An Approach Towards Disease Detection in Potato Leaf – A Review

Vyanktesh Rampurkar [1], Kunika Shah[2], Mansi Todmal[2], Akanksha More[2], Indira Lakhotiya[2]

[1]*Asst. Prof., Department of Computer Engineering, VPKBIET, Baramati ,Pune, India.*

[2]*Student, Department of Computer Engineering, VPKBIET, Baramati, Pune, India.*

Corresponding Author: kunikass06@gmail.com

Abstract: - Potatoes are well known all over the world's people. But the fact is in the last few years the export and produce level is decreasing because of some serious disease of potato leaf. Manual detection of this diseases requires huge amount of time and manpower. The farmers also have to suffer for this reason. To distinguish these disorders from potato leaves, a deep learning model will be offered and this will be very beneficial for farmers. This study is primarily focused on images, so that a large number of images are required. Different deep learning and machine learning models are used for this research to detect the disease from images of the potato leaf. In this paper we review different models based on deep learning and machine learning for leaf disease detection.

Key Words: — ***VGG19, Logistic Regression, Early Blight, Late Blight, Transfer Learning, Neural Networks.***

I. INTRODUCTION

India is one of the rapidly growing countries, where agriculture is the backbone of the country's growth. As the global population grows at a rapid rate. agriculture is struggling to meet its needs. Furthermore, importance of knowledge of cultivation in the minds of younger generation must be instilled. There is a trait to food security due to climate change, pollinator declined , crops pest, lack of irrigation, etc. Both quality and quantity of the food produce reduces due to crop disease. These crop diseases have an effect on global food security and also has a negative impact on the small-scale farmer whose living depends on the safe cultivation.

Crop diseases can be monitored with our naked eyes as soon as they appear on the crops. Today this problem has an effective solution thanks to the advent of the internet and the field of the computer vision.

Manuscript revised December 16, 2022; accepted December 17, 2022. Date of publication December 18, 2022.

This paper available online at www.ijprse.com

ISSN (Online): 2582-7898; SJIF: 5.59

A mistaken diagnosis of plant disease can result in loss of production, time, resources and product quality. It's important to identify the state of the plant, whether it's critical or not for its effective cultivation.

There are different environmental anomalies such as fungi, water shortage, insects and weeds which can affect the crops. In order to boost productivity of crops, it is required for the farmer to take preventive steps.

This research aids in focusing on the visually targeted quality of crop. Today Artificial Intelligence advances have made it possible to identify plant diseases automatically from raw images. Here, we can use the concept of the Deep Learning, where Deep Learning learns system based on the neural networks. Advantages of Deep Learning is that it can automatically extract features from photos. This is beneficial for the purposed system. While preparation, neural network learns how to extract features.

A Famous deep learning model which is used in the research is CNN(Convolutional Neural Network) ,a multilayer feed forward neural network. The motivation for using CNN Model is that it has recently been primarily used in agriculture, With CNN's growing popularity and success in solving many problems related to agriculture and the fact that various research uses CNN to discuss various agricultural problems which exists today. Therefore, CNN is most popular and commonly used approach in agriculture research today. The

proposed system uses the trained model like logistic regression , to extract the relevant structure of the data-set .By doing this , we leverage previous learning and avoid starting from the scratch. There are different trained models available VGG16 , VGG19,Inception,Alexnet,etc.We Extracted features from images by passing them into the model and then the extracted features are now passed as input to various classifiers such as SVM, KNN, logistic Regression. After evaluating result from above approaches we found that logistic regression gave the highest accuracy.

II. LITERATURE SURVEY

2.1 POTATO LEAF DISEASE DETECTION USING DEEP LEARNING

As agricultural technology advances and artificial intelligence is used to diagnose plant diseases, it will be important to conduct relevant research for sustainable agricultural development. Various diseases such as late blight have a great impact on potato quality and quantity and the manual interpretation of these leaf diseases is very time consuming and tedious. As this requires an enormous amount of expertise, efficient and automatic detection of these diseases at the germination stage can help improve potato yields. Various models have been previously proposed to detect multiple plant diseases.

The paper, introduces a model, which uses pretrained models such as VGG19 for fine-tuning (transfer learning) to extract -related features from the dataset. The results were then observed using several classifiers. Among them, logistic regression outperformed the others with a significant difference in classification accuracy, reaching 97.8% on the test dataset. The paper uses transfer learning Technique to develop an automated system to diagnose and classify different potato leaf diseases which achieves classification accuracy of 97.8% over test dataset. The accuracy is increased by 2.8% over CNN model combined with multiclass classifier. This technique helps farmers in early detection of diseases on leaves which eventually helps in better yield production.

