Data Augmentation for Crop Condition Assessment

[1]Nivetha D, [2]Pavitra Kailash, [3]Poonguzhali R, [4]Vijayabhanu A
[12] Final Year Students, [3] Assistant Professor(SS), [4]Deputy Director, Information System, ISRO
[1]Computer Science and Engineering,
[1]Periyar Maniammai Institute of Science and Technology, Tamilnadu, India

Abstract : Rice is considered to be a vital crop growing in India. The environmental changes caused due to pollution, global warming, etc are affecting crops to a great extent. Thus to overcome this problem, there is a need for the development of system that identifies the healthy and diseased paddy crops from the given image dataset. The system is fed with image dataset of paddy leaves. But the main problem of implementing this system is that very less amount of images are available. To increase the number of images and to provide good accuracy, data augmentation is used. Data augmentation increases the size of the dataset by using methods like flip, rotate and translate. These images can then be processed by methods like segmentation and Machine Learning algorithms..

keywords- machine learning, data augumentation, dataset, paddy crop.

I. INTRODUCTION

Agriculture is considered to be the growing field of interest and covers about 60% of the total land. Factors like global warming, pollution and natural disasters have affected farming to a great extent. As a result, modern agriculture or modern farming is used. Analyzing the soil conditions of the fields, measuring climate temperatures, suggestion of pesticides for crops diagnosis of diseases, etc are done using the modern techniques.

Techniques like Machine Learning, Deep Learning and Image Processing are used for plant disease diagnosis. Machine Learning plays a crucial role in this field where it takes leaves as a whole or a part of the leaf for assessment. Image Processing techniques are used for processing the images of crops. 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 technique teaches the system automatically using the given information. It does not have to be explicitly programmed unlike other languages. There are two steps: training and testing. Training Data is provided by the user so that it gets trained. The more amounts of data, the better results. The testing data is then given to the trained system.

Diseases diagnosis of crops is done with the help of images of crops which can be laboratory-based or field-based images. Machine Learning algorithms do the work of processing these images. The system developed identifies diseased and healthy crops.

II.LITERATURE SURVEY

There is lot of studies done on data augmentation techniques for crop assessment. They are as follows:

Radhika Deshmukh and Manjusha Deshmukh[1] studied detecting paddy leaf diseased using k-means algorithm and Artificial neural network algorithm. They mainly focused on Brown Spot disease. The flow of this paper is Image Acquisition, Image Preprocessing and Segmentation, Feature Extraction and Classification using ANN. Segmentation of images is done using K-Means Clustering. The main aim of this paper is accurate results and fast disease detection using the proposed system.

A remote sensor is used to process the paddy leaf images captured by using cameras. Brown Spot, Rice Sheath Blight and Rice Blast diseases are taken here. Y Sanjana[2] et al. captured 500 diseased paddy leaf images where they are processed using methodologies like 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. It concludes that this method is simple and least expensive.

Amrita A Joshi and B D Jadhav[3] stated that diseases like rice bacterial blight, rice blast, rice brown spot and rice sheath rot affect diseases to a great extent. The methods used here are Minimum Distance Classifier (MDC) and k-Nearest Neighbor classifier (k-NN). 200*200 size image (115 in number) collected. The images collected are RBG in nature and are converted into YCbCr image. Later feature extraction is completed. 70:30 ratio of images are taken for training and 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. K Jagan Mohan and M Balasubramanian[4] used SVM for feature extraction. A total of 120 images were captured, of which 90 for training and 30 for testing. Accuracy obtained from this method is 97.73%

Jia Shijje et al.,[5] proposed in their paper that it is necessary to perform data augmentation methods on image datasets. Here AlexNet is used as a pre-training network model and datasets such as CIFAR10 and ImageNet are chosen as original datasets. The

data augmentation methods used are GAN/WGAN, Flipping, Cropping, Shifting, PCA jittering, Color jittering, Noise, Rotation, and some combinations. This paper concludes that compared to other enhancement models, WGAN, Cropping, Rotation, Flipping are more effective.

Blast disease, Brown Spot disease, Narrow Brown Spot are recognized in the paper published by R P Naramda and G Arulvadivu[6]. The images are first processed using image processing techniques for a better quality process and then K-Means Clustering for segmentation. Methods like digital image processing for detecting, diagnosing, recognizing of paddy diseases are used. K-means clustering algorithm is used for automatically recognition of the disease for more accuracy.

