Cotton Plant Leaf Detection and Diagnosis System Using Deep Neural Network.

Dilip S. Kale
Asst. Professor
Department of Computer Engineering
Rajiv Gandhi Institute of Technology

G. P. Bhole
Asst. Professor
Department of Computer Engineering
Veermata Jijabai Technological Institute

Abstract **- The cotton agriculture industry faces the economic losses due to the pest infections, bacterial or viral contagions, the farmers lose nearly 10-20% of the total profit on an average annually in India. The paper proposes a solution to the agricultural problem, which involves cotton leaf disease recognition by using machine learning and deep learning. In this paper, the study sets out to classify cotton crop images into classes, whether the crop is infected by a disease or not. Also, we endeavor applications that give farmer readily available means to identify the diseases on their crop and take appropriate damage control actions. The dataset used to train the model was user created (mobile capture images with high-resolution camera) from a farm located in Maharashtra region at Morshi Taluka, Survey No.162, district Amravati. Two variety of cotton plants are utilised for training, SP 7517 BG II (Cotton hydride seeds with Bollgard Technology) and Kavari 3028, with Dimension 4160 x 3120, Horizontal and Vertical resolution 96 dpi, bit depth 24, of 3611 images, containing four classes namely Rust, Mosaic Virus, Woolyaphids and Healthy plants. The trained models have provided a best performance reaching a 79.53% success rate in identifying the corresponding cotton plant disease with a combination of a healthy plant. The model used in the study delivers significant accuracies of classification on the dataset used by employing Dense Neural Network. The model is very useful advisory or early warning tool for the farmers for identification of diseases in the early stage**

Keywords - **Cotton plant leaf, Dense Neural Network, image processing.**

I. Introduction

Agriculture is the primary occupation in India. Cotton is also referred as 'White Gold' and it has a major use in the textile industry today. Cotton, as a major agricultural crop in India, has a dominant impact on the overall Indian agriculture sector. Nowadays, a tremendous loss in the quality and quantity of cotton yield is observed, subject to various diseases in the plant. Plant disease classification is a critical step, which can be useful in early detection of pest, insects, disease control, increase in productivity, among other examples. This study looks forward to predict crop diseases by processing the images of the crop. For this, Image Processing techniques are used for the very fast, accurate and appropriate classification of diseases. Symptoms of diseases in cotton predominantly come out on leaves of plants. Farmers recognise disease manually with foregoing symptoms of plants, and with experts, whereas the actual diseases are hard to distinguish with naked eyes, and it is time-consuming to predict whether the crop is healthy or not. Cotton plant leaf disease diagnosis is very difficult through observation to find the symptoms on plant leaves, incorporates it's a part of a high degree of complexity. Due to complexity and a large number of cultivated plants and their existing phytopathological problems, even experienced agronomists and plant pathologists often fail to successfully diagnose specific diseases and are consequently led to mistaken conclusions and treatments. The existence of an automated system for the detection and diagnosis of plant diseases would offer a support system to the agronomist who is performing such diagnoses through observation of leaves of infected plants.

The existing techniques for disease detection have utilized various image processing methods followed by various classification techniques. Crop Yield Forecasting has been an area of interest for producers, agricultural-related organizations. Timely and accurate crop yield forecasts are essential for crop production. The proposed system uses an artificial neural network to classify the health of a cotton leaf plant.

The flow diagram of the proposed system given in Fig. 1. consists of steps used to acquire the desired output.

