Hybrid Classification Method for Plant Disease Detection

Damini Bhardwaj[1], Vishal Kalra[2]

Research Scholar, Guru Nanak Institute of Technology, Mullana

Assistant Professor, Guru Nanak Institute of Technology, Mullana

Abstract - Image processing is majorly utilized for diagnosing the diseases occurred on plants due to the deployment of complex data in input. Various stages are executed to detect the infections in plants. This task is accomplished using diverse algorithms. The former work suggests a SVM (Support Vector Machine) model for diagnosing the infection. This research work projects a voting system to improve several evaluation components like accuracy, precision and recall generated through the former work. MATLAB is executed to simulate the projected model. An analysis is conducted on the results with respect to some metrics. The outcomes indicated that the projected model performed more effectively in contrast to the conventional techniques on the basis of accuracy, precision and recall.

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

I. INTRODUCTION

In the agriculture field, the information regarding the amount and nature of production of crops is obtained through the leaves. A number of factors such as climate changes, existence of weed and soil infertility lay impact on the yield of food. Another factor is the infection occurred on the plants and leaves that affects the growth of farming production. In case of failure of detecting such disorders or bacteria, the pesticide or fungicide is applied in insufficient manner. Therefore, the technical experts focus on plant infection and the biological attributes of disorders. Accurate farming aims to deploy the progressive expertise during the optimization of the process to make decisions. Most frequently, the experts and biological review detect the plant for the visual evaluations when it is needed [1]. However, this method leads to utilize much time and proved more expensive. The significant task of innovative and intelligent approaches is to diagnose the infections of plant to tackle such problems. Diverse studies make the exploitation of conventional ML (Machine Learning) models to accomplish the farming tasks. However, DL (Deep Learning) techniques which are a subpart of ML, are effective for detecting, identifying and classifying the objects accurately at present. Therefore, farming sector emphasizes on DL based approaches as they generate optimal results to carry out the agricultural tasks for distinguishing the crop or weed, harvesting the fruit and recognizing the plant. Meantime, the fundamental goal of the present studies is on addressing the issues of recognizing the diseases occurred on plants. Numerous traditional DL (Deep Learning) are presented to classify the disorders. In addition, the authors concentrate on extending the existing algorithms to improve the efficacy to classify the disorders in various plant species.

CNN (Convolutional Neural Network) is a kind of NN (Neural Network). This algorithm is effective in the field of recognizing and classifying the image. The faces, objects and traffic signs are identified and the vision in robots and automatic vehicles is supported using CNN. Over the last years, the technique based on CNN (Convolutional Neural Network) has generated promising outcomes. Its effective outcomes arouse the attention of researchers towards multi-layered supervised network. This algorithm emphasizes on extracting the attributes from the input picture. Small filters are employed for preserving the spatial relation among pixels after learning the attributes of an image [2].

Diverse stages to diagnose the diseases of plants are defined as:

a. Image Acquisition: It is a primary stage executed to attain images of several plants from diverse datasets. PlantVillage dataset is considered that is available publically. There are54,323 images included in this dataset. The selection of several classes is done for every specie. The organized circumstances are considered to capture images. The testing dataset is generated from the Google in which several images are comprised. These pictures offer additional plant anatomy, background data and variable phases of disease.

b. Pre-Processing: The efficacy of the framework is computed with the process used to pre-process the image. The complex task is of differentiating the viral, bacterial, and fungal ailments. In general, this process displays an overlay of symptoms including quantifiable variation in color, shape, or function. These symptoms are taken place after the reaction of plant to the pathogen. Thus, the RGB images are employed to address this issue. The images obtained in this phase are free of noise and clear, whose time usage is higher in comparison with the greyscale image in the training stage. Though, these images are effective in models of recognizing the disorders in plants. These models work reliably on small sized datasets [3].

c. Segmentation: This approach is employed to split a leaf from its background. The process of segmenting the images is significant in various conditions in which classifiers requires the complete information regarding the scene. To illustrate, the level of pathogen harm around the diseased tissue is understood in this stage. The segmentation is adopted to classify the disease since 1990s and optimal outcomes are obtained at initial phases. The earlier work attains good accuracy for examining the limitations. These works indicate the inefficacy of the technique in enhancing the image quality. Thus, the major focus is on gathering the data and to pre-process the data. The research is enhanced when it is integrated with specific imagery.

