

LEAF DISEASE DETECTION USING IMAGE PROCESSING TECHNIQUES AND CNN

Prof. S.A.Dhole¹, Divya Bharti², Krishna Jaiswal³, Kalpana Kasabe⁴

¹Prof. Dept. of Electronics and Telecommunication Engineering, Bharati Vidyapeeth College of Engineering for Women, Pune, India

^{2,3,4} Students, Dept. of Electronics and Telecommunication Engineering, Bharati Vidyapeeth College of Engineering for Women, Pune, India

Abstract - Even though India is agricultural country lot of challenges are faced by farmer. Plant disease reduces product of farmer both in quality and quantity. So quick detection and identification of disease plant are of more importance. It also needs continuous monitoring of farm. To overcome above problem it is necessary to develop such system which continuously monitor the farm and detect the disease as quick as possible. In this project we gave brief idea to solve this problem by detecting the leaf diseases of plant and process it using image processing techniques. Image processing gives fast, automatic and accurate solution to user.

1. INTRODUCTION

The world textile industries are being ruled by “King Cotton”. India thus enjoys the distinction of being the earliest country in the world to domesticate cotton and utilize its fiber to manufacture fabric. India accounts for approximately 25 per cent of world's cotton area and 16 per cent of total cotton production. Maharashtra is the important cotton growing state in India. The 2nd largest producer of cotton in the world. About 3 million farmers are engaged in cotton cultivation in the state mostly in backward region of Marathwada and Vidarbha. Disease on the cotton is the main problem that decreases the productivity of the cotton. The main source for the disease is the leaf of the cotton plant. About 80 to 90 % of disease on the cotton plant is on its leaves.

So, for that our study of interest is the leaf of the cotton tree rather than whole cotton plant the cotton leaf is mainly suffered from diseases like fungus, Foliar leaf spot of cotton, Alternaria leaf spot of cotton. The machine vision system now a day is normally consisted of computer, digital camera and application software. Various kinds of algorithms are integrated in the application software. Image analysis is one important method that helps segment image into objects and background. One of the key steps in image analysis is feature detection. Study of diseases on the cotton leaf can robustly studied by the image processing toolbox and also the diagnosis by using MATLAB helps us to suggest necessary remedy for that disease arises on the leaf of cotton plant. We know that perception of the human eye is not so much stronger that he can differ minute variation in the infected part of image because that minute variation pattern of colour can be a different disease present on the leaf of cotton. Our software can provide the exactly differentiate the variation of colour present on these leaves and depending upon that variation the further compare with database stored image features related to the colour.

Image Definition - Image is a collection of pixels or dots which are stored in rectangular array. Each individual pixel is having certain kind of color. We can measure the size of the image by counting the no of pixels in that particular image. Different type of images are there such as Black and White and Grey scale images. Both types vary from each other. In black and white image each dot or pixel is either black or white, therefore only one bit is needed per pixel. Whereas Grey scale images uses 8 bits per pixel. For color images things gets slightly difficult. In color images number of bits at every dot termed as the height of image. It is also referred as the bit plane. For bit plane consisting of x, 2x color are possible. Different methods are available to store the color information of image. One of the method is RGB image also termed as true color image. For every pixel red, green and blue component is stored in three-dimensional array.

Semantic Network - Semantic network also termed as neural networks consist of layer of connected nodes like neurons in brain used for computing purposes. These networks has the ability to learn the data and get trained over that data, in order to identify patterns and also classification. Neural network or semantic network has the advantage that they can solve the complicated problem very easily. To train the neural networks or to implement them in MATLAB we have NN toolbox. Different commands are there to operate with NN toolbox. Different functions are there in NN toolbox for fitting, recognizing the pattern and clustering.