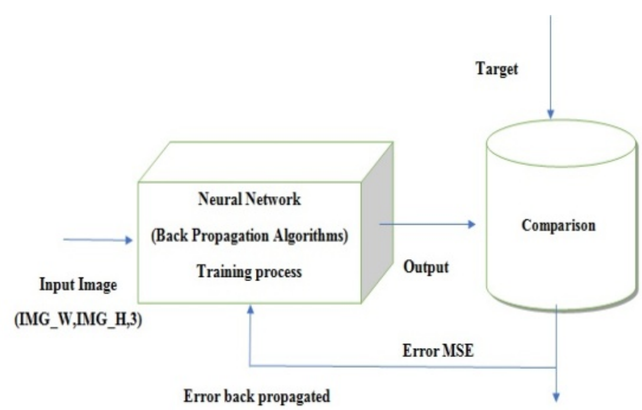

Fig 1. Flow Diagram of Proposed System

II. Methodology

A. *Input images*

For this initial process images of high resolution (4160 x 3120) are taken from datasets as input by setting IMG_W, IMG_H with 3 channels (RGB), for better visibility, go with > 180. The features that are used for classification of the images. The images are foremost pre-processed into a 4160x 3120, RGB format with pixel values ranging from 0 to 255. The feature normalization used in the study is the min-max normalization. It is the ratio of the difference between the instance's feature value and the minimum value of a feature in the instance to the difference between the maximum and minimum values of features in the instance.

B. *Training and Testing*

The tensor flow framework is used for training and testing. The model which is employed is the DNN Classifier. The DNN Classifier which is created has a seven-layer neural network. The activation function used is ReLU which is used for each of the hidden layers and the SoftMax function used in the last layer. Gradient descent optimizer is used for optimization. The dataset is divided into a train test split of 70-30%. compare the result and errors back-propagated.

C. *Feature Extraction*

The DNN classifier used was fed with features (pixels of image instances). DNN classifier is used with 5 hidden units with 100 nodes each and a 6th hidden unit with 50 nodes. The images were pre-processed through resizing. The images with given height, width and channels are fed to the DNN. ReLU (Rectified Linear Unit) activation function was used.

D. *Dataset and their labels.*

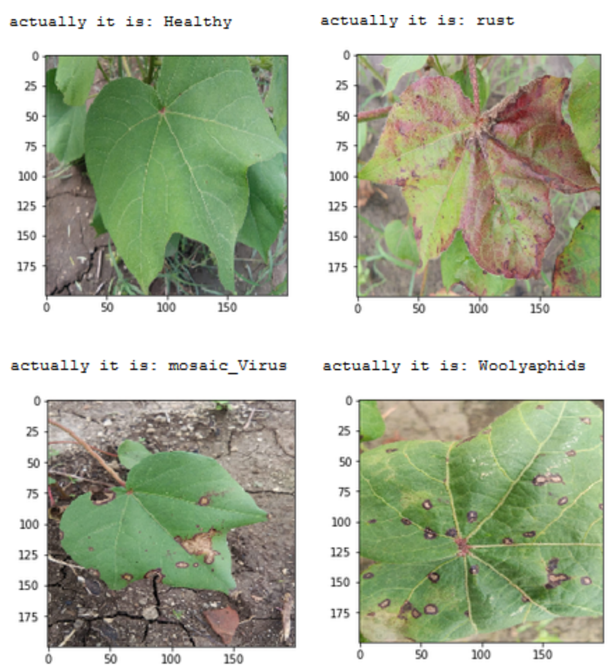

Fig 2. Dataset Examples and their labels

III. Literature Review

The existing techniques for disease detection have utilized various image processing methods followed by various classification techniques. Crop Yield Forecasting has been an area of interest for producers, agricultural-related organizations. Timely and accurate crop yield forecasts are essential for crop production. According to the literature available on crop yield forecast, following techniques are available for consideration of usage on crop data:

- Disease Detection using Image classification and Regression Techniques.
- Disease Detection Using Segmentation, Texture Pattern, Feature Extraction analysis, and color transformation.
- Disease Detection Using IoT.
- Disease Detection using Deep Neural Network.
- Disease Detection Using Convolutional Neural Network.

