred as image acquisition.


**Figure 2 Image processing steps**

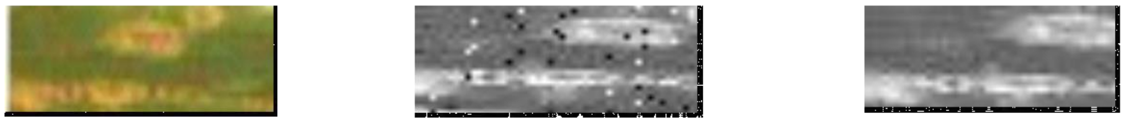
## Data collection

An appropriate dataset is required at all stages of object recognition research, starting from the training phase to evaluating the performance of recognition algorithms. A total of 1000 images are collected from different sources, such as the Plant Village and Google websites, including different periods of occurrence of rice leaf diseases, sample of rice leaf disease shown in figure 3. These images are divided into 2 different categories healthy leaves and diseased leaves.

**Figure 3 Data sample of rice leaf disease**

## Pre-processing

Noise reduction techniques are used in order to remove unwanted noise from an image. Image clipping is also conducted to get the desired effect. Image enhancement techniques are also used to improve the contrast. The histogram equalization technique is also used to distribute the intensities of an image shown in figure 4.



color Image Noisy Image Noise free Image

**Figure 4 Noise removals in pre-processing steps**

## Region of Interest (ROI)

Almost 1000 images of rice leaves are selected as dataset for training and testing purpose. Firstly, preprocess the selected image shown in figure 5 and then perform the segmentations. We mask out the background as well as the green region of the leaves. Thus we extract our region of interest (ROI) that only contains visible diseases symptoms shown in figure 6.

RGB Image HSI Image HSI Equalized Image

**Figure 5 HSI Color equalization**

Here the given image is segmented in the 4 different clusters

Image in cluster1 Image in cluster 2 Image in cluster 3 Image in cluster 4

**Figure 6 Region of interest (ROI) Image segmentation**

## Feature extraction

Feature extraction is a method of extracting visual content from an image. Usually, it involves extracting the various features of an image, such as shape, color, and texture. This technique helps minimize the amount of processing time needed to get the best feature from a large data set. It does so by combining the various variables into features that can describe the data set. Feature extraction is also an important step in the process of classifying an image.

This paper discussed how these features can be used to identify the plant leaf of a classification system. It presents an approach that combines the gray level co-occurrence matrix with statistical texture features to extract second order features of an image. Statistical texture analysis is a process that involves analyzing the distribution of intensities in an image. The GLCM is a method that can extract second-order statistical textures. A number of features can be deriving using several statistics, such as the contrast, correlation, energy, homogeneity, and Entropy, was computed using GLCM [15].

## Disease classification

The SVM classification tool is used to identify diseases that can be easily detect early growth of the disease. It is a powerful machine learning tool that can perform various tasks related to binary data classification. The concepts of support vector machines were proposed to classify different types of plant diseases and their effects on the plant leaf.

