Volume 7, Issue 6, June – 2022 International Journal of Innovative Science and Research Technology
ISSN No:-2456-2165

Analysis of Feature Extraction Methods for Predicting Plant Diseases

K. Sree Lekha
Department of Electronics and Computer Engineering,
Sreenidhi Institute of Science and Technology,
Hyderabad, India

Chinthala Karthik
Department of Electronics and Computer Engineering,
Sreenidhi Institute of Science and Technology,
Hyderabad, India

Jakka Manideep Reddy
Department of Electronics and Computer Engineering,
Sreenidhi Institute of Science and Technology,
Hyderabad, India

Katkuri Manoj
Department of Electronics and Computer Engineering,
Sreenidhi Institute of Science and Technology,
Hyderabad, India

Abstract:- Plant diseases have a catastrophic impact on the food production industry. Plant diseases lead to reduced quality and quantity of the produce. It also leads to huge losses for the farmers as well. There are various types of plant diseases which affect different plants in different ways. Countries where these plant diseases were not identified at an early stage have been heavily affected in the past. Thus, a quick and automatic identification of these plant diseases is highly desired. Quick identification will help in the appropriate diagnosis and will help reduce any loss. Thus, the automatic identification and diagnosis of plant diseases are highly desired in the field of agricultural information. Thus, a quick and automatic identification of these plant diseases is highly desired. Quick identification will help in the appropriate diagnosis and will help reduce any loss. Many methods have been proposed for solving this task, where machine learning is becoming the preferred method due to the impressive performance. In this work, we study the use of GLCM features extraction technique followed by application of various machine learning techniques.

Keywords:- *Augumentation; Filtering; Segmentation; Feature Extraction; GLCM.*

I. INTRODUCTION

The occurrence of plant diseases has negative effects on agricultural production, and if the plant diseases are not detected in time, there will be an increase in food insecurity. The early warning and forecast are the basis of effective prevention and control for plant diseases. They play crucial roles in the management and decision-making for agricultural production. Automatic recognition of plant diseases is an essential research topic, as it may prove benefits in monitoring large fields of crops.

II. LITERATURE SURVEY

The crucial role to prevent losses in agricultural product yield and quantity is early identification of plant diseases. It necessitates a great deal of effort, as well as knowledge of plant diseases and long processing times. As a consequence, image processing is used to identify plant diseases. Picture collection, image pre-processing, and disease detection are all stages in the disease detection process. It has been clarified how texture statistics can be used to detect plant leaf disease. Since HSV is a strong colour descriptor, the RGB structure is first transformed into HSV space using colour transformation. Green pixels are masked and extracted using a pre-calculated threshold stage. Then, using a 32X32 patch size, segmentation is performed and useful segments are obtained. The colour co-occurrence matrix is used to evaluate the texture of these segments. Finally, when the texture parameters of the regular leaf are compared. This research is focused on Image Edge Detection Segmentation techniques, in which the collected images are first processed for image enhancement. Then, to achieve target regions, R, G, and B colour attribute image segmentation is performed (disease spots). Following that, image features such as border, form, colour, and texture are extracted for disease spots in order to identify diseases and control pests. In this project, we implement and test different image processing techniques for automatically detecting and classifying plant diseases using image processing. Farmers in India have less access to agricultural specialists. Farmers sometimes receive expert answers to their questions too late because they are delayed. Most of the previous techniques Detection and classification technique of diseases in plant leaves using k-means clustering, Naïve Bayesian classifier, Color Co-Occurrence method[2]. In many cases the existing system even follows SVM Classifier using K-means Clustering[3]. Some of the projects even used a genetic algorithm, Arduino, Masking the green pixel and color co-occurrence method. Some papers even referred to FUZZY Classification, Support Vector Machine (SVM), K-Means algorithm color co-occurrence method[4]. for all these algorithms the accuracy is less and efficiency is not so good. Thus, we have implemented feature extraction followed by classification using various classification algorithms. Although having higher efficiency, some models cannot detect images which are brighter than usual.[6]. The goal of this project is to analyze various feature extraction image processing algorithms that can help or enrich the ease of automatically detecting diseases or other factors that may impact crops from color, texture and shape, and deliver rapid and accurate remedies to farmers.