Image classification and regression techniques play a very important role because it allows identifying, group, and properly of organisms from a standardized system. Pinto et al. [2] present an application of machine learning to agriculture, solving a particular problem of diagnosis of crop disease based on plant images taken with a smartphone, the author presents a classification system that trains a 5-class classification system to determine the state of disease of a plant. The 5 classes consist of one health class and four disease classes. Crop disease classification using texture analysis [3], uses Image processing to detect and classify sunflower crop diseases based on the image of their leaf, using k-means clustering to get the diseased part of the leaf. These are then run through the various machine learning algorithms and classified based on their color and texture features.

We apply an algorithm for image segmentation technique on data for automatic detection and classification of plant leaf diseases. Sandika et al. [4] and Shah et al. [5] use segmentation and texture pattern feature, color transformation methods to classify images. A Novel Approach to classifying individual pixels in crop diseased images taken in the field as diseased or healthy. The approach is based on the machine learning algorithm linear discriminant analysis (LDA) and color transformation. Five color spaces were applied and compared over diseased images infected by four diseases commonly observed in cucumber crops - target spot, angular leaf spot downy mildew, and powdery mildew.

Deep learning is a set of learning methods attempting to model data with complex architectures combining multiple non-linear transformations. The element of deep learning is the neural networks that are combined to form the deep neural networks. These techniques have enabled significant progress in the fields of image processing and image classification.

There exist several types of architectures for neural networks:

1. The multilayer perceptions, that are the oldest and simplest ones,
2. The Convolutional Neural Networks (CNN), particularly adapted for image processing
3. The recurrent neural networks and dense neural network, used for sequential data such as text or times series.

IV. Common Diseases in Plants

The diseases on cotton are fungal, bacterial, viral diseases which not only appears on a leaf of the plant but also cotton balls, cotyledons, seedling, root, etc. To be familiar with the diseases found in the cotton we may able to know the sign and symptoms of the disease and get most extreme thoughts concerning the control measures and administration of infection found in cotton, recognizable about the pesticides like product at various circumstances applying the diverse technique.

A few cotton diseases usually occur on cotton are:

- Angular Leaf Spot or Black Arm Disease
- Vascular Wilt Disease
- Grey Mildew or Dahiya Disease
- Anthracnose Disease
- Root Rot Disease.
- Boll Rot Disease
- Leaf Spot or Blight Disease.

Category: Angular Leaf Spot

Fig 3. Angular Leaf Spot (1)

Symptoms

Small, circular brown lesions on cotyledons and seedling leaves which expand and develop a concentric pattern; necrotic areas coalesce and often have a purple margin; centers of lesions may dry out and drop from the plant creating a "shot-hole" appearance on the leaves.

Cause: Fungus

Comments: Plants stressed by drought, nutrient deficiency, and other pests are more susceptible to the disease; fungus spreads rapidly in dense canopies, especially during periods of warm, wet weather.

Management: Plow crop residue into the soil to reduce inoculums levels; provide plants with adequate irrigation and nutrients, particularly potassium; applications of appropriate foliar fungicides may be required on susceptible cultivars.

Angular Leaf Spot

The diseases considered in the study are the viral disease Mosaic, Rust which is a fungal disease and Woolly aphids which are insects that feed on plant sap. The mosaic virus causes wrinkled or curled leaves with yellow, green or white streaks on foliage. A short period after the plant gets infected with Rust, it gets covered with reddish-orange spore masses. Later these may turn yellow-green and eventually black. Woolly aphids produce a filamentous waxy white covering which resembles cotton or wool.