2.2 DETECTION OF POTATO DISEASES USING IMAGE SEGMENTATION AND MULTICLASS SUPPORT VECTOR MACHINE

Potato is one of the most cultivated crops in India. The most common disease observed in Potato leaves can be Phytophthora infestans (late blight) and Alternaria solani (early blight), if not treated on time cane result in substantial yield loss. Over these years the most common practices method is of naked eye observation by experts or the farmers. This method can be unfeasible due excessive requirement of time and labor. Hence an alternative is proposed, which is of image analysis tools.

This method turns out to be effective for continuous monitoring of plant status. It analyses the visible pattern of leaves. The image analysis combines with machine learning techniques which offers solution to issue of agricultural productivity and ensures food security. So, this approach is to develop an imaging and machine learning based effective and error free plant disease detection system for plant. The dataset taken for the paper is of Plant Village ((www.plantvillage.org).

This dataset is publicly available image dataset, which consists of 54306 images of diseased and healthy plant leaves of 14 crop species which are collected under controlled conditions. 300 images of potato leaves are analyzed which commonly has a spread of 3 different class labels assigned to them - 1) Late Blight affected potato leaf.2) Early Blight affected potato leaf. 3) Healthy or non-diseased potato leaf. Segmentations on leaf images are performed to reveal the latent characteristics of potato leaves for easy to understand. Masking out the background as well as the green region of the leaves. So that we can extract our region of interest (ROI) that only contains visible disease symptoms. Different diseases have specific color and texture properties for different regions. By training our multiclass model Support vector machine (svm) with these properties we detect and distinguish the disease. The proposed system is as follows according to flowchart.

Image Segmentations: Techniques was chosen based on set of masks generated by the analysis of color and luminosity components of different regions of the images in L*a*b color spaces. The region containing disease symptoms has significant difference in color and texture than the surrounding regions, we try to isolate them. The background pixels which add noise and redundancy to our region of interest are removed . Therefore, to extract leaf only images by masking out the background. Feature extraction: The three fundamental pattern elements used by human interpretation of images are spectral, textural, and contextual features. In this work, extracted 10 features from every leaf image of the dataset.

Database consists of 100 health leaves and 200 diseased leaves. Dataset is divided into two sets training set containing 180 images 60% and testing set containing 120 images 40% . For classification purpose multiclass SVM with

linear Kernal was utilized. For performance evaluation of the classification model, performance parameters such as accuracy, sensitivity, recall, and F1 score were calculated. At 60%-40% train test split , testing accuracy of the classification model is 95%, moreover to make the model more robust ,5 fold cross verification technique was applied while 93.7% accuracy was achieved.

2.3 Leaf Disease Detection: Feature Extraction with K-means clustering and Classification with ANN

India is the second largest producer of different types of crops such as rice wheat pulses and spices the production in India just for these crops is yield up to 30Due to multiple diseases affecting the plant crop production have turn to be a nightmare. It is important and required to detect automatically the symptoms of the diseases as early as possible as soon as they appear on the plant leaves. Of the major disease on a crop is a leaf spot. These leaves spots are caused due to bacteria fungi viruses etc. Therefore, detection of these plant diseases in an early stage is highly recommended for improving the economy of the farmer and cultivating high quality of crops and also for food production. Monetary crops from remote places for detecting the diseases play a key role for successful cultivation. Image processing with machine learning is one technique used for automatic diseases detection so the proposed system used the clustering techniques and artificial intelligence for automatic disease detection. The objective of this paper is detection and classification of cotton and tomato leaf. Different stages followed are: Stage one acquiring image from the database Stage 2 segmenting the image into cluster using k means clustering Stage 3 extracting the features from the segmented clusters Stage 4 with the help of extracted features the disease detected and classified While using image processing we start from image acquisition it is the process of collecting the images from the data set. the images are in the form of dot png.jpg.gpig.bmp.gif and so on. After the image acquisition part is done it can be segmented into different parts that is be perform clustering by using key means clustering method in this method the RGB colors pace transformation into *a*b colors space is done.

K means cluster algorithm in our system can be used to separate the stain part of the leaves and the healthy leaf region. Firstly, we check that the image is loaded into the matlab from the database and then we convert the RGB image into L*a*b Colors space. L * represent the lightness ,a * and b* represent the chromaticity layers. All the color information is in the a*,b* layer. Next step followed is clustering the variant colors. The selected image is classified into three regions by relocating each pixel into its nearest cluster which reduces the sum of distance is calculate the centroid of the clusters. Specific cluster consists of specific segments of leaf image. 3 clusters have index values which are used to label every pixel in the image using results from k means. The next step followed is creating a blank cell array to store the result of the clustering.

