Crop/Plant Disease Detection Using Deep Learning Algorithms

Harshpal Singh, Dr. Anand Khandare, Amit Vishwakarma, Wilfred Almeida ,Rishabh Shukla
Department of Computer Engineering, Thakur College of Engineering &Technology, Mumbai, India
harshpal2026@gmail.com, anand.khandare@thakureducation.org
, amitvishwakarma11102001@gmail.com, almeidawilfred642@gmail.com, rishabhvinodshukla18@gmail.com

Abstract—**The very foundation of the mankind relies on the quality of foods and grains consumption and the plants /crops from where they originate. To ensure, the well being of the crop and save subsequent health hazards we have come up with the most ideal way to categorize the infected crop from the non infected ones which uses deep learning Convolution neural network as a classification model and the features of the crops are perceived through computer vision and image processing techniques.**

An ideal CNN model would be a cumulative result of all these techniques and to provide a significant accuracy of the crop for the farmer or whomsoever it may concern as manual prediction of the infected crop happens to be a tedious and a time consuming task. Moreover, chances of accuracy may seem to vary on a significant scale. To reduce such complexities, CNN yields an ample amount of accuracy and can be easily relied on to ensure safe and non hazardous cultivation.

Keywords)- — ***classification model, Convolution Neural Networks,*** **computer vision , *infected crop.***

I. INTRODUCTION

Agriculture and farming has been one of the oldest and most profitable professions until today's date. Computer and Technology has made an evolutionary change year by year in order to yield out maximum profit and also to reduce the human effort. Significant changes in farming techniques has been observed with increase in technology and have also led to the ease of farmers.

Although, there are still some factors leading to loss of cultivation and not only affecting the farmers wages and salaries but is also contributing to disruption of the eco- system. Some crop diseases are contagious and may also lead to infertility of the soil upon which the crop's foundation is based. These menace would gradually lead to poor fertility of the land and may no longer be hold enough moisture and other constituent nutrients required for an ideal growth. A huge amount of effort has been put in to reform the cultivation methods once again and make the work of the famers convenient to some extent this may not only inhibit the soil for withstanding the cultivation but also damage the plants in the nearby vicinity.

Infected plants can be hazardous to the consumer in a certain way wherein, they may not be fit enough for the consumption, these lost soil nutrients do not replenish quickly. Hence, not making them fit enough for harvesting throughout the year.

For the farmers, the ones who are experienced tend to solve the solution by adding appropriate fertilizers or any other remedy which would not be suffice the situation significantly and the crop has already been infected and these infected crops are sold which in turn breeds numerous diseases.

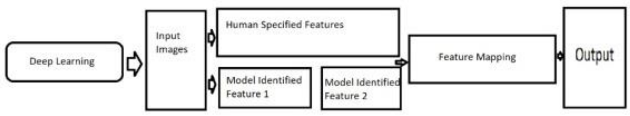

Figure 1: Deep Learning Approach

Therefore, to tackle this situation a technology has been introduced which would classify the infected crops using Convolution Neural Networks which comprises of several layers and each subsequent layer treats the previous layer's result as an input. This filtration goes through multiple stages before the final result is generated a diagrammatic representation i.e. Fig 1 illustrates the procedure.

Till now many researchers have shared multiple techniques for detecting disease from plant leaf using CNN. In the below section we have reviewed the others paper work and we have tried out to explain the recent advancement which are done in the CNN technique for disease detection. We hope our this work will be helpful for the people who are looking forward for the exploring of this field more.

Benefits of Convolution Neural Network for detection of crop disease:

1. ***Works well for internal representation of two dimensional images.***
2. ***It is self reliant and works well without any human intervention.***

We upon reviewing multiple researcher's paper have come up with few incites which would help us to identify which one is the best method of detection of crop disease.

II. LITERATURE SURVEY

In the paper[1], First, leaf images are captured, then image processing techniques are applied to extract features useful for disease detection.

A. Image analysis

To detect the disease, the image of the infected leaf needs to be examined in a series of ways. The input image is preprocessed and its features are extracted according to the dataset. Then, in order to classify the disease according to a specific data set, it is necessary to use some classification method.

