ESP-JETA
ESP Journal of Engineering & Technology Advancements
ISSN: 2583-2646 / Volume 3 Issue 1 February 2023 / Page No: 98-106
Paper Id: JETA-V3I2P105 / Doi: 10.56472/25832646/JETA-V3I2P105

Original Article

Classification of Potato Plant Leaf Diseases Using Convolution Neural Networks

Umesh D. Shirale[1], Dr. Ajij Sayyad[2]

[1,2]*Electronics & Telecommunication Engineering, Maharashtra Institute of Technology, Aurangabad, India.*

Received Date: 07 February 2023 Revised Date: 20 February 2023 Accepted Date: 27 February 2023

Abstract: *One of the most common food crops grown worldwide is the potato. Numerous diseases prevent potato plants from growing properly. This plant's leaf area has observable illnesses. Early Blight (EB) and Late Blight are two prevalent leaf ailments that affect potato plants (LB). To improve the yield of this crop, it would be very beneficial if these diseases were discovered early on. Image processing is the greatest solution for resolving this issue by identifying and assessing these disorders. The system to identify and categorize potato leaf diseases proposed in this paper is based on deep learning and image processing. The 2152 images of healthy and diseased potato leaves utilized in this study were collected from the publicly accessible Plant Village database. Convolution neural networks were employed for object recognition, and logistic regression was used to accurately classify diseased from healthy leaves. Our suggested strategy results in a path for automatic plant leaf disease detection.*

Keywords: *Potato diseases, Late blight, Early blight, CNN*

I. INTRODUCTION

Agriculture is the main source of food, raw material, and fuel which contributes to the economic development of a nation. Nearly 66% of the population depends on agriculture directly or indirectly. As there is rapid growth in the global population, agriculture is struggling to fulfill its necessity. Food security remains threatened by various circumstances including climate change, the decline in pollinators, crop diseases, lack of irrigation, etc. Crop disease decreases both food quality and yield.

Potato (Solanum tuberosum) is the third most important food crop in the world, after cereals and rice, which is used by various food industries. Overall production exceeds beyond 300 million metric tons and is an important nutrition and calorie provider for humanity. According to the Agricultural and Processed Food Products Export Development Authority (APEDA), Uttar Pradesh produces more than 30.33% of all the potatoes in India. Cotton and worsted are sized using potato starch (farina) in the textile business. There is a heavy loss in potato production due to diseases despite their importance. Microorganisms, genetic abnormalities, and infectious agents like bacteria, fungi, and viruses are the main causes of these disorders. The majority of bacteria and fungi cause diseases in potato leaves. While soft rot and common scab are bacterial diseases, late blight, and early blight are fungal diseases [3]. Early blight and late blight are common diseases in potatoes that primarily affect leaves and stem. These diseases impact crops, resulting in significant losses to farmers and agricultural productivity. Crop diseases have negative effects on small-scale farmers whose livelihoods depend on proper cultivation, in addition to having an impact on global food security. The ability to control agricultural diseases by spotting them as soon as they appear on crops is a benefit. Therefore, it is very essential to recognize diseases early so that the plants can be safeguarded resulting in an increase in income and productivity.

Initially, people used to follow the instructions made by experts to identify the disease and prevent them, but this process takes a long time to identify diseases in a large field, also it is very expensive. Diseases have adverse implications on agricultural and plant lands. This vital vegetation's exposure to these diseases inspires us to develop an automated approach that might improve crop yield, maximize farmer profit, and make a bigger contribution to the nation's economy. There are several ways to spot plant diseases, and various researchers have offered different ways to spot diseases of the potato leaf in their studies.

[1] In proposed a work to detect and identify diseases in tomato leaves with the help of the model called LeNet of Convolutional Neural Network (CNN). Automatic feature extraction technique is used by the neural network model, which

This is an open access article under the CC BY-NC-ND license (https://creativecommons.org/licenses/by-nc-nd/2.0/)

gives aid to the classification. Their proposed system has achieved an average accuracy of 94% in identifying and detecting the leaves, which indicates the feasibility of the neural network.