Identifying Patterns - In order to detect the same kind of pattern different pattern recognition techniques are used in MATLAB. Using these techniques we recognize the similar kind of the pattern in the problem. When same kind of pattern are detected then these can be used to generate outputs or solve the problems more efficiently. In order to recognize the pattern, we need to train the machine. For this first we need to classify the data. The data is classified using the key features. For classifying the data we have different type of learning modules is there such as supervised learning and unsupervised learning modules. Both of these modules are used to identify the patterns. In supervised learning module we train the machine by recognizing the patterns in the dataset and the results which are generated are applied to the testing data set. We train the machine over the training dataset and test it over the testing data set. In unsupervised learning module, there are no visible pattern the dataset, so with the help of the some algorithm we try to catch the patterns. Clustering algorithm, classification algorithm such as Markov Model (MM) is there. For recognizing the patterns we identify we have different techniques such as preprocessing,

Extraction of features and classification. In preprocessing we try

to filter out, smooth the data by normalizing in more ordered way. Filtering such as noise filtering is there. Feature extraction is usually done using the software which collect the information from the data. Sensors area also used for this purpose and the final phase is the classification.

2. LITERATURE REVIEW

Various papers describing to detecting the diseases and methods suggesting the implementation ways as illustrated and discussed here.

[1] S. Arivazhagan, developed processing scheme in which a color transformation structure for the input RGB image is created, and then the green pixels are masked. Shape and texture features are extracted. The classification is done by using the Minimum Distance Criterion and Support vector machines (SVMs).

[2] Dheeb Al Bashish, et al., have proposed A Framework for Detection and Classification of Plant leaf and Stem Diseases in which the images at hand are segmented using the K Means technique, RGB input images are converted into HIS color space. Then calculating color and texture based features. Neural network classifier that is based on statistical classification is used for classification.

[3] Elham Omrani, have proposed Potential of radial basis function-based support vector regression for apple disease detection, detection of leaf diseases has been used method is threefold: First the leaf images were stored in RGB – a device dependent color space. To segment the images, they had to be transferred to the device-independent color space CIELAB. Second in order to extract the infected area, region-based segmentation was employed. K-means cluster is an important technique implemented in the segmentation phase. Third to extract the color, shape and texture-based features. They are normally used for region description. To obtain the texture features, wavelet and gray level co-occurrence matrix techniques were employed. A support vector regression (SVR) technique to classify apple leaf diseases.

[4] Haiguang Wang, et al have proposed Image Recognition of Plant Diseases Based on Principal Component Analysis and Neural Networks in which 21 color features, 4 shape features and 25 texture features were extracted from the images of wheat and grape diseases. This features are 12 extracted by principal component analysis (PCA). neural networks including backpropagation (BP) networks, radial basis function (RBF) neural networks, generalized regression networks (GRNNs) and probabilistic neural networks (PNNs) were used as the classifiers to identify wheat diseases and grape diseases, respectively.

[5] Santanu Phadikar, have proposed Rice Disease Identification using Pattern Recognition Techniques describes a software prototype system for rice disease detection based on the infected images of various rice plants. Using digital camera images of infected rice plants are captured and using image growing, image segmentation techniques to detect infected parts of the plants. Then the classification of infected part of leaf is done by neural network.

[6] P. Revathi, V, have proposed Cotton Leaf Spot Diseases Detection Utilizing Feature Selection with Skew

Divergence Method in this work enhanced PSO feature selection method adopts skew divergence method and to extract Edge, CYMK color feature, GA feature, Color, Texture variances features. The extracted feature was input to the SVM, Back propagation neural network (BPN), Fuzzy with Edge selection and classification is done by them.

From above literature survey found that the methods are used by different researchers for leaf disease detection & analyses are following:

1. Otsu segmentation, K-Means clustering & Neural Network.
2. Color transformation, segmentation, computing texture features & Support vector machines for developing classification and regression.
3. Otsu segmentation, Sobel operation & Grading method.
4. Fuzzy c-means clustering in hue-saturation space
5. Image clipping, filtering & thresholding.
6. Comparison of two techniques: Otsu threshold and K-Means clustering

3. PLANT DISEASE FUNDAMENTALS

The diseases on the cotton leaves are classified as

- a) Bacterial disease: e.g. Bacterial Blight, Crown Gall, Lint Degradation.
- b) Fungal diseases: e.g. Anthracnose, Leaf Spot.
- c) Viral disease: e.g. Leaf Curl, Leaf Crumple, Leaf Roll.
- d) Diseases Due To insects: e.g. White flies, Leaf insects.

Out of the above types of disease these are dramatically affect the leaf of cotton plant and its leaves. We go through the selective type of diseases on the cotton leaves. And further we discuss the CNN image segmentation method to detect the diseases on cotton plant by scanning of cotton leaves through our portable dedicated scanner.