B. Image capture

Image capture is the process of capturing and converting it to the desired output format. In this application, the analog image is first captured, then converted to a digital image for further processing.

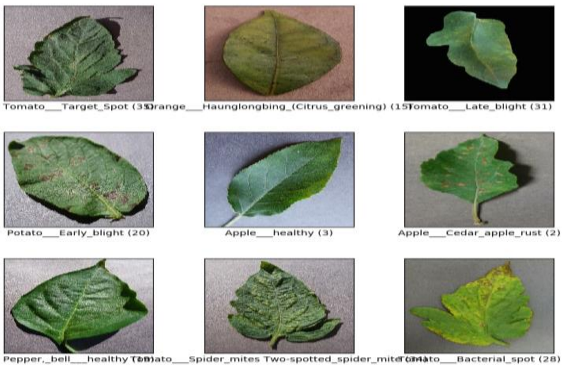

Figure 2: Training Dataset

C. Image preprocessing

The following pre-processing steps are performed on the captured image. Freeze frame or use active binarization, look-up table, or image plane separation to increase image contrast. Reduces image resolution degradation through binning. Rotate Image Converts a color image to a grayscale image.

D. Feature extraction

The goal of this phase is to extract features such as color and shape. Two geometric features, such as area and perimeter, are extracted from the binary segmentation image. Color feature is extracted from the color segmentation image. The color characteristics include the average gray value of the R /G/ B component, the distribution of the gray value of the R / G / B component, the R / G / B component, the color ratio in RGB color model, -H. / S / V component gray value average, H / S / V- gray value distribution H / S / V- component gray value component and skewness.

E. Leaf image classification

In the classification between the affected leaf, the classifier relied on Bayes' theorem, and SVM was used to distinguish between the classification and the affected leaf.] First, the captured images are classified into infected leaves and unaffected leaves. The color distribution is the same for the unaffected leaves, but the color distribution for the affected leaves is not uniform. This is because the pixel values of the affected leaves were completely different from the pixel values of the normal leaves.

The image quality of the leaf is improved by applying an averaging filter after the where image segmentation is performed by Otsu's threshold algorithm. After extracting features from a given leaf image, a recognizer is required to recognize the disease in the leaf image from the stored database. In this paper, we proposed a detection method using the back propagation network (BPN). Back propagation can train multilayer feed forward networks which consist of a forward pass and a backward pass as shown in Fig. 3. In the forward pass outputs are calculated and compared with preferred outputs.

Errors from preferred and actual output are calculated. In the backward pass this error is used to alter the weights in the network in order to reduce the dimension of the error. Forward and reverse paths are repeated until error is low. Normally, the user sets the value of the accepted error. When the trains the NN, it feeds the network a set of sample with inputs and desired outputs. If you choose a learning rate, your swing will help you adjust your weight.

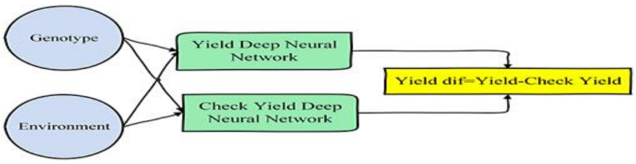

Figure 3: Crop Neural Network

The output layer contains neurons. Support Vector Machine (SVM) is a supervised machine learning algorithm that can be used for classification. This algorithm plots each data item as a point in n dimensional space n is the number of features you have with a feature value of for each leaf.

An approach based on image processing technology has been proposed in to help detect plant diseases. Disease detection is primarily the purpose of the proposed approach and can detect leaf disease with little computational effort. This proposed approach consists of four phases.

Accuracy is improved by using various image processing techniques such as image analysis, preprocessing, feature extraction and classification. Speed and accuracy are two main features of plant disease detection using deep learning techniques that must be achieved. With the proposed method, accuracy of up to 92% can be achieved. With the more featured SVM classifier, you can improve the recognition accuracy by significant amount. This approach can be used for applications such as disease classification of plant parts such as leaf .