[2] Back Propagation Neural Network algorithm for identifying and classifying the disease in the leaf image and obtained a classification accuracy of 92%. They used an approach of segmentation using K Means Clustering on various features of Potato leaf image samples such as colour, texture, area, etc.

[3] Potato plant diseases such as Late blight and Early blight may be accurately identified in 95% of the 300 publicly available images.

[4] Detection of diseases on citrus trees, including canker and anthracnose leaf attacks on grapefruit, lemons, limes, and oranges. The experiment's results were genuinely recognized 95% of the time. diseases of grape plants Both Powdery Mildew and Downy Mildew are detected with an average accuracy of 88.89%.

[5] Feed forward back propagation algorithm was used to evaluate a proposed work for the detection of plant diseases, and it performed well with a precision of about 93%. They performed experiments on a treatment for the plant diseases early scorch, cottony mould, late scorch, and small whiteness.

[6] In proposed algorithm for disease detection in sugarcane cultivation. Algorithms for image processing was employed in feature extraction. For detecting Leaf Scorch Disease in sugarcane leaf, it achieved an accuracy of 95%.

[7] In this research work, Fuzzy classifier technique was used to detect infection of wheat crops. with the dataset of healthy and unhealthy leaves, 88% of healthy and unhealthy leaves were correctly classified, while 56% of diseases were correctly identified.

[8] In proposed a model to use the CNN classification technique to distinguish between healthy leaves and 13 different diseased leaves of peach, cherry, pear, apple, and grapevine. More than 30000 photos were included in the dataset, with accuracy ranging from 91% to 98% for each class tests and an average accuracy of 96.3%.

[9] In this work, method was suggested to identify plant leaf diseases using deep learning. A deep CNN was developed to classify 26 diseases and 14 crop kinds using a dataset of 54,306 images from the PlantVillage dataset.

[10] In this work a straightforward cotton plant disease detection system that makes use of the plant's leaf image. A image of the diseased leaf was taken. Then, a distinction between healthy and diseased samples was made utilizing a variety of image processing algorithms including Artificial Neural Networks (ANN). The accuracy of this ANN classification is 80%.

II. MATERIALS AND METHODS

In this paper the implementation is done in different phases in the following manner: collecting the dataset, pre-processing the dataset, splitting the dataset, training the Convolutional Neural Network (CNN) model to identify the type of crop disease, training CNN model to detect the disease, validation of model through obtained results. Fig. 1 represents the block diagram of the proposed potato disease prediction method. Firstly, the data is load and then split into train, validation and test set with different proportion and then preprocessed by resizing the input images and further a NumPy array is created for the same image. Next the convolution neural network with transfer learning for object recognition is build. The model had been trained on a specific data set consisting of images of the different diseased plant leaves which are considered for this study.

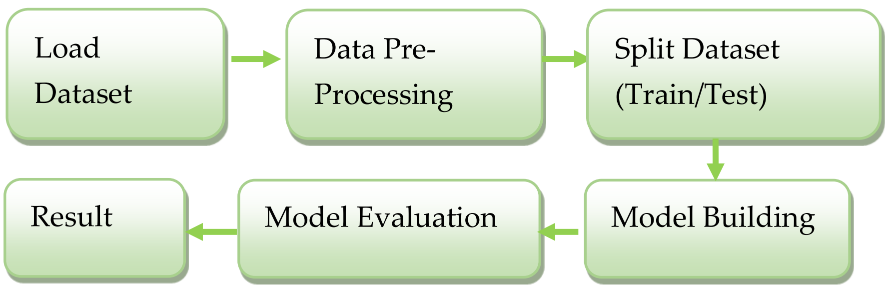


Figure 1: Block Diagram of Proposed Scheme

A. Dataset

The dataset that is used in this paper is Plant Village Dataset that is taken from Kaggle. There is a different directory for each plant and each disease corresponding to that plant has separate folder. The subset of the dataset we are using is the Potato plant dataset. In this dataset there are about 2152 images of leaves of potato plants. This dataset is a combination of healthy and diseased leaves. The diseased leaves are divided into three categories namely early blight, late blight and healthy. In this dataset there are 1000 images of early blight leaf images, 1000 images of late blight and 152images of heathy leaves. Some images from the dataset are shown below:

(a) ***(b)*** ***(c)***

Figure 2: Potato leaf disease (a) Early blight, (b)Late Blight and (c) Healthy

a) Late Blight (Phytophthora infestans)

All parts of a potato plant, including leaves, stems, and tubers, are affected by the disease. It starts as tiny, light green patches that grow into big, water-soaked sores on leaves. On the underside of leaves, a white mildew ring (cottony growth) develops around the dead spots. Water-soaked regions turn necrotic brown in dry weather. Light brown, elongated lesions that may encircle stems develop on them. Dry rot lesions that range in depth from shallow to deep appear on tubers. The afflicted tuber flesh has a "caramelized" texture that resembles sugar. Frequently, the edges of the afflicted tissue take on a metallic colour. The true carriers of the disease are tubers, which act as the disease's source the next season. Although infected seed tubers develop into healthy plants, the disease infects the stem and lower leaves when the environment is favorable for its development (10–12°C and RH > 80%).

b) Early blight (Alternarial solani)

Mostly leaves and tubers are affected by the disease. The symptoms initially appear as small (1-2 mm) circular to oval brown patches on the lower and older leaves. Later on, these lesions have a propensity to grow large and angular. On leaves, mature lesions appear dry and papery and frequently have the concentric rings that resemble a bullseye. On the tuber, the symptoms include brown, rounded to irregular, depressed lesions, and underneath, the flesh becomes dry, brown, and corky. During storage, lesions often become larger, and the damaged tubers subsequently become shriveled.

B. Data Preparation

In this paper the implementation is done in different phases in the following manner: collecting the dataset, pre-processing the dataset, splitting the dataset, training the Convolutional Neural Network (CNN) model to identify the type of crop disease, training CNN model to detect the disease, validation of model through obtained results.

C. Platform Utilized

The hardware used in the paper is a laptop with the specification of 8 GB RAM (1 x 8 GB) with 11th Gen Intel(R) Core TM) i3-1115G4 @ 3.00GHz 3.00 GHz and 1 TB 5400 rpm SATA hard drive. For data visualization we have used an open-source Python package used for plotting is called Matplotlib. TensorFlow and the Keras library were used for machine learning.

D. Convolution Neural Network Architecture

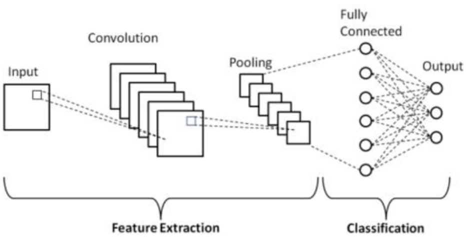


Figure 2: Convolution Neural Network Architecture

This research has the proposed model for convolution neural network (CNN) architecture in identifying diseases in potato leaves is shown in fig. 2. Convolution neural network architecture model build using 3 convolution layers and 3 max pooling layers.

a) Convolution layer:

It is the first layer for dimension extraction from any input image. The convolution layer consists of filters that help to extract particular characteristics, which results in a feature map of the input images. This is a mathematical operation that receives two inputs.

b) Pooling layer:

The pooling layer functions in such a way that a 2D filter slides over every channel of the feature map and conveys the features lying within the area enveloped by the filter. Given a specified dimension of any feature map the pooling layer output dimension is expressed as follows

nh – Feature map height
nw – Feature map width.
nc- Number of channels included in each feature map.
f- Filter size
s- Length of stride

c) Max pooling layer: It is that feature map region from where the maximum number of elements are selected and hidden by the filter. As a result, the feature map produced by the max-pooling layer contains the most noticeable features from the earlier feature map.

