International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 07 Issue: 07 | July 2020 www.irjet.net p-ISSN: 2395-0072

Recognition of Leaf Disease Using Convolution Neural Network

G K Chaithanya Krishna[1], Harshil Y K[2], Darshan B J[3], Akash Ashok Patil[4], Dr. Vijayalakshmi G V[5]

[1-4]U.G. Student, Dept. of Electronics and Communication Engineering, BMSIT Bangalore, Karnataka, India
[5]Associate Professor, Dept. of Electronics and Communication Engineering, BMSIT Bangalore, Karnataka, India

---

Abstract - *Crop production problems are common in India which severely effect rural farmers, agriculture sector and the country's economy as a whole. In Crops leaf plays an important role as it gives information about the quantity and quality of agriculture yield in advance depending upon the condition of leaf. Plant leaf diseases and destructive insects are a major challenge in the agriculture sector. Faster and an accurate prediction of leaf diseases in crops could help to develop an early treatment technique while considerably reducing economic losses. Modern advanced developments in Deep Learning have allowed researchers to extremely improve the performance and accuracy of object detection and recognition systems. In this paper, we have proposed a deep-learning-based approach to detect leaf diseases in many different plants using images of plant leaves. Our goal is to find and develop the more suitable deep learning methodologies for our task. Therefore, we are considering the Alex Net model, which was used for the purpose of this work. The proposed system can effectively identified different types of diseases with the ability to deal with complex scenarios from a plant's area.*

Key Words: **CNN, Alex Net, SGD, Machine Learning, Disease Detection**

1. INTRODUCTION

India is an agricultural country and about 70% of the population depends on agriculture. Farmers have large range of diversity for selecting various suitable crops and finding the suitable pesticides for plant. Disease on plant leads to the significant reduction in both the quality and quantity of agricultural products. The studies of plant disease refer to the studies of visually observable patterns on the plants. Monitoring of health and disease on plant plays an important role in successful cultivation of crops in the farm. In early days, the monitoring and analysis of plant diseases were done manually by the expertise person in that field. This requires tremendous amount of work and also requires excessive processing time. The image processing techniques can be used in the plant disease detection. In most of the cases disease symptoms are seen on the leaves, stem and fruit. The plant leaf for the detection of disease is considered which shows the disease symptoms. The rest of the paper is organized as follows. Section II provides Literature survey, the methodology is explained in section III, the results are discussed in section IV and finally section V concludes the paper.

2. LITERATURE SURVEY

In this section, various methods of image processing for plant disease detection is discussed. Identification of the plant diseases is the key to prevent the losses in the yield and quantity of the agricultural product. The studies of the plant diseases mean the studies of visually observable patterns seen on the plant. Sachin, Patil [1] discussed the various techniques used to segment the disease part of the plant. It also discusses some Feature extraction and classification techniques to extract the features of infected leaf and the classification of plant diseases. The use of ANN methods for classification of disease in plants such as self-organizing feature map, back propagation algorithm, SVMs etc. can be efficiently used.

The symptoms of the attacks are usually distinguished through the leaves, stems or fruit inspection. Zulkifli et al [2] discusses the effective way used in performing early detection of chili disease through leaf features inspection. Leaf image is captured and processed to determine the health status of each plant. Leaves images captured are processed to determine the healthiness of each plant. By using leaf recognition technique, it does will identify the potential problems to the chili plants before its goes seriously damage for all chili plants. With this method, the use of harmful chemicals on plants can be reduced and hence ensure a healthier environment and may be even lowering the production cost of the maintenance and producing a high quality of chili.

Pooja et al [3] purposes a disease detection and classification technique with the help of machine mechanism and image processing tools. Initially, identifying and capturing the infected region is done and latter image processing is performed. Further, the segments are obtained and the area of interest is recognized and the feature extraction is done on the same. Finally the obtained results are sent through SVM classifier to get the results. Pooja et al [3] mainly focuses on the plant disease detection and through the application of various methodology. Usage of various feature extraction techniques and a stable, sufficient data set have facilitated in obtaining satisfactory experimental results. The usage of classifier Support Vector Machines (SVM) have enhanced the performance of the system which provides better results.

