Detection of Disease and Damage Control for Crops using Convolution Neural Networks

Garlapati Pranay
School of Computer Science and Engineering, Vellore Institute Of Technology, Vellore, India

Aditya Kuruba
School of Computer Science and Engineering, Vellore Institute Of Technology, Vellore, India

Boda Bhagya Rishiroop
School of Computer Science and Engineering, Vellore Institute Of Technology, Vellore, India

Mythili Thirugnanam
Associate Professor (Senior), School of Computer Science and Engineering, Vellore Institute Of Technology, Vellore, India

Abstract— **The growth in the economy of a country majorly depends on the agriculture of the country. In India, the agricultural sector is facing a rigorous downside to increase crop yield and crop quality. With the rise of population globally global warming is increasing exponentially, which leads in the climatic conditions and other environmental changes either occurred naturally or man-made has become a major threat for agricultural sector. Several Data mining techniques and Machine Learning models are required for the problems to be solved out. The prediction of crop yield is based on beforehand available data like weather conditions, soil type, previous crop yield. The varying conditions climatic conditions and environmental changes not to affect the crop yield but also lead to various crop diseases thus resulting in the deterioration of crop quality. We in our work are looking forward to using image processing techniques like color histogram, canny, Sobel edge detector, and many others to detect the feature disease in a crop leaf. These diseases can be analyzed and classified using several classification algorithms to classify them into various diseases. Once the reason for the disease is identified, the required treatment can be followed up.**

Keywords— Convolution Neural Network, Machine Learning, Crop Disease prediction, Image Processing, Remedies.

I. INTRODUCTION

Agriculture which are also known as farming and cultivation, is the fundamental source of many countries. In India about 55 percent of total population comes under agri sector. Agriculture sector in the country covers the most employment, nearly half of the country. Agriculture contributes around 18% GDP to the country. The country's food grain production is rapidly growing yearly. India, one of the top countries in the world that produces variety of crops. Not only the production that gets increasing day by day, but the diseases, wastages of crops are also increasing rapidly. Crop mainly damages by pathogens and pests. It is a worldwide problem. The excessive use of pesticides to kill the micro-organisms lead to pollution of water, soil and air. Due to the issues which leads to damage of the crops, many farmers in the country are taking off their lives, due to poverty. The Solution for these problems is to build a proper system that monitors the condition of the crop at each and every stage of development. The earlier system and most followed system is to detect the condition of the crop by direct eye contact. Later many methods have been developed using Regression, DBSCAN, K-means clustering and artificial neural networks. These techniques didn't provide efficient results in order to predict the condition of crop. Hence, In this project we used Convolution Neural Networks. which helps us to give a high range of accuracy to predict the position of the crop at different levels of development. A User friendly GUI is developed with help of python. The system is developed in such a way that even an uneducated one can also understand and proceed with ease. The dataset is collected by capturing images. Then CNN helps to identify the disease or the position of the crop and gives a detailed instruction how to proceed in order to save the crop with less damage. This type of system with high accuracy helps many farmers, agriculture workers and many industries produce crops with high yield and less damage which helps poor people a lot.

II. LITERATURE REVIEW

Data mining, one of major techniques used in the agri field which has lots of issues and troubles. Predicting the yield of the crop using several regression techniques we can predict the several number. They have used regression techniques such as pure quadratic, polynomial, and quadratic among this they have selected a model with help of mean percentage prediction error or root means square error. There are no remedies for diseases found and for the regression model used it assumes a linear relationship and does not depend on multiple only a single line is relationship is used and the dataset size is of few records only [1].

The wide-extending mixture of consumptions for regarding this situation of involving items in computerized depictions makes it troublesome for stakeholders to outlook every plausible ways available in the writing, which allows likely solutions for complicated issues to be unexploited. for this specific situation, this research article tried to familiarize an exhaustive assessment on the subject, targeting is a foundation stage for those directing exploration on this issue.
Because of the vast number of orientations, the depictions are brief, by providing a quick review of the opinions basic every one of the activities It is the key task to showcase the work on a specific field is not limited to what was shown here. papers are supposed to consider the different problems **[2].**