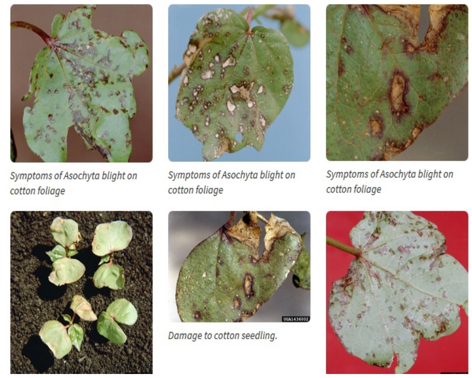

Fig 4. Angular Leaf Spot (2)

Disease	Visual Cue	No of Images
Healthy	Green leaves, no spots on leaf	911
Mosaic	Wrinkled leaves, yellow, green or white streaks	900
Rust	Reddish-orange spore masses	890
Woolly aphids	Waxy white covering	910

Table 1. Visual Cue of Diseases

V. MATERIAL AND METHODS

A. Neural Networks

An artificial neural network is an application, nonlinear with respect to its parameters θ that associates to an entry x an output y=f(x, θ). For the sake of simplicity, we assume that y is uni-dimensional, but it could also be multidimensional. This application f has a particular form that we will precise. The neural networks can be used for regression or classification. As usual in statistical learning, the parameters θ are estimated from a learning sample. The function to minimize is not convex, leading to local minimizes. The success of the method came from a universal approximation theorem due to Cybenko (1989) and Hornik (1991). Moreover, Le Cun (1986) proposed an efficient way to compute the gradient of a neural network, called backpropagation of the gradient, that allows to obtain a local minimizer of the quadratic criterion easily. An artificial neuron is a function fj of the input x= (x1, . . ., xd) weighted by a vector of connection weights wj = (wj, 1, ..., wj, d), completed by a neuron bias bj, and associated to an activation function φ, namely.

$$yj = fj(x) = \varphi(\langle wj, x \rangle + bj)$$

Several activation functions can be considered

- The identity function,

$$\phi(x) = x.$$

- The sigmoid function (or logistic),

$$\varphi(x) = \frac{1}{1+e^{-x}}$$

- The hyperbolic tangent function (tanh),

$$\varphi(x) = \frac{e^{x}-e^{-x}}{e^{x}+e^{-x}} = \frac{e^{2x}-1}{e^{2x}+1}$$

- The hard threshold function,

$$\varphi(x) = 1 \quad , x \geq \beta$$

- The Rectified Linear Unit (ReLU) activation function,

$$\varphi(x) = \max(0, x)$$

Here is a schematic representation of an artificial neuron where

$$\Sigma = \langle wj\ x \rangle + bj$$

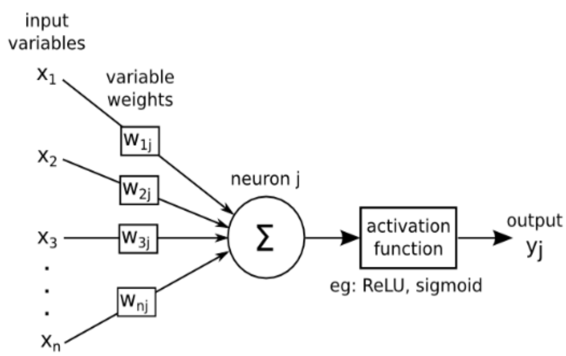

Fig 5.Artificial Neuron Model (Andrew James)

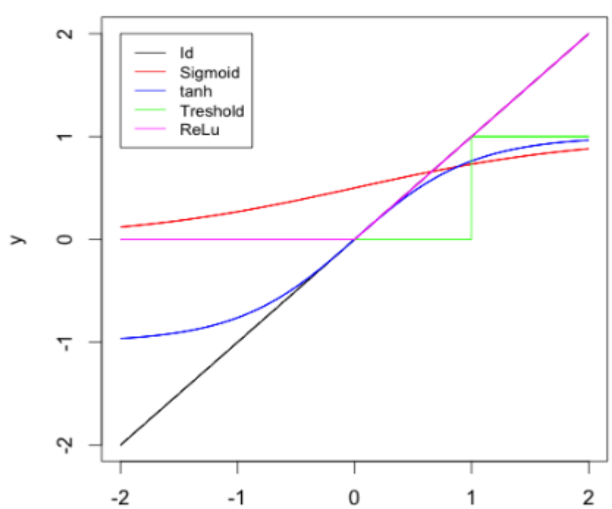

Fig 6. Activation Function

Historically, the sigmoid was the mostly used activation function since it is differentiable and allows to keep values in the interval [0,1]. Nevertheless, it is problematic since its gradient is very close to 0 when |x| is farther away 0.