III. ARCHITECTURE

Initially we took 700 images of real time field visited pepper plant leaves. The developer augments these images into 14,000 images consisting of diseased and healthy plant leaves collected under controlled conditions. Then the developer uses various filtering techniques to 9 these augmented images for tasks like noise reduction. After applying various filtering techniques, we implement segmentation techniques. Machine Learning Developer implements feature extraction for the collected data in order to reduce or remove unnecessary data. Feature extraction is generally used for dimensionality reduction. And then ML developers build the model and train the data using different algorithms and predict whether the plant is healthy or diseased.

A class diagram is of static type. The first row describes the class name. The second row describes properties of the class which are also called attributes. The third row describes methods and operations to be performed. It shows how various classes of a system are related to each other. There are 5 classes: Wrangle data, Filtering methods, Segmentation methods, Feature Extraction and Model.

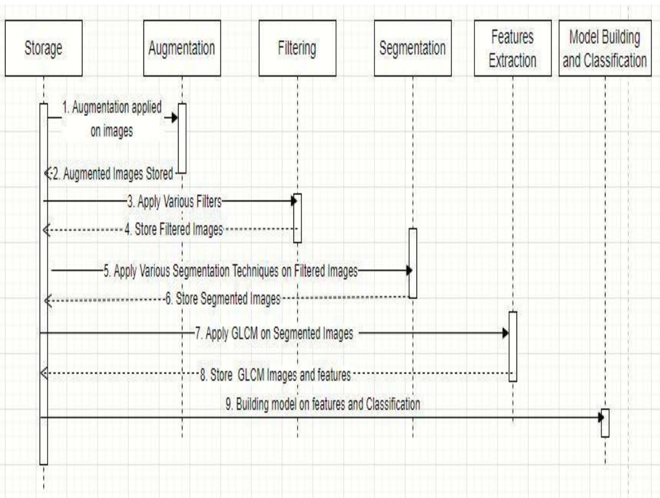

Fig 1 Class Diagram

A sequence diagram explains the relation between various objects in an orderly and timely manner. Initially there are 700 images of real time field visited pepper plant leaves. We augment these images into 14,000 images and the augmented images are stored. Now we perform various filtering techniques and again the filtered images are stored. The best filter is selected based on image quality metrics values. After selecting the best filter we implement segmentation techniques and segmented images are stored. We select the best segmentation technique based on the PSNR values. After applying feature extraction and saving images to storage and then performing classification.

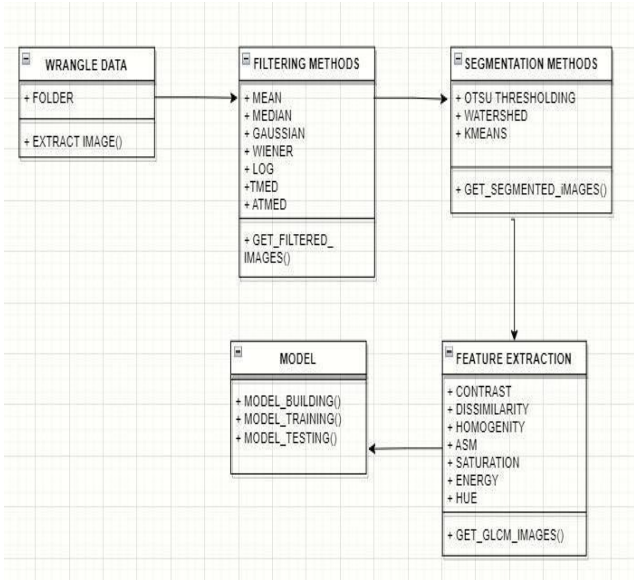

Fig 2 Sequence Diagram

The control flow of the system is represented by an activity diagram. First step is reading the images from the dataset. The dataset is the real time field visited pepper plant leaves and public dataset from Kaggle. Then applying various filtering techniques and computing various image quality metrics. The best filter is selected based on the image quality metrics and implements the image segmentation process. Again compute the image quality metrics and finalize the segmentation algorithm. After that applying GLCM feature extraction and then applying various ML classification algorithms.

