Hybrid Classification Approach with Textural Feature Analysis for Plant Disease Detection

Neeraj Kumar[1], Mrs Kirandeep Kaur[2]
[1]*Research Scholar, Sant Kabir Institute of Technology and Management, Bahadurgarh, Haryana*
[2]*Assistant Professor, Sant Kabir Institute of Technology and Management, Bahadurgarh, Haryana*

Abstract - The significant task in image processing is to detect the diseases in plants as the data utilized for input is complex in nature. The infected plants are diagnosed in diverse phases. For this purpose, various algorithms are available. The prior work presented SVM (Support Vector Machine) algorithm in order to detect the disorder. This research work introduces a voting classification system for enhancing the diverse metrics such as accuracy, precision and recall obtained from the earlier work. The MALALB is applied to deploy the introduced and traditional technique. Some parameters are considered for analyzing the outcomes. The results of analysis depicted that the introduced approach outperforms the traditional approach concerning accuracy, precision and recall.

Keywords - Plant Disease, GLCM, K-mean, SVM, Voting Classifier

I. INTRODUCTION

In agricultural crops, leaves act significantly for offering information related to the amount and nature of horticultural yield. The productivity of food is affected due to various factors including climate changes, occurrence of weed and soil infertility. Another factor that has a great impact on the productivity of food is the occurrence of disease in plant and leaf due to which progress of various agricultural products are affected. If these infections or bacteria are not detected in plants consequently, the utilization of pesticide or fungicide is done inadequately. Thus, plant diseases are taken into consideration by the technical organization, with a focus on the biological attributes of diseases. Precision agriculture makes the implementation of the progressive expertise in order to optimize the decision-making. The experts and biological review often conduct the visual assessments using the plant diagnosis at the time of their necessity [1]. But, this technique consumes much time and it is not cost effective. The detection of diseases in plant is required with the help of advanced and intelligent methods for dealing with these issues. A number of studies utilize the traditional ML algorithms for carrying out the agricultural operations. But, DL methods as a subset of Machine Learning (ML) are exploited to detect, recognize and classify the objects efficiently in recent times. Thus, agricultural research takes the direction towards the Deep Learning-based solutions. The promising outcomes are acquired using the Deep Learning (DL) methods for executing the agricultural operations such as to differentiate the crop or weed, harvest the fruit and identify the plant. Likewise, other imperative agricultural issues related to identify the disease in plant is also considered in recent studies. Several conventional DL models are implemented for classifying the diseases occurred on plant for which popular Deep Learning (DL) architectures are deployed. Furthermore, modified versions of DL algorithms are also put forward by the researchers for enhancing the performance while categorizing the disease in numerous plant species.

Convolutional Networks belong to the family of Neural Networks. CNNs have proved their efficacy in the fields of image recognition and classification. ConNets have been successfully applied for recognizing faces, objects, and traffic signs in addition to strengthening vision in robots and automatic vehicles. In the last decade, CNN based plant disease classification has obtained remarkable outcomes. The continuing attainment of eye-catching results has increased the popularity of multi-layered supervised network among researchers. The main objective of Convolutional Network is the extraction of attributes from the input image. Convolution makes use of small filters to preserve the spatial association among pixels for which the attributes of an image are learned [2].

All steps mentioned in the above figure have been explained below:

a. Image Acquisition: This task is concerned with the acquirement of different plant imagery from different datasets. One such dataset is 'The PlantVilllage Dataset'. It is an open-access repository containing 54,323 images in total. Multiple classes are selected per specie. The images are generally captured in controlled environmental conditions. This might lead to model bias. A test dataset comprising a number of images can be obtained from Google as well to access this. These images comprise supplementary plant anatomy, in-field background data and variable disease stages.