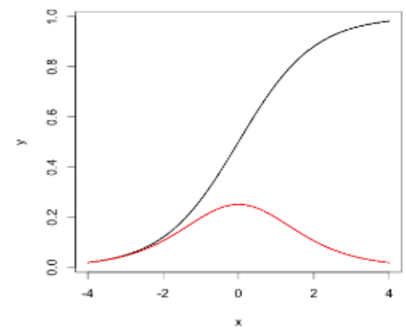

Fig 7. Sigmoid function (black) and its derivative (red)

The Fig 7. represents the Sigmoid function and its derivative. With neural networks having a higher number of layers (which is the case for deep learning), this causes troubles for the back-propagation algorithm to estimate the parameter (back propagation is explained in the following). This is why the sigmoid function was supplanted by the rectified linear function. This function is not differentiable at 0 but in practice this is not really a problem since the probability to have an entry equal to 0 is generally null. The ReLU function also has a scarification effect. The ReLU function and its derivative are equal to 0 for negative values, and no information can be obtained in this case for such a unit, this is why it is advised to add a small positive bias to ensure that each unit is active. Several variations of the ReLU function are considered to make sure that all units have a non-zero gradient and that for x <0 the derivative is not equal to 0. Namely $\phi(x) = \max(x,0) + \alpha \min(x,0)$ where α is either a fixed parameter set to a small positive value, or a parameter to estimate.

B. *Multilayer perceptron*

A multilayer perceptron (or neural network) is a structure composed by several hidden layers of neurons where the output of a neuron of a layer becomes the input of a neuron of the next layer. Moreover, the output of a neuron can also be the input of a neuron of the same layer or of neuron of previous layers (this is the case for recurrent neural networks). On the last layer, called output layer, we may apply a different activation function as for the hidden layers depending on the type of problems we have at hand: regression or classification. The Fig 8. represents a neural network with four input nodes, three output nodes, and two hidden nodes.

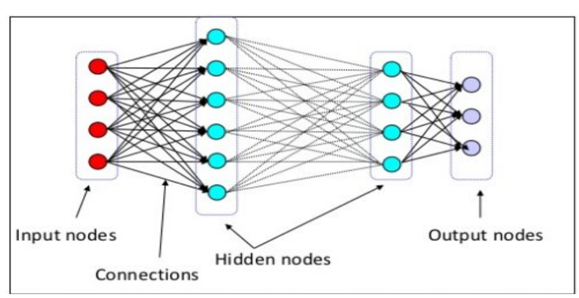

Fig 8. A basic Neural Network

Multi layers perceptions have a basic architecture since each unit (or neuron) of a layer is linked to all the units of the next layer but has no link with the neurons of the same layer. The parameters of the architecture are the number of hidden layers and number of neurons in each layer. The activation functions for each layer are subject to choice of every individual. For the output layer, as mentioned previously, the activation function is generally different from the one used on the hidden layers. In the case of regression, we apply no activation function on the output layer. For binary classification, the output gives a prediction of P(Y= 1/X); since this value is in[0,1], the sigmoid activation function is generally considered. For multi-class classification, the output layer contains one neuron per class, giving a prediction of P(Y = I/X). The sum of all these values has to be equal to1. The multidimensional function SoftMax is generally used as SoftMax.

$$\sigma(z)_j = \frac{e^{zj}}{\sum_{k=1}^{K} e^{zk}}.$$

Let us summarize the mathematical formulation of a multilayer perceptron with L hidden layers. We set h(0)(x) = x.