d) Fully Connected layer: In the CNN, the fully connected (FC) layer represents the input feature vector. It includes important details regarding the input. This feature vector is subsequently used for classification, regression, etc. during network training. Additionally, it serves as an encoded vector. This is used to calculate the loss during training and helps in network training. Prior to the FC layers, the convolution layers include crucial data regarding the local properties of the input image, such as edges, blobs, shapes, etc. Multiple filters that each represent one of the local characteristics are present in each convolution layer. The most critical and comprehensive information from all convolution layers is collected at the FC layer. Model summary with various parameter values can be seen in Table I

Table 1: Convolution Neural Network (CNN) Model Summary

Layer	Output Shape	Param #
sequential (Sequential)	(32, 256, 256, 3)	0
conv2d (Conv2D)	(32, 254, 254, 32)	896
max_pooling2d (MaxPooling2D)	(32, 127, 127, 32)	0
conv2d_1 (Conv2D)	(32, 125, 125, 64)	18496
max_pooling2d_1 (MaxPooling2D)	(32, 62, 62, 64)	0
conv2d_2 (Conv2D)	(32, 60, 60, 64)	36928
max_pooling2d_2 (MaxPooling2D)	(32, 30, 30, 64)	0
flatten (Flatten)	(32, 57600)	0
dense (Dense)	(32, 64)	3686464
dense_1 (Dense)	(32, 3)	195

Total params: 3,742,979

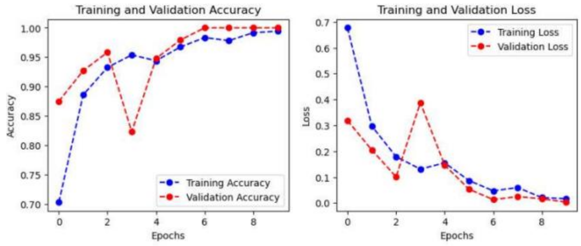
Trainable params: 3,742,979

Non-trainable params: 0

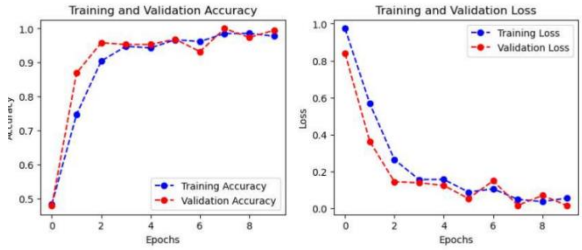
III. RESULTS

The results show that, as expected, using more data during training phase than testing improves classification and lowers error rates (see figure 3). With 90% of the dataset used for training and 10% for testing, the best trained model correctly identified 99% of the images. As the training set size lowers, accuracy for 7 of the 9 training sets remains above 90%, according to the results. Training and validation accuracy shown in Table II and Table III