b. Pre-Processing: Image pre-processing is effective technique in deciding the efficacy of a system. It is quite challenging to differentiate viral, bacterial, and fungal disorders, and generally, an overlay of symptoms appears. Such signs are available in the form of quantifiable variation in color, shape, or function that occurs as the plant reacts to the pathogen. This criticality might be overcome by using RGB images. It generates clear, de-noised images that take more time in contrast to the greyscale image in the training phase. However, generally are more appropriate in plant disease recognition frameworks. The reliability of a model is sensitive to smaller datasets or unvaried data [3].

c. Segmentation: Segmentation is the process of separating a leaf from its background. It is also possible to apply technique this approach in conditions where the classification model needs the knowledge of scene. For instance, this approach can be used to understand the level of pathogen harm around the diseased tissue, contrasted with only the diseased tissue. Segmentation is an old concept. This approach has been implemented for classifying disease since the 1990s. This approach provided good results even at this early stage. The work carried out in the past provided efficacy to recognize the drawbacks. These works show that the method is incapable to improve the image quality. Therefore, it is important to carefully collect data, and performing its pre-processing. Combining this with particular imagery has huge research potential.

d. Feature Extraction: Generally, color, contour, texture, and shape of plant leaves are used for classifying plants. The grayscale image is transformed into a binary image by binarization, and then contour is extracted. The features of the contour line are used to extract features. The machine learning based classification of these features provides recognition rate of 90%. It is quite challenging to perform classification based on just features as similarity in the shape of the leaf outlines occur generally.

$G\_rau=0.299\times I\_r+0.587+0.114\times I\ldots(1)$

Apart from this, it is possible to apply brightness or shape transformation with cumulative histogram operations.

e. Classification by CNN: The plant disease detection based on CNN is divided into three stages. All of these stages have been explained below [4]:

Stage 1-Trialling of Image size: the first stage is concerned with investigating the effect of image size the model performance. Overall, the testing of 5 images of diverse size is performed. The initial task is to download the pre-trained weights namely Resnet34. All the layers excluding final two layers are frozen as a default of transfer learning. The novel weights are comprised in these layers and considered suitable to accomplish the task of classifying the plant disease. These layers are facilitated to be disease separately trained in freezing without any back propagation of gradients. Similar to this way, the final layers are trained using one-cycle policy. The completion of this process leads to release the remaining layers. The fine-tuning procedure is completed with the generation and analysis of plot that displayed the learning rate and loss. An appropriate learning is chosen from this and the model is executed. The recreation of model is carried out to the additional four image sizes using the results. There is not any change in every trial such as the learning rate.

Stage-2 Optimization: The ResNet34 model is optimized by the most appropriate image size. The performance of the model is further optimized by adding more augmentation settings. After this, the last 2 layers are separated and their training is done at the default learning rate. When it is completed, fine tuning is carried out in which many trials are run for testing a sequence of learning rates and amount of epochs.

Stage-3 Visualizations: A number of visualizations are generated on the basis of validation and test datasets for understanding purpose. Moreover, the model is devised for creating a web application. This is carried out by storing the important files which are accomplished in a GitHub repository and the system is transferred as a pickle file. This system is deployed by connecting the repository to the united platform called Render. This task is completed by using the ‘Render Examples’ GitHub repository as a guide.

II. LITERATURE SURVEY

Punam Bedi, et.al (2021) investigated an innovative hybrid model in which CAE (Convolutional Autoencoder) network was integrated with CNN (Convolutional Neural Network) for detecting the diseases on plant in automatic way [7]. The Bacterial Spot disease which affected the peach plants was detected using this model considering their leaf images. Though, this model was capable of detecting any plant disease. PlantVillage dataset was applied to conduct the experiments. The suggested model yielded accuracy up to 99.35% in training phase and 98.38% in testing phase. Moreover, the suggested model utilized only least amount of metrics in comparison with other methods.