For k = 1 → L (hidden layers),

$a^{(k)}(x) = b^{(k)} + W^{(k)}h(k-1)(x)h(k)(x) = \varphi(a(k)(x))$

For k = L+ 1(output layer),

$a(k)(x) = b^{(k)} + W^{(k)}h(L)(x)h(L+1)(x) = \psi\ (a(k)(x)) := f(x,\theta)$

Where, φ is the activation function and ψ is the output layer activation function (for example SoftMax for multiclass classification). At each step, $W^{(k)}$ is a matrix with number of rows the number of neurons in the layer k and number of columns the number of neurons in the layer k−1.

C. Back propagation Algorithms

D. Dataset, Preprocessing and Features

The dataset contains a total of 3611 images. This is constituted by 911 healthy plants images, 890 having the Mosaic virus, 910 plants being infected by Rust and 900 samples of plants having Woolly aphids. Table 2 shows the composition of dataset of leaf images with class name and the number of plant leaf samples.

Image Class	Number of Samples
Healthy	911
Mosaic virus	890
Rust	910
Woolly aphids	900
Total	3611

Table 2. Amount of data samples per class

VI. Discussions and Results

The study was set out to classify if the crop is infected by a disease or is healthy. The DNN classifier used was fed with features (pixels of image instances). DNN classifier is used with 5 hidden units with 100 nodes each and a 6th hidden unit with 50 nodes. Figure 3. shows the Graphs of execution each hidden layer with a fraction of zero values (0-5 hidden layer) and figure 4. shows the histograms of each hidden layer with a fraction of zero values.

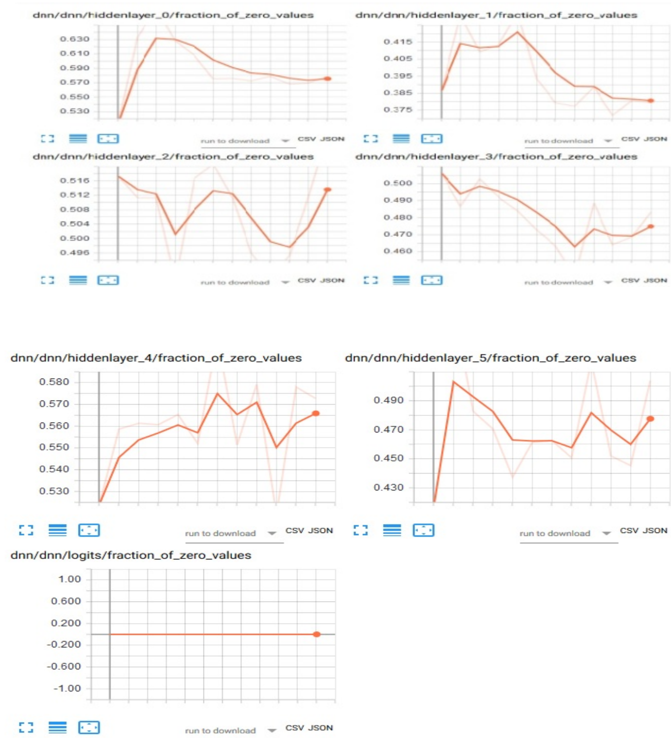

Fig 9. Execution of different hidden layer with fraction of zero values.

