Crop Condition Assessment using Machine Learning

R. Poonguzhali, A.Vijayabhanu

Abstract: ***Agriculture is considered to be the backbone of our country. Crops play an important role in our daily routine providing us with nourishments. Due to environmental conditions, crops are getting affected with many diseases. Farmers are not able to detect these diseases at an early stage. Thus, assessment of crop condition is vital. The growing technology plays a major role and techniques like Machine Learning, Deep Learning are used. This paper focuses on the assessment of the crop condition with the help of their leaves. Healthy as well as diseased leaves are captured using cameras from real-time environments. The captured images undergo processes like preprocessing and segmentation. K-means clustering is used for segmentation. After segmentation, they undergo classification using Machine learning algorithms in which healthy and diseased leaves are detected. Thus this system helps to reduce the difficulties faced by the farmers during crop cultivation which helps in increasing the crop yield.***

Keywords: Crop cultivation, Machine Learning, Deep Learning, K-means clustering.

I. INTRODUCTION

The growing field of interest in today's world is agriculture. It is one of the important occupations practiced in India and covers about 60% land. With the change in the environment due to factors like pollution, global warming, natural disasters, etc farming has become difficult for farmers. Thus modern agriculture or modern farming is used nowadays for profitable farming. The latest techniques used in agriculture helps to analyze the soil condition of the fields, temperature, the pesticides that is best suited for specific crops, disease diagnosis, water level to be used etc.

Plant disease diagnosis uses techniques like Machine Learning Algorithms, Image Processing Techniques, Deep Learning etc for identification of healthy and diseased crops. Machine Learning plays a crucial role in this field. The crop as a whole or the leaves of the plants are taken into consideration and analyzed by these techniques. The symptoms present on the leaves or the crops are processed with the help of image processing. These are the main source for the detection of diseases. They can be of various types depending on the stages of the disease. Thus, detection of the diseases at any stage should be possible.

Machine Learning is a technique in which the system learns automatically using the information provided. It improves with experience without being explicitly programmed unlike other languages. The process is divided into two types; training and testing. The user provides the training data to the system which helps it to get trained. Large amount of training data gives good accuracy of the results.

Revised Manuscript Received on March 26, 2019

R. Poonguzhali, Assistant Professor, Department of Computer Science and Engineering,

A.Vijayabhanu, Deputy Director Information System, EDPO, ISRO

The testing data is given to the trained system. The result depends on how much the system has been trained. When the system is given images as the input, the images have to be processed with the help of image processing techniques. Disease diagnosis of crops requires images of the plants in real-time conditions. These images can be either laboratory-based or field-based images. The images captured are processed and classified using machine learning algorithms. Paddy leaf images are classified as training data and testing data. The system is trained such that it distinguishes between healthy and diseased leaves.

II. LITERATURE SURVEY

There are various studies and research done on crop condition assessment using machine learning. Radhika Deshmukh and Manjusha Deshmukh [1] showed that paddy leaf diseases are detected using K-means algorithm and Artificial neural network algorithm. The disease mainly focused here is Brown Spot. The flow of this paper is Image Acquisition, Image Preprocessing and Segmentation, Feature Extraction and Classification using ANN. K-means algorithm is used for segmentation of images. This papers results with fast disease detection.

Paddy leaf images captured by using cameras are processed in remote server. The diseases considered here are Brown Spot, Rice Sheath Blight and Rice Blast. Y. Sanjanaet al.[2] captured 500 diseased paddy leaf images. The methodology used here are Image Capture, image selection, zoom and crop, share image with expert crop and receive notification from central server. The expert groups perform image processing steps for both training and analysis stages. This method is simple and least expensive

Amrita A Joshi and B D Jadhav[3] have proposed "Monitoring and Controlling Rice Diseases Using Image Processing Techniques". In this system, the rice plant gets attacked by diseases like rice bacterial blight, rice blast, rice brown spot and rice sheath rot. Methods like Minimum Distance Classifier (MDC) and k-Nearest Neighbor classifier (k-NN) are used. 115 paddy images of size 200*200 are collected. RGB images are converted into YCbCr images. After conversion of the diseases to YCbCr, feature extraction is completed. 70% data is used for training and 30% data is used for testing. Accuracy with K-NN is 87.02% and accuracy with MDC is 89.23%.