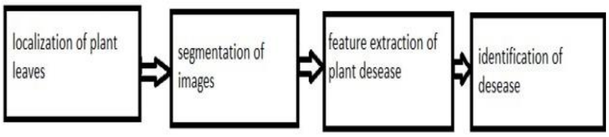

Figure 4: CNN Approach

III. PROBLEM STATEMENT

India being one of the major cultivating centres. Farmers grow varieties of crop in order to make out the most profit. But there is one of the factors which are yielding out low profit and that is crop disease which is not only hazardous but also contagious to the crops in the nearby vicinity. Hence spoiling the quality of the soil foundation .Hence a solution has been drawn out by creating an android application making using of neural networks and yielding out results for the same also making use of machine vision simultaneously.

Making use of CNN and also considering Vgg architecture with depths of 16 and 19 but it was found out to be vgg uses 3X3 convolution layers stack on top of each other and training those models where not feasible because of the issues with respect to convergence and training deeper networks Hence it was slow to train and consumed a lot of bandwidth .Hence to overcome these issues ResNet comes in ResNet does this by connecting nth layer to (n+x)th layer. This implies that Training of this form of network is easier than the other form of networks and also solves the problem of error rate as it decreases the error rate.

In [2] For early detection of disease, capture images taken days or weeks before plant disease becomes visually observable and train Model to detect plant disease early. It has been proposed to be used for. When implementing this framework, it is necessary to experiment with Combinations of different image bands, including bands beyond the visible spectrum, to determine the optimal band combination that can be used to model the deep learning train that can detect Plant diseases even before they are visible to the naked eye.

This is based on the assumption that the disease can be detected before it becomes visually observable, as the spectral reflectance of diseased leaves and canopies is known to be more pronounced in the invisible part of the spectrum. (Campbell & Wynne, 2011). In fact, recent research results show that infections in tomato Botrytis cinerea leaves can be detected 9 hours after infection (long before visual symptoms appear) using near-infrared and red-rimmed sensors. (Fahrentrapp et al., 2019). As shown in 3.2, pre- illness Images should be categorized according to the time steps (e.g., days) captured before the illness becomes visually observable. This shows how healthy and diseased crops are displayed in natural colors and in a virtual false color composite of bands extracted from aerial images.

The first row (RGB view 1) shows an image of a healthy crop, and the Image in the second row (RGB view 2) also shows a healthy crop, except for the last image (TN). As Shows, diseased crops are displayed in natural colors (RGB). The third line, (False color view), is an example of what the second line looks like. A combination of virtual false color bands in which existing disease classes exist. In this view, the first three images (T0 to T2) represent a view of a healthy herd of, and the next three Images (T3 to T5) are displayed in false colors, thus indicating the disease class. Represented, is the last image. (TN) displays the same diseased crop on the second line as it appears in this fictitious false color view. As an example of, only one pre-disease class is shown here, but many other classes can be identified during model experimentation and training.

The trained disease Classification model should be continually updated to include "newly discovered" ones. Early stage illness class. 3.3 shows the key components of the proposed framework for early detection and continuous monitoring of plant health. Rounded rectangles indicate common Machine learning tasks, and connection arrows indicate user workflows.

The green Rounded rectangle represents the new disease early detection task developed by this Study. This framework is intended for use by, for example, regional and national Government agencies to monitor the health of a wide range of plants. Such institutions are expected to monitor this The Seasonal crops are in the growing season and the perennial crops are annually maintained in the health of the plants of various crops.

The agencies should use trained crop Classification models at the beginning of the growing season to identify and map the Areas where specific crops are sown, and then deploy trained crop diseases classification Models to images covering only those areas planted with the identified crops.

In [3] The basic “structure of plant ailment detection system” includes numerous stages defined ultimately as: photo acquisition, labelling and pre-processing of statistics accompanied with the aid of using statistics augmentation. After this section education and trying out of dataset executed thru numerous getting to know strategies and class thru deep Convolutional neural networks.

Image Acquisition: In photo acquisition step, pix are obtained thru excessive decision virtual digital digicam or from numerous reassets of internet .Few fashionable datasets like Plant Village, APS etc. also are to be had for the researchers The pix are amassed beneath wild, managed or out of control conditions.