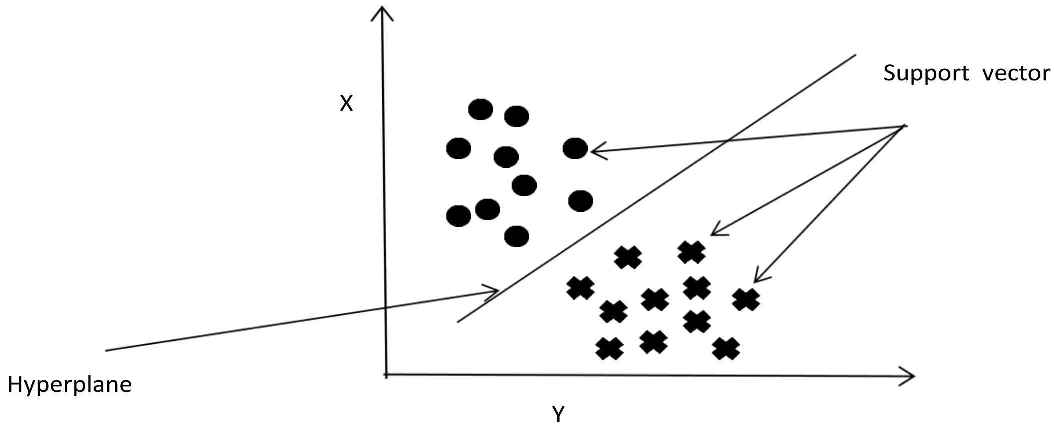


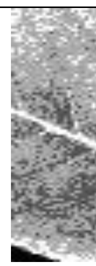
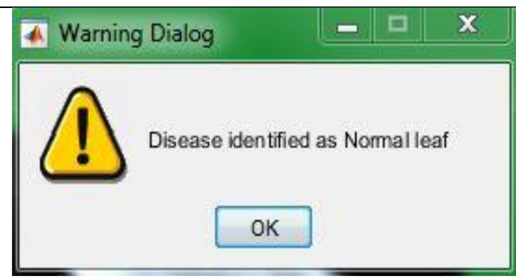

**Figure 7 Disease classification using SVM**

This step can identify objects in predefined classes by analyzing the patterns shown in the figure 7. It uses a statistical machine learning algorithm to classify the objects. It does so by creating a linear boundary between the two classes. The features that are used in the SVM are then fed to it. The results are far better than those derived from traditional methods.

## Result

The GLCM method can be used to extract various statistical texture from an image. It was used to extract the features of an image. Then recognize detected portion of leaf classified through SVM. Results successfully validate the healthy and unhealthy leave images.

**Table 1 sample of healthy rice leaf disease and leaf spot rice disease detection**

| S.no | Leaf disease name | Leaf sample image | Identified disease | Result derived from the application tool developed |
|---|---|---|---|---|
| 1 | Leaf spot |  |  |  |
| 3 | Healthy leaf |  |  |  |

The accuracy of an algorithm is a crucial stage in the development of research model. This test involves measuring the performance of a SVM algorithm with the help of GLCM feature extraction method.

A classification algorithm's performance is evaluated using a confusion matrix shown in the table 2. This table summarizes the basic characteristics to define the measurement metrics and various aspects of an algorithm's classification. Accuracy is to represent ratio of correctly classified images to the total number healthy classified images and total number unhealthy images.

The following equation can be used to determine the accuracy of an algorithm model.

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

Table 2 Basic characteristics to define the measurement matrix

| | |
|---|---|
| **TP (True Positive)** | Total number of unhealthy leaves who have been properly Classified to have infectious, meaning they have the disease. |
| **TN(True Negative)** | Total number of correctly classified leaves who are healthy. |
| **FP (False Positive)** | Its represents the number of misclassified but actually they are healthy. |
| **FN (False Negative)** | Its represents the number of unhealthy as healthy but actually they are suffering from the leaf spot disease. |

Almost 1000 images were collected from different sources, such as the Plant Village and Google depository; it consist healthy images data set as well as leaf spot disease data set. To classify the disease data classification techniques is used and its accuracy can be achieved using above equation shown in the table 3.The ratio of correctly classified images to the total number of images can be obtained using confusion matrix.

Table 3 healthy and disease leaf count using performance matrix

| **Rice leaf Images** | **Total no of images** | **Total Correctly classify images** |
|---|---|---|
| Healthy | 100 | 100 |
| Disease | 900 | 810 |

Here in the Table 4 training and testing ratio were used to simulate the model and performance measure gives 91% accuracy using SVM classification Approach.