d. Feature Extraction: In general, the plants are classified according to the color, contour, texture, and shape of plant leaves. The binarization is adopted to convert the grayscale image into a binary image and to extract the contour later on. The attributes are extracted using the features of the contour line. These attributes offer the recognition rate up to 90% based on ML (machine learning). The complex task is to classify the image on the basis of features due to the occurrence of the similarity in the shape of the leaf outlines.

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

Furthermore, the brightness or shape transformation is applied easily with cumulative histogram operations.

e. Classification by CNN: The diseases of plants are diagnosed on the basis of CNN (Convolutional Neural Network) into 2 phases. These phases are discussed as [4]:

Stage-1 Trialing of Image size: The fundamental objective of this phase is to discover the impact of image size on the efficacy of the system. Generally, 5 images having dissimilar size are considered in the testing phase. Primarily, this phase emphasizes on downloading the weights namely Resnet34 whose pre-training is done. This approach leads to freeze all the layers without considering last 2 layers as a default of TL (transfer learning). These layers are consisted in the new weights and effective for a task to classify the plant disorders. These layers are useful for training the disease at individual level when the images are frozen without applying BP (back propagation) of gradients. One-cycle policy is employed to train the final layers. This procedure results in releasing the rest of the layers. The plot is created and analyzed to accomplish this process for illustrating the learning rate and loss. This process is assisted in selecting a suitable learning and implementing an algorithm. Moreover, the outcomes are considered to reconstruct the system on the extra 4 images sizes. There is no change found in every trial like learning rate.

Stage-2 Optimization: The suitable image size is exploited for the optimization of the ResNet34 algorithm. More settings of augmentation are inserted for optimizing the efficacy of the system. Thereafter, this approach concentrates on splitting last 2 layers and to train them at the default learning rate. These layers are adjusted to run several trials to test a sequence of learning rates and amount of epochs.

Stage-3 Visualizations: More information is obtained by performing various visualizations according to the datasets executed to test and validate the system. Furthermore, a web application is developed through a framework. For this, the considerable files are stored in a GitHub repository and the transmission of a system is done as a pickle file. A link is established amid the repository and the united platform known as Render to implement the presented framework. The ‘Render Examples’ GitHub repository is utilized as a guide to accomplish this task.

II. LITERATURE SURVEY

Punam Bedi, et.al (2021) studied a pioneering hybrid model integrating a CAE (conventional autoencoder) network with a CNN (conventional neural network) to automate disease detection on plants [7]. Based on the photographs of leaves, this architecture identified Bacterial spot disorder deteriorating peach plants. However, this model was able to detect all types of plant diseases. This architecture tested on the PlantVillage dataset. The architecture achieved accuracy of 99.35% and 93.8% in the training and testing phases, respectively. In addition, the new framework used minimal metrics compared to other PDD architectures.

Xulang Guan, et.al (2021) fostered an effective strategy for diagnosing plant disease. This strategy integrated and implemented 4 ConvNet algorithms namely, Inception, ResNet, InceptionResNet and DenseNet [8]. This work set forth a stacking method to process the results of these algorithmic schemes. An open-source dataset comprising 36258 photographs was used for the experiment. The accuracy obtained with the new method was counted as 87%, which was more than the accuracy of a single framework. This accuracy rate confirmed the suitability of the new framework implementing the staking method. This approach was further adapted to real-world farming conditions as an advanced warning mechanism.

Melike Sardogan, et.al (2018) hypothesized an approach based on the ConvNet and the LVQ scheme to identify and distinguish disease on the leaves of tomato plants [9]. Features were retrieved and classified by CNN algorithmic modeling. This work used colour info to study diseases on the leaves of plants. This architecture applied filtering on 3 channels based on RGB essentials. The in the learning vector quantization algorithm trained the network using output feature vector of the convolution segment. The new methodology was evaluated on a dataset consisting of 500 images of tomato leaves with 4 traits. The new method fruitfully identified various diseases according to the results obtained.