Preprocessing: Preprocessing is a way of casting off noises or distortions in a photo to enhance its quality. Various preprocessing techniques consists of Contrast Stretching, Noise filtering and Histogram adjustments etc. "Various filters like low pass, excessive pass, etc. are implemented to cast off numerous kinds of noises

Data augmentation: Since maximum of the datasets aren't to be had easily. In order to create huge datasets, statistics augmentation strategies implemented at the smaller datasets for enhancement. Big datasets reduces the hassle of over fitting.

Training and Testing of statistics: Training and trying out of statistics consists of splitting of datasets into numerous train-check splits. The version is educated through numerous getting to know strategies like baseline education, great tuning or switch getting to know.

Classification: Classification of illnesses is primarily based totally at the predefined dataset values. "For class section, unique type of system getting to know strategies are being hired to allocate a category to a fixed of unclassified statistics". The fundamental strategies used for class are Support Vector Machine (SVM), Neural Networks, Fuzzy Classifier, Linear Discriminant Analysis, KNN etc. Deep getting to know is a rising studies region with inside the area of system getting to know in addition to in convolutional neural network. It is a subset of system getting to know strategies. Various deep getting to know processes had been used for photo class and recognition. In deep getting to know, CNN is one of the maximum outstanding approach. It includes middle structures: a convolutional layer and a pooling layer. This paper concentrates in particular at the plants that have been categorized thru deep getting to know convolutional neural networks.

In [4] author has developed the model based on convolution neural network (CNN) for detecting the diseased plant images. They have taken the public dataset which is freely available on PlantVillage. 54,306 images of different plant including health and diseased dataset of 14 crop species on 26 diseased have been collected in this dataset.

When model was tested it gives the accuracy of 99.35% which is considered to be good accuracy overall. But when model was tested under different dataset compare with normal dataset it performs accuracy of 31.4% which was still good compare with traditional random selection.

The model was developed using two different deep learning CNN architectures that is AlexNet and GoogleNet. These architecture are most popular among CNN architectures for image processing and video processing. For getting the best accuracy different methods of training were also implemented that is TRAINING FROM SCRATCH for AlexNet and TRANSFER LEARNING for GoogleNet architecture. The transfer learning outperforms over training from scratch.

Three versions of plant image dataset was taken (color, greyscale, segmented) to test the model. And the 30 epochs(parameters) where used for feature extraction. The colored dataset gives the best accuracy compare with greyscale and segmented but to overcome the inherent biases of neural network in lightening conditions the model was tested on greyscale and segmented versions also. To check the role of a background on image dataset segmented version was taken. The segmented version performs good and gives greater accuracy as compared to grayscale version but perform less with colored version. And the grayscale version of dataset perform worst among all three version of dataset.

To test the models performance to any unseen data the four different training-testing approaches were implemented. That is 20%-80%(20% for training and 80% for testing), 40%-60%(40% for training and 60% for testing), 50%-50%(50% for training and 50% for testing), 60%-40%(60% for training and 40% for testing) and 80%- 20%(80% for training and 20% for testing). the accuracy was found out for AlexNet architecture is 97.36%, 98.60%, 98.96%, 99.07%, 99.27% respectively and for the

GoogleNet it was is 98.20%, 99.14 %, 99.16%, 99.24%, 99.34% respectively for the Colored version of dataset. We can see here as the training size increases the accuracy is also increased. And from here it can be seen that the GoogleNet architecture gives better result over the AlexNet architecture.

When the model was evaluated on the dataset other than PlantVillage it gives less accuracy of 31.4%. since the PlantVillage dataset taken under ideal condition the real leaf image will be directly taken in different circumstances in more complex environment. While the accuracy of 31.4% is still good with traditional random selection approach but more diverse image data can help to train the model. But when this model will be applied to different datasets then this will eventually increase it's performance and the accuracy will also increase. and another limitation is model only performs on the single leaves images on the homogeneous background only while it should also perform on the plant leaf itself.

In [4] the author has proposed a model based on RPN(reason propose network), chen-vese algorithm and transfer learning for developing the model for disease detection. RPN was used to localizing the leaves into complex background. Chen-vese algorithm for segmentation of images from RPN algorithm input and the transfer learning model which is pre-trained model on multiple diseased dataset with normal background for disease recognition. This model archives accuracy of 83.57%.