Table 4 Accuracy analysis for Leaf Spot rice disease using GLCM & SVM

| **S.NO** | **Training & Testing** | **Healthy and Unhealthy (Leaf spot) leaves** | | | |
|---|---|---|---|---|---|
| | | **Precision** | **Recall** | **F-1 Score** | **Accuracy** |
| 1 | 90:10 | 0.90 | 0.90 | 0.90 | 0.911881 |
| 2 | 85:15 | 0.91 | 0.91 | 0.91 | 0.91841 |
| 3 | 80:20 | 0.9 | 0.90 | 0.90 | 0.901786 |
| 4 | 75:25 | 0.91 | 0.91 | 0.91 | 0.913479 |
| 5 | 70:30 | 0.93 | 0.93 | 0.93 | 0.934768 |
| 6 | 65:35 | 0.91 | 0.91 | 0.91 | 0.91409 |
| 7 | 60:40 | 0.88 | 0.88 | 0.88 | 0.888819 |
| 8 | 55:45 | 0.87 | 0.87 | 0.87 | 0.87022 |
| 9 | 50:50 | 0.89 | 0.89 | 0.89 | 0.896341 |
| **Average** | | 0.90 | 0.90 | 0.90 | 0.91 |

The graph shown in the figure 8 finds accuracy analysis for the given model and simulates the same.

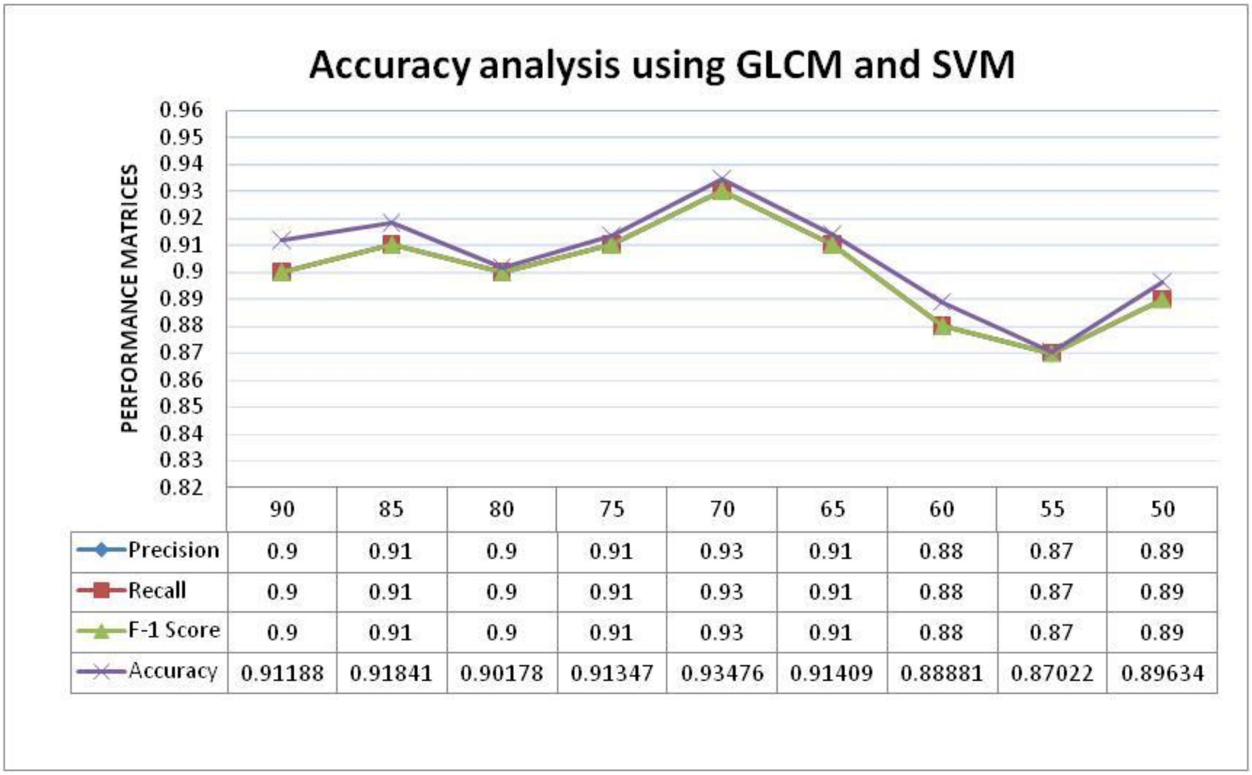



Figure 8 Accuracy analysis using GLCM and SVM

Here this graph simulate the propose work to identify leaf spot rice disease detection using GLCM feature extraction method and SVM classification techniques and finds the 91% of accuracy.