Histogram Oriented Gradient Features is used for the recognition of paddy diseases. Diseases like brown spot, bacterial blight and leaf blast are considered.

The Histogram of Oriented Gradient (HOG) is a feature descriptor used in computer vision and image processing for the purpose of object detection. K Jagan Mohan and M Balasubramanian[4] have extracted the features using Support Vector Machine algorithms. Out of 120 images, 90 images were used for training and remaining 30 images for testing. Accuracy obtained from this method is 97.73%.

R.P Naramda and G.Arulvadivu[5] recognized diseases like Blast disease, Brown Spot Disease, Narrow Brown Spot disease. The images are first processed using image processing techniques for a better quality process. They are segmented using K-means clustering. This paper evaluates using digital image processing for detecting, diagnosing, recognizing of paddy diseases. K-means clustering algorithm is used for automatic recognition of the disease for more accuracy.

Harshadkumar B. Prajapati et al.[6] system detects and classifies rice plant diseases. Three diseases namely, Bacterial leaf blight, Brown spot, and Leaf smut are detected. Four techniques of background removal and three techniques of segmentation are considered. K-means clustering is used for accurate extraction of features. HSV masks are used for background removal. After segmentation using K-means and removal of green region, features like color, shape and texture are extracted. SVM is used for classification that consists of modified images. Training accuracy achieved is 93.33% and testing accuracy is 73.33%.

Amandeep Singh, Maninder Lal Singh[7] system automates Blast Disease Detection from Paddy Plant Leaf - A Color Slicing Approach. In this paper, they detect the paddy leaf by applying color slicing approach. They take real time of paddy images from field. The workflow of this paper is to collect the images, and then it extracts R.G.B to H.S.I followed by color slicing approach. Finally the testing process is done. They use MATLAB application and the result is 96.6%.

III. METHODOLOGY

There are 5 components in the system. They are (i) Image Acquisition (ii) Image Preprocessing (iii) Image Segmentation (iv) Feature Extraction (v) Classification.

In Image Acquisition, real-time paddy leaf images from the fields nearby are captured. It consists of both healthy leaf images as well as diseased leaf images. Diseases like Brown Spot and Leaf Blast are considered. The captured images are cropped to a specific size. The cropped RGB images are converted to grayscale in Image Preprocessing. Image Segmentation is the third component. It consists of segmenting the converted grayscale images using K-Means Clustering. This helps to get rid of problems like backgrounds, illumination of light, etc. Feature Extraction is extracting or showing the diseased portion of the leaf so that classification becomes easy. The last module includes the classification in which Tensor Flow and ANN algorithm is used. Fig 1 represents the flow work of the project describing the stages of processing.

