Leaf Disease Detection Using Hybrid Feature Extraction Techniques And Machine Learning

Shafiulla Shariff
Dept. of CSE,
Jain Institute of Technology
Davanagere, India
shafiullashariff@jitd.in

Saraswathi H S
Dept. of CSE,
Jain Institute of Technology
Davanagere, India
Saraswathi@jitd.in

Panchami M P
Dept. of CSE,
Jain Institute of Technology
Davanagere, India

Vaishnavi S
Dept. of CSE,
Jain Institute of Technology
Davanagere, India

Sumayya Banu
Dept. of CSE,
Jain Institute of Technology
Davanagere, India

Abstract—**Standard agricultural output has a significant impact on India's economy. The two most significant cash and fiber crops are cotton and tomatoes. There are various illnesses that injure the plant, and one of the most important ones is leaf disease, which significantly reduces the production of cotton and tomatoes. There is a need for disease detection techniques that can be applied to lessen the illness's impact on plant losses. By categorizing good and ill leaf photos using pre-trained CNN, transfer learning, and KNN machine learning algorithms, this study attempts to identify the cotton and tomato leaf disease. The two models, InceptionV3 and Inception ResNetV2, were trained using leaf pictures. The most significant findings indicated that the 50% dropout rate InceptionV3 model and the Inception ResNetV2 model.**

Keywords—Detection, deep learning, CNN, KNN inception V3, inception ResNetV2, transfer learning

I. INTRODUCTION

In moment's world, Cotton is more extensively used than any other fiber and thus cotton place a superior part in the agrarian and artificial frugality of the country. Tomato's are one of the most widely cultivated and consumed vegetables in the world. Currently the loss in quality and volume of the cotton and tomato yield is increased immensely due to the colorful factory condition. The cotton and Tomato factory is susceptible to infections by pathogens like fungi and bacteria and its product is cursed due to the multitudinous diseases affecting the factory. Traditionally, the conditions were detected by the planter with the bared eyes. And discovery of any complaint by our naked eye is veritably less accurate. In cotton and tomato shops the science of numerous conditions can be seen substantially on the leaves. To exclude this con its necessary to carry out the automatic complaint discovery which can be more precise and error free. Machine learning is a growing technology that allows computers to automatically learn from one data. In this work, sententia of semi supervised learning-grounded system known transfer learning is used to classify healthy and unhealthy cotton and tomato leaves. This system sticks to the classical machine learning set up but surmises only a limited number of labeled samples for training. The type of complaint would be further linked by K-Nearest Neighbor and Convolutional Neural Network for cotton and tomato factory independently.

II. OBJECTIVES

The major goals of agricultural research are to improve food production while lowering costs and improving food quality. The goal of this project is
- To detect leaf illness
- To boost the model's accuracy
- To cut down on errors.

III. LITERATURE SURVEY

The team looked for and reviewed various patents, research papers, documents, newspapers, and magazine articles from diverse scenes for this project's literature review.

According to [1], The rover's controller is an Arduino. The rover moves thanks to Arduino. The servo motor in the rover rotates the pi camera to collect images of the plants from both the right and left sides. The background python code can compare the acquired image with the database image after the images have been taken. If neither of them is appropriate for his or her approach, the image and revered issue are sent to the farmer cellphone.

A study [2] stated that, Using the Adaptive Neuro Fuzzy Inference System, shape and texture feature extraction is followed by categorization of the plant diseased leaf image. Here, the Root Mean Square error is used to identify the discrepancy between the actual characteristics and the expected qualities. The system developed for identifying leaf infections extracts features from diseased leaves and will be used to classify leaf diseases using adaptive neuro fuzzy classification.

A study [3], is focused on developing a CNN-based model to improve the identification of pests and diseases affecting cotton leaves. The researchers have done this by using prevalent pests and diseases that affect cotton leaves,

188

including bacterial blight, spider mites, and leaf miners. The dataset splitting technique of K-fold cross-validation was used to increase the generalization of the CNN model. The model used supervised learning on 2400 datasets using a process for extracting the four key features.