3.1 Types of Diseases

Various diseases are found on the cotton plant out of this we discuss the disease some of the major diseases which are often found on the leaves of cotton, that are viz.

3.1.1 Foliar leaf spot on cotton



As shown in above figures the, the disease is known as foliar disease arises due to potassium deficiency. The early stage of this disease is as shown in fig, now if the more spots of this disease results into the final stage of this plant where the plant leaf is get

fall so it is called as Foliar disease of the cotton plant as shown in fig. The leaf is having multiple no of spots which clearly denotes more potassium deficiency in the plant.

3.1.2 Curl Gemini virus



Cotton leaf curl Gemini virus (CLCuV) causes a major disease of cotton in Asia and Africa. Leaves of infected cotton curl upward Fig. and bear leaf-like enations on the underside along with vein thickening Fig. Plants infected early in the season are stunted and yield is reduced drastically. Severe epidemics of CLCuV have occurred in Pakistan in the past few years, with yield losses as high as 100% in fields where infection occurred early in the growing season. Another cotton Gemini virus, cotton leaf crumple virus (CLCrV), occurs in Arizona, California, and Mexico. CLCrV symptoms are distinguishable from CLCuV symptoms in that infected leaves curl downward accompanied by interveinal hypertrophy and foliar mosaic Fig 3.2, both CLCrV and CLCuV infect dicotyledonous plants and are whitefly-transmitted (Brown et al., 1983; Mansor et al., 1993). Previous studies (Brown and Nelson, 1984; 1987; Hameed et al., 1994; Mansor et al., 1993) suggested that they belong to the subgroup III Gemini viruses. However, little information is available on the relationship of these two viruses with each other and with other subgroup III Gemini viruses.

3.1.3 Bacterial Blight



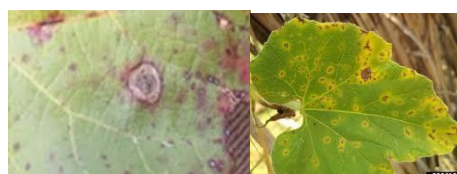
Xanthomonas campestris pv. *Malvacearum* Bacterial blight starts out as angular leaf spot with a red to brown border. The angular appearance is due to restriction of the lesion by fine veins of the cotton leaf. Spots on infected leaves may spread along the major leaf veins as disease progresses, leaf petioles as shown in Fig. The angular leaf spot, results in premature defoliation and stems may become infected resulting in premature defoliation.

3.1.4 Alternaria Leaf Spot



As shown in Fig 6, small, pale to brown, round or irregular spots measuring 0.5 - 3 mm in diameter and cracked centers appear on the affected leaves of the plant. Affected leaves become dry and fall off. The disease may cause cankers on the stem. The infection spreads to the bolls and finally falls off.

3.1.5 Cercospora leaf Spot



The disease affects older leaves of mature plants. The spots are round or irregular in shape yellowish brown, with purple, dark brown or blackish borders and white centers. Affected leaves become pale in color and finally fall off as shown in Fig.

4. METHODOLOGY

The step-by-step procedure as shown below:

Image acquisition: Image acquisition is the first phase which includes capturing of infected leaf images to build database. The RGB color images of infected cotton leaves are captured using digital camera in JPEG format from cotton farm in Pune district with required resolution of 250 x 190 which gives good quality images for disease detection. The database of 100 images are collected. During this step, noise is generated which is reduced by preprocessing.

Image Pre-processing: In preprocessing step to improve image data that removes background, noise and also suppress undesired distortions. The aim of pre-processing is improvement of an image data that suppress unwanted distortions and enhances some image features important for further processing. The preprocessing methods incorporate image enhancement, color conversion, resizing and filtering of an images. In present system resizing and median filter is adopted for pre-processing to expel noise and to get good quality image. First the