The work which was proposed mainly focuses on the pattern identification system for detection and classifying the available cotton leaf diseases i.e. Bacterial Blight and Alternaria. Pictures which are necessary for research are collected from many arenas at the CICR i.e Central Institute of Cotton Research Nagpur. Contour model is primarily used for image segmentation where they applied same in their research and for the model training in this system they followed adaptive neuro-fuzzy system. Here the images of cotton leaf which are diseased are classified by the Backpropagation neural network from the available 3 kinds of leaf images training is done by involving the major seven invariant moments by extracting them. The accuracy for the model bild to classify is 85.52%. Using a snake segmentation algorithm is an resourceful technique to identify the spot which is diseased is a sluggish process. which directly impacts in extensive testing and as well as training phase for the developed Other features can also be included for extracting the feature for assemble the system much efficient. which results in performance of the system better but still accuracy is low being a small dataset **[3].**

The approach which was discussed in this paper by the author is innovative to grade disease automatically on a plant leaf. Majorly many plant pathologists depend on the naked eye for classifying and predicting the disease by scoring scale which specifically tells how far it is infected to grade the disease. This work majorly focuses on profit maximization, reduction of environmental damage, and input rationalization by modifying the agricultural practices and following new approaches for the site demands. The work that is followed by this author is divided into categories: (1) Image acquisition (considering an image) (2) Image Pre-processing(adjusting the borders and values using operators) (3) Calculating AD and AT (5) Disease grading is done by Fuzzy Logic is not always accurate, so The results are produced based on the assumption by the author, so it is not widely accepted by the stakeholders. This system proposed by the author don't have the ability of machine learning type pattern recognition **[4].**

This proposed system by the author deals with the use of an Artificial Neural Network. Many factors decide the productivity. So parameters that are included as the input based on the preprocessing with these inputs, the required output should be produced with model provided. There are many parameters like pH, nitrogen and etc. Also, the system tells using some fertilizers or some pesticides for the crops leads to improve productivity. The way of improving the proposed system by the author involves the certain processes: 1. Data collection (acquisition, collecting)/ Preparation 2. To Build the Model 3. Classifying it 4. providing suggestion for respective crops using pesticides. developed System should be efficient and it should be easy for stakeholder to be used, the system is developed as an Application which is used on mobile phone being portable we can use it anywhere. The accuracy gained through this proposed system is 90%. When ANN produces a result, it will not provide any clue how it happens and why it has happened. So basically ANN requires hardware components and processor so in order to achieve great accuracy

proceeding with neural networks would be great **[5].**

Paper provides an algorithm for the method of Segmentation technique of image which used for classification and automatic detection of plant leaf diseases as well as for the evaluation of different disease class techniques for leaf diseases. Image classification, which is a vital factor in diagnosing plant diseases, is done using a genetic algorithm. Agricultural is the backbone of India's economy. This is one of the reasons why the discovery of plant diseases plays an chief role in the agricultural sector, as plant diseases are very natural. Properly neglected in this area it causes adverse impact on the plants as well as result affects the quality of the product, quantity or product. Detection of plant diseases by means of a systematic process is helpful as it reduces the amount of vigilance on large plant farms, and in its early detection of disease signals means it comes from plant leaves there are 4 steps mainly for the advanced processing, where the first, RGB input image, color change layout is made, Since this RGB is used majorly for color production and a modified or updated RGB image, i.e., HSI is for color dictionary. For the second process, using the limit value, the green pixels are hidden and detached. For the Third, by involving the calculated level, the green pixel extraction and concealment are made for the vital needed parts first extracted from this process, while the picture is separated. And in the final step the separation is completed **[6].**

Smart Farming framework utilizing an important basis is an imaginative modernization that advances the quality and number of horticultural creation in the country including tomato. The ongoing improvements in PC visualization made conceivable by deep Machine learning made ready for disease finding for this tomato vegetable. This examination built up imaginative arrangement that gives productive disease identification for tomato plants. A measured engine image holding box for capturing was made to hold to cover all sides of tomato plant to perceive diseases, the task of framework is to recognize the diseases in Leaf Miner. Utilizing a dataset of 4,923 pictures of diseased and sound plant leaves gathered under precise conditions, we train using convolutional neural system to recognize 3 diseases or. This framework follows CNN so called Convolutional Neural System to differentiate the diseases is present on the researched plants of tomato. So F-RCNN prepared detection show formed a approximate score of 80 % when the Exchange Learning disease acknowledgment a precision of 95.75 %. The computerized image framework was executed and enlisted a 91.67 % precision in the nod of the tomato plant leaf diseases. It will aid in limiting the man mediation required in checking diseases regarding tomato plants and it moreover causes in acquainting horticulture with progressively complex gadgets that can be utilized in cultivating **[7].**