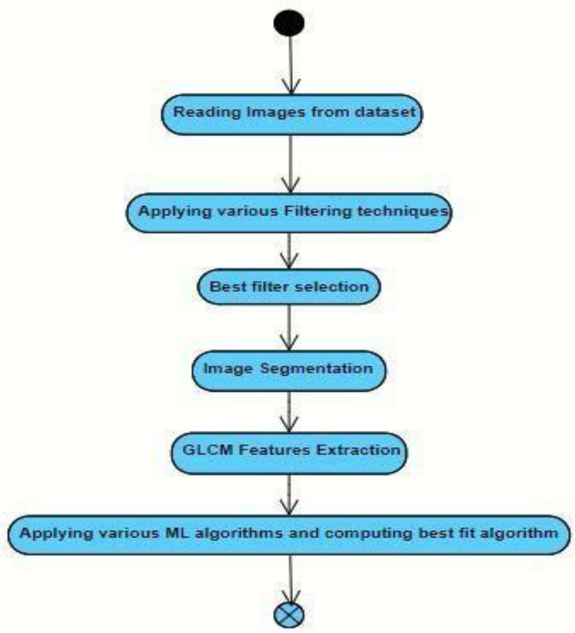

Fig .3 Activity Diagram

IV. IMPLEMENTATION

Image Data augmentation techniques are used to increase the training dataset artificially by modifying the versions of images in the collected data. Here we took 700 images of real time field visited pepper plant leaves and then augmented them to 14,000 images. Using 14,000 images consisting of diseased and healthy plant leaves collected under controlled conditions.. Then we implemented various filtering techniques on it Any image has a chance of having a noise. Even in the images that we have collected, there might be noise for some or the other reason. There are many image pre-processing techniques that help in removing the noise and give good performance by reducing noise. In total we have used 15 image pre-processing filters. Among those filters we select the best filter based on image quality metrics. We got best results for guassain and unsharp filters. These image processing filters help in restoring better features and reduce unwanted data/ noise in the image.

Gaussian Filter is a linear, non-uniform low pass filter used to eliminate noise that is drawn from a normal distribution. Gaussian function is given by Where σ is the standard deviation which is directly proportional to the amount of blurring, the mean of the distribution is assumed to be zero and the weight is chosen depending on the shape of the function. 2D gaussian function required to work with images is obtained by multiplying two 1D functions. This process is done by the image with a discrete approximation of 2D gaussian distribution function. It theoretically needs an infinite large kernel but it can be limited to contain only 23 values as most of the distribution falls within three standard deviations. Kernel is such a way that its coefficients have higher values in the center and decrease as it goes outwards and gets close to zero at the edge of the mask. These coefficients vary with the value of σ to maintain its gaussian nature.

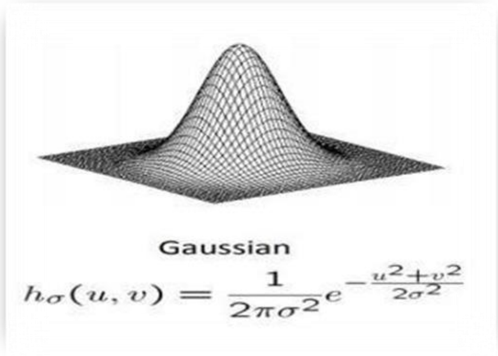

Fig 4 Gaussian Filter

Unsharp Masking Filter is an application of linear filtering which enhances the image by sharpening, i.e., to increase fine details (differences with respect to local averages), accentuating edges. The simplest way to design a filter that achieves this goal is by interpreting the output image as the sum of the input and a scaled detail image, the latter consisting in the details that are removed by smoothing fdetail = f − hσ * f, that is, The sharpening filter has the impulse response hsharp = δ + α (δ − hσ) = (1 + α) δ − αhσ, (α > 0) where δ (Delta) is implemented with a matrix that is zero everywhere except at the central entry, which is 1. The process includes creating a negative image to create a mask and then is combined with the original image to get an output image which is less blurry compared to the initial one.