Alex et al [4] worked on the test data, they achieved top-1 and top-5 error rates of 37.5% and 17.0% which is considerably better than the previous state-of-the-art. The neural network, which has 60 million parameters and 650,000 neurons, consists of five convolutional layers, some of which are followed by max-pooling layers, and three fully-connected layers with a final 1000-way softmax. To make training faster, we used non-saturating neurons and a very efficient GPU implementation of the convolution operation. To reduce overfitting in the fully-connected layers we employed a recently-developed regularization method called "dropout" that proved to be very effective. We also entered a variant of this model in the ILSVRC-2012 competition and achieved a winning top-5 test error rate of 15.3%, compared to 26.2% achieved by the second-best entry.

Enhanced images have high quality and clarity than the original image. Color images have primary colors red, green and blue. It is difficult to implement the applications using RGB because of their range i.e. 0 to 255. Hence they convert the RGB images into the grey images. Then the histogram equalization which distributes the intensities of the images is applied on the image to enhance the plant disease images. Monica Jhuria et al [5] used image processing for detection of disease and the fruit grading. They have used artificial neural network for detection of disease. They have created two separate databases, one for the training of already stored disease images and other for the implementation of the query images. Back propagation is used for the weight adjustment of training databases. They consider three feature vectors, namely, color, textures and morphology and it was found that the morphological feature gives better result than the other two features.

Mrunalini and Deshmukh [6] compare the Otsu threshold and the k-means clustering algorithm used for infected leaf analysis. They have concluded that the extracted values of the features are less for k-means clustering. The clarity of k-means clustering is more accurate than other method.

These methods discussed earlier concentrated on the traditional machine learning approaches where the features were handcrafted and carefully picked to be provided for the algorithm to generate the optimized model that further does the classification. Selecting the significant features is a complex task as they need to be robust and reliable. Thus this paper proposes to use one of the deep learning techniques, Convolutional Neural Network (CNN) that combines the feature extraction and classification process. The CNN automatically detects the important features without any human supervision and is more efficient as it reduces the number of parameters.

CNN takes advantage of local spatial coherence in the input images, which allow them to have fewer weights as some parameters are shared. This process, taking the form of convolutions, makes them especially well-suited to extract relevant information at a low computational cost.

3. METHODOLOGY

This section provides an insight to the proposed model using CNN technique. The CNN architecture used in the work is Alex Net [4]. The methodology is as illustrated in Fig -1 that involves (i) Data acquisition (ii) Pre-processing (iii) CNN training and (iv) Classification.

The images required for the study were collected from Plant Village dataset [7]. The dataset has 54,000 number of images. The images span 14 crop species: Apple, Blueberry, Cherry, Corn, Grape, Orange, Peach, Bell Pepper, Potato, Raspberry, Soybean, Squash, Strawberry, and Tomato. It contains images of 17 fungal diseases, 4 bacterial diseases, 2 mold diseases, 2 viral disease, and 1 disease caused by a mite. 12 crop species also have images of healthy leaves that are not visibly affected by a disease. The sample of the dataset with Tomato leaf disease is displayed in Fig -2.

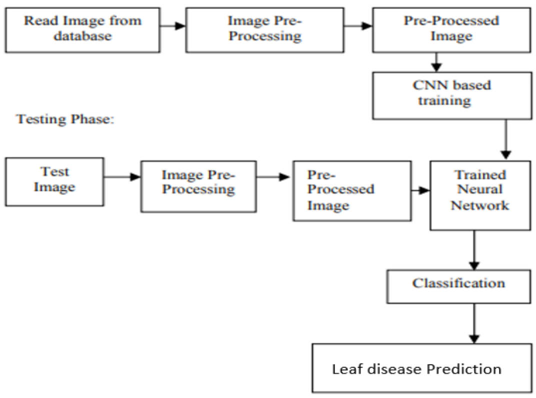

Fig -1: Model of the system

Fig -2: Sample of the dataset with Tomato leaf diseases (a) Bacterial sport (b) Leaf mold (c) Mosaic Virus and (d) Late Blight

The main aim is to design a system which is efficient and which provides disease name. Thus for validating our work, the entire dataset is divided into 2 folders in the ratio of 80:20 for training and testing.

(i) Training folder:

This folder [Tr] contain 43,394 leaf images with 14 different crop species where each class of species will contain healthy leaves as well as different disease which may occur to that particular class.

(ii) Testing folder:

This folder [Te] contain 10,913 leaf images with 14 different crop species which are similar to training folder but the images used in this will be different from the training folder.

