A Study on Various Preprocessing Methodologies For Plant Disease Detection Using Leaf Image

Arya Rajagopal[1], Lekshmi R Nair[2], Dr. Priya S[3]
[1, 2, 3] Dept of Computer Science
[1, 2, 3] College of Engineering Cherthala, Kerala

Abstract- *This survey paper mainly the research-oriented methods which detects the leaf diseases to the earliest and image processing methods have been proposed by the authors which enhances the crop yield thus in controlling various bacterial, virus, fungal and nematodes. Summary of various documents have been drawn, which gives us clear idea on how to improve other parameters which in turn helps us to stop the diseases. In this paper, we studied and evaluated existing techniques for detection of plant diseases to get clear outlook about the techniques and methodologies followed.*

Keywords- Image processing, Detection, Methodologies.

I. INTRODUCTION

Agriculture is just not helpful for human feeding or earning it is much more like energy and global warming. Leaf disease has been affecting many aspects in the field of agriculture mainly they are production, quality and quantity. India is a country which is dependent on agriculture. Leaf disease detection can be helpful for the farmers. Research works in smart computing surrounding to identify the disease using the pictures of leaves. Several problems are to be identified which are given as follows. Detecting the diseased leaf, to measure area affected by the disease, identifying the boundary of affected area by disease, finding out the color of the affected area and what exactly causes the disease i.e., by insects, rust, nematodes etc., Diseases on the leaves are mainly viral, bacterial, fungal. Plant disease is one of the important factors which causes significant reduction in the quality and quantity of plant production. Detection and classification of plant diseases are important task to increase plant productivity and economic growth. Detection and classification are one of the interesting topics and much more discussed in engineering and IT fields. Identification of the plant diseases is the key to preventing the losses in the yield and quantity of the agricultural product. The studies of the plant diseases mean the studies of visually observable patterns seen on the plant. Health monitoring and disease detection on plant is very critical for sustainable agriculture.

It is very difficult to monitor the plant diseases manually. It requires tremendous amount of work, expertise in the plant dis- eases, and also require the excessive processing time. Hence, image processing is used for the detection of plant diseases. Disease detection involves the steps like image acquisition, image pre-processing, image segmentation, feature extraction and classification. There are various techniques emerged to detect the plant disease such as thresholding, region growing, clustering, watershed, etc. To detect plant disease the image should go through pre-processing, segmentation, feature ex- traction and classification processes. The pre-processing is an improvement process of image data to suppresses unwanted distortion or enhances some image features important for further processing

II. DATA PREPROCESSING

There are various techniques emerged to detect the plant disease such as thresholding, region growing, clustering, trashed, etc. To detect plant disease the image should go through pre-processing, segmentation, feature extraction and classification processes. The pre-processing is an improvement process of image data to suppresses unwanted distortion or enhances some image features important for further processing

.

The segmentation process is to partition an image into meaningful regions and it is vital process through which image features are extracted. There are various features of an image such as grey level, color, texture, shape, depth, motion, etc. Classification process is used to classify the given input data into number of classes and groups. It classifies the data based upon selected features.

III. DATA SETDESCRIPTION

A. Specifications of the Dataset

We analyze 54,306 images of plant leaves images from the publicly available plant village dataset which have a spread of 38 class labels assigned to them. Each class labels are a crop disease pair, and we make an attempt to predict the crop disease pair just the image of the plant leaf. We resize the image to 256 x 256 pixels and we perform both the model optimization and predictions on these downscaled images.

Across all our experiments we use three different versions of the whole plant village dataset. We start with the plant village dataset as it is, in color shown in fig 1, Then we experiment with a gay scaled version of the plant village dataset, and finally we run all the experiments on the version of plant village dataset where the leaves were segmented.

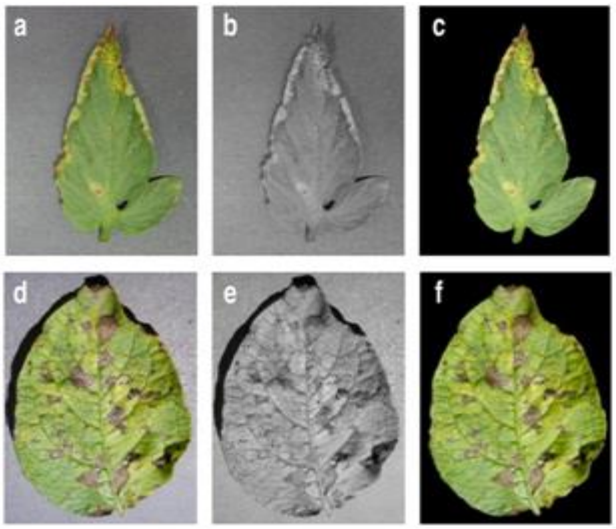

Fig.1.Example Of Dataset Image.