According to Sandeep Kumar et al. [4] states that, The development of image processing algorithms for cotton harvesting robot in-field boll detection under natural lighting circumstances has been made. For the real-time segmentation of cotton bolls in outdoor natural light, a total of four image processing algorithms were created. Three of these algorithms, based on the color difference method, color component ratio method, and chromatic aberration method, were created using the RGB color model, and one algorithm was created using the YCbCr color model. The chromatic aberration technique showed favorable performance for cotton bolls detection in the field in settings of natural daylight, confirming its suitability for robotic cotton harvesters.

In [5], diseases in the initial stage were addressed by pre-processing the image. Followed by these processing, k-means clustering is carried out in images. Then the features are extracted and then the green masking pixel is done by using the thresholding constant value 0-255.The performance measures are taken into consideration for better results by comparing the two algorithms Principal Component Analysis (PSA) and Support Vector Machine (SVM). Naseeb Singh et al.

According to Mohanty et al. [6] Using pre-trained deep learning models, 14 crop species and 26 illnesses were identified in the Plant Village open dataset; the highest degree of classification accuracy was 99.35%.

According to Chen et al. [7] created VGGNet to train around 1000 pictures of five illnesses of rice and four diseases of maize utilizing both general data and data gathered for diseases of rice.

According to Fan et al. [8] By combining the characteristics extracted using transfer learning with the features extracted using conventional techniques, the Inception V3 network was improved for the detection of apple and coffee leaf diseases.

According to Yu et al. [9] employed a deep learning model to identify pests affecting tomatoes. And Karthik et al. [10] Tomato leaf tissue was used to determine the kind of infection using the CNN model.

According to Abbas et al. [11] The results of training a deep learning network to identify tomato leaf diseases were satisfactory. CNN is also an effective method for identifying plant diseases.

Naseeb Singh et al. [12] puts forward a framework of deep learning to detect the disease. DL method helps in detecting the disease's severeness. The Artificial Neural Network is trained using Stochastic Gradient Descent on the training set. The training in this research is carried on the base of small convolutional networks with various depth from scratch and finely tuned 4 stages of cutting-edge models like VGG16,VGG19, ResNet50, and Inception-v3.When comparing the data acquired from the deep model the performance of the network can be enhanced in certain

datasets. The result showed that the VGG16 gives 90% accuracy.

IV. METHODOLOGY

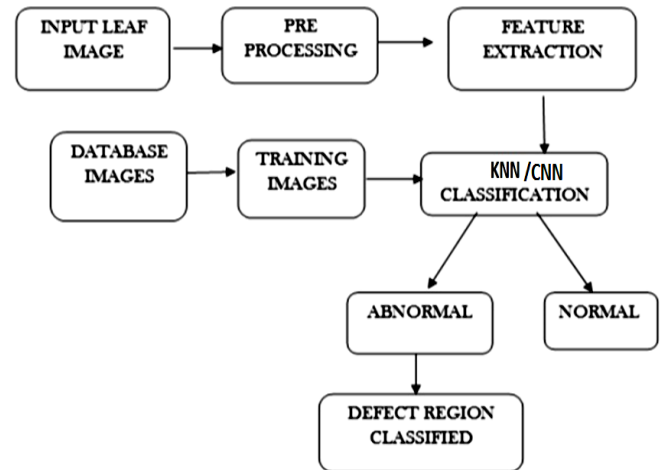The Figure 1 below shows the block diagram of proposed method to detect leaf disease

Fig.1.Block diagram of proposed system

Using Python, this work procedure is broken down into several parts, including image acquisition, image pre-processing, disease identification, disease classification, and offering cures.