Xulang Guan, et.al (2021) devised an effectual approach such that the disease occurred on plants was diagnosed. This approach focused on integrating 4 CNN algorithms such as Inception, Resnet, Inception Resnet and Densenet and implementing them [8]. A stacking technique was presented to process the outcomes of these algorithms. A publicly available dataset containing 36258 images was executed in the experimentation. The accuracy acquired from the presented technique was calculated 87% which was found higher in contrast to individual model. This accuracy rate demonstrated the appropriateness of the introduced approach via staking technique. This approach was further expanded for the practical cultivation conditions as an enhanced warning tool.

Melike Sardogan, et.al (2018) projected an approach on the basis of CNN model and LVQ system in order to detect and classify the diseases on leaf tomato plant [9]. The attributes were extracted and classified by modeling CNN algorithm. An analysis was performed on the plant leaf disease based on color information. This model deployed filters on 3 channels on the basis of RGB elements. The output feature vector of convolution part was employed in Learning Vector Quantization algorithm for training the network. The projected approach was quantified on dataset in which 500 images of tomato leaves having 4 symptoms were comprised. The experimental outcomes revealed the effectiveness of the projected approach for recognizing 4 diverse kinds of diseases.

Hilman F. Pardede, et.al (2018) developed an unsupervised feature learning algorithm that exploited the CAE for diagnosing the disorders of plant [10]. First of all, the manual attributes were not required in this algorithm because of its potential to generate the discriminative features. This algorithm had not assigned labels to data. After that, the SVM algorithms made the deployed of output of the autoencoder as inputs for detecting the diseases of plant automatically. In the end, the results exhibited the supremacy of the developed algorithm over the traditional algorithm.

Hui Fuang Ng, et.al (2021) designed a mobile application so that the plant disease was detected and classified with the help of DL (deep learning) model of detecting the object [11]. Faster R-CNN detector was implemented with Inception-v2 for providing robustness and efficiency for detecting the diseased plants. The experimental results indicated that the designed application offered accuracy up to 97.9%. This application assisted the farmers having no knowledge in detecting and controlling the diseases in plants at initial phase. This resulted in alleviating the losses and preventing the further spread of the disease.

N Radha, et.al (2021) presented a model in order to monitor the plant and detect the plant disease at initial phase [12]. The automatic methods employed to detect the plant diseases were effective for detecting the signs of disease in advance. This model was evaluated using a dataset in which images related to affected and normal leaves were included. CNN (Convolutional Neural Network) model was adopted to train the model for detecting the plant diseases. The presented model was able to attain the accuracy of 85% for detecting the diseases of plants and negligible loss was found 0.25 in the process when the data was trained.

Husnul Ajra, et.al (2020) investigated a technique for detecting and preventing the diseases occurred on plants leaf in the farming sector in which image was processed and 2 CNN models namely AlexNet and ResNet-50 were deployed [13]. At first, Kaggle dataset was implemented to simulate this method with the purpose of recognizing the symptoms of affected leaf. Thereafter, the attributes were extracted and classified from the images while detecting leaf diseases. The outcomes of experiment depicted that the investigated method was efficient and the initial model offered the accuracy 97% and ResNet-50 offered 96.5% accuracy for detecting the plant diseases.

III. RESEARCH METHODOLOGY

This research work is focused on recognizing the diseases in plants. The entire cycle to detect the infection contains different functions that are discussed as:

1. Pre-processing: - The image is pre-processed in this function with the objective of diagnosing the diseases from leaves of plants. The images are captured to be fed in input. A consistent data source is considered to collect these images. A public dataset called Plant Village is generated in which all the collected are comprised. The Plant Village is a website that is helpful for deriving the information about the plant and its disease’s kind. The images taken from wheat are involved in the dataset. This dataset consists of 3 portions in which images normal leaves, leaves suffered from early blight and the images of late blight disease, are inserted. This process also focuses on transforming the input images into gray scale.