Hence removing all the extra background information, which might have the potential to introduce some inherent bias in the dataset due to the regularized process of data collection. We chose a technique based on set of masks generated by analysis of the color, lightness and saturation components of different parts of the images in several color space.

IV. COMMON DISEASES IN LEAF

A. Leaf Blotch

It has small oval and rectangular or irregular brown spots appear on leaves and it will become dirty brown as shown in Figure 2. The disease is controlled by use of Mancozeb pesticides.

B. Leaf spots

It causes greyish or whitish spots with brown boundary of different sizes which appear on the upper surface of leaves and the spots are greyish or whitish dark in the center. Due to leaf spot, leaves will get dry and died as shown in Figure 3. The disease is controlled by use of Zineb or Bordeaux pesticides.

C. Target Spot

This disease produces tan to brown color spot that have concentric rings like a bull's-eye as shown in Figure 4. Infected plant may look healthy from the top, so it is important to check lower leaves, where the first spot usually appears. It will start with only a few spots but after that, the disease will progress with more infection, and it does not take so much time spread on plant.

V. LEAF DISEASE DETECTION ARCHITECTURE

To get a sense of how our approaches will perform on new unseen data and also to keep a track of if any of our approaches are overfitting. we run all our experiments across

Fig.2.leaf affected with Blotch.

Fig.3.leaf affected with Spots.

whole range of train test splits namely 80-20(80It must be noted that in many cases, the plant village dataset has multiple images of same leaf (taken from different orientation). We have mappings of such cases for 41,112 images out of 54,306 images and during all these test train splits, we make sure all the images of same leaf go either training or testing set.

A. Steps Involved in Supervised Learning

- First Determine the type of training dataset.
- Collect/Gather the labelled training data.
- Split the training dataset into training dataset, test dataset, and validation dataset.

Fig 4. leaf affected with Target Spot.

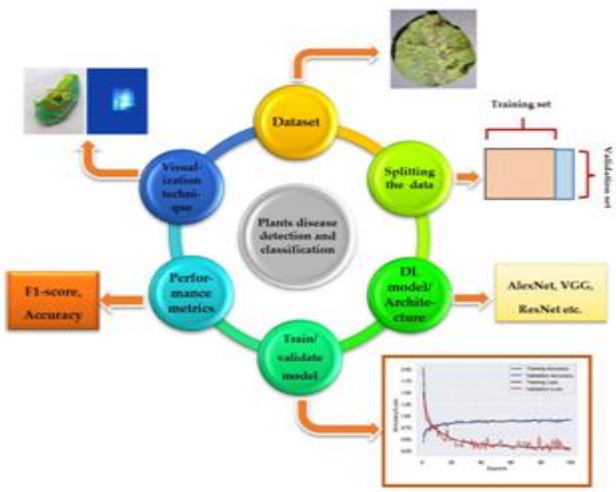

Fig 5. various steps involved in disease detection Architecture..

- Determine the input features of the training dataset, which should have enough knowledge so that the model can accurately predict the output.
- Determine the suitable algorithm for the model, such as support vector machine, decision tree, etc.
- Execute the algorithm on the training dataset. Sometimes we need validation sets as the control parameters, which are the subset of training datasets.
- Evaluate the accuracy of the model by providing the test set. If the model predicts the correct output, which means our model is accurate.

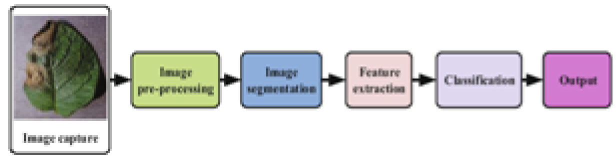

Fig. 6. flow diagram showing the process involved in leaf disease detection.

VI. IMAGE PREPROCESSING TECHNIQUES

- Grayscale conversion
- Normalization
- Data Augmentation
- Image standardization

A. Gray scale conversion

Grayscale conversion Grayscale is simply converting images from colored to black and white. It is normally used to reduce computation complexity in machine learning algorithms. Since most pictures don't need color to be recognized, it is wise to use grayscale, which reduces the number of pixels in an image, thus, reducing the computations required.

B. Normalization

Also referred to as data re-scaling, it is the process of projecting image data pixels (intensity) to a predefined range (usually (0,1) or (-1, 1)). This is commonly used on different data formats, and you want to normalize all of them to apply the same algorithms over them.