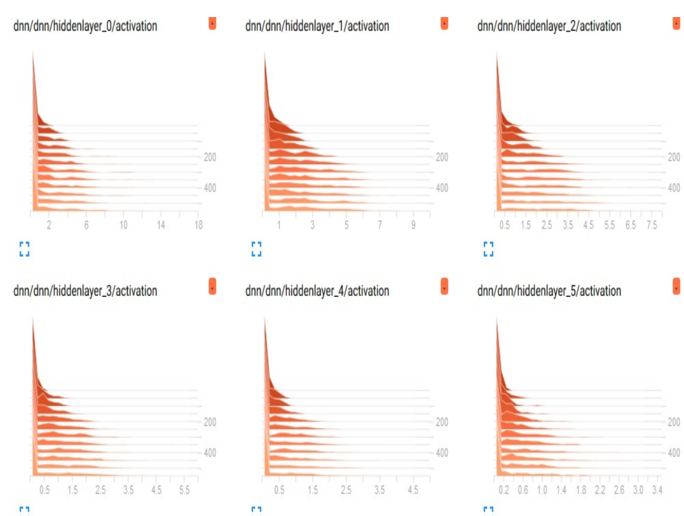

Fig 10. Histograms of a different hidden layer with a fraction of zero values

The images were pre-processed through resizing. The images with given height, width and channels are fed to the DNN. ReLU (Rectified Linear Unit) activation function was used.

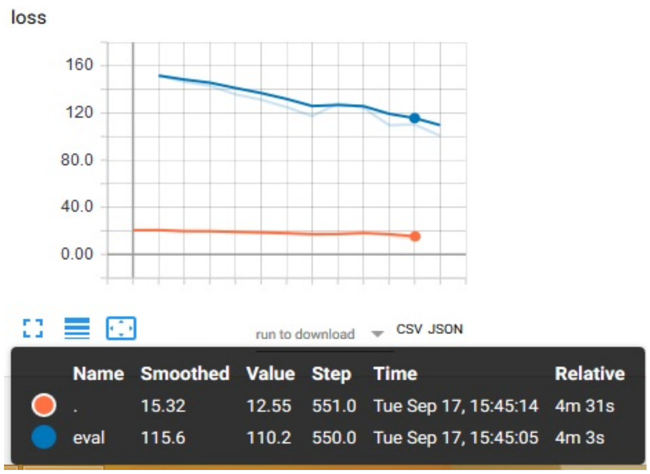

Fig 11. Shows Line Plot of accuracy using Rectified Linear Activation function

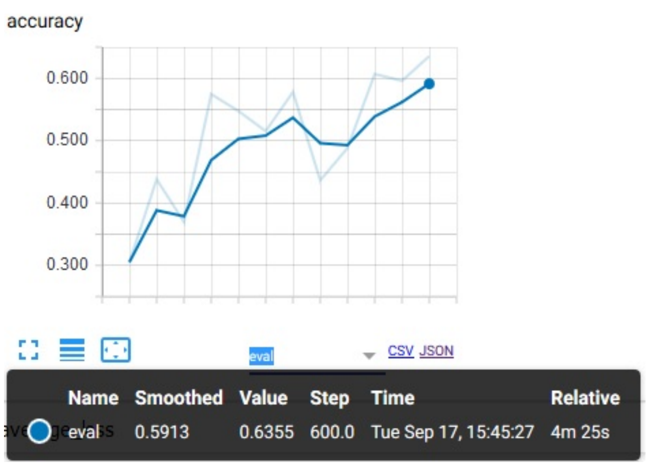

Fig 12. Shows Line Plot of average loss using Rectified Linear Activation function

Fig 11. Line Plot of average loss during training period

Gradient descent algorithm is used as an optimization function to converge to the optimized classes quicker. The number of true values predicted correctly are represented as T, several false values predicted correctly are represented as F. The error was calculated using the error function mentioned.

$$\text{Error} = \left(\frac{F}{F+T}\right) * 100$$

VII. Conclusions and Future Scope

In this paper, we consider cotton crop as it is the most important and cash crop in India, it affects major Indian economy due to plant diseases. In India, normally Orange rust, Mosaic virus, white wooly aphids are the hazardous diseases in the cotton crop in our country. Here, we consider Deep Neural Network for cotton disease recognition using leaf images for classification. There are several methods in computer vision for plant disease detection and classification process, but still, this research field is lacking and nobody explodes their idea throughout available and reach to farmer house. Besides, there are no commercial solutions available in the market, dealing with plant recognition based on the leaves images is to use a new approach of deep learning method which automatically classifies and detect plant diseases from leaf images.