Around 1000 leaf image dataset in complex background is taken from the plant photo bank of China(PPBC) including many plant at various growth stage. And the PlantVillage public dataset is also being used here. And the transfer learning is pre-trained on 537 black rot, 1032 bacterial plaque disease, 293 rust and 2852 healthy leaves.

The plant disease recognition in the complex environment is performed with three steps.

1.desease leaves image segmentation.
2.feature extraction
3. disease identification

1. image segmentation : the segmentation of images in the complex background is a tedious task. So the author has also reviewed multiple approaches to overcome with this challenge.
delta E-color difference algorithm can be applied for separating the infected area on diseased leaves.

some of the researchers have used the region of interest(ROI) for segmentation of images. And another approach would be the leaves can be divided into multiple parts with diseased spots as a parameter for the segmentation process.

feature extraction: different different images have different texture, shape, motion-related attribute, color of leaves this puts up challenge for feature extraction of all the images. But several methods are easy to use and we also gets desired outputs.

- color and texture feature
- Dempster Shafer (D-S) theory for feature extraction and for prediction the result
- improved local binary pattern (LBP) for Greyscale image and processing over it.

Disease Identification: researchers have developed multiple methods to detect the disease info with images.

- Hybrid clustering For Detection
- CBIR(content-based image retrieval) for feature extraction and then SVM for classification.
- novel approaches such as with use of mobile device for image segmentation and them improved k-means clustering for disease information.
- trained CNN with 1632 corn images for detecting the disease of a corn leaf.
- improved CNN with AlexNet and VCG-16 net architecture for detecting the disease on maize leaves.

This is the general flow of model which is suggested for detecting the disease on plant.

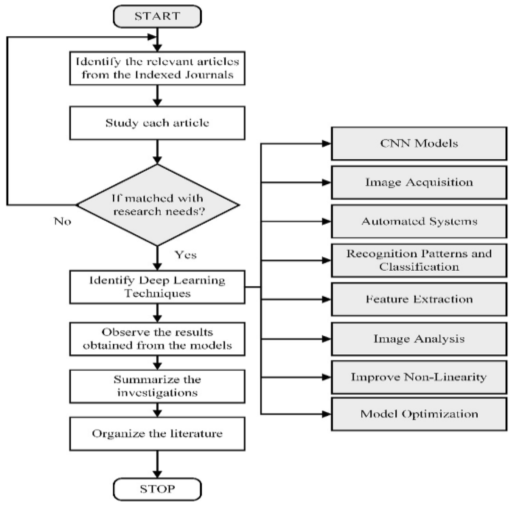

Figure 5: CNN Approach

Leaf localization: RPN algorithm is trained with convolution neural network to classify the boundary of image. Classification neural network is used to detect the leaf image present in picture is boundary or an object. Basically it uses intersection over union for detecting the boundary. If IOU is 0.5 then the image is object and for IOU greater then 0.5 will be background.

Leaf Segmentation: chen-vese algorithm is used here for the segmentation of leaf images.

Disease Identification: the model which is used for detecting the disease information is transfer learning. First it is pre- trained with the diseased images on simple background and then output of the previous method is tested for finding out the correct disease information.

RPN algorithm is used for the training of image dataset in complex environment and the CNN will be used for retrieving the diseased plant leaves and these inputs is given to chen-vese algorithm for the segmentation purpose and finally the output of these all steps is test under the transfer learning model for identifying the disease which is present on the leaf or plant.

In [6] the author has reviewed multiple research papers and presented current trends and difficulties which are present in the plant image disease detection with CNN.

In another method for the same apple leaf detection the BP neural network model was performed for classification and accuracy was found around 92.6%.with 63 parameter selected for feature extraction of rice leaf. And the Bayesian discriminant method was used for classification of diseases. The model has perform the classification with 97.2% of accuracy.