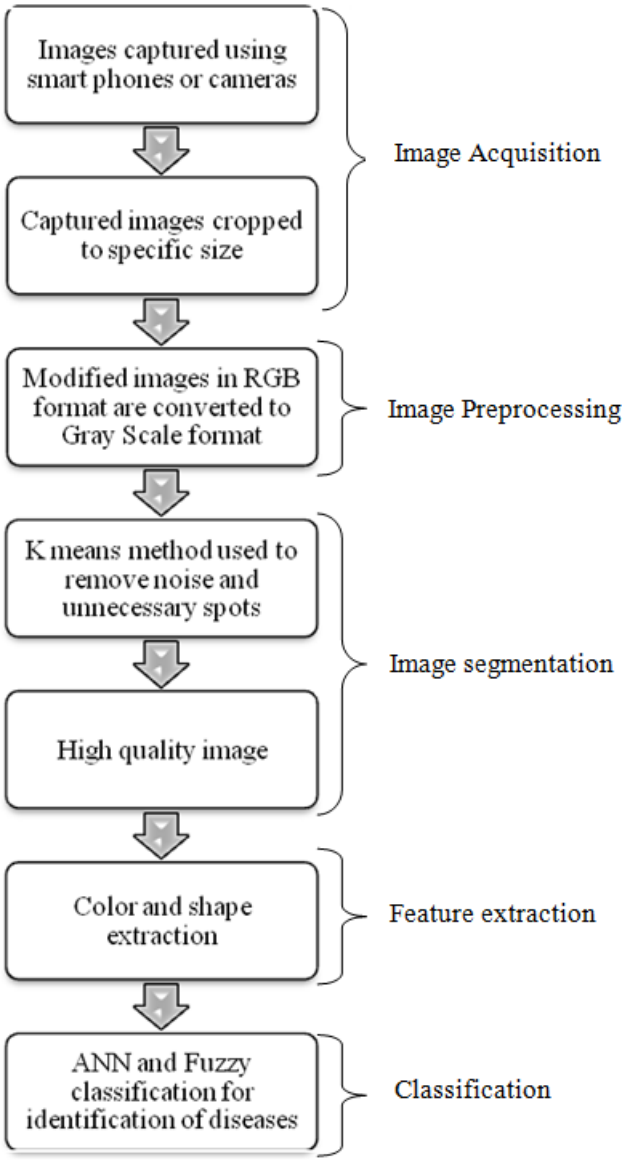

Fig 1. Architecture

A. Image Acquisition

Image Acquisition is the process of collection of images. We had performed a refined search on the dataset of paddy leaves but could not find one that matches our requirements. Thus to overcome this problem, real-time images were taken from Kurumkudi village, Thanjavur.

Two categories of leaves were considered, Healthy and Diseased. These images were captured using mobile phone cameras from the paddy areas. Healthy leaves are plain in texture without any spots on them where as diseased leaves show the presence of spots or damaged areas on the leaf. The images are captured when the age of the paddy crop was 60 days old. 50 images of each type are collected and separated into two folders; Healthy and Diseased. Healthy images are shown in Fig 2.

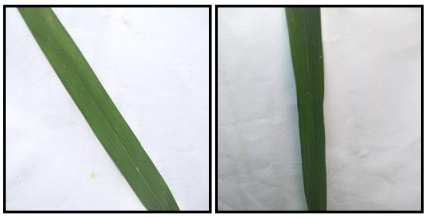

Fig 2. Healthy Leaf

The various diseases that are considered are Brown Spot and Rice Blast. Brown Spot disease shows the presence of brown spots on the surface of the leaf. This disease is caused due to fungus attack. The spots differ in shape and size. Brown spot occurs in all stages of crop, but it is most critical in maximum tillering up to the ripening stage. This causes 50% yield reduction. Nutrition-deficient soil and un-flooded soils are the main source for this disease. Rice Blast shows the presence of white to gray-green lesions or spots with dark green borders. A leaf blast infection can kill seedling up to tillering stage. It reduces the leaf area for grain fill, reducing the grain yield. The spots enlarge as the disease progress. Diseased images with brown spots are shown below in Fig 3.

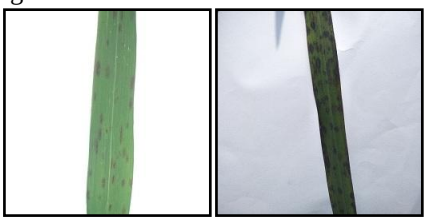

Fig 3. Diseased Leaf

The camera captures the full size of the leaves. Using these images for processing is a very crucial task. Thus, these images are cropped to a specific size; i.e. 255*255 mm. This helps to highlight the area of the diseased portion which helps in easy classification. Cropping the images to a specific size helps increase the accuracy.

Good accuracy helps in a better system. This system is a great benefit for farmers as identification of crops is essential. Healthy and diseased crops have to be identified correctly so that the yield production of farmers does not get disturbed and preventive measures can be taken at the earliest.

B. Image Preprocessing

Image preprocessing includes converting RBG images into Grayscale images using Python. An RGB image means the images present with its original colors. Grayscale images have the combination of black and white. Conversion of RGB to grayscale is done for enhancing the dataset available. Converting the images to grayscale helps in improving the accuracy of the result. Grayscale images help to reduce noise and also make the background neutral (Fig 4). It also helps to improve brightness of the image.