Normalization is usually applied to convert an image's pixel values to a typical or more familiar sense.
Its benefits include:

- Fairness across all images - For example, scaling all images to an equal range of [0,1] or [-1,1] allows all images to contribute equally to the total loss rather than when other images have high and low pixels ranges give strong and weak loss, respectively.
- Provides a standard learning rate - Since high pixel images require a low learning rate and low pixel images high learning rate, re-scaling helps provide a standard learning rate for all images.

C. Data Augmentation

Data augmentation is the process of making minor al iterations to existing data to increase its diversity without collecting new data.

It is a technique used for enlarging a dataset. Standard data augmentation techniques include horizontal vertical flipping, rotation, cropping, shearing, etc.

Performing data augmentation helps in preventing a neural network from learning irrelevant features. This results in better model performance.

Standard data augmentation techniques include horizontal vertical flipping, rotation, cropping, shearing, etc.

There are two types of augmentation:

Fig.7. leaf image BRG TO RGB.

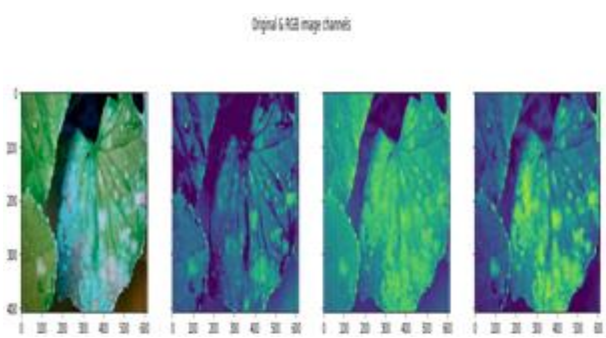

Fig. 8. LEAF IMAGE THROUGH DIFFERENT RGB CHANNELS.

1. Offline augmentation - Used for small datasets. It is applied in the data preprocessing step. We will be covering this augmentation in this tutorial.
2. Online augmentation- Used for large datasets. It is normally applied in real-time.

D. standardization

Standardization is a method that scales and preprocesses images to have similar heights and widths. Its rescales data to have a standard deviation of 1 (unit variance) and a mean of 0. Standardization helps to improve the quality and consistency of data.

VII. LITERATURE SURVEY

[1]. Amar Kumar Deyaa, Manisha Sharma, M.R. Mesh ram, 2015. This paper provides a survey to study in different image processing techniques used for studding leaf diseases. This is an innovative approach ever done for extracting disease features of the leaf. The methodology uses a blend of machine

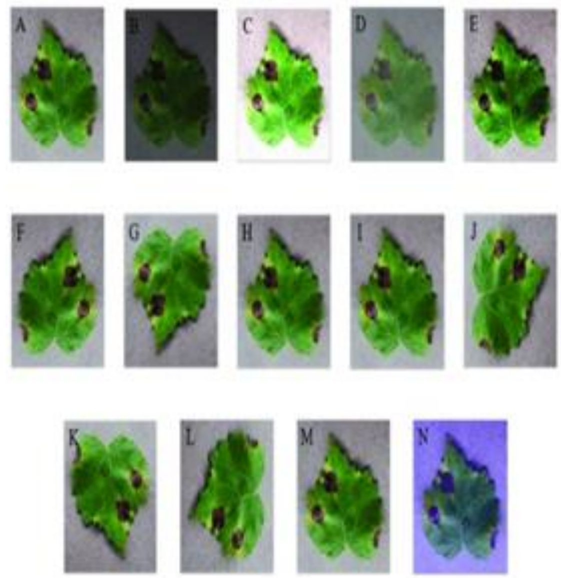

Fig. 9. Data augmentation of grape leaf disease images. (A) Original image; (B) low brightness; (C) high brightness; (D) low contrast; (E) high contrast; (F) vertical flip; (G) horizontal flip; (H) low sharpness; (I) high sharpness; (J) 90 • rotate; (K) 180 • rotate; (L) 270 • rotate; (M) Gaussian noise; (N) PCA jittering.

vision and machine intelligence for precision agriculture. In machine vision part, image processing is used where the leaf details, the disease infected area will be extracted. This is a small contribution towards agriculture and growing this medicinally valued precious plant species, to boost up the national economy as well as the national employment generation through proper exploitation of betel vine crop

[2]. Prof. Sanjay B. Dhaygude, Mr. Nitin Kumbha, 2011 The detection of plant leaf is a very important factor to prevent serious outbreak. Automatic detection of plant disease is essential research topic. Most plant diseases are caused by fungi, bacteria, and viruses. Fungi are identified primarily from their morphology, with emphasis placed on their reproductive structures. Bacteria are considered more primitive than fungi and generally have simpler life cycles. With few exceptions, bacteria exist as single cells and increase

in numbers by dividing into two cells during a process called binary fission Viruses are extremely tiny particles consisting of protein and genetic material with no associated protein.