A. Image Acquisition

To start with image accession, formerly available datasets are used. There are two sets of splint images, one for training and another for confirmation. A aggregate of roughly 2000 images are named for training. Different angles of the leaves are considered for training. Brochure consists of images in the rate 80 to 20 for training and confirmation independently, for 2 different orders i.e., Bacterial scar and coil virus.

B. Image pre-processing

In this, images are formatted before they are used by model training. This includes color correlation, resizing, orienting, etc. Image preprocessing is required to clean image data for more input.

Fig 2: Original image

189

Fig.2 shows the original size of the image from the dataset. The images in the dataset are different in size. So, it is necessary to resize all the images in the same size to increase the efficiency and maintain uniformity.

Fig 3: Resized Image

Fig.3 shows the resized image. Resizing of images is done in the ratio 691*691.

C. Disease Detection

The factory conditions are classified using Transfer of knowledge. Transfer learning (TL) is an exploration problem in machine learning (ML) that focuses on preserving knowledge obtained while working on one problem and transferring it to another problem that is unrelated to the original but still affects it. Commencement models are used to categorize the factory leaves as a residual network. The splint complaint is described using the KNN and CNN algorithms.

D. Disease classification:

The unhealthy leaves are farther classified into different conditions using the KNN machine learning algorithm after they've been classified as healthy or unhealthy. A popular supervised learning and machine learning algorithm is K-Nearest Neighbor. The KNN method places the new case in the order that is analogous to the being orders on the premise that the new case and being cases are similar. The data points were based on how closely they resembled data. This means that using the KNN algorithm, fresh data can be quickly categorized into an appropriate order. Working of KNN:

- Step 1: Define value of 'k'.
- Step 2: Apply Euclidean distance on all the records.
- Step 3: Sort the distances.
- Step 4: Select nearest neighbor depending on the value of 'k'.
- Step 5: Classify the test point to categorize them into affected or unaffected leaf.

KNN determines the distances between a query and each example in the data, chooses the K instances that are the closest to the query, and then votes for the label with the highest frequency. K has been set to have a value of one. The confusion matrix indicates that our model KNN is more

accurate at K=1 due to its higher accuracy. The KNN recognizes the leaf picture from either of two illnesses, i.e., bacterial blight or curl, if the ResNet50 result of the leaf is unhealthy. Then, utilizing Python programming, treatments for the ailment are sent to farmers in text format. The distance metrics used in KNN is

$$\left(\sum_{i=1}^{n} |x_i - y_i|^p \right)^{1/p} \qquad (1)$$

To find picture features, a CNN uses layers for convolution, batch normalization, activation, and pooling. Additionally, it has dropout and thick layers that are used to classify images based on automatically generated characteristics. CNN architecture is as shown below.

Following equation describes CNN layers.

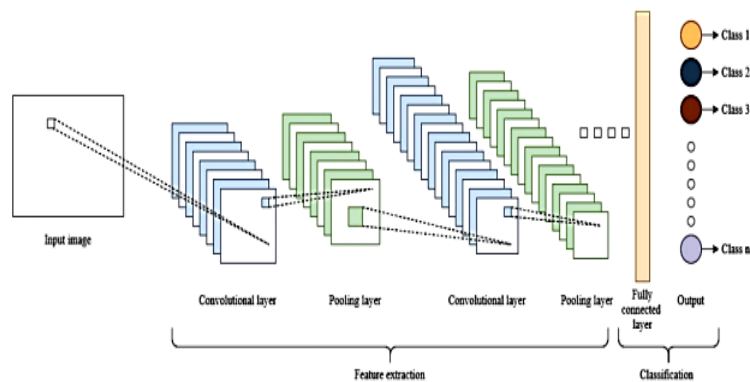$$C(h,d) = (k * f)(h,d) = \sum_i \sum_j k(h-i, d-j) f(i,j) \qquad (2)$$

Fig 4: CNN Architecture

Then, CNN is trained by figuring out the neural network's ideal weights to lessen the tolerances between the dataset's actual input and expected output. Backpropagation is a common method for training neural networks, and it is crucial to consider loss and optimization functions. In the identification of plant diseases, it has been shown that CNN performs better than conventional feature extraction techniques.