After creating the training and testing datasets, the entire process is carried out in two phases:

(a) Training phase:

In this phase the entire training folder with the class labels [Tr, y] where y indicates the label of the diseases is given to the CNN where each image will undergo image pre-processing and training. In image pre-processing each image will be resized to the target size of 224 X 224 pixels by using Image Data Generator function.

As mentioned, Alex Net Architecture [4] is used in this work and based on this architecture the model was created. The Alex Net Architecture is used because of the following advantages: (i) It uses Relu activation function instead of Tanh to add non-linearity which accelerates the speed by 6 times at the same accuracy. (ii) It uses dropout instead of regularization to deal with overfitting. (iii) Overlap pooling to reduce the size of the network.

The model was optimized using stochastic gradient descent (SGD) method that uses batches of data to approximate gradient descent. It is an iterative method for optimizing a function with suitable smoothness properties. It can be regarded as a stochastic approximation of gradient descent optimization, since it replaces the actual gradient by an estimated value. This reduces the computational burden, achieving faster iterations in trade for a slightly lower convergence rate. The first step of SGD is to randomize the whole training set. Then, for updating every parameter we use only one training example in every iteration to compute the gradient of cost function. As it uses one training example in every iteration this algorithm is faster for larger data set. In SGD, one might not achieve accuracy, but the computation of results is faster. Stochastic gradient descent maintains a single learning rate (termed alpha) for all weights and the learning rate remains constant during training.

(b)Testing phase:

In this phase testing folder [Te] is used whose images will fit to the CNN network. The testing dataset provides an unbiased evaluation of a model fit on the training dataset while tuning the model's hyper parameters. Testing datasets can be used for regularization by early stopping (stopping training when the error on the testing dataset increases, as this is a sign of overfitting to the training dataset). A test dataset is a dataset that is independent of the training dataset, but that follows the same probability distribution as the training dataset. If a model fit to the training dataset also fits the test dataset well, minimal overfitting has taken place. A better fitting of the training dataset as opposed to the test dataset usually points to overfitting. Thus the performance of the classifier is assessed using the testing data set.

4. RESULT

The functionality of the proposed work was divided into training and testing phases. During training, the parameters of the CNN were trained using SGD and modified by setting the epoch limit of 100. The network converged for 20 epochs generating the suitable model. The training and validation accuracy plots obtained is as shown in Fig -3. Also the training and validation loss plot obtained is displayed in Fig -4. The generated model was further tested using the testing images which provided an accuracy of 96.974%. Also various other performance measures were computed using the Confusion matrix for assessment.

The confusion matrix provided by the model for the testing images that we had chosen to predict the disease is shown in Fig -5. From the confusion matrix, the performance metrics: Precision, Recall and F1 score were computed. The model report obtained with the performance measures for each class is shown in Fig -6. In our work we have created a sample website which runs on the localhost using python where user can avail the service, the model which we have created using the above methodology has been included in the source code of the website for identification of the leaf disease. The home page of the website has the option to choose the file of the leaf which he want to predict the disease is shown in Fig -7. The Fig -8 shows the folder where the user can select the required image of the leaf. After selecting the image once he clicks on the predict button, the leaf name along with the disease name and the image selected by user will be displayed as shown in the Fig -9.