Harshadkumar B. Prajapati et al.[7] considered three diseases namely, Bacterial leaf blight, Brown spot, and Leaf smut. 4 methods of Background techniques and 3 techniques of segmentation are included. K-means clustering is used for segmentation. After segmentation using K-means and removal of green region, features like color, shape and texture are extracted. Training accuracy achieved is 93.33% and testing accuracy is 73.33%.

Amandeep Singh, Maninder Lal Singh[8]. detected paddy leaf by apply color slicing approach. Real- time field images are taken. The images are first collected and then extracted 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%.

WhereverTimesisspecified,TimesRomanorTimesNewRomanmaybeused.Ifneitherisavailableonyourwordprocessor,pleaseusethefontclosestinappearancetoTimes.Avoidusingbit-mappedfonts.TrueType1orOpenTypefontsare required.Pleaseembed all fonts, in particular symbolfonts,aswell,formath,etc.

II. METHODOLOGY

In **Image Acquisition**, paddy leaf images are captured from nearby fields. It is a combination of both healthy and diseased images Brown Spot and Leaf Blast diseases are taken into considered. These images are cropped. Conversion of RGB images to grayscale are done in **Image Preprocessing**. The captured and converted dataset is improved using **Data Augmentation** by methods like flip, transform and rotate.

There are 3 modules:

1. **Image Acquisition**
2. **Image Preprocessing**
3. **Data Augmentation**

.

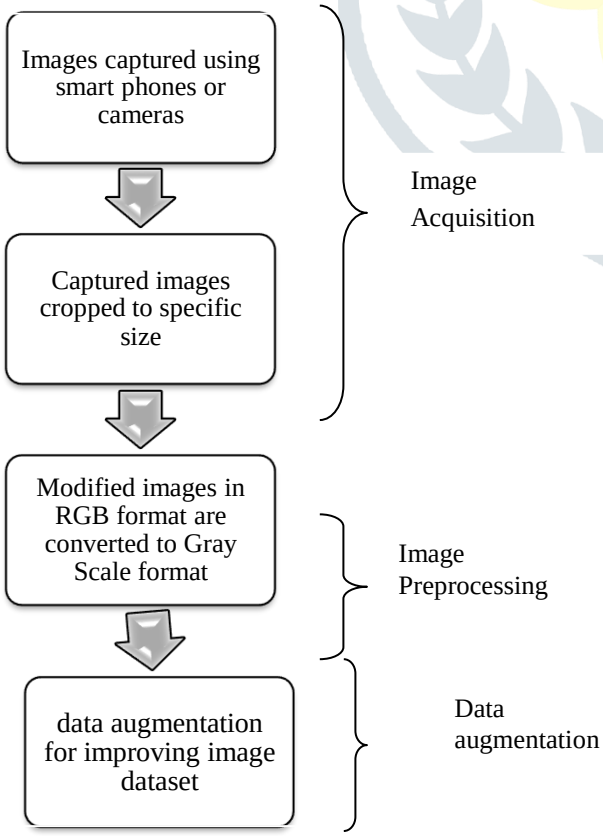

Fig 1. Architecture

2.1 IMAGE ACQUISITION

Image Acquisition is capturing of images from real-time fields as online dataset for paddy images is not available. These images are taken from KURUMKUDI village in THANJAVUR.

Two categories of leaves Healthy and Diseased are considered. Mobile phones are used for capturing of images. Healthy leaves are plain in texture where as diseased leaves show spots on the leaf. **50 images** are collected and separated into two folders; Healthy and Diseased. Healthy images are shown in Fig 2.

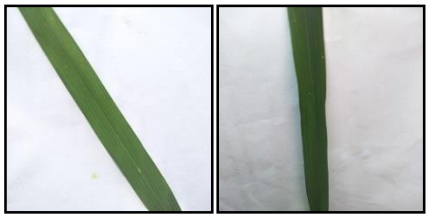

Fig 2. Healthy Leaf

Two diseases considered are Brown Spot and Rice Blast. **Brown Spot** shows brown spots on the surface and is caused by funguses. They are of various shapes and sizes. Brown spot is worst in maximum tillering up to the ripening stage which causes 50% yield reduction. **Rice Blast** shows white lesions or spots with dark green borders. It reduces the leaf area for grain fill, reducing the grain yield. Diseased images with brown spots are shown below in Fig 3.