V. RESULTS

The execution of the proposed system is carried out using visual studio. Various libraries and models required for the execution such as OS, NumPy, pandas, TensorFlow were imported. The expected output is shown in a website which is developed using Flask python framework.

Testing and training are the two key components needed for classification. The collection comprises of pictures of both healthy and diseased cotton leaves. Transfer learning RESNET50 model is trained using different parameters like dataset color, no of epochs and optimizer. The images are resized to the dimension of (691, 691). This establishes consistency across all images.
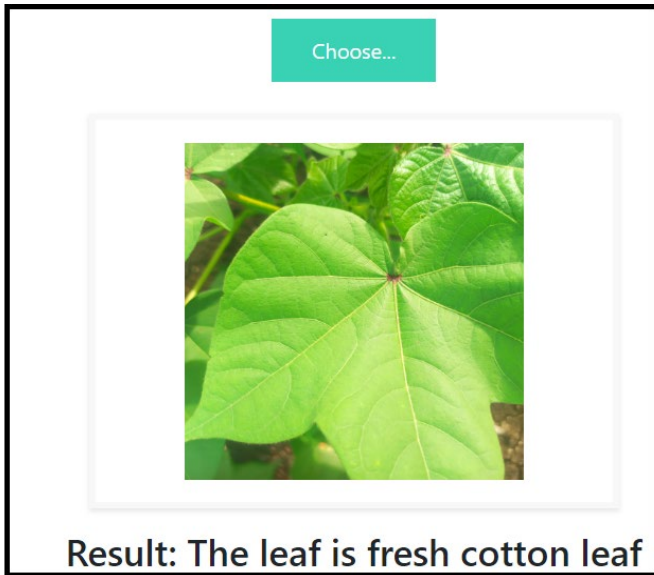
190

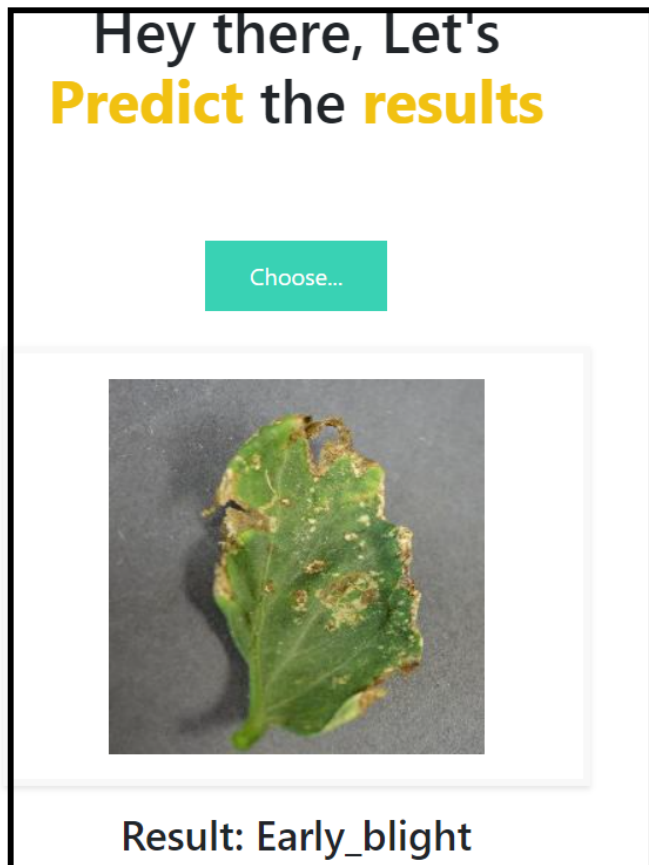Fig.5. GUI of system shows the results of a related work.