2. Segmentation: - The second function aims to split a digitized image into distinct portions. The process to segment an image is adopted to identify the objects and retrieve the information from the images. The images are easily analyzed in this stage. The technique of segmenting an image is put forward to locate the RoI and bounding line of pictures. every pixel is marked with a label. Every pixel having similar label leads to strengthen several features. This work adopts KMC (K-means clustering) for segmenting the pictures of plant leaves. Moreover, the samples are gathered into various clusters according to the distance using this algorithm. Two points having least distance amid them are assisted in offering compressed and independent cluster as a closing target. The 3 is taken as the optimal value for the input. The value of k is considered to segment an image. After that, this phase selects the essential portion from the part of input leaf which suffers from infection.

3. Feature Extraction: - The prior stage offers the output in the form of RoI (region of interest). This function is carried for retrieving the attributes from the required region. This process is effective to extract a set of values known as attribute. The further processing is done on the basis of attributes having information regarding the images. A number of elements namely color, texture, morphology and color coherence vector help in detecting the infections in plants. Several methods are

adopted for extracting the attributes from the images. A diseases diagnosis system is constructed on the basis of these features. GLCM, SGLDM, and HOG algorithm are useful approaches for extracting the attributes. The texture features are classified via GLCM.

4. Classification: - The final stage is carried out for generating a classification algorithm in order to diagnose the infected plants. There are 2 sections of entire dataset. The amount of data in the training section is higher in contrast to the testing section. This work projects KNN (K-Nearest Neighbor) algorithm to classify the data. The unknown samples have association with the known ones or the similarity functions are employed to represent this algorithm. The projected algorithm is trained and tested at the same time. This algorithm is efficient for investigating K nearest centers and allocating the square of the greater part to the unknown instance. The majority voting and its k neighbors are considered to classify the data. Moreover, an efficient and robust ML (machine learning) algorithm called RF (Random Forest) is implemented in which a variety of tree predictors are involved. This algorithm is capable of proving optimal outcomes and handling the large volume of data. The construction of RF is done on the basis of integrating various DT (decision trees) to yield the results at superior accuracy. Every DT contains a set of rules in accordance with the input features values. The optimization of these attributes is done to classify all the components. This algorithm faces an issue related to overlapping of RTs (random trees). A random subset of attributes comprised in the dataset assists in making the results more promising. The voting algorithm is fed with the output generated through RF (Random Forest) and KNN (K-Nearest Neighbor). The introduced algorithm is adaptable to cast its vote to either of the both algorithms and achieve the final predictive outcomes.

IV. RESULT AND DISCUSSION

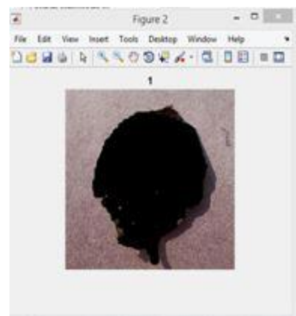
This work considers Plant Village website to gather a dataset. A set of images wheat is taken in this dataset so that diverse kinds of infections of plants are diagnosed. The execution of every stage is discussed as:

1. Input Database:- The collection of data is done from the plant village dataset. The sample images are represented as:

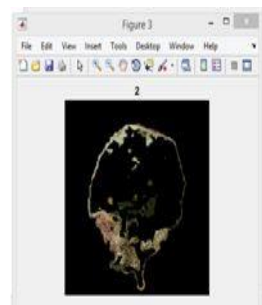
Figure 1: Sample Images

Figure 1 depicts the representation of the sample images utilized to accomplish the further processing.

2. Segmentation:- The K-Means algorithm is adopted for segmenting the images into specific portions. The images whose segmentation is done are defined in figure 2.



a. Segment 1



b. Segment 2

c. Segment 3

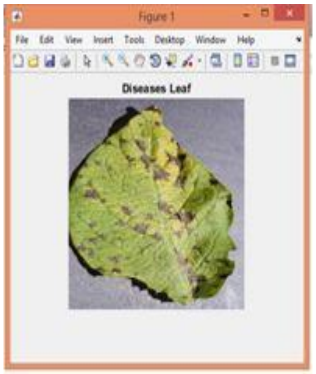
Figure 2: Region Based Segmentation

Figure 2 depicts the implementation of the technique of k-mean to segment the images based on region. the experiments take 3 for the k value. Thus, the number of formed segments is three. The a, b, and c illustrates the 3 segments respectively.