A. Feature extraction: feature extraction is performed on the newly formed segmented cluster and four steps are followed which are: 1. Entering the cluster number of the disease affected leaf part only .2. converting it into grey scale if image is creating the grey level ko occurrence crisis deriving statistics that is features from glcm. The derive features of the model are contrast correlation energy humidity mean standard deviation and variance. These features can help us as given as input to the classifier for the processing. the features are extracted from the disease affected leaf cluster. These features are used to detect and classify the leaf disease.

B. Classification: in this paper the classification part is done by using neural network tool. 7 extracted features such as contrast correlation energy mean standard deviation etc are given as an input to our neural network and the target data is given to the neural network as class vector. Simultaneously in this back propagation neural network was used to classify the data. Neural networks can give the performance plot confusion Matrix error histogram, call precision if one score after completion of training of the neural network.

In this proposed method the part of image processing is done using the matlab 2016 software. Image processing in our method is performed on the leaf images and features are extracted from the disease affected cluster with the help of k means clustering algorithm. In this paper the data is considered for cotton and tomato leaf. 20 cotton samples are taken for diseases that is bacterial leaf spot and target spot also 20 tomato samples are taken for septoria leaf spot and leaf mold disease. Cotton and tomato leaf diseases are classified with the help of ann.

Accuracy for four different bacterial diseases leaves spot target spot septoria leaves spot and leaf mold are 90% 80% 100% respectively and its average classification accuracy is 92.5%.

2.4 Plant Disease Detection using Machine Learning

The agriculturists in different parts of worlds may think that it is hard to differentiate different diseases occurring on their crops and it is not feasible to always go to agriculture office and then discovering what the disease is. Main objective of this system is to differentiate the illness occurred in plant by

looking at its morphology by image data handling and machine learning.

Distinguishing different plant diseases is sometimes tricky and troublesome. Techniques which are based on leaf image classification has shown impressive results in this domain.

In this paper random forest model is used for classifying images into diseased or healthy category. There are various phases of implementation which are dataset creation, data preprocessing, feature extraction, training the classifier and classification.

Random forests are a learning method which is used for regression, classification and some other tasks. Random forests are basically forests created by building multiple decision trees. It overcomes decision tree's disadvantage of overfitting training data set and handles both numerical and categorical data. The histogram of oriented gradients(HOG) is used for feature extraction. HOG is an element descriptor which is used as a part of image processing. Proposed system uses three component descriptors - Hu moments which is basically used to extract shapes of leaves, haralick texture which is used to extract texture of leaves and color histogram is used for identifying distribution of colour on leaf image. To find out whether the leaf is healthy or diseased there are certain steps that needs to be followed which are preprocessing, feature extraction, training the classifier and classification. Preprocessing an image is bringing all the images to uniform size.

The proposed method uses HOG for extracting features of a preprocessed image. It is a feature descriptor which is used for object detection. There are three descriptors which have been used in this system they are:

Hu moments: Hu moments helps in describing particular outline of leaf. Hu moments are calculated over a single channel hence first we need to convert RGB image to grayscale image. This will give an array of shape descriptors.

Haralick Texture : Usually healthy leaf and diseased leaf have different textures. Haralick texture is used for this purpose. For working of haralick texture it is needed to image be in the grayscale format. Haralick texture works on concept of adjacency matrix.

Color Histogram : Color histogram is used to identify distribution of color pixels in given image. RGB image needs to be converted into HSV because, HSV model closely aligns to how human eye will discern an image. Histogram plot gives us information about different pixels present in given colour ranges.

Among multiple machine learning algorithms random forest outperforms other like logistic regression, SVM, linear discriminant analysis while having a smaller number of image data set. The labeled datasets are partitioned as training and testing datasets. This method applies HOG feature extraction on training dataset and then random forest model is trained on it, later feature extraction is applied on test images and this feature vectors are applied to random forest model in order to test the classifier.

Main objective of this system was to detect abnormalities in leaf image. This model is trained over 170 images of Papaya leaves and achieves approximate 70\% accuracy. This accuracy can be increases by training the model on vast number of images and using local features together.

III. CONCLUSION

From below table it can be concluded that highest accuracy can be obtained when we are using convolutional neural network combined with logistic regression. Improvement in classification accuracy can help to farmers in predicting diseases early which can enhance crop production.