Fig.6. shows the disease in tomato leaf

This project helps the farmers to detect the leaf disease in the early stage. We had done this project by taking 2000 leaf images of both fresh and defected leaves. In this project we are using transfer learning for the classification of the leaves and algorithms like KNN and CNN are used to detect the disease in the leaves.
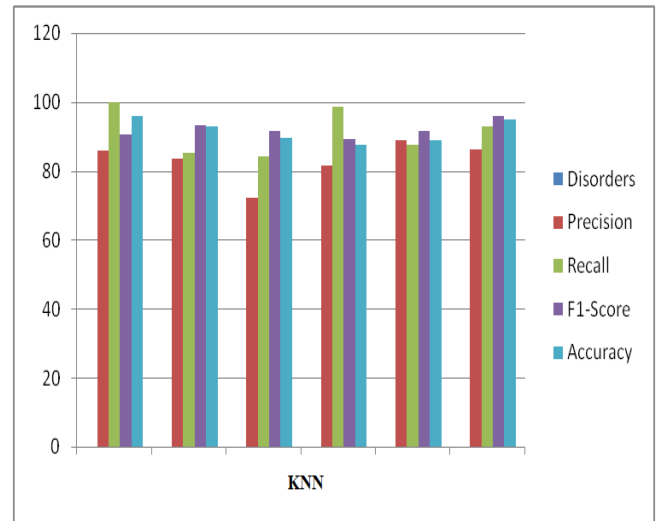
Performance of KNN is as shown below

Fig.7. Performance of KNN

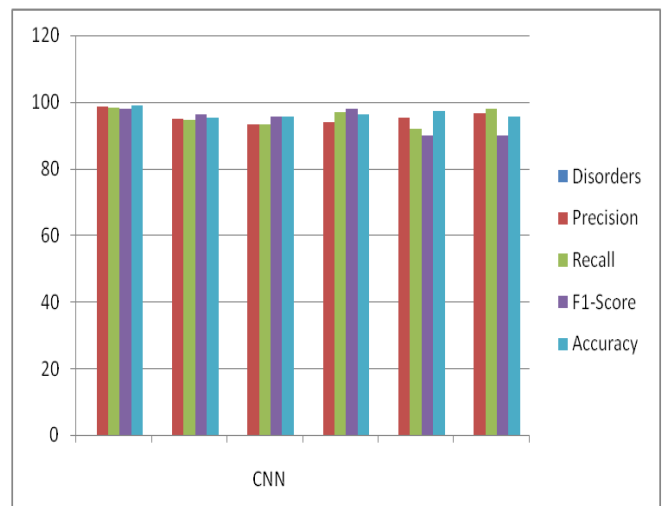Performance of CNN is as shown below

Fig.8. Performance of CNN

VI. CONCLUSION

In order to maintain uniform dimensions, leaf photos are cropped and scaled in this study. The HE method is used to improve the image quality. K-means clustering is used for segmentation. The boundary of the leaf image is extracted using the counter tracing approach. To extract meaningful features from leaves, DWT, PCA, and GLCM are employed as a number of descriptors. Finally, KNN and CNN are used to classify the retrieved features. The proposed approach performs with an accuracy of 86%.

FUTURE SCOPE

In the future, a mobile application may be created, and the farmers could receive the right treatment for the condition via email or a message service.

REFERENCES

[1] Azath M. ,1 Melese Zekiwos,2 and Abey Bruck1 "Deep Learning-Based Image Processing for Cotton Leaf Disease and Pest Diagnosis", Volume 2020

[2] K. SaiManoj, Detection of Plant Disease Using Deep Learning Techniques, International Journal of Advanced Research in Engineering and Technology (IJARET), 12(1), 2021, pp. 911-924.

[3] J. Karthika et al. 2021 J. Phys.: Conf. Ser. 1916 012224, Disease Detection In Cotton Leaf Spot Using Image Processing.