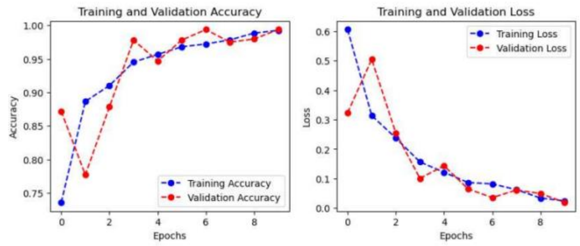
Train-Test Split Set = 90% - 10%



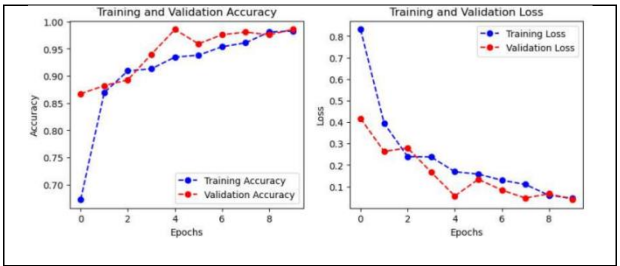
Train-Test Split Set = 80% - 20%



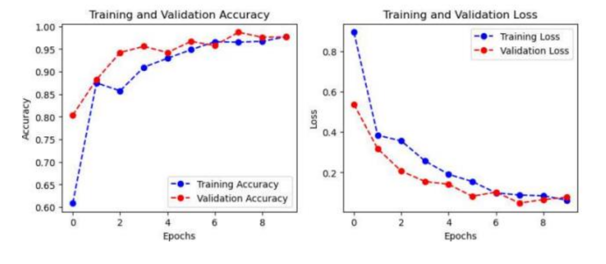
Train-Test Split Set = 70% - 30%



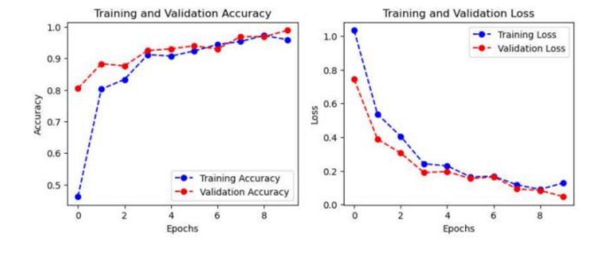
Train-Test Split Set = 60% - 40%



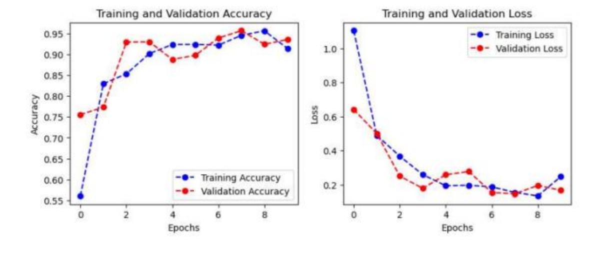
Train-Test Split Set = 50% - 50%



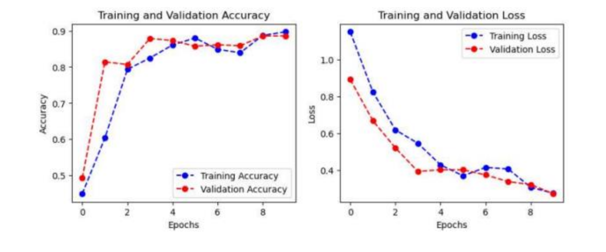
Train-Test Split Set = 40% - 60%



Train-Test Split Set = 30% - 70%



Train-Test Split Set = 20% - 80%



Train-Test Split Set = 10% - 90%

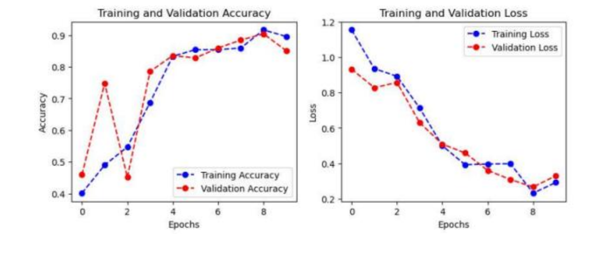


Figure 3: Training and Validation Accuracy & Loss

In these 9 training sets, the average difference between the best training set's error rates (90 % train, 10 % test) and the poorest training set's (20 % train, 80 % test) was 3.11%. When the CNN was trained on 10% of the dataset and tested on 90% of it, there was a noticeable decline in performance. Compared to the classifier's performance with 20% train and 80% test, the correct classification for this training set fell to 87.39% from 93.87%. The difference in training and testing accuracy was minimum for the training set (70% train, 30% test) can be seen in figure 3 so for the further process we use the training set (70% train, 30% test). Table III shows the performance metrics for the potato leaf disease detection and the figure 4 shows the actual and predicted diseases results of potato leaf.