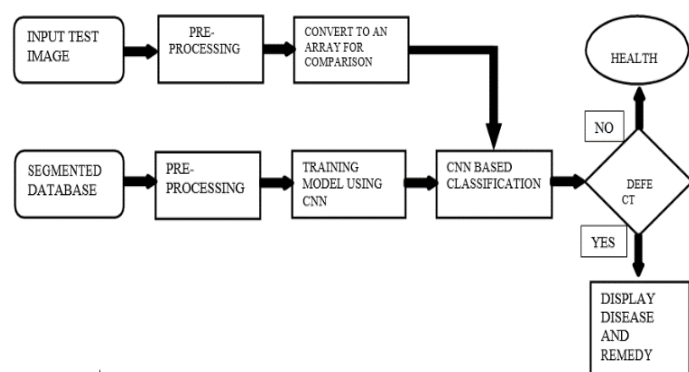
image is read and then resized into 250×250 pixel and then median filter is applied. Median filter is used as it is more accurate as compared to mean filter and preserve edges while removing noise.

Image Segmentation: Image segmentation is process used the representation of an image into something that is more meaningful object of interest from background. The suitable color group numbers lead to the better color extraction. The input RGB color image is converted into other color spaces such as HIS and CIELAB. Because RGB is color dependent space model but HSI and CIELAB are color independent space model and this are also derived from human perception. In segmentation step to find out the infected region. Segmentation mostly can be done by k-mean clustering, edge detection algorithm

Feature Extraction : After segmentation the infected region various features are extracted to describe the infected region. Color, texture and shape based features are normally used for region description. Color features are important to sense image environment, recognize objects and convey information. Texture is one of the most important feature which can be used to classify and recognize objects. It is a powerful regional descriptor that helps in the image retrieval process. Contrast, Homogeneity, Dissimilarity, Energy and Entropy features are intended to describe texture. Shape is one of the primitive features for image content description. In this project for feature extraction we are using CNN model. Layers of CNN model is do process of feature extraction for input image to achieve result.

Classification : It is final stage in disease detection. It is identifying a rule according to selected features and assigning each disease to any one the predetermined classes.

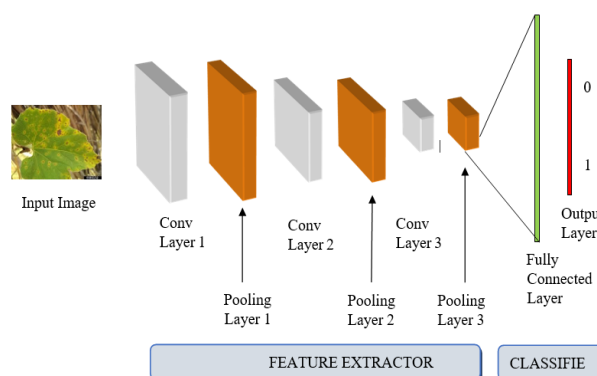
5. BLOCK DIAGRAM OF SYSTEM



1. The input test image is acquired and pre-processed in the next stage and then it is converted into array form for comparison.
2. The selected database is properly segregated and pre-processed and then renamed into proper folders.
3. The model is properly trained using CNN and then classification takes place.
4. The comparison of the test image and the trained model take place followed by the display of the result.
5. If there is a defect or disease in the plant the software displays the disease along with the remedy .

6. CONVOLUTION NEURAL NETWORK (CNN)

It is a Deep Learning algorithm which can take in an input image, assign importance (learnable weights and biases) to various aspects/objects in the image and be able to differentiate one from the other. CNNs are used for image classification and recognition because of its high accuracy. It follows a hierarchical model which works on building a network, like a funnel, and finally gives out a fully-connected layer where all the neurons are connected to each other and the output is processed.



CNN is a widely-used image recognition model that has been shown to attain greater than 78.1% accuracy on the Image Net dataset. The model is the conclusion of many ideas developed by many researchers over the years. An 256x256x3 input representing a visual field of 256 pixels and 3 color (RGB) channels. Five convolution layers, with a few interspersed max-pooling operations. Successive stacks of “CNN Models”. A softmax output layer at the end at an intermediate output layer just after the mixed layer. Steps involved in CNN are Convolution layer in CNN is performed on an input image using a filter. Relu (Rectified Linear Unit) which simply converts all of the negative values to 0 and keeps the positive values the same. Pooling layer is used to reduce the spatial size of the Convolved Feature. They are of two types such as Max Pooling and Average Pooling. Fully Connected layers in a neural networks is a layer where all the inputs from one layer are

connected to every activation unit of the next layer. These networks are commonly trained under a log loss (or cross-entropy) system, giving a non-linear variant of multinomial logistic regression.