[4] Sandeep Kumar Department of ECE Sreyas Institute of Engineering & Technology, Hyderabad, India drsandeep@sreyas.ac.in, Leaf Disease Detection and Classification based on Machine Learning. IEEE 2020

[5] N Nandhini and KB Srisathya 2021 J. Phys.: Conf. Ser. 1916 012008, Identification of Plant leaf diseases using Adaptive Neuro Fuzzy Classification.

[6] Wspanialy, P.; Moussa, M. A detection and severity estimation system for generic diseases of tomato greenhouse plants. Comput. Electron. Agric. 2020.

[7] Lins, E.A.; Rodriguez, J.P.; Scoloski, S.I.; Pivato, J.; Lima, M.B.; Fernandes, J.M.; da Silva Pereira, P.R.; Lau, D.; Rieder, R. A method for counting and classifying aphids using computer vision. Comput. Electron. Agric. 2020.

[8] Wu, F.; Duan, J.; Chen, S.; Ye, Y.; Ai, P.; Yang, Z. Multi-Target Recognition of Bananas and Automatic Positioning for the Inflorescence Axis Cutting Point. Front. Plant Sci. 2021.

[9] Ahmed, I.; Yadav, P.K. Plant disease detection using machine learning approaches. Expert Syst. 2022.

[10] Andayani, P.Y.; Franz, A.; Nurlaila, N. Expert System for Diagnosing Diseases Cocoa Using the Dempster Shafer Method. Tepian 2020.

[11] Tan, H.Y.; Goh, Z.Y.; Loh, K.-H.; Then, A.Y.-H.; Omar, H.; Chang, S.-W. Cephalopod species identification using integrated analysis of machine learning and deep learning approaches. PeerJ 2021.

[12] Usama Mokhtar, Nashwa El-Bendary, "SVM-Based Detection of Tomato Leaves Diseases," Springer, 2015.

[13] Melike Sardogan, Adem Tuncer, "Plant Leaf Disease Detection and Classification based on CNN with LVQ Algorithm," IEEE, 2018.

[14] Jobin Francis, Anto Sahaya Dhas D, Anoop B K, "Identification of Leaf Diseases in Pepper Plants using Soft Computing Techniques," IEEE, 2016.

[15] Vijai Singh, Varsha, "Detection of the unhealthy region of plant leaves using Image Processing and Genetic Algorithm," ICACEA, IEEE, 2015.

[16] Mrunmayee Dhakate, Ingole A. B., "Diagnosis of Pomegranate Plant Diseases using Neural Network," IEEE, 2015.

[17] Sachin D. Khirade, A. B. Patil, "Plant Disease Detection Using Image Processing," IEEE, 2015.

[18] S. Yun, W. Xianfeng, Z. Shanwen and Z. Chuanlei, "Pnn based crop disease recognition with leaf image features and meteorological data", *International Journal of Agricultural and Biological Engineering*, vol. 8, no. 4, pp. 60, 2015.

[19] A. Caglayan, O. Guclu and A. B. Can, "A plant recognition approach using shape and color features in leaf images", *International Conference on Image Analysis and Processing*, pp. 161-170.

[20] X. Zhen, Z. Wang, A. Islam, I. Chan and S. Li, "Direct estimation of cardiac bi-ventricular volumes with regression forests", *Accepted by Medical Image Com- puting and Computer-Assisted Intervention-MICCAI 2014*, 2014d.

[21] P. Wang, K. Chen, L. Yao, B. Hu, X. Wu, J. Zhang et al., *Multimodal classification of mild cognitive impairment based on partial least squares*, 2016.

[22] Shrivastava, V.K.; Pradhan, M.K. Rice plant disease classification using color features: A machine learning paradigm. J. Plant Pathol. 2021, 103, 17–26.

[23] Agarwal, M.; Gupta, S.K.; Biswas, K. Development of Efficient CNN model for Tomato crop disease identification. Sustain. Comput. Inform. Syst. 2020, 28, 100407.