A. Accuracy:

Classification accuracy is the simplest metric to use and implement and is defined as the number of actual predictions divided by the total number of predictions, multiplied by 100. We can implement this by comparing actual value and predicted value.

$$Accuracy = \frac{(TP + TN)}{(TP + TN + FP + FN)}$$

B. Precision:

Precision is the ratio of true positives and total positives predicted. The precision metric focuses on Type-I errors (FP). A Type-I error occurs when we reject a true null Hypothesis (H^0). So, in this case, Type-I error is incorrectly labeling infected plant as healthy. A precision score towards 1 will signify that model didn't miss any true positives, and is able to classify well between healthy and infected labeling of potato diseases.

$$Precision = \frac{TP}{(TP + FP)}$$

Table 3 Training and Validation Accuracy & Loss

Train-Test Split (%)	Training Accuracy (TrA)	Validation Accuracy	Test Accuracy (TsA)	Difference = |TrA – TsA|
90 – 10	0.9943	1.00	0.9844	0.0099
80 – 20	09777	0.9948	0.9961	0.0184
70 – 30	0.9926	0.9937	0.9886	0.0040
60 – 40	0.9828	0.9856	0.9833	0.0005
50 – 50	0.9770	0.9769	0.9651	0.0119
40 – 60	0.9595	0.9886	0.9911	0.0316
30 – 70	0.9141	0.9354	0.9387	0.0246
20 – 80	0.8966	0.8857	0.8739	0.0227
10 – 90	0.8958	0.8510	0.8650	0.0308

C. Recall:

The recall is the ratio between the number of actual values correctly classified as actual to the total number of actual values. It measures the model's ability to detect actual values. The higher the recall, the more actual values are detected. The recall metric gives type-II errors (FN). A type-II error occurs if we accept a false null hypothesis (H^0). So, in this case, the type-II error is incorrectly labeling healthy plants as diseased. Recall towards 1 will signify that model didn't miss any true positives and is able to classify well between correctly and incorrectly labeling of diseased plants.

$$Recall = \frac{TP}{(TP + FN)}$$

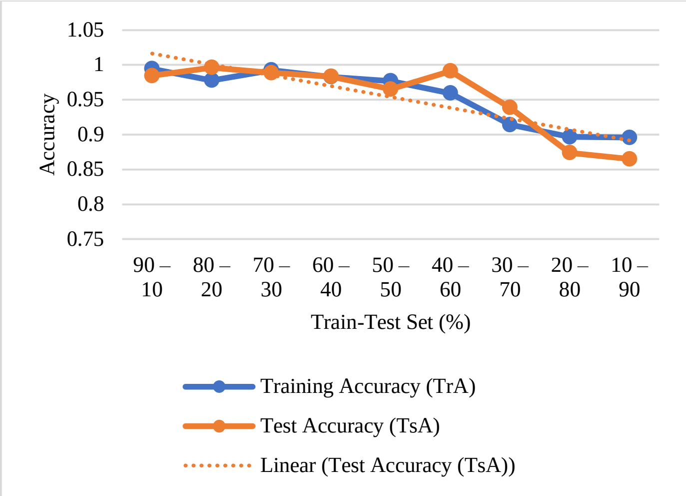


Figure 4: Training and Test Accuracy of Each Train-Test

D. F1-Score:

The F1 score is a performance metric, indicating the harmonic mean of Precision and Recall. The range for F1 Score is (0, 1). High precision but lower recall, gives you an extremely accurate, but it then misses a large number of instances that are difficult to classify. The greater the F1 Score, the better the performance of the model.

$$F1 = 2 * \frac{1}{\frac{1}{precision} + \frac{1}{recall}}$$ potato

Table 2: Performance Metrics of Proposed Methodology

Class	Precision (%)	Recall (%)	F1-Score (%)	Support
Potato Early blight	100	100	100	18
Potato Late blight	100	90	95	10
Potato Healthy	67	100	80	2
Accuracy			97	30
Macro Average	89	97	92	30
Weighted Average	98	97	97	30

Figure 5: Actual and Predicted Results of Potato Leaf Diseases

In this paper the implementation is done in different phases in the following manner: collecting the dataset, pre-processing the dataset, splitting the dataset, training the Convolutional Neural Network (CNN) model to identify the type of crop disease, training CNN model to detect the disease, validation of model through obtained results.