Traditional methods can detect the diseases with good accuracy but they lack in detecting with more diverse dataset and also early detection of diseases was not specified in the traditional approaches. instead of taking whole leaf images the single spots or lensions can be used with augmented data. And when this applied on GoogleNet architecture the model is able to detect disease with accuracy of 94% which is greater than when using whole images 82%. another method was to focuses on diseases diseased area. And when Mask R-CNN model with [ResNet 50 or ResNet 101 feature extraction] was used to detect the diseases. the accuracy was found out to 92.01% on test dataset.transfer learning gives more accuracy over training from scratch on AlexNet architecture for detecting the diseases correctly. The accuracy can reach upto 95.53%.

Corn leaf diseases detection in the complex background with small sample dataset with CNN (VGG 16) on transfer learning model performs 95.33% of accuracy.

ResNet 50 pre-trained on ImageNet dataset when used for identifying the apple leaf diseases gives accuracy of 97% for simple background and accuracy of 51% for a complex background images.

New DL architectures for leaf detection

When multiple CNN classifiers was combined and used the accuracy was high compare with single and double classifiers. And the accuracy was 97.8%.

for some tomato leaf with problems of having shadow, occlusion and light integrity on images the Deconvolutional Guided VGG network can identify plant diseases with 99.10% accuracy.

In ResNet18 architecture when convolutional 1 layer of ResNet18 is combined with multiple convolution kernel & SETNet module addition the model is able to solve low recognition issue of grapes due to different degree of diseases and this model achieves accuracy of 90.83%.another model which is developed on CNN with adding 3 convolutional layer, 3 pooling layer, 2 fully connected layer it is able to detect different diseases on tomato leaves. And this developed model has an accuracy of 91.2%.

IV. ACKNOWLEDGEMENT

We the students of TCET, with a collaborative effort with Dr. Anand Khandare we have been successful in designing a full length research paper with topic named **Crop Disease Detection Using Deep Learning.** We feel delighted and would also like to thank all the concerned faculties for teaching us the basic structure and overview of the Deep Learning models and convolutional neural network with the help of which we were able to draw out meaningful conclusions of the numerous authors around the world.

V. REFERENCES

[1] Leaf Disease Detection and Classification using Neural Networks V. Ramya1, M. Anthuvan Lydia

[2] Deep Learning for Early Detection, Identification, and Spatiotemporal Monitoring of Plant Diseases Using Multispectral Aerial Imagery Joseph Kimani Mbugua Claremont Graduate University

[3] Classification of Plant leaf diseases: A Deep Learning Method Taruna Sharma, Ruchi Mittal

[4] Sharada Prasanna Mohanty1,2, David Hughes3,4,5, and Marcel Salathé1,2,6 1Digital Epidemiology Lab, EPFL, Switzerland; 2School of Life Sciences, EPFL, Switzerland; 3Department of Entomology, College of Agricultural Sciences, Penn State University, USA; 4Department of Biology, Eberly College of Sciences, Penn State University, USA; 5Center for Infectious Disease Dynamics, Huck Institutes of Life Sciences, Penn State University, USA; 6School of Computer and Communication Sciences, EPFL, Switzerland

[5] [2] Yan Guo,1,2 Jin Zhang,3 Chengxin Yin,4 Xiaonan Hu,1 Yu Zou,1 Zhipeng Xue,1 and Wei Wang 3 1 College of Information Engineering, Sichuan Agricultural University, Ya'an, Sichuan, China 2 Key Laboratory of Agricultural Information Engineering of Sichuan Province, Sichuan Agricultural University, Ya'an, Sichuan, China 3 College of Management, Sichuan Agricultural University, Ya'an, Sichuan, China 4 College of Management, Chengdu Aeronautic Polytechnic, Chengdu, Sichuan, China.

[6] LILI LI1 , SHUJUAN ZHANG 2 , AND BIN WANG 2 1College of Information Science and Engineering, Shanxi Agricultural University, Jinzhong 030800, China 2College of Agricultural Engineering, Shanxi Agricultural University,Jinzhong 030800, China.

[7]Crop Disease Detection using Deep Convolutional Neural Networks Nikhil Patil Department of Information Technology Vidyalanakar Institute of Technology, Mumbai, India Vaibhav Wankhedkar Rajab Ali Department of Information Technology Vidyalankar Institute of Technology, Mumbai, India Prof. Deepali Nayak Department of Information Technology Assist. Professor, Department of Information Technology.