A Review Paper on Plant Disease Detection System

Prof. Pallavi Sambhare[1] Shreeya Dumore[2] Shweta Khobragade[3] Shrutika Matte[4]
[1]Assistant Professor
[1,2,3,4]Department of Information Technology
[1,2,3,4]G.H.Raisoni College of Engineering, Nagpur, India

Abstract — More than 50% of Indians are employed in agriculture. And any disruption in this sector affects a large part of the population. Some of these disturbances are climate change, infertile soils, and plant diseases. Leaf diseases are the major problem of crops and identification of leaf diseases is very difficult. If these diseases are not detected, it can lead to great losses. In this work, the disease of a plant leaf is detected using images. In this work, the process of dataset creation, feature extraction, classifier training, and classification feature algorithm is followed. In this work, many machine learning and deep learning algorithms are used to find out which are best for leaf disease detection, such as Random Forest Classifier (RFC), Support Vector Machine (SVM), K-Nearest Neighbors (KNN), Convolutional Neural Network. It was obtained that the Random Forest algorithm was the most suitable for classification. This algorithm can help us classify the various kinds of diseases that would occur in the plant leaves.

Keywords: Image Classification, Plant Disease Detection, Machine Learning, Deep Learning, Neural Networks, SVM, KNN, CNN, ANN

I. INTRODUCTION

Agriculture is considered a primary and profitable activity in India due to the abundance of cultivable land. Since India is a developing country, most families derive their income solely from agriculture. And the soils in India are diverse and fertile, which allows the cultivation of different varieties and better quality of crops, bringing in profits. But these profits turn into losses as plant diseases destroy crops. Among these plant diseases, one of the most harmful diseases that occur in leaf.

The detection of plant diseases is necessary for managing and producing crops. Although it needs a high level of knowledge and competence, it may be effectively accomplished by monitoring and assessment of changes in plant leaves by scouting experts. Artificial intelligence (AI)-based data analysis techniques used in advancements can increase the reliability of diagnoses and as a result, be include in solutions for effective treatment. It is difficult to recognize plant diseases using visual features. Also, certain weather circumstances can also lead to such disease in the plants.

The evaluation of plant disease severity is a crucial first step in developing accurate and efficient crop management techniques for the generation of better yield in the agricultural sector.

A leaf is an essential part of the plant. Leaves use the process of photosynthesis to provide energy for the plant which is necessary for the growth of the plant. Leaf diseases damage the leaf and the ability to photosynthesize, which affects the growth of the plant and puts the plant at risk. These diseases affect not only the quantity but also the quality of the plant products. These diseases are mostly caused by bacteria, and viruses. Detection of these diseases is very important for the plants. If this disease is not detected quickly, it can spread to other leaves and destroy the plant.

Due to lack of awareness of leaf disease, sometimes there is a huge loss of yield, which leads to shortage of food. When these detections are done by the human eye, it will consume a huge amount of labor, time and also money. Moreover, it is difficult for farmers to identify all the diseases, and hiring a large number of experts is not possible due to financial constraints. To make this easier, many approaches have been developed that use artificial intelligence and machine learning. Leaf disease detection programs are developed to detect and eradicate the problem.

This model analyzes the provided image and provides an output after applying the algorithm. To make a program accurate, training and testing is an important part for which selecting the right dataset is very essential. This work uses the "PlantVillage" dataset which consists of 54303 different images from 38 different categories. Since this dataset contains a large amount of data, it helps in training and testing algorithms and provides more accuracy. In this work, many algorithms were tested like Support Vector Machine (SVM), Random Forest Classifier, K-Nearest Neighbors, Convolutional Neural Network and contrast the output and the algorithm with best accuracy was used, namely Random Forest algorithm and CNN to detect the diseased leaves of the plant.

II. LITERATURE SURVEY

S Ramakrishnan et al [1], this study discusses the automated method used for plant disease detection that would be highly optimal as the time required to observe the large crop farms would be much less and also find out plants having disease in a very early stage.The work gives an approach focused on the identification and categorization of leaf diseases in plants using an image segmentation procedure. It also provides us with a summary of several disease classification techniques that could be useful to the detection of plant leaf diseases. It also informs us that the image classification and genetic algorithm are very important steps for leaf disease detection.

Aakanksha Rastogi et al [2], this study provides a quick and efficient computational technique for grading and identifying leaf diseases utilizing modern image processing and machine learning. This method is split into two parts. The first part involves the identification of the plant based upon the distinct characteristics its leaves have. This part involves pre-processing leaf images, their feature extraction and training and using artificial neural networks for classification. In the second part, the model figures out the disease on the plant. This procedure comprises feature extraction from the defective region, segmentation of the affected area using K-Means, and disease classification using an ANN (artificial neural network). The severity of the disease would be then graded based on how much of it is visible on the leaf.