Paper proposes plants recognition on features of leaf, with the pre-processing of leaf images too, and feature extraction is by ANN so called Artificial Neural Network-based classification and training for identification of leaf. Later, the identified disease which is available on the leaf is processed to classify, using the K-Means based clustering of diseased area, extraction of features from the defected portion, and the classification for disease. Here the grading of disease is on criteria of the quantity of disease currently present in leaf. The worked aimed at developing a system for automatic leaf recognition and leaf disease grading for various leaves of plants. The work included several stages in the training and testing phases that comprise of several data mining techniques and image processing techniques. They used the Fussy Logic approach to find the grade of infection on leaves. Grid partitioning for is one of the major problems faced for data sets with so many parameters **[8].**

An utilization of surface investigation in identifying and characterizing plant leaf illnesses has been clarified during this research. Subsequently, the proposed calculation was tried on 10 types of plants specifically lemon and sapota. The illnesses explicit to those plants were adopted for strategy. Results from the tests show the proposed way can cluster the leaf sicknesses. With this approach the plant sicknesses can be identified at the low. The explanations overdue misclassification are: the indications of the sick plant leaves swing (at the starting, small, dull earthy colored to dark spots, at future time, it has the wonders of shriveled leaf, dark or leaf cancellation), additionally the considered element ID vectors need

to become advanced. So to improve sickness ID rate at different phases, the grounding tests can be prolonged and shape highlight and shading highlight among the ideal highpoints can be given as data state of malady distinguishing proof **[9].**

The author for this paper targeted on the analyzing the data of agriculture for identifying best factors in order to increase the production of crop effectively using data mining techniques and Multiple Linear Regression. The data is collected from the official websites provided by the Government of India. updated DBSCAN method is utilizing for grouping the data based on unique characters like soil type. data mining techniques are used to make cluster on based data from where production of crop is maximum. Regression method is followed to predict the yield annually for the crop. The quality prediction of several clustering methods requires usage of internal quality metrics. But in this paper the authors have used just the external metrics to predict and PAM resulted in comparatively low clustering quality **[10].**

III. METHODOLOGY

1) Convolutional Neural Network

We used the Convolution Neural Networks for predicting the type of disease the crop has. The Convolution Neural Networks (CNNs) are very effective and efficient in the process of recognizing the image and further classification. These type of Neural networks have a Unique architecture compared to the regular Neural Networks.

Just like humans learn to recognize the objects they see with their eyes, the computers recognize images that are numbers represented in a 2-D array, known as pixels. We can train the machines to recognize various patterns just like we recognize them. For making this happen, we use an Artificial Neural Network known as convolution neural networks (CNN). The CNN are inspired by the way the brain works.

Regular Neural Networks remodel an enter through placing it through a chain of hidden layers. Every layer is made of a fixed set of neurons, wherein every layer is attached to the neurons with inside the earlier than layer. The final layer (output layer) is eventually fully-linked that indicates predictions.