Table.1. Comparison of Classification Accuracy

Paper Name	Methods	Accuracy
Potato Leaf Disease Detection using Deep Learning [1]	VGG19 + Logistic Regression	97.8%
Detection of potato diseases using image segmentation and multiclass support vector machine [2]	CNN + Multiclass SVM	95%
Leaf Disease Detection: Feature Extraction with K means clustering and classification with ANN [3]	ANN + K-means Clustering	92.5%
Plant Disease Detection using Machine Learning [4]	Random Forest Model	70%

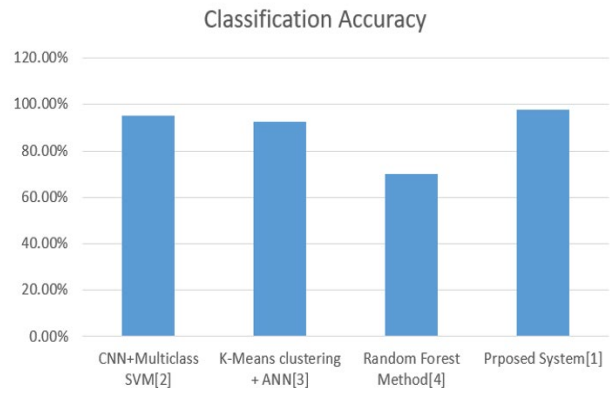

Fig.1. Bar Chart for Comparison of Classification Accuracies

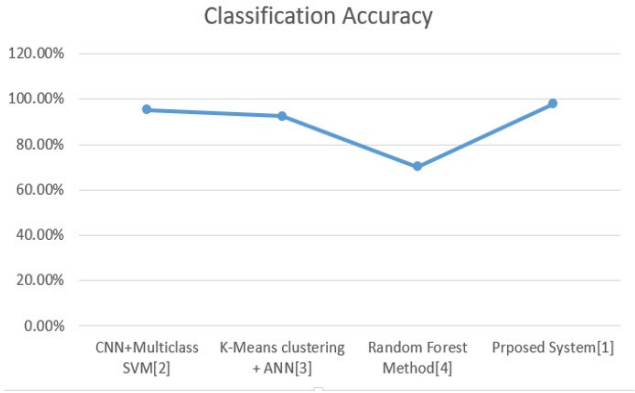

Fig.2. Line Chart for Comparison of Classification Accuracies

REFERENCES

[1]. D. Tiwari, M. Ashish, N. Gangwar, A. Sharma, S. Patel and S. Bhardwaj, "Potato Leaf Diseases Detection Using Deep Learning," 2020 4th International Conference on Intelligent Computing and Control Systems (ICICCS), 2020, pp. 461-466. M. Young, The Technical Writer's Handbook. Mill Valley, CA: University Science, 1989.

[2]. M. Islam, Anh Dinh, K. Wahid and P. Bhowmik, "Detection of potato diseases using image segmentation and multiclass support vector machine," 2017 IEEE 30th Canadian Conference on Electrical and Computer Engineering (CCECE), 2017, pp. 1-4.

[3]. C. U. Kumari, S. Jeevan Prasad and G. Mounika, "Leaf Disease Detection: Feature Extraction with K-means clustering and Classification with ANN," 2019 3rd International Conference on Computing Methodologies and Communication (ICCMC), 2019, pp. 1095-1098.

[4]. Ramesh Maniyath, Shima & V, Vinod & M, Niveditha & R, Pooja & N, Prasad & N, Shashank & Ram, Hebbar. (2018). "Plant Disease Detection Using Machine Learning." 41-45.

[5]. Chowdhury, Muhammad & Rahman, Tawsifur & Khandakar, Amith & Ibtehaz, Nabil & Khan, Aftab & Khan, Muhammad Salman & Al-Emadi, Nasser & Reaz, Mamun Bin Ibne & Islam, Mohammad & Ali, Sawal. (2021). "Tomato Leaf Diseases Detection Using Deep Learning Technique." 10.5772/intechopen.97319.

[6]. S. Ahmed, M. B. Hasan, T. Ahmed, M. R. K. Sony and M. H. Kabir, "Less is More: Lighter and Faster Deep Neural Architecture for Tomato Leaf Disease Classification," in IEEE Access, vol. 10.

[7]. H. Ajra, M. K. Nahar, L. Sarkar and M. S. Islam, "Disease Detection of Plant Leaf using Image Processing and CNN with Preventive Measures," 2020 Emerging Technology in Computing, Communication and Electronics (ETCCE),2020.

[8]. S. Barburiceanu, S. Meza, B. Orza, R. Malutan and R. Terebes, "Convolutional Neural Networks for Texture Feature Extraction. Applications to Leaf Disease Classification in Precision Agriculture," in IEEE Access, vol. 9, pp. 160085-160103, 2021.

[9]. V. K. Gajjar, A. K. Nambisan and K. L. Kosbar, "Plant Identification in a Combined-Imbalanced Leaf Dataset," in IEEE Access, vol. 10, pp. 37882-37891, 2022.

[10]. J. M. Al-Tuwaijari, M. A. Jasim and M. A. -B. Raheem, "Deep Learning Techniques Toward Advancement of Plant Leaf Diseases Detection," 2020 2nd Al-Noor International Conference for Science and Technology (NICST), 2020, pp. 7-12.