Usha Kumari et al [3], Following study proposed four stages which include image capturing,followed by image segmentation, extraction of vital features, and classification are discussed in this disease detection approach. Using K-means clustering approach, features are generated from the disease-affected cluster for picture segmentation. Extracted characteristics include correlation, contrast, homogeneity, mean, energy, and variance. As classifier inputs, the disease is classified using the characteristics that were collected from the disease cluster. In this study, we consider a tomato leaf and Neutral Networks were utilized for classification. The accuracy rates for cotton leaf diseases such as bacterial leaf spot and target spot are found to be 90% and 80%, respectively.

Santhosh S et al [4], this study focuses on discovering the different kinds of diseases that could occur in different kinds of plants.The main aim of the following study is to detect such diseases affecting plants and classifying them accordingly using various computational techniques.

U. Shruthi et al [5], the study focuses on increasing the crop yield and to achieve that we need to detect the diseases occurring in the plants. This study compares several machine learning based classification algorithms for the plant disease detection and rectifies the steps of a comprehensive detection system. This study showed that a convolutional neural network would yield good accuracy and can identify a larger variety of diseases in plants.

Mohit Agarwal et al [6], it mainly focuses on disease detection in tomato crops. The numerous diseases cause the tomato crop's quality and yield to decline. Therefore, a deep learning-based methodology for disease detection is explored in the study. An approach based upon convolution neural networks (CNN) is mainly used to identify and classify diseases in leaves. In [6],there are three convolutional layers and three max pooling layers precede the next two fully linked layers in this model. The experimental findings of this study demonstrates that their proposed model's superior performance over models such as VGG16, MobileNet and InceptionV3, three pre-trained models. In the study, it suggested that the model's average classification accuracy for the distinct disease classes and one class (healthy) is 91.2%, which varies from 76% to 100% depending on the class.

X.E.Pantazi et al [7], the proposed technique illustrates a computational method for plant disease diagnosis on distinct leaf sample images matching different crops through Local Binary Patterns (LBPs) for the extraction of feature and for classification, One Class Classification. According to their suggested technique, the health status of each plant, such as black rot, healthy, downy mildew, and powdery mildew, would have its own Classifier. Extremely high performance was demonstrated when trained on vine leaves, while the testing was performed on different crops when uncertain samples of data that can belong to one or more conditions, an optimal method that proposes conflict resolution between One Class Classifiers identifies the right condition. For the combination of 46 plant-conditions that were evaluated, a 95% overall positive rate was obtained.

Md. Nabobi et al [8], this review study has examined both earlier and more recent efforts to identify plant leaf diseases. Because it is unexpected and due to the inconsistency, the conventional manual visual quality inspection could not be specified systematically. In addition to the excessive processing durations, it would also require a significant level of skill in the field of phytopathology that involves plant disease diagnosis. As an outcome, digital image processing has been widely used to identify certain plant diseases. There are three key sections to this study.. The first section offers a thorough analysis based on algorithms, comparing the most important algorithms used in both image processing and artificial intelligence research. The second section examines the ideas and contrasts them with earlier efforts.

III. EXISTING SYSTEM

The existing system includes the simple naked-eye observation that experts can use to detect plant diseases. Plants must be constantly monitored by a team of experts, which drives up costs for large farms.

Identifying plant diseases in laboratories can be a hectic task, and it can also be less accurate because the area of investigation is limited. [1] Image processing for plant disease detection uses algorithms such as K-means for color segmentation and GLCM for disease classification

In [2], the author used K-means to segment the disinfected area. GLCM for texture feature extraction and fuzzy logic for disease classification. An ANN is used to classify images of diseased leaves. K-means Algorithm along with ANN would be used to generate a better model with higher accuracy that can classify the diseases in the plants.

IV. PROPOSED METHODOLOGY

This study would recommend the use of photos of various sorts of plant parts mostly taken using a camera or a device capable of taking images, and those images are then analyzed to spot the damaged areas in the parts of the plant. Then, the images are analyzed using image-processing methods.

To determine if the leaf is healthy or diagnosed there are certain steps that need to be followed which include pre-processing, feature extraction, classification, and classifier training. Pre-processing would involve converting all the image sizes to a single uniform size.