CNN LAYER:

Conv2D: It is the layer to convolve the image into multiple images activation is the activation function.

MaxPooling2D: It is used to max pool the value from the given size matrix and same is used for the next 2 layers.

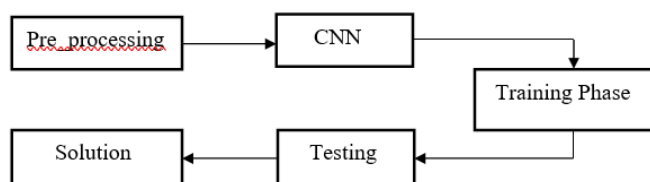
Dropout: It is used to avoid over fitting on the dataset and dense is the output layer contains only one neuron which decide to which category image belongs.

Fully Connected: It has neurons that are fully connected to the neurons in previous layer. This FC layer is often kept as the final layer of a CNN with “SOFTMAX” as its activation function for multi-class classification problems. The FC layer is responsible to predict the final class or label of the input image. Thus, it has an output dimension of $[1 \times 1 \times N]$ where N denotes the number of classes or labels considered for classification.

Epochs: It tells us the number of times model will be trained in forward and backward pass.

7. DETECTION AND IDENTIFICATION

Once the model is trained with the available dataset, it is then tested. Input the Image (filename.jpg) for the prediction. Predict function is called on the loaded pickle module to classify the image as healthy or not.

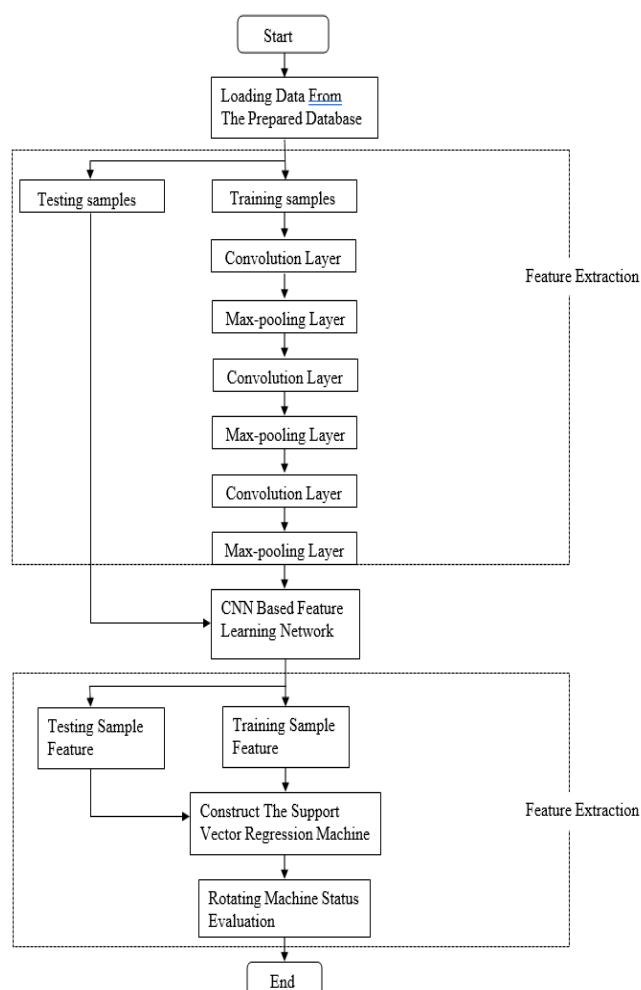


The final stage is the detection of the diseases and with the help of disease classify the plants with the disease matches with the given dataset. For the disease detection and classification, we are implementing the deep learning algorithm. Deep learning algorithm CNN is used to classify the specified image into appropriate disease hence it will be easy to detect the disease and find out the remedy over the disease. Deep learning algorithm is the part where we are finding out the relevancy count of the pixels by

comparing the images with the data set. According to the relevancy count we will find out the matched disease.