Hilman F. Pardede, et.al (2018) formulated an unsupervised feature learning algorithm based on CAE to diagnose plant infections [10]. First, this algorithm did not require manual features because of its ability to produce discriminatory attributes. This algorithm did not assign labels to the data. Thereafter, the SVM algorithm inputted the output of the autoencoder into the framework to automate disease detection on plants. In conclusion, the results demonstrated the dominance of the formulated algorithm over the conventional approach.

Hui Fuang Ng, et.al (2021) developed a mobile application to detect and classify plant diseases by virtue of a DL (deep learning) architecture of object detection [11]. In order to detect plant diseases with good efficiency and high robustness, this work applied Faster R-CNN detector with Inception-v2 architecture. The developed architecture achieved about 97.9% accuracy in the experiment. This application facilitated the farmers who did not know how to detect and control plant diseases in the early stage. This minimized the damage and prevented further spread of the disease.

N Radha, et.al (2021) fabricated an architecture for plant monitoring and detection of plant disease at an early stage [12]. Automated techniques of plant disease detection effectively and timely detected disease symptoms. This architecture was evaluated using a dataset consisting of images relating to contaminated and heathy leaves. ConvNet (Traditional Neural Network) model was used to detect plant diseases by training the model. The fabricated architecture achieved 85% accuracy for the detection of plant diseases and a small loss of 0.25 in the process of training the data.

Husnul Ajra, et.al (2020) studied a methodology that applied two ConvNet architectures namely AlexNet and ResNet-50 and pre-processed photographs to detect and prevent plant leaf diseases in the field of cultivation [13]. Initially, this work used the Kaggle dataset for simulations aimed at identifying affected leaf traits. Subsequently, diseases were identified on the leaves by extracting and classifying the characteristics from the photographs. The results of the tests demonstrated that the method tested was efficient and that the fabricated architecture delivered 97% accuracy and ResNet-50 delivered 96.5% accuracy for the detection of plant disorders.

III. RESEARCH METHODOLOGY

This research concentrates on identifying diseases in plants. The whole cycle of infection detection consists of various operations which are described as follows:

1. Pre-processing: - This step applies pre-processing operation on images to detect diseases caused on the plants' leaves. Photographs are acquired to be inpythe input. A coherent data source is assumed to collect these images. This work uses an open-source dataset named Plant Village which includes all the aggregates. Plant Village as a website helps getting information about the plant and its disease types. Images acquired of wheat plants are included in the dataset. This dataset has three sections in which images of normal leaves, early blight and late blight are inserted. This process is aimed at converting the input photographs to gray scale.

2. Segmentation: - The second step is to split a digitized photograph into separate parts. The process of segmenting a photograph is applied to classify objects and extract information from the photographs. This step simplifies the image analysis operation. This step uses an image segmentation technique to find the interest areas and the bounding line of the images. A specific label if assigned to each pixel. Each pixel with the same label strengthens multiple features. This work segments the photographs of leaves of plants by implementing KMC (K-Means Clustering). Furthermore, using this algorithm the samples are collected into different clusters according to the distance.The two points with the shortest distance help to offer a narrower and independent cluster as an end target. The finest value for the input is 3. The value of k is supposed to segment a photograph. Thereafter, this step picks the required part from the fragment of the input leaf that is prone to infection.

3. Feature Extraction: - The previous step provides the output as interest area. The main objective of this step is to extract features from the interest area. This operation effectively extracts a set of values called features. More processing is done based on the characteristics that carry information about the photographs. Several features such as colour, texture, morphology and color coherence vector help identifying plant disorders. There are several methods used to extract features from images. By virtue of these characteristics a diagnostic system is designed. The GLCM, SGLDM, and HOG algorithms are the most common features extraction algorithms. GLCM is used to classify texture features.