The following would describe about the pipeline of the model proposed and various algorithms that are used to detect the diseases in plants:

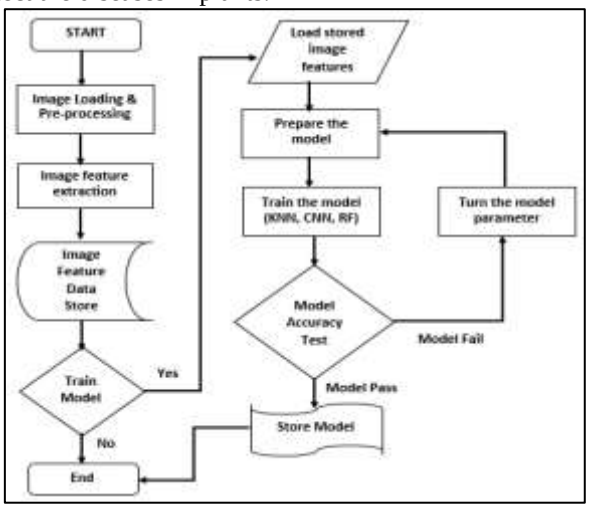

Fig. 1: Flowchart of Proposed Model

The following algorithm are included in this study for the image detection and classification for detection the diseases in plants:

A. Convolution Neural Network:

A convolutional neural network (CNN) is used for the classification of the image dataset. It takes an input image, assigns meaning to different aspects of the image, and can distinguish between them. The components of CNN include input layer, hidden layer and output layer. Sigmoid or Softmax functions are used to convert each class into the probability values of each class. The layers used to build the CNN model are the convolutional layer, the activation function layer, and the pool layer.

The various components that are associated with a CNN are listed below:

1) Convolution Layer: This layer is used to extract features from the input images. The output consists of information about the edges and corners of the images, called a feature map.
2) Pooling Layer: Reduction of computational cost is done by reducing size of folded feature map. There are different types of pooling operations depending on the method used.
3) Fully connected layer: consists of weights, biases and neurons. The input image from the previous slice is smoothed and a mathematical function operation is performed.
4) Dropout: If all the features are associated with the layer FC, it may lead to overfitting. To solve this problem, a dropout layer is used where some neurons are omitted, resulting in a reduction in the size of the model. This leads to an improvement in performance.
5) Activation Functions: The most commonly used activation functions are Sigmoid and Softmax for binary classification models and Softmax is used for multi-class classification.

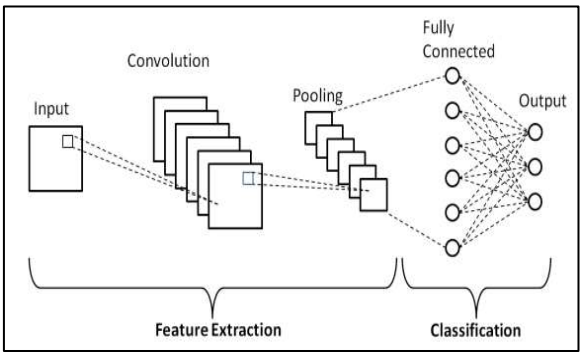

Fig. 2: Representation of a CNN

B. K-Nearest Neighbors:

K-Nearest Neighbors learns from a labeled training set and hence is a supervised machine learning algorithm. It consists of training data, the model trains it, and predicts the output as the class with the majority according to distance metric. The model takes an image for testing, calculates the k training images close to the test image, and predicts the output.

The algorithm for our model can be given as:

– Step-1: Gathering the image dataset regarding plant leaf-based diseases.
– Step-2: Splitting the dataset into testing and training for our model
– Step-3: Using KNN algorithm for training our model.
– Step-4: Evaluating the model performance and accuracy

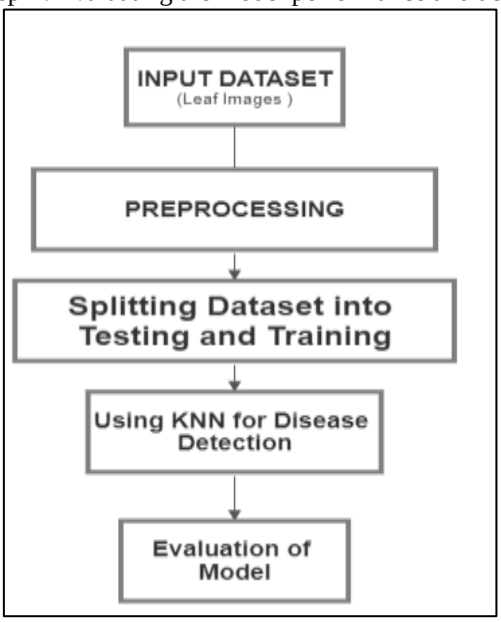

Fig. 3: Implementing KNN for Leaf Disease Detection

C. Random Forest Algorithm:

Random Forest is a machine learning algorithm which can be used for solving regression problems and classification problems. The random forest consists of decision trees trained using the bagging method. The bagging method is used to improve the overall result.