## Conclusion

In this work approximately 1000 images dataset of healthy leaf images and leaf spot rice disease were tested. The tool was developed in the MATLAB using GLCM feature extraction and SVM classifier to automate the system. This automatic technique gives almost 91% of accuracy in detection of the rice disease.

## Future work

The same work can be further enhanced by treating the images through multiple intensity based multi-fractal fractal dimension (ILMFD) method by splicing number of images having different intensity value to develop automated system to identify crop disease detection.

**References**

[1] Flood, J., 2010. The importance of plant health to food security. Food Security, 2(3), pp.215-231.

[2] Chaerle, L., Leinonen, I., Jones, H. and Van Der Straeten, D., 2006. Monitoring and screening plant populations with combined thermal and chlorophyll fluorescence imaging. Journal of Experimental Botany, 58(4), pp.773-784.

[3] Eerens, H., Haesen, D., Rembold, F., Urbano, F., Tote, C. and Bydekerke, L., 2014. Image time series processing for agriculture monitoring. Environmental Modelling & Software, 53, pp.154-162.

[4] Camargo, A. and Smith, J., 2009. An image-processing based algorithm to automatically identify plant disease visual symptoms. Biosystems Engineering, 102(1), pp.9-21.

[5] He, Q., Ma, B., Qu, D., Zhang, Q., Hou, X. and Zhao, J., 2013. Cotton Pests and Diseases Detection Based on Image Processing. TELKOMNIKA Indonesian Journal of Electrical Engineering, 11(6).

[6] Sai,, S., Gopichand, G., Reddy, C. and Mona Teja, K., 2019. High Accurate Unhealthy Leaf Detection. Computer Vision and Pattern Recognition, [online] Available at: <https://arxiv.org/abs/1908.09003> [Accessed 11 August 2022].

[7] Woods, K. and Marco, G., 2007. Genetic Algorithms: Colour Image Segmentation Literature Review. [online] Available at: <https://www.researchgate.net/publication/228536624_Genetic_Algorithms_Colour_Image_Segmentation> [Accessed 11 August 2022].

[8] Sekulska-Nalewajko J, Goclawski J (2011) A semi-automatic method for the discrimination of diseased regions in detached leaf images using fuzzy c-means

clustering. In: Seventh IEEE international conference on perspective technologies and methods in MEMS design (MEMSTECH), Polyana, 11-14 May. pp. 172–175

[9] Narvekar PR, Kumbhar MM, Patil SN (2014) Grape leaf diseases detection & analysis using SGDM matrix method.International Journal of Advance Engineering and Research Development, 2014.. 1(09).

[10] Sekulska-Nalewajko J, Goclawski J (2011) A semi-automatic method for the discrimination of diseased regions in detached leaf images using fuzzy c-means clustering. In: Seventh IEEE international conference on perspective technologies and methods in MEMS design (MEMSTECH), Polyana, 11-14 May. pp. 172–175

[11] Huang, K., 2007. Application of artificial neural network for detecting Phalaenopsis seedling diseases using color and texture features. Computers and Electronics in Agriculture, 57(1), pp.3-11.

[12] ACM other conferences. 2022. Segmentation and classification of tobacco seedling diseases | Proceedings of the Fourth Annual ACM Bangalore Conference. [online] Available at: <https://dl.acm.org/doi/10.1145/1980422.1980454> [Accessed 11 August 2022].

[13] Google Books. 2022. Proceedings of 3rd International Conference on Computing Informatics and Networks. [online] Available at: [Accessed 11 August 2022].

[14] Rothe, P. and Kshirsagar, R., 2014. Adaptive neuro-fuzzy inference system for recognition of cotton leaf diseases. 2014 Innovative Applications of Computational Intelligence on Power, Energy and Controls with their impact on Humanity (CIPECH),.

[15] Ijsrp.org. 2022. [online] Available at: <https://www.ijsrp.org/research-paper-0513/ijsrp-p1750.pdf> [Accessed 11 August 2022].