$$f_{\text{sharp}} = \underbrace{f}_{\text{input}} + \alpha \underbrace{(f - h_\sigma \star f)}_{\text{detail}}.$$

Fig 5 Unsharp Masking Filter

After selecting the best image pre- processing filters based on the image quality metrics, we do image segmentation using various image segmentation techniques. We used 2 image segmentation techniques which are Otsu Thresholding, K Means. Among this segmentation we select the best segmentation technique based on the image quality metrics and used for further process. Otsu thresholding is an adaptive image binarization technique. The main objective of otsu thresholding is to find an optimum threshold value. K-means is based on clustering technique. It will help to identify the various segments by clustering into groups of similar pixel values. K defines the number of clusters. To start the algorithm a random K value and data points are assigned. Then all the other data points are assigned to a particular cluster. This process goes on until the values in the cluster keep repeating. Then on that basis the image is divided into segments. The higher the K value, more will be the number of clusters. Then we implemented the feature extraction process. Results are followed by different machine learning algorithms for prediction of disease.

Since we have huge dataset, applying feature extraction will reduce the redundant data and build more efficient model with most relevant features. By feature selection process, computational burden on the machine are going to be reduced, hence computational efficiency is increased. Gray level cooccurrence matrix is formulated to get statistical texture features. A number of texture features could also be extracted from the GLCM. Only four second order features namely angular moment , correlation, inverse difference moment, and entropy are computed. This method may be a way of extracting second order statistical texture features. The approach has been used by variety of applications, Third and better order textures consider the relationships among three or more pixels. These are theoretically possible but not mostly implemented due to calculation time and interpretation difficulty.

V. RESULTS

Analysis of various image processing techniques such as filters for noise reduction and segmentation are applied on both the public and private datasets. Public dataset consists of two classes: Diseased and Healthy. Various filtering techniques were applied on the public dataset to know their effect or efficiency, image quality metrics such as psnr, mse, ssim are calculated. Applying filters on Public Dataset: The mean Image quality metrics values obtained by comparing the output image after applying filtering techniques and original image are tabulated. Taking into regard the PSNR values for both Diseased and Healthy data, Unsharp masking filter and Gaussian filter values are highest denoting best performance. Similarly based on SSIM values: unsharp masking, Adaptive median and Gaussian filters in the corresponding order are most suitable. Taking into consideration all the three image quality metrics, Unsharp masking filter and Gaussian filter give the best performance.

Accuracy of the RF on test set: 0.862

	precision	recall	f1-score	support
0	0.85	0.85	0.85	385
1	0.87	0.87	0.87	442
accuracy			0.86	827
macro avg	0.86	0.86	0.86	827
weighted avg	0.86	0.86	0.86	827

Table 1 Filtering Results

Applying Segmentation on Public Dataset The Public Dataset consists of Healthy and Diseased data. Image segmentation methods such as Otsu thresholding and K-Means clustering are applied on the Unsharp masking and gaussian filtered data and their PSNR values is tabulated.

PUBLIC DATA SET			
FILTER	**PSNR**	**SSIM**	**MSE**
guassian	30.905	0.659	212.480
unsharp	33.701	0.937	97.383

Table 2 Segmentation Results

Then we performed feature extraction on which machine learning algorithms are applied for prediction. Later using gridsearchcv by applying to best accuracy classifiers so that we can get the best parameters. So after doing that we got best accuracy for random forest i.e. 86.2% compared to gradient boot i.e. 86.2%. So we considered random forest classifier is the best fit for this problem.