IV. CONCLUSION

A convolution neural network's efficiency for classifying diseased potato image patches into early blight and late blight disease classes and an uninfected class was investigated with classification accuracy of 97%. The 2,152 images were classified by the trained CNN model using different training sets to select the right train-test. The results show that for fully trained CNN models, the correct classification ranges from 89% for the model trained on the least amount of data, to 99% for the model trained on 90% of the data. It is sufficient to employ 20% of the images to get classification rates higher than 90%.

V. REFERENCES

[1] P. Tm, A. Pranathi, K. SaiAshritha, N. B. Chittaragi and S. G. Koolagudi, "Tomato Leaf Disease Detection Using Convolutional Neural Networks," 2018 Eleventh International Conference on Contemporary Computing (IC3), Noida, 2018, pp. 1–5.

[2] Arpita Patel, Mrs. Barkha Joshi, "A Survey on the Plant Leaf Disease Detection Techniques", International Journal of Advanced Research in Computer and Communication Engineering, Vol. 6, Issue 1, ISO 3297:2007

[3] Monzurul Islam, Anh Dinh and Khan Wahid, "Detection of potato Diseases Using Image Segmentation and Multiclass Support Vector Machine", IEEE 30th Canadian Conference on Electrical and Computer Engineering (CCECE), Canada 2017, pp. 1-4.

[4] Kiran R. Gavhale, Ujwalla Gawande and Kamal O. Hajari, "Unhealthy region of citrus leaf detection using image processing techniques", IEEE International Conference on Convergence of Technology (I2CT), Pune 2014, pp. 1-6.

[5] Dheeb Al Bashish, Malik Braik and Sulieman Bani- Ahmad, "A Framework for Detection and Classification of Plant Leaf and Stem Diseases", IEEE International Conference on Signal and Image Processing (ICSIP), Chennai 2010, pp. 113-118.

[6] Umapathy Eaganathan, Jothi Sophia, Vinukumar Lackose, Feroze Jacob Benjamin, "Identification of Sugarcane Leaf Scorch Disease using K-means Clustering Segmentation and KNN based Classification", International Journal of Advances in Computer Science and Technology (IJACST), Vol. 3, No. 12, Special Issue of ICCEeT, Dubai , 2014, pp. 11- 16

[7] Diptesh Majumdar, Arya Ghosh, Dipak Kumar Kole, Aruna Chakraborty and Dwijesh Dutta Majumder, "Application of Fuzzy CMeans Clustering Method to Classify Wheat Leaf Images based on the presence of rust disease", Proceedings of the 3rd International Conference on Frontiers of Intelligent Computing: Theory and Applications, Vol. 327, 2015, pp. 277-284.

[8] Srdjan Sladojevic, Marko Arsenovic, Andras Anderla, Dubravko Culibrk and Darko Stefa-novic, "Deep Neural Networks Based Recognition of Plant Diseases by Leaf Image Classification", Computational Intelligence and Neuroscience, Article ID 3289801, 2016.

[9] S. P. Mohanty, D. P. Hughes, and M. Salathé, "Using deep learning for image-based plant disease detection," Frontiers in Plant Science, vol. 7, pp. 1–10, Sep. 2016, doi: 10.3389/fpls.2016.01419

[10] Malvika Ranjan, Manasi Rajiv Weginwar, NehaJoshi, A.B. Ingole, "Detection and Classification of Leaf Disease Using Artificial Neural Network," International Journal of Technical Research and Applications, 2015, pp. 331–333.