3. Feature Extraction:- This stage is executed to extract the features. The GLCM (gray-level co-occurrence matrix)

algorithm is adopted to retrieve 13 features to classify the images.

4. Classification:- The final stage emphasized on predicting the disease. The figure 4 defines the predicted disease image.



a. Input Image

b. Detected Disease

Figure 3: Disease Detection

Figure 3 represents that the image a is utilized for the input and image b is used to illustrate the detected image which contains infection.

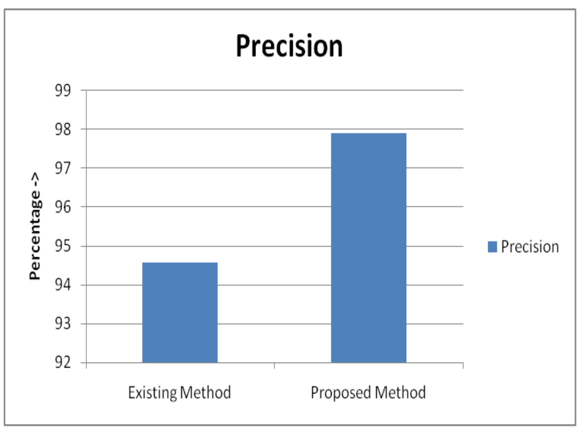


Figure 4: Precision Comparison

Figure 4 demonstrated that the introduced system offers more efficiency as compared to the existing system with regard to the precision.

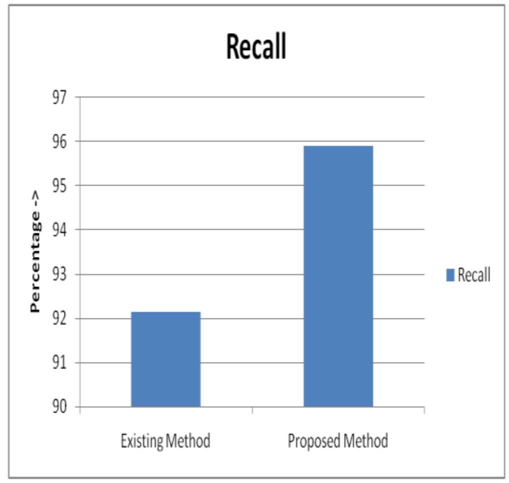


Figure 5: Recall Comparison

Figure 5 exhibits that the introduced system is better in contrast to the existing one with regard to recall value.

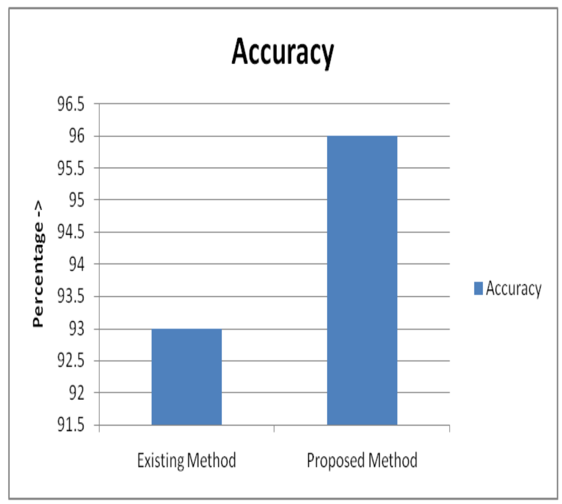


Figure 6: Accuracy Comparison

Figure 6 indicates that the introduced system outperformed the existing system concerning accuracy value.