[3]. Meghna's, Nivedita C.R, 2012 In this paper detection of leaf diseases has been used method is threshold: 1) identifying the infected object based upon k- means clustering; 2) extracting the features set of the infected objects using color co-occurrence methodology for texture analysis; 3) detecting and classifying the type of disease using NNs, moreover, the presented scheme classifies the plant leaves into infected and not-infected classes.

[4] Piyush Chaudhary, Anand K Chaudhary, 2010 In this paper a comparison of the effect of CIELAB, HSI and YCbCr color space in the process of disease spot detection is done. All these color models are compared and finally 'component of CIELAB color model is used.

[5] Arivazhagan, S. and Ligi, S. V. in 2018 proposed a framework based on automated deeplearningfor the recognition and classification of various diseases inmango plants. The dataset utilized for this framework consists of 1200 images which in- clude both diseased and healthy leaves of mango. The accuracy obtained from the proposed framework is 96.67 %Oppenheim,

[6] D. and Shani G. in 2017 proposed a framework based on convolutional neural network architecture forthe recogni- tion and classification of various diseasesinpotato plants. The dataset utilized for this framework consists of 2465 potato images.barbedo,J.G.A in 2018

[7] barbedo,J.G.A in 2018 investigated and identified the pros and cons through various factors that affect the model and effeciency of deep learning model which are used for the recognition as well as the classification of various plant diseases.The investigation carried out with the image database consist of of 50000 images of various plants diseases.

[8] Brahmi,M,et al.in 2017proposed a framework based on convolutional neural network for the detection o and classifi- cation of various diseases in tomato crop. The dataset utilized for this framework consists of 14828 tomato leaf images with almost nine diseases from the plant village database. the proposed framework able to achieve an accuracy of 99.18 % [9]Shrivastava,Vk,et al .in 2019 focused on the detection and classification of vvarious diseases in the rice plants using a framework with the aid of cnn architecture along with SVM.The framework was implemented on the dataset that consists of 619 rice plants leaf images with all four categories of diseases.the accuracies are evaluated for various proportions of training and testing dtaset and the maximum accuracy is achieved is 91.37%

[10].Ozguven,M.M and Adem,K .in 2019updated an ex- isting faster region based CNN architecture by varying the parameters for the identification of disease affected regions in the case of sugar beet. The dataset consist of 155 sugar beet images and an accuracy rate of 95.48 % achieved using the proposed framework.

[11] Uguz, S. and Uysal, N. in 2020 considered a comparison of a transfer learning scenario with CNN architectures such as VGG-16 and VGG-19 along with proposed CNN architectures in case of Olive plant diseases. The framework implemented on the dataset consists of 3400 Olive plant leaf images. In this framework, a data augmentation methodology was implemented for improving the size of the dataset. Before data augmentation, the accuracy attained about 88%

VIII. CONCLUSION

This study summarizes various image preprocessing methodologies used for identification of leaf disease detection such as data augmentation, standardization etc. This approach can significantly support an accurate detection of leaf disease. There are five steps for the leaf disease identification which are said to be image acquisition, image pre-processing, segmen tation, feature extraction, classification. By computing amount of disease present in the leaf, we can use sufficient number of pesticides to effectively control the pests in turn the crop yield will be increased. We can extend this approach by using different algorithms for segmentation, classification. By using this concept, the disease identification is done for all kinds of leafs and also the user can know the affected area of leaf in percentage by identifying the disease properly the user can rectify the problem very easy and with less cost.

IX. FUTURE SCOPE

In future, the existing algorithms can also be utilized in outdoor requirements along with the combination of leaf front and leaf backs into a common data set Further research can be devoted to model the detection and recognition techniques under mixed lighting conditions. It is of great significance to confirm that the accuracy of classification can be held under these conditions or not.

An unexplored mixture of feature extraction, selection and learning methods can also be used to increase the effectiveness of detection and classification techniques The future work can also be dedicated to the automatic estimation of severity of the detected diseases The instant solutions can be made available to the farmers by designing the mobile based applications. Online solutions related to plants diseases can be provided by using web portals. It is also required to increase the number of data for training and testing purposes to achieve better accuracy. Existing work can also be extended to achieve high speed and accuracy by developing the advanced algorithms.