4. Classification: - The last step is to devise a classification architecture for the diagnosis of diseased plants. The whole dataset has two parts. The training section uses more data than

the testing section. This step classifies data using KNN (K-Nearest Neighbor) algorithm. Indefinite patterns are correlated with known patterns or similarity functions are used to signify this algorithm. The training and testing of the newer algorithm is done simultaneously. This algorithm can efficiently examine the K nearest centers and assign the class of the larger section to the unknown case. Majority of votes and its k neighbors are the basis for data classification. In addition, this work enforces an effectual and powerful ML (Machine Learning) algorithm named RF (Random Forest) which includes different types of tree predictions. This algorithm has potential to deliver optimal results and manage large amounts of data. RF is formulated by integrating many DTs (Decision Trees) to obtain results with improved accuracy. Each DT has a set of rules as per the values of the input features. These attributes are optimized to classify all elements. This algorithm encounters a problem associated with the overlapping of RTs (Random Trees). A random subset of the features included in the dataset helps to yield outstanding results. The output generated by RF (Random Forest) and KNN (K-Nearest Neighbor) is inputted into the voting classifier. The formulated framework is optimized for either of the two algorithms to cast their vote and obtain the final prognosis result.

IV. RESULT AND DISCUSSION

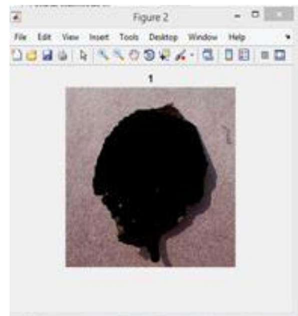
This work uses Plant Village website for collecting a dataset. This dataset is consisted of a set of images of wheat for diagnosing distinct disorders of plants. Every phase is explained as:

1. Input Database: - The given dataset is employed to gather the data. The sample images are illustrated as:

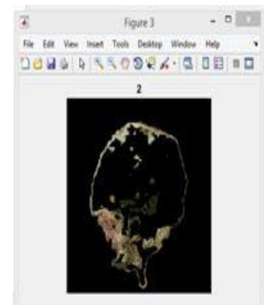
Figure 1: Sample Images

Figure 1 demonstrates the sample images which are useful to further process the data

2. Segmentation: - This stage emphasizes on implementing KMC (K-Means Clustering) algorithm so that the images are segmented into certain sections. Figure 2 explains the images after segmenting them.



a. Segment 1



b. Segment 2

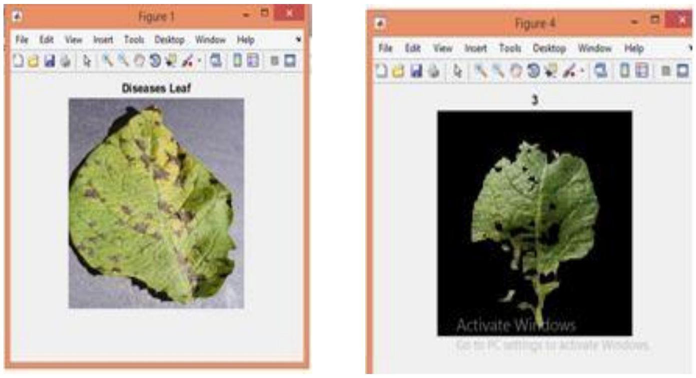
c. Segment 3

Figure 2: Region Based Segmentation

Figure 2 exhibits that the KMC algorithm is deployed for segmenting the images on the basis of region. Three is assigned for the k value in the experimentation. This implies that there are three segments whose formation is done using this algorithm. The a, b, and c are utilized to signify these 3 segments individually.

3. Feature Extraction: - In this phase, a variety of attributes are retrieved. The images are classified on the basis of attributes which are extracted from the GLCM (gray-level co-occurrence matrix) algorithm.

4. Classification: - This phase is carried out to predict the disease. The predictive disease image is illustrated in the given figure.



a. Input Image b. Detected Disease

Figure 3: Disease Detection

Figure 3 demonstrates the deployment of image ‘a’ to be inserted in the input and image b helps in representing the detected image related to disorders of plants.