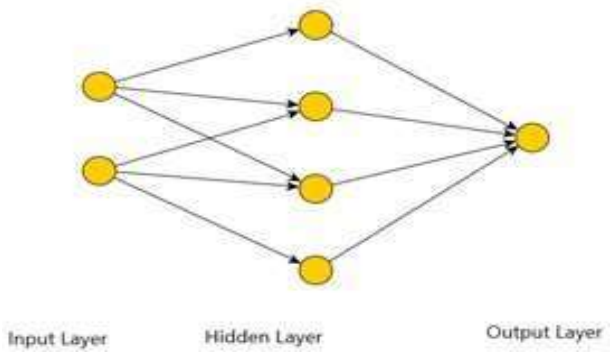

Fig 1. Neural Network

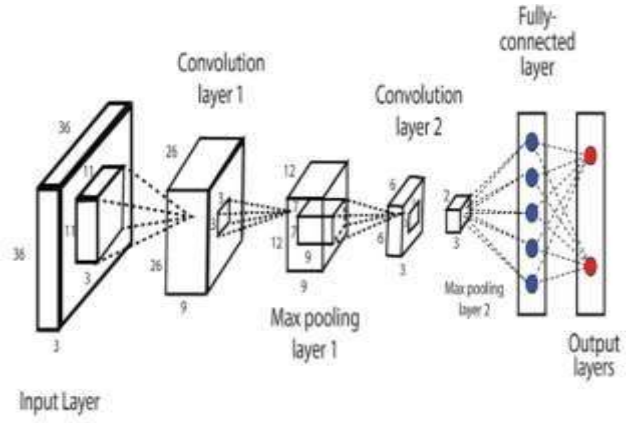

Fig.2 Convolution Neural Network (CNN)

The Convoluted neural networks have different layers organized in three different dimensions. The three dimensions of them are width, height and depth respectively. The neurons present in a single layer do now no longer connect with all neurons with inside the subsequent layers as an alternative best to a small location of it. The output achieved is reduced into a solo vector that contains probability values, organized alongside the depth dimension. These Convoluted Neural networks have two parts knows as the hidden layer and the classification layer. In Fig 2, Layers named Convolution Layer 1 are hidden layers and the layers named Convolution layer 2 are the Classification layers. The method for extraction of the image completes the process of convolution.

Convolution is one of the essential and main building blocks of the CNN. Convolution refers back to the mathematical aggregate of capabilities of two functions to provide a 3rd function. Two sets of information may be merged the usage of convolution. In case of a CNN, the convolution is performed on input with the usage of a filter or kernel to provide a feature map.

Training the dataset for the CNN is finished using the method backpropagation or the gradient descent. As there are mathematical operations associated with convolutions, this turns into a bit complicated. CNN is specific from different Neural networks in which a filter slides over the input and merges the input value alongside the filter value on the feature map.

2) Architecture

1. Collection of DataSet -> 2. Training -> 3. Python GUI -> 4. Testing

PHASE 1. Collection of DataSets Publicly available datasets : We collect different datasets each for plant type, weeds, disease, pests etc that are publicly available. Resizing images for efficient storage and prediction.
Datasets: Horticulture crops taken : Corn Peach Disease categories: 1 healthy for each plant crop 1 diseases for Peach 2 diseases for Corn
PHASE 2: Training Vectorize each image of dataset when loaded. Train a CNN (YOLO architecture) on different categories of datasets using keras with tensorflow backend. Save the weights
PHASE 3: Creating the GUI for python so that its easy to work on
PHASE 4: Testing on a leaf image having disease taken from internet.