The data set of size 3611 images is collected from my farm with different diseases. The plant leaf images are considered for the training and testing purpose of the network. Initially, with the use of Gradient Descent and Backpropagation algorithm classification are performed and it gives the prediction of diseases with 87% efficiency.

Furthermore, Convolution Neural Network (CNN) is used for better classification accuracy. The main aim is to detect the plant leaf diseases from the database and train the images in such a way that the trained model gives the solution to farmers. The proposed model can recognize 11 different types of plant diseases. Here we consider plant stream and affected area by the disease boundaries, color variation, size and shape of plant leaves.

The future work of this project is to develop a complete system consisting of server-side components containing a trained model and an application for smart mobile devices with features such as displaying recognized diseases in plants, based on leaf images captured by the mobile phone camera. This application will serve as an aid to farmers, enabling fast and efficient recognition of plant diseases and facilitating the decision-making process when it comes to the use of chemical pesticides.

Acknowledgement

References

[1] E. Mwebaze and G. Owomugisha, "Machine Learning for Plant Disease Incidence and Severity Measurements from Leaf Images," 2016 15th IEEE International Conference on Machine Learning and Applications (ICMLA), Anaheim, CA, 2016, pp. 158-163.

[2] L. S. Pinto, A. Ray, M. U. Reddy, P. Perumal, and P. Aishwarya, "Crop disease classification using texture analysis," 2016 IEEE International Conference on Recent Trends in Electronics, Information and Communication Technology (RTEICT), Bangalore, 2016, pp. 825-828.

[3] M. K. Tripathi and D. D. Maktedar, "Recent machine learning-based approaches for disease detection and classification of agricultural products," 2016 International Conference on Computing Communication Control and Automation (ICCUBEA), Pune, 2016, pp. 1-6.

[4] B. Sandika, S. Avil, S. Sanat, and P. Srinivasu, "Random forest-based classification of diseases in grapes from images captured in uncontrolled environments," 2016 IEEE 13th International Conference on Signal Processing (ICSP), Chengdu, 2016, pp. 1775-1780.

[5] J. P. Shah, H. B. Prajapati and V. K. Dabhi, "A survey on detection and classification of rice plant diseases," 2016 IEEE International Conference on Current Trends in Advanced Computing (ICCTAC), Bangalore, 2016, pp. 1-8.

[6] B. S. Prajapati, V. K. Dabhi and H. B. Prajapati, "A survey on detection and classification of cotton leaf diseases," 2016 International Conference on Electrical, Electronics, and Optimization Techniques (ICEEOT), Chennai, 2016, pp. 2499-2506.

[7] U. Ayub and S. A. Moqurrab, "Predicting crop diseases using data mining approaches: Classification," 2018 1st International Conference on Power, Energy and Smart Grid (ICPESG), Mirpur Azad Kashmir, 2018, pp. 1-6.

[8] T. Li, J. Yang, X. Peng, Z. Chen and C. Luo, "Prediction and Early Warning Method for Flea Beetle Based on Semi-supervised Learning Algorithm," 2008 Fourth International Conference on Natural Computation, Jinan, 2008, pp. 217-221.

[9] A. AlSuwaidi, B. Grieve and H. Yin, "Feature-Ensemble-Based Novelty Detection for Analyzing Plant Hyperspectral Datasets," in IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, vol. 11, no. 4, pp. 1041-1055, April 2018. (JOURNAL)

[10] Rajathi Natarajan, Jayashree Subramanian, Elpiniki I. Papageorgiou, "Hybrid learning of fuzzy cognitive maps for sugarcane yield classification", Computers and Electronics in Agriculture, Volume 127,2016, Pages 147-157.