V. CONCLUSION

This work concludes that the introduced approach is applicable to diagnose the diseases occurred on plants. This technique detects the infected plants in diverse phases. The K-Means is adopted to segment the plant image. The law textural technique is exploited to extract the attributes. The disease is predicted using RF (Random Forest) and DT (Decision Tree) algorithms. A comparative analysis is conducted on the introduced approach against the SVM (Support Vector Machine) for diagnosing the ailment of plant. The introduced approach offers more effectiveness in comparison with the conventional technique concerning accuracy, precision, and recall to detect the infected plants.

VI. REFERENCES

[1] Aravindhan Venkataramanan, Deepak Kumar P Honakeri, Pooja Agarwal, “Plant Disease Detection and Classification Using Deep Neural Networks”, 2019, International Journal on Computer Science and Engineering (IJCSE), Vol. 11 No. 08

[2] Nikhil Patil, Rajab Ali, Vaibhav Wankhedkar, Prof. Deepali Nayak, “Crop Disease Detection using Deep Convolutional Neural Networks”, 2019, International Journal of Engineering Research & Technology (IJERT)

[3] Shivani Machha, Nikita Jadhav, Himali Kasar, Prof. Sumita Chandak, “Crop Leaf Disease Diagnosis Using Convolutional Neural Network”, 2020, International Research Journal of Engineering and Technology (IRJET), Volume: 07 Issue: 02

[4] Rajashree Patil, Dr. Sampada Gulvani, “Plant Disease Detection using Neural Network: a review”, 2019, Journal of Emerging Technologies and Innovative Research (JETIR), Vol. 6, Issue 2

[5] Divya Verma, Gurpreet singh, Hatesh Shyan, “Multilayer Convolutional Neural Network for Plant Diseases Detection”, 2020, International Journal of Engineering and Advanced Technology (IJEAT), Volume 9 Issue 5

[6] R.Sangeetha, M. Mary Shanthi Rani, “Tomato Leaf Disease Prediction using Convolutional Neural Network”, 2019, International Journal of Innovative Technology and Exploring Engineering (IJITEE), Volume 9 Issue 1

[7] Punam Bedi, Pushkar Gole, “Plant disease detection using hybrid model based on convolutional autoencoder and convolutional neural network”, 2021, Artificial Intelligence in Agriculture

[8] Xulang Guan, “A Novel Method of Plant Leaf Disease Detection Based on Deep Learning and Convolutional Neural Network”, 2021, 6th International Conference on Intelligent Computing and Signal Processing (ICSP)

[9] Melike Sardogan, Adem Tuncer, Yunus Ozen, “Plant Leaf Disease Detection and Classification Based on CNN with LVQ Algorithm”, 2018, 3rd International Conference on Computer Science and Engineering (UBMK)

[10] Hilman F. Pardede, Endang Suryawati, Rika Sustika, Vicky Zilvan, “Unsupervised Convolutional Autoencoder-Based Feature Learning for Automatic Detection of Plant Diseases”, 2018, International Conference on Computer, Control, Informatics and its Applications (IC3INA)

[11] Hui Fuang Ng, Chih-Yang Lin, Joon Huang Chuah, Hung Khoon Tan, Kar Hang Leung, “Plant Disease Detection Mobile Application Development using Deep Learning”, 2021, International Conference on Computer & Information Sciences (ICCOINS)

[12] N Radha, R Swathika, “A Polyhouse: Plant Monitoring and Diseases Detection using CNN”, 2021, International Conference on Artificial Intelligence and Smart Systems (ICAIS)

[13] Husnul Ajra, Mst. Khairun Nahar, Lipika Sarkar, Md. Shohidul Islam, “Disease Detection of Plant Leaf using Image Processing and CNN with Preventive Measures”, 2020, Emerging Technology in Computing, Communication and Electronics (ETCCE)