Fig.3 Architecture Diagram

3)Flow chart

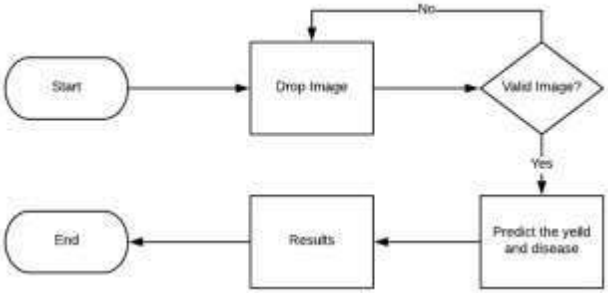

Fig.4 Flowchart of the system

IV. RESULTS

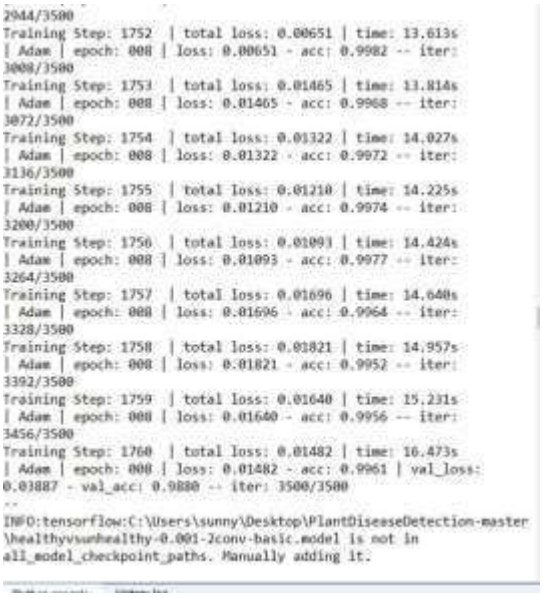

Fig.5 Values generated between the iteration

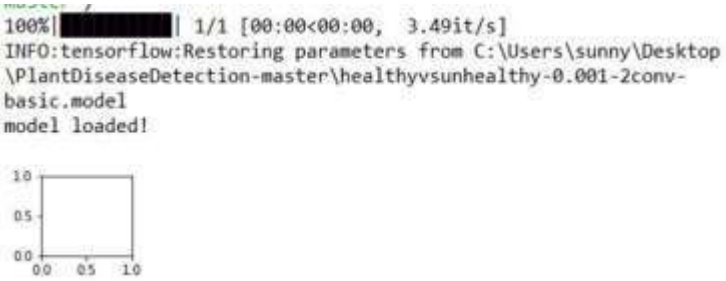

Fig.6 Confusion Matrix

We trained over model with around 4000 images and tested it on 2000 images with which we achieved accuracy over 98 Percent with detection of disease and providing remedies for the diseased leaves. he diseases were detected and accordingly remedies were suggested to the farmers. This in turn helped the yield to be high.

V. CONCLUSION

Agriculture or crop losses due to the disease is very catastrophe, mainly keeping in mind of farmer's efforts. The infrastructure and supply chain in this field are very backward. Even though, fertilizers and toxins are used to disintegrate pests, it leads to severe health problems for consumers who consume. In many countries factors like poverty, having no proper disease management, no adequate technical support and no proper support from government due to population, problems like these arise in agriculture sector leads to many disasters.

We desire, the product or system we developed will be useful to reduce the impact of troubles caused by diseases to crops. In addition, to provide a system which is user friendly even to the uneducated, at less cost.

ACKNOWLEDGEMENT

This following research on the project helped us an opportunity to gain a greater understanding of this field and to explore more. On this platform, we want to express our wholehearted respect to our faculty, for all the support, guidance and encouragement from beginning to end of this project.

REFERENCES

[1] A. Shastry, H. Sanjay, and E. Bhanusree, "Prediction of crop yield using regression 21 techniques," International Journal of Soft Computing, vol. 12, pp. 96–102, 2017.

[2] J. Garcia, "Digital image processing techniques for detecting, quantifying and classifying plant diseases," SpringerPlus, vol. 2, p. 660, 2013.

[3] P. Rothe and R. Kshirsagar, "Cotton leaf disease identification using pattern recognition techniques," in IEEE, 2015, pp. 1–6.

[4] Sannakki, Sanjeev S, Rajpurohit, Vijay S, V. Nargund, A. Kumar, and Yallur, Prema S, "Leaf disease grading by machine vision and fuzzy logic," Int J, vol. 2, pp. 1709–1716, 2011.

[5] G. Ravichandran and R. Koteeshwari, "Agricultural crop predictor and advisor using ANN for smartphones," in IEEE, 2016, pp. 1–6.

[6] V. Singh and A. Misra, "Detection of unhealthy region of plant leaves using image processing and genetic algorithm," in IEEE, 2015, pp. 1028–1032.

[7] L. Robert, E. P. Dadios, and Bandala, Argel A, "Automated image capturing system for deep learning-based tomato plant leaf disease detection and recognition," in IEEE, 2018, pp. 1414–1419.

[8] A.Rastogi, R. Arora, & S. Sharma (2015). Leaf disease detection and grading using computer vision technology & fuzzy logic. 2015 2nd International Conference on Signal Processing and Integrated Networks (SPIN).

[9] S. Arivazhagan, S. R. Newlin, S. Ananthi, and V. S. Vishnu, "Detection of unhealthy region of plant leaves and classification of plant leaf diseases using texture features," Agricultural Engineering

[10] J. Majumdar, S. Naraseeyappa, and S. Ankalaki, "Analysis of agriculture data using data mining techniques: application of big data," Journal of Big data, vol. 4, p. 20, 2017.