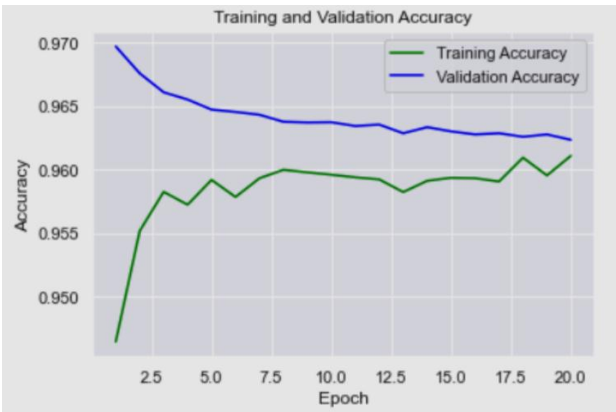

Fig -3: Training and validation accuracy plot

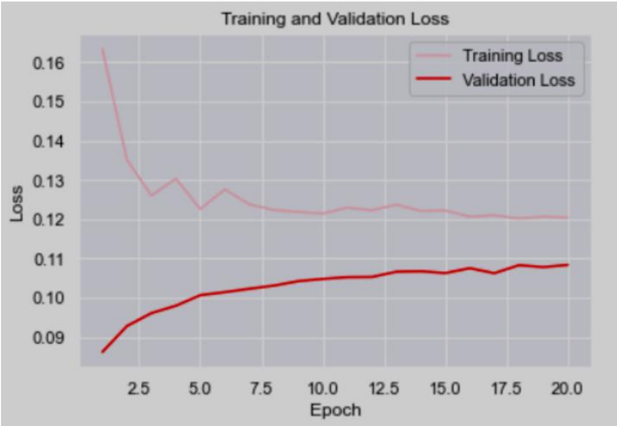

Fig -4: Training and validation loss plot

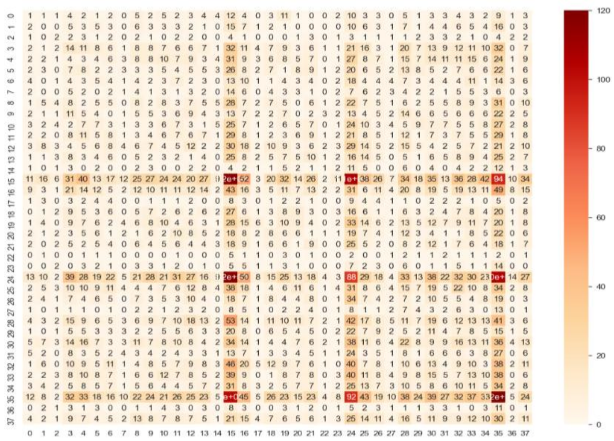

Fig -5: Confusion Matrix

	precision	recall	f1-score	support
Apple___Apple_scab	0.03	0.03	0.03	120
Apple___Black_rot	0.02	0.02	0.02	132
Apple___Cedar_apple_rust	0.00	0.00	0.00	54
Apple___healthy	0.04	0.04	0.04	324
Blueberry___healthy	0.03	0.03	0.03	300
Cherry_(including_sour)___Powdery_mildew	0.01	0.01	0.01	228
Cherry_(including_sour)___healthy	0.01	0.01	0.01	168
Corn_(maize)___Cercospora_leaf_spot Gray_leaf_spot	0.00	0.00	0.00	108
Corn_(maize)___Common_rust_	0.04	0.04	0.04	240
Corn_(maize)___Northern_Leaf_Blight	0.03	0.03	0.03	200
Corn_(maize)___healthy	0.01	0.01	0.01	235
Grape___Black_rot	0.02	0.02	0.02	245
Grape___Esca_(Black_Measles)	0.02	0.02	0.02	285
Grape___Leaf_blight_(Isariopsis_Leaf_Spot)	0.02	0.02	0.02	215
Grape___healthy	0.00	0.00	0.00	85
Orange___Haunglongbing_(Citrus_greening)	0.11	0.11	0.11	1101
Peach___Bacterial_spot	0.04	0.04	0.04	460
Peach___healthy	0.01	0.01	0.01	72
Pepper,_bell___Bacterial_spot	0.01	0.00	0.01	203
Pepper,_bell___healthy	0.02	0.02	0.02	296
Potato___Early_blight	0.03	0.03	0.03	200
Potato___Late_blight	0.01	0.01	0.01	200
Potato___healthy	0.00	0.00	0.00	31
Raspberry___healthy	0.00	0.00	0.00	75
Soybean___healthy	0.10	0.10	0.10	1020
Squash___Powdery_mildew	0.03	0.03	0.03	367
Strawberry___Leaf_scorch	0.03	0.03	0.03	222
Strawberry___healthy	0.01	0.01	0.01	91
Tomato___Bacterial_spot	0.05	0.05	0.05	425
Tomato___Early_blight	0.05	0.05	0.05	200
Tomato___Late_blight	0.05	0.05	0.05	381
Tomato___Leaf_Mold	0.04	0.04	0.04	191
Tomato___Septoria_leaf_spot	0.02	0.02	0.02	355
Tomato___Spider_mites Two-spotted_spider_mite	0.03	0.03	0.03	336
Tomato___Target_Spot	0.03	0.03	0.03	281
Tomato___Tomato_Yellow_Leaf_Curl_Virus	0.10	0.10	0.10	1072
Tomato___Tomato_mosaic_virus	0.00	0.00	0.00	75
Tomato___healthy	0.03	0.03	0.03	320
accuracy			0.05	10913
macro avg	0.03	0.03	0.03	10913
weighted avg	0.05	0.05	0.05	10913