Data augmentation is a way of creating new data which has benefits like the ability to generate more data from limited data and it prevents over fitting. This method

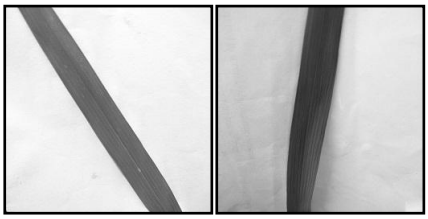

Fig 4. Grayscale converted leaf image

C. Image Segmentation

Image segmentation breaks the image down into meaningful regions. It divides digital image into multiple segments. The goal is to simplify or change the representation into more meaningful image. It differentiates between the objects we want to inspect further and the other objects or their background. It consists of segmenting the converted grayscale images using K-Means Clustering.

K-means clustering is done to get rid of problems like backgrounds, illumination of light etc. K-means is a least-squares partitioning method that divide a collection of objects into K groups. The algorithm iterates over; a) Compute the mean of each cluster, b) Compute the distance of each point from each cluster by computing its distance from the corresponding cluster mean. Assign each point to the cluster it is nearest to c) Iterate over the above two steps till the sum of squared within group errors cannot be lowered any more. Fig 5 represents segmented images.

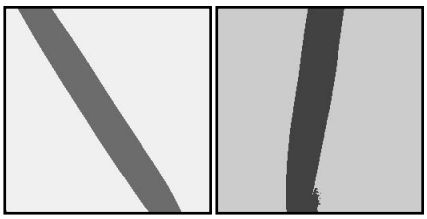

Fig 5. Segmented Images

D. Feature Extraction

Feature extraction is extracting or showing of the diseased portion of the leaf so that classification becomes easy. Features are extracted in order to differentiate between the images. Features extraction is used in almost all machine vision algorithms. The common goal of feature extraction and representation techniques is to convert the segmented objects into representations that better describe their main features and attributes. Here, shape of the spots or the diseased area of the leaf is extracted.

E. Classification

The last module includes the classification in which Tensor Flow and Machine Learning algorithm will be used. Tensor Flow is a Python-friendly open source library for numerical computation that makes machine learning faster and easier. Tensor Flow allows developers to create dataflow graphs - structures that describe how data moves through a graph, or a series of processing nodes. Each node in the graph represents a mathematical operation, and each connection or edge between nodes is a multidimensional data array, or tensor.

IV. RESULT

The collected images are resized to 255*255 and converted to grayscale. These images are then classified as training and testing images, where the training dataset is trained using Machine Learning algorithm. The training dataset consists of both healthy as well as diseased images. Convolutional Neural Network (CNN) algorithm is used to train the images. The images are successfully trained and classified. The system detects whether the image is healthy or unhealthy.

V. CONCLUSION

Many systems used images dataset for identifying the condition (healthy and bad) of the crops. Images with better quality produce better results and leads to the development of more accurate image analysis tools. Here, machine learning algorithms provide good and better results. By identifying the crop condition, appropriate pesticides can also be recommended and this in turn increases the crop cultivation in agriculture.

REFERENCES

1. Radhika Deshmukh, ManjushaDeshmukh, “Detection of Paddy Leaf Diseases”, in: International Conference on Advances in Science and Technology, 2015.
2. 2. Y.Sanjana, AshwathSivasamy, SriJayanth, “Plant Disease Detection Using Image Processing Techniques”, in: International Journal of Innovative Research in Science Engineering and Technology, 2015.
3. Amritha A Joshi, B D Jadhav, “Monitoring and Controlling Rice Disease Using Image Processing Techniques”, in: International Conference on Computing, Analytics and Security Trends (CAST), 2016.
4. K Jagan Mohan, N Balasubramanian, “Recognition of Paddy Plant Diseases Based on Histogram Oriented Gradient Features”, in: International Journal of Advanced Research in Computer and Communication Engineering, 2016.
5. R.P Narmada, G Arulvadivu, “Detection And Measurement of Paddy Leaf Disease Symptoms using Image Processing” in: International Conference on Computer Communication and Informatics, 2017.
6. Harshadkumar B Pragapati, “Detection and Classification of Rice Plant Diseases”in: Intelligent Decision Technologies11, 2017.
7. Amandeep Singh, Maninder Lal Singh, “Automated Blast Disease Detection from Paddy Plant Leaf - A Color Slicing Approach”, in: International Conference on Industrial Technology and Management, 2018.