	PUBLIC DATASET	
Segmentation	**Guassian**	**Unsharp**
Otsu Thresholding	30.21	30.13
K-Means	31.22	31.42

Table 3 Prediction Results

VI. CONCLUSION AND FUTURE SCOPE

This project gives us an insight on the various filtering and segmentation techniques which can be used in detecting Plant diseases (Pepper Plant). We have used various Filtering techniques and analyzed them with the help of the various image quality metrics. Later we applied GLCM Feature Extraction to get the features and then performed Classification. We were able to select the best filtering technique and apply segmentation techniques on them. After Classification, we got best accuracy for Random Forest. Among the 15 filtering techniques we have used to process the image and denoise them, Gaussian and Unsharp mask filters were the best on all the data. Some filters performed well on a particular kind of data, but failed to perform on others. These filters have performed well and uniformly on both the Public and Private Datasets on all the classes present in them. Gaussian and unsharp masked data on both the Public and Private datasets is then followed by segmentation. Two segmentation techniques Otsu thresholding and K-Means clustering are used. Segmentation technique and the dataset that is being used for segmentation and the PSNR values (mean of all classes) are compared. It is quite evident that for both the public and private datasets if the filtering technique used is Gaussian then K-Means performs well. If the filtering technique used is Unsharp then Otsu thresholding performs well when looked at on an overall basis. However, the Unsharp masking filter stands at top and both segmentation techniques perform well on it. Suppose we want to extract features based on Haralick texture for example, then in that case gaussian blur is a standard filter which performs well universally. Therefore based on the next steps such as feature extraction and model building which could be performed on data the combination of which filter and segmentation to use depends. Therefore, for the dataset which we have used we can state that, We were able to select the best filtering technique and apply segmentation techniques on them. This will help in easier feature selection as the images have lesser gaussian, salt and pepper noise.

Leaves are the most vulnerable portion of the plant, they are the first to develop disease symptoms. From the beginning of their life cycle to the moment they are ready to be harvested, the crops must be monitored for illness. Image augmentation, pre-processing, segmentation are all processes in the disease detection process. The approaches for plant disease detection using photographs of the leaves were explained in this project. This project also covered certain segmentation algorithms for plant leaves. Image Analytics has been increasingly becoming popular and is being employed in various domains. One such domain is Agriculture. It can simplify various tasks involving human effort like weeding, disease identification, identification of plant growth stage etc. The scope of our project is that it can be used for automatic plant disease detection and this algorithm can be fed to an automated robotic system that can carry out the task. This Feature Extraction of Plant diseases data can be extended by using Feature Extraction methods which use neural network to detect and obtain features and thereby building machine learning or deep learning models. This model can be further deployed in Robotic systems such as drones, Field robots which detect plant diseases on field. It can be even deployed in a mobile application which will be helpful for farmers. It can reduce some of the struggles of the farmers and this system has a scope to become more and more efficient with large amounts of image data.

REFERENCES

[1]. S Jeyalakshmi, R Radha, “An effective approach to feature extraction for classification of plant diseases using machine learning”, Indian Journal of Science and Technology, Volume: 13, Issue: 32,September 2020

[2]. N. Nandhini, R. Bhavani, “Feature Extraction for Diseased Leaf Image Classification using Machine Learning”, IEEE , 2020 International Conference on Computer Communication and Informatics (ICCCI).

[3]. Ch. Usha Kumari, S. Jeevan Prasad, G. Mounika, "Leaf Disease Detection: Feature Extraction with K-means clustering and Classification with ANN", IEEE, August 2019.
[4]. Faye Mohameth, Chen Bingcai, Kane Amath Sada, "Plant Disease Detection with Deep Learning and Feature Extraction Using Plant Village", Scientific Research, Journal of Computer and Communications , Vol.8 No.6, June 2020.
[5]. Gurleen Kaur Sandhu, Rajbir Kaur, "Plant Disease Detection Techniques: A Review", IEEE, July 2019.
[6]. Nisar Ahmad, Hafiz Muhammad Shahzad Asif, Gulshan Saleem, Muhammad Usman Younus, Sadia Anwar , Muhammad Rizwan Anjum, "Leaf Image-Based Plant Disease Identification Using Color and Texture Features", SpringerLink, 2021.