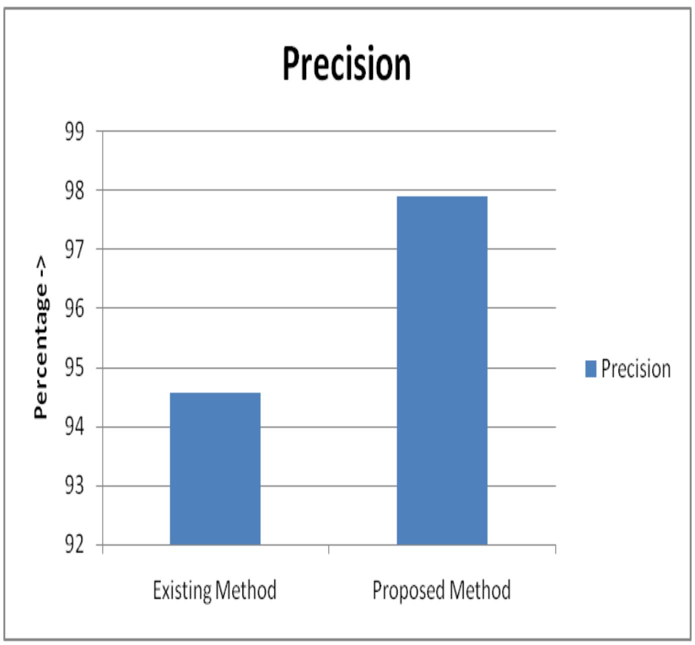


Figure 4: Precision Comparison

Figure 4 depicts the supremacy of the projected approach against the traditional approach concerning precision.

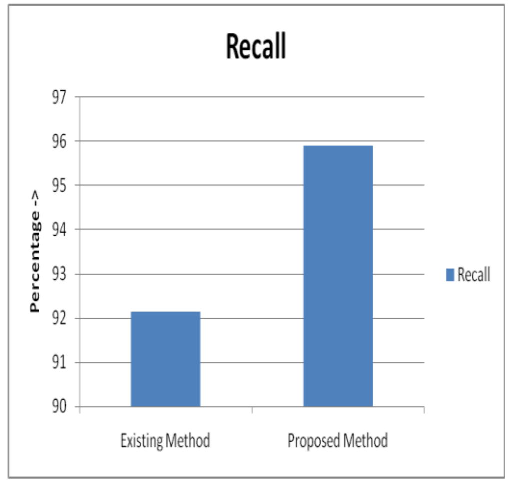


Figure 5: Recall Comparison

Figure 5 exhibits that the projected approach yields higher recall in comparison with the traditional technique.

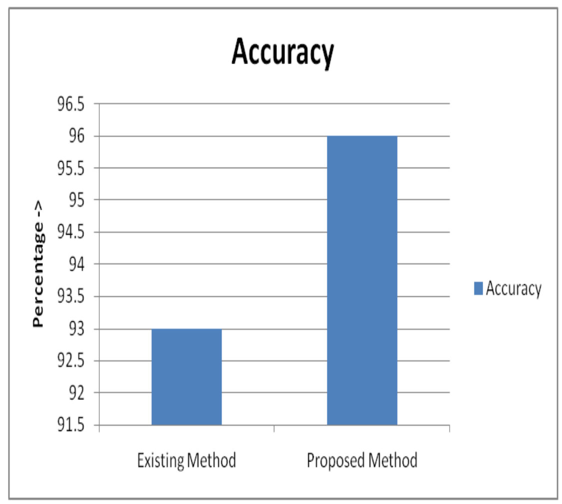


Figure 6: Accuracy Comparison

Figure 6 denotes that the accuracy obtained from the projected approach was higher in comparison with the traditional approach.

V. CONCLUSION

The conclusion drawn from this work is that the formulated architecture is appropriate to the diagnosis of plant diseases. This framework distinguishes infected plants at different stages. The image of the plant is segmented using K means algorithm. Features are extracted by applying law textural method Disease prediction is done using RF (Random Forest) and DT (Decision Tree) algorithms. The performance of the formulated architecture is compared with SVM (Support Vector Machine) for plant disease diagnosis. The newer architecture provides greater efficacy than the state-of-the-art methodology with regard to accuracy, precision and recall for identifying diseased 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)