Fig -6: Model report with performance measures

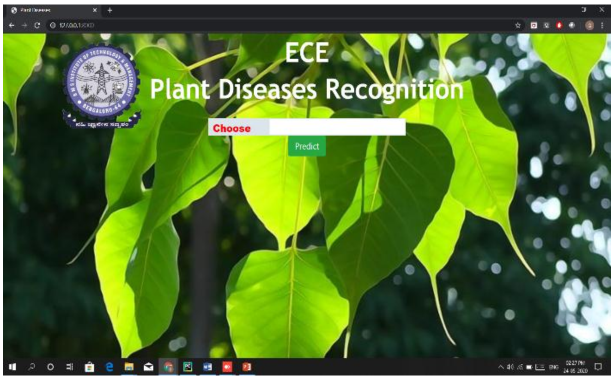

Fig -7: Home page of the website

Fig -8: Choosing the Image from Testing Folder

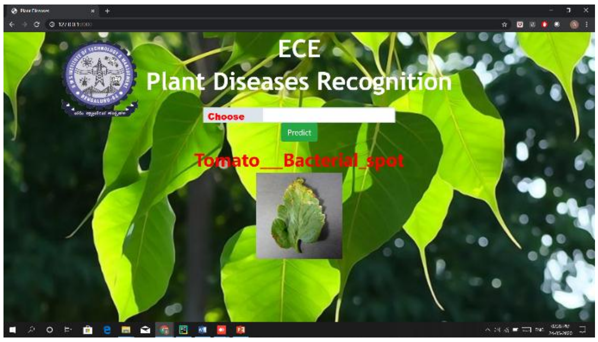

Fig -9: shows the disease name along with selected Image

5. CONCLUSIONS

The trained network is used to re-train a CNN Classifier. Several images were used to test the network and obtain the probabilities of the network to determine the right class.

The results have indicated that for the image database of 54000 images, the accuracy rate is found to be around 96.974%. The error rate is found to be around 3% due to small number of images present in the database for training and testing. As the number of images in the database are increased, the accuracy rates can also be drastically improved.

The project depicts how neural networks and image processing can be used to avoid uncertainties in the detection of a disease and hard work that goes in supervising a large plot of land as our farmers cannot be on the field 24x7.

The proposed system was developed taking in mind the benefits of the farmers and agricultural sector .The developed system can detect disease in plant. The proposed system is based on python, the accuracy and speed can be increased by use of Googles GPU for processing. The system is designed in such a way that the farmer can take the snapshot of the leaf using his phone and get the health status of leaf using his smartphone if his phone is connected to our network. By proper knowledge of the disease, the farmer can take appropriate measures to resolve that disease by applying required chemicals or organic materials to the soil so that the plant can restore the required nutrients.

REFERENCES

[1] Sachin D. Khirade, A. B. Patil, “Plant Disease Detection Using Image Processing”, International Conference on Computing Communication Control and Automation, 2015.
[2] Zulkifli Bin Husin, Abdul Hallis Bin Abdul Aziz, Ali Yeon Bin Md Shakaff, Rohani Binti S Mohamed Farook, “Feasibility Study on Plant Chili Disease Detection Using Image Processing Techniques”, Published in 2012 Third International Conference on Intelligent Systems Modelling and Simulation.
[3] Pooja V, Rahul Das, Kanchana V Department of computer science Amrita University Mysore, Karnataka “Identification of Plant Leaf diseases using Image Processing Techniques”, Published in IEEE International Conference on Technological innovation in ICT for Agriculture and rural Development 2017.
[4] Alex Krizhevsky, Ilya Sutskever, Geoffrey E. Hinton, University of Toronto, proposed “ImageNet Classification with Deep Convolutional Neural Networks”, Advances in neural information processing systems, 2012
[5] Monica Jhuria, Ashwani Kumar, and Rushikesh Borse, “Image Processing For Smart Farming: Detection Of Disease And Fruit Grading”, Proceedings of the 2013 IEEE Second International Conference on Image Information Processing (ICIIP-2013)
[6] Mrunalini R. Badnakhe, Prashant R. Deshmukh, “Infected Leaf Analysis and Comparison by Otsu Threshold and k-Means Clustering”, International Journal of Advanced Research in Computer Science and Software Engineering, Volume 2, Issue 3, March 2012.
[7] https://www.kaggle.com/vipoooool/new-plant-diseases-dataset