The fundamental idea of Random Forest Algorithm consists of combining several weak classifiers into a single powerful classifier. All nodes are either labeled as test or train set which helps in separating the input data. Each internal node represents a weak classifier that splits the samples according to a certain property. The most advantageous outcome determined by voting on all classification trees becomes the ultimate decision result of Random Forest Classifier.

– Step-1: Select samples randomly from a specified data set.
– Step 2: A decision tree will be built by the algorithm for every data set.
– Step-3. The Decision tree will then conduct voting.
– Step 4: Then the final prediction result is selected who got the most votes.

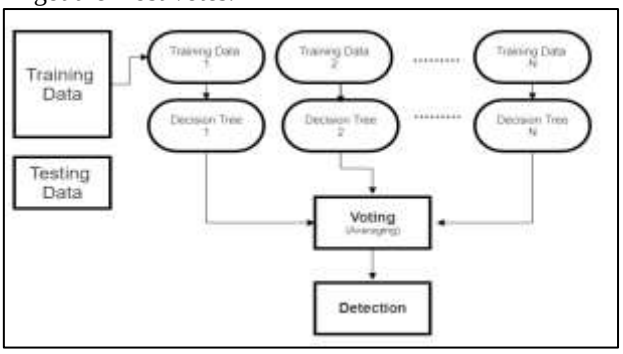

Fig. 4: Implementing Random Forest for our model

V. CONCLUSION

The technique that is being suggested for the plant leaf disease detection system focuses on developing a cutting-edge, effective system that makes it easier for farmers to produce crops with high yields. The project intends to use image processing techniques within upbringing technology to figure out the prominent diseases that affect plant leaves. In simpler terms, This will help farmers identify the disease just with the help of an image of that diseased plant.

Image Processing methods and Machine Learning algorithms help in identifying diseases. Different features are retrieved from the portion of the image that exhibits the leaf disease. The various techniques discussed in this study for the disease detection and classification in leaf include KNN, CNN and Random Forest Algorithm.

This study shows the comparison between our system and systems that already exist along with appropriate techniques and execution. It attempts to simplify farmers' lives. The technology might assist the agricultural industry since it improves agricultural production and yield, which is important because the agricultural industry is important source of income for Indian people.

REFERENCES

[1] Vijai Singh, A.K. Misra: "Detection of plant leaf diseases using image segmentation and soft computing techniques" Information Processing in Agriculture, Volume 4, Issue 1, 2017, Pages 41-49
[2] Aakanksha Rastogi, Ritika Arora and Shanu Sharma: "Leaf Disease Detection and Grading using Computer Vision Technology & Fuzzy Logic" 2nd International Conference on Signal Processing and Integrated Networks (SPIN)2015.
[3] Ch. Usha Kumari; S. Jeevan Prasad; 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)
[4] Santhosh S. Kumar; B.K. Raghavendra "Diseases Detection of Various Plant Leaf Using Image Processing Techniques: A Review" 2019 5th International Conference on Advanced Computing & Communication Systems (ICACCS)
[5] U. Shruthi; V. Nagaveni; B.K. Raghavendra: "A Review on Machine Learning Classification Techniques for Plant Disease Detection" 2019 5th International Conference on Advanced Computing & Communication Systems (ICACCS)
[6] Mohit Agarwal, Abhishek Singh, Siddhartha Arjaria, Amit Sinha, Suneet Gupta: "ToLeD: Tomato Leaf Disease Detection using Convolution Neural Network" International Conference on Computational Intelligence and Data Science (ICCIDS 2019)
[7] X.E.Pantazi,D.Moshou,A.Tamouridou: "Automated leaf disease detection in different crop species through image features analysis and One Class Classifiers" Aristotle University, School of Agriculture, Agricultural Engineering Laboratory, Thessaloniki 54124, Greece
[8] Md. Nabobi Hasan, Mufrad Mustavi, Md. Abu Jubaer, Md. Tanvir Shahriar, Tanvir Ahmed:"Plant Leaf Disease Detection Using Image Processing: A Comprehensive Review": A Comprehensive Review", Malaysian J. Sci. Adv. Tech., vol. 2, no. 4, pp. 174–182, Oct. 2022.