Fig 3. Diseased Leaf

The camera takes full size images which are then cropped to a specific size; i.e. 255*255mm. This highlights the diseased portion which is easy in classification. Cropping the images to a specific size helps increase the accuracy.

Good accuracy helps in a better system. Crops have to be correctly identifies so that the yield production of farmers is better.

2.2 IMAGE PREPROCESSING

Image preprocessing includes converting RBG images into **Grayscale images** using Python coding. An RGB is the presence of original colors.

Conversion of RGB to grayscale is done for enhancing the dataset available. This helps to reduce noise and also make the background neutral (Fig 4). It also helps to improve brightness of the image.

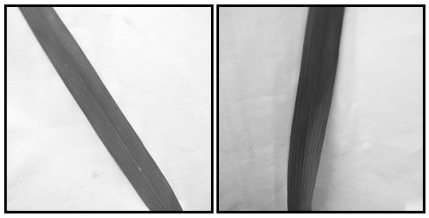

Fig 4. Grayscale converted leaf image

2.3 DATA AUGMENTATION

Data augmentation creates new dataset and has benefits like ability to generate more data from limited data. This method increases the numbers of images in the dataset by methods like **flip, rotate** and **transform.**

Flip is a method that forms a mirror image of the original image. **Rotate** method is used for the rotation of the images. They can be rotated by specific ways like, 45° and 90°. **Transform** is shifting the images;

III. FUTURE WORK

Assessing the images of paddy to identify whether it is diseased or healthy gives good results only with good dataset. Thus, the images in the dataset have to be further processed using methods like K-Means clustering for segmentation and Machine Learning algorithms like CNN, ANN for classification.

IV. CONCLUSION

Data Augmenting helps increase and improve the dataset for better accuracy. This method not only increases the number of images, but also improves the quality of the images. These images are then processed using techniques like segmentation and classification.

V.ACKNOWLEDGMENT

We wish to thank our esteemed guide Ms.R.POONGUZHALI, B.E, M.E., (Ph.D.), ASSISTANT PROFESSOR(SS), COMPUTER SCIENCE AND ENGINEERING, for their valuable guidance, spontaneous help and assistance in providing the continuous encouragement for the successful completion for the project.
We wish to thank MS.A.VIJAYABHANU, DEPUTY DIRECTOR INFORMATION SYSTEM, EDPO, ISRO, for guiding us this project and to work under the guidance

VI. REFERENCES

[1] Radhika Deshmukh, ManjushaDeshmukh, “Detection of Paddy Leaf Diseases”, 2015 International Conference on Advances in Science and Technology.
[2] Y.Sanjana1, AshwathSivasamy2, SriJayanth2, “Plant Disease Detection Using Image Processing Techniques”, 2015 International Journal of Innovative Research in Science Engineering and Technology.
[3] Amritha A Joshi, B D Jadhav, “Monitoring and Controlling Rice Disease Using Image Processing Techniques”, 2016 International Conference on Computing, Analytics and Security Trends (CAST).
[4] K Jagan Mohan, N Balasubramanian, “Recognition of Paddy Plant Diseases Based on Histogram Oriented Gradient Features”, 2016 International Journal of Advanced Research in Computer and Communication Engineering.
[5] Jia Shijie, Wang Ping, Jia Peiyi and Hu Siping, “Research on Data Augmentation for Image Classification Based on Convolution Neural Networks”, 2017 Chinese Automation Congress.
[6] R.P Narmada, G Arulvadivu, “Detection And Measurement of Paddy Leaf Disease Symptoms using Image Processing” 2017 International Conference on Computer Communication and Informatics.
[7] Harshadkumar B Pragapati, “Detection and Classification of Rice Plant Diseases”, 2017 Intelligent Decision Technologies11.
[8] Amandeep Singh, Maninder Lal Singh, “Automated Blast Disease Detection from Paddy Plant Leaf - A Color Slicing approach”, 2018 International Conference on Industrial Technology and Management