8. FLOWCHART AND ALGORITHMS

1. Image Pre-processing
2. Loading Images.
3. Building CNN model.
4. Training the model.
5. Classification of the image.



9. CONCLUSIONS

There are many methods in automated or computer vision plant disease detection and classification process, but still, this research field is lacking. In addition, there are still no commercial solutions on the market, except those dealing with plant species recognition based on the leaves images. In this paper, a new approach of using deep learning method was explored in order to automatically classify and detect plant diseases from leaf images. The developed model was able to detect leaf presence and distinguish between healthy leaves and 5 different diseases, which can be visually

diagnosed. The complete procedure was described, respectively, from collecting the images used for training and validation to image preprocessing and augmentation and finally the procedure of training the deep CNN and fine-tuning. Different tests were performed in order to check the performance of newly created model. New plant disease image database was created, containing more than 90 original images taken from the available Internet sources and extended to more than 30,000 using appropriate transformations. The experimental results achieved precision between 91% and 98%, for separate class tests. The final overall accuracy of the trained model was 96.3%. Fine-tuning has not shown significant changes in the overall accuracy, but augmentation process had greater influence to achieve respectable results. As the presented method has not been exploited, as far as we know, in the field of plant disease recognition, there was no comparison with related results, using the exact technique. In comparison with other techniques used and presented, comparable or even better results were achieved, especially when taking into account the wider number of classes in the presented study. An extension of this study will be on gathering images for enriching the database and improving accuracy of the model using different techniques of fine-tuning and augmentation.

10. ACKNOWLEDGEMENTS

On the submission of our seminar report on “Leaf disease detection using image processing techniques”, we would like to extend our gratitude and sincere thanks to our HOD **Prof. (Dr.) S. R. Patil** and guide **Prof. Dr. S. A. Dhole**, Project Co-ordinator, Department of Electronics and Telecommunication for her constant motivation and support during the course of our work. We truly appreciate and value her esteemed guidance and encouragement from the beginning. We are indebted to her for having helped us shape the problem and providing insights towards the solution. And for providing a solid background for our studies and research thereafter. She has been a great source of inspiration to us and we thank her from the bottom of our heart. Above all, we would like to thank all our friends whose direct and indirect support helped us.

11. REFERENCES

- [1] Mr. V. A. Gulhane & Dr. A. A. Gurjar, “Detection of Diseases on Cotton Leaves and Its Possible Diagnosis” Indian journal of Psychology (IJIP), Volume 5: Issue (5): June 2011.
- [2] Mr. P. Revathi & M. Hemalatha “Classification of Cotton Leaf Spot Diseases Using Image Processing Edge Detection Techniques” Institute of Electrical and Electronics Engineers ISBN : 978-1-4673-5144-7/12/\$31.00 © 2012 (IEEE).
- [3] Kamaljit Singh Kailey, Gurjinder Singh Sahdra, “Content-Based Image Retrieval (CBIR) For Identifying Image Based Plant Disease,” Int. J. Computer Technology & Applications, Vol 3 (3), 1099-1104: May - June 2012.
- [4] Prof. Sanjay B. Dhaygude, Mr. Nitin P. Kumbhar, “Agricultural plant Leaf Disease Detection Using Image Processing,” International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering, Vol. 2, Issue 1, January 2013.
- [5] Applications of Perceptual Hash Algorithm in Agriculture Images, XimLiu, QianZhang, RuPengLuan, Feng Yu, Institute of Electrical and Electronics Engineers 2013 (IEEE).
- [6] Evy kamilah Ratnasari, Mustika Mentari, Ratih kartika Dewi, R. V. Hari Ginardi “Sugar Cane Leaf Disease Detection and Severity Estimation Based On Segmented Spots Images” Institute of Electrical and Electronics Engineers 978-1-4799-6858-9/14/\$31.00 2014 (IEEE).
- [7] S. B. Borate, P. V. Mulmule “study and analysis of methodologies for Leaf Disease detection using image processing” IJRECE Volume 3 Issue 1 Jan-March 2015.
- [8] Mr. Pawan P. Warne, Dr. S. R. Ganorkar “Detection of Diseases on Cotton Leaves Using K-Means Clustering Method” International Research Journal of Engineering and Technology (IRJET). Volume: 02 Issue: 04 July 2015.