REFERENCES

[1] Piyush Chaudhary, Anand K. Chaudhari, Dr. A. N. Cheeran and Sharda Godara, "Color Transform Based Approch For Disease Spot Detection on Plant Leaf", International Journal of Computer Science and Telecom- munications, 2012, pp. 65-70.

[2] Ms. Rupali S. Zambre, "Classification of Cotton Leaf Spot Disease Using Support Vector Machine", International Journal of Engineering Research and Applications, 2014, pp. 92-97.

[3] Kandiannan, Thankamani, Srinivasan, Rajeev, "Turmeric Plant", Indian Institute of Spices Research, Indian Council of Agricultural Research, Kerala, 2008,3-7.

[4] P. Revathi and M. Hemalatha, "Advance Computing Enrichment Eval- uation of Cotton Leaf Spot Disease Detection Using Image Edge detection", IEEE, 2012.pp. No 88-89

[5] Amar Kumar Deya*, Manisha Sharmaa , M.R.Meshram' A Survey on Plant Disease Identification',IJCST, Dec 2015.

[6] Johannes, Alexander, Artzai Picon, Aitor Alvarez-Gila, Jone Echazarra, Sergio Rodriguez-Vaamonde, Ana D´ıez Navajas, and Amaia Ortiz- Barredo. "Automatic plant disease diagnosis using mobile capture devices, applied on a wheat use case." Computers and electronics in agriculture 138: 200-209.2017

[7] M. Young, The Technical Writer's Handbook. Mill Valley, CA: Univer- sity Science, 1989.

[8] Nagasubramanian, Koushik, Sarah Jones, Asheesh K. Singh, Arti Singh, Baskar Ganapathysubramanian, and Soumik Sarkar. "Explaining hy- perspectral imaging based plant disease identification: 3D CNN and saliency maps." arXiv preprint arXiv: 1804.08831 2018.

[9] Slodojevic, Srdjan, Marko Arsenovic, Andras Anderla, Dubravko Culi- brk, and Darko Stefanovic. "Deep neural networks based recognition of plant diseases by leaf image classification." Computational intelligence and neuroscience 2016.

[10] Ehler, L. E. (2006). Integrated pest management (ipm): definition, historical development and implementation, and the other ipm. Pest Manag. Sci. 62, 787–789. doi: 10.1002/ps.1247,

[11] Deng, J., Dong, W., Socher, R., Li, L.-J., Li, K., and Fei-Fei L. (2009). "Imagenet: A large-scale hierarchical image database," in Computer Vision and Pattern Recognition,2009. CVPR 2009. IEEE Conference on. (IEEE).

[12] Erhan, Dumitru, Christian Szegedy, Alexander Toshev, and Dragomir Anguelov. "Scalable object detection using deep neural networks." In Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 2147-2154. 2014.

[13] Kaiming He, Xiangyu Zhang, Shaoqing Ren, Jian Sun, "Deep Residual Learning for Image Recognition", 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), DOI:10.1109/CVPR.2016.90. b14 A. G. Howard et al. MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications. Cornell University Library. 17 Apr 2017

[14] Mark Sandler, Andrew Howard, Menglong Zhu, Andrey Zhmoginov, Liang-Chieh Chen, "MobileNetV2: Inverted Residuals and Linear Bot- tlenecks", 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, DOI: 10.1109/CVPR.2018.00474 b16Ross Girshick, Jeff Donahue, Trevor Darrell, Jitendra Malik, "Rich Feature Hierarchies for Accurate Object Detection and Semantic Segmentation", 2014 IEEE Conference on Computer Vision and Pattern Recognition

[15] joseph Redmon, Ali Farhadi, "YOLO9000: Better, Faster, Stronger", 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), DOI:10.1109/CVPR.2017.690.

[16] Tsung-Yi Lin, Piotr Dollar, Ross Girshick, Kaiming He, Bharath Hariha- ran, Serge Be- ´ longie, "Feature Pyramid Networks for Object Detec- tion", 2017 IEEE Conference on Computer Vision and Pattern Recogni- tion (CVPR), DOI:10.1109/CVPR.2017.106. b19Jonathan Huang, Vivek Rathod, Chen Sun, Menglong Zhu,Anoop Korattikara,Alireza Fathi, Ian Fischer, Zbigniew Wojna, Yang Song, Sergio Guadarrama, Kevin Murphy, "Speed/Accuracy Trade-Offs for Modern Convolutional Object Detectors",2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), DOI:10.1109/CVPR.2017.351