International Journal of Innovative Technology and Exploring Engineering (IJITEE)
ISSN: 2278-3075 (Online), Volume-9 Issue-1, November 2019

Tomato Leaf Disease Prediction using Convolutional Neural Network

R.Sangeetha, M. Mary Shanthi Rani

Abstract* — Plant diseases are the common cause of the reduction in yield eventually resulting in low income to farmers. Researchers are at their best efforts to find a solution for the detection of plant diseases to increase farm productivity. In this paper, a novel approach of disease detection and prediction for tomato plant leaves has been proposed using deep learning techniques. Training of the models was performed with the use of an open database of 13,848 images, which included 7 distinct classes of [plant, disease] combinations, including healthy tomato crops. Convolution Neural Network which is well-suited for detection and prediction problems has been used for predicting healthy and unhealthy leaves affected by two types of diseases septoria spot and bacteria spot. Experiments are conducted using plant village dataset comprising of 4930 images including healthy and unhealthy leaves. The performance of the model is evaluated using precision, recall and F1-score and the model has achieved the highest accuracy 94.66 %. The significantly high success rate makes the model a very useful advisory or early warning tool.

***Keywords*—Plant diseases, Prediction, Tomato leaves**

I. INTRODUCTION

India is a nation that depends strongly on the agricultural sector for a majority of the population. India's most popular vegetable is tomato. The three leading antioxidants, vitamin E, vitamin C, and beta-carotene are contained in tomatoes. They are also rich in potassium, for excellent health a very significant mineral. The cultivation region of tomato crops in India covers approximately 3,50,000 hectares and the manufacturing quantities amount to approximately 53,00,000 tonnes, making India the world's third-biggest tomato producer. The sensitivity of crops coupled with climatic conditions has made diseases common in the tomato crop during all the stages of its growth. Plants impacted by disease make up 10-30% of the total crop loss. Monitoring the plant diseases manually is a difficult task due to its complex nature and is a time-consuming process.

This paper deals with the creation of the model for leaf disease identification based on image classification, using deep convolutional networks. In the field of image classification, the latest generation of convolutional neural networks (CNNs) has achieved impressive results. To facilitate the fast and simple execution of the system in practice, new training and methodology are used.

Revised Manuscript Received on November 30, 2019.
* Correspondence Author

R.Sangeetha*, Department of Computer Science and Applications, The Gandhigram Rural Institute - Deemed to be University, Gandhigram, Dindigul, Tamilnadu - 624 302

M. Mary Shanthi Rani, Department of Computer Science and Applications, The Gandhigram Rural Institute - Deemed to be University, Gandhigram, Dindigul, Tamilnadu - 624 302

A. *Deep Learning*

Machine Learning is an artificial intelligence application in which the system learns and improves itself from the previous experience without needing it to be programmed.

It focuses on the evolution of the computer program so that the accessed data can be used for self- learning.

Deep Learning is a subdivision of machine learning which is excellent at understanding pattern based on a large amount of data. Recognizing objects from the images is done using three or more layers of artificial neural network where more than one feature of the image is extracted from every layer.

Neural Networks are proved to efficient method for Deep Learning. It is a computational model that works in a similar way to the neurons in the human brain. Each neuron takes an input, performs some operation and passes results as the output of the following neurons.

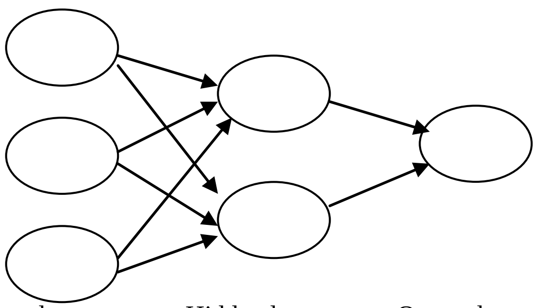

Fig. 1: Neural network

It consists of a large number of highly interconnected processing elements called neurons, each producing a sequence of real-valued activations. By providing input to initial neurons, they get activated and eventually activating the neurons in the next layer through weighted connections between successive layers. It may require several links and computational stages depending on the application.

Deep Learning is all about accurately assigning weights across all stages in the network. A neural network with multiple hidden layers is called Deep Neural Network. With the evolution of Deep Learning, it has progressed rapidly in the image classification field. The objective of deep learning methods is to learn the feature hierarchy automatically, thereby high-level features are obtained from the structure of the low-level feature. Deep Learning has emerged as a hot research field with the availability of powerful computing processors and storage. This paper is organized as follows: Section II focuses on the major work done in the field concerned. Section III sets out the methodology proposed and the model used along with the steps taken to achieve the necessary results. Section IV covers the findings and evaluation of the suggested methodology. Section V presents the paper's conclusion and provides the scope for a future job.

II. RELATEDWORK

In machine learning and deep learning, an extensive study was performed to compare disease detection and classification methods. We studied techniques of classification of Artificial Neural Network (ANN) and Convolutional Neural Network Classification used in the identification and efficiency of plant leaf diseases.

Artificial Neural Network is a commonly used computational model for machine learning and pattern recognition, using classifier ANN to identify plant leaf disease.

Dheeb Al Bashish et al [1] evaluated the proposed work for crop leaf disease recognition using feed-forward propagation algorithm and well performed with approximately 93 percent efficiency. Early scorch, cotton mold, late scorch, and small whiteness diseases that affect crops were monitored for a solution.

Keyvan Asefpour Vakilian et al [2] have been developed to extend the precision in the identification of two kinds of fungal diseases and the experimental result demonstrates approximately 92 % accuracy.

Mrunmayee Dhakate et al [3] proposed a system to recognize and classify diseases like leaf spot, bacterial blight, fruit spot and fruit rot diseases of pomegranate plant using back-propagation algorithm and the experimental result shows around 90% accuracy.

Ramakrishnan M et al [4] proposed work on the identification of groundnut plant disease Cercospora (leaf spot) using Backpropagation method. The experimental results and observation shows out of 100 sample diseased leaf images they classified four types of diseases and secured 97.41% of accuracy.

RashmiPawar et al [5] developed a technique for detecting the pomegranate plant disease and observed using 40 images with an accuracy of 90%.

Srdjan Sladojevic et al [6] proposed a technique to detect healthy leaves and 13 different diseased leaves of peach, cherry, pear, Apple, and Grapevine using CNN classification technique. More than 30000 images used in the dataset, achieved accuracy between 91% and 98% for separate class test and average accuracy 96.3%.

Sharada P. Mohanty et al [7] elaborate a technique for plant disease detection using public dataset 54306 images of 14 crops and 26 diseases with accuracy 99.35% using 20% of Testing information and using 80% of test information for 98.2%.

Serawork Wallelign[8] model designed to detect Septoria, Frogeye and Downy Mildew crop diseases using CNN classifier. They used the dataset containing 12673 four-class leaf images and precision of 99.32 percent.

Konstantinos P et al [9] CNN classification method has been created for the recognition of crop diseases. The dataset includes 87848 images of 25 distinct crops in a collection of 58 diseases and 99.53% accuracy.

III. METHODOLOGY

The primary goal of this study research is to construct a Convolution Neural Network to efficiently predict tomato leaf diseases. The seven common diseases that affect tomato leaves include target spot, mosaic virus, yellow leaf curl virus, bacterial spot, early blight, late blight and septoria leaf spot.

The proposed technique involves four major phases, namely: data acquisition, pre-processing, training, and data prediction. Figure 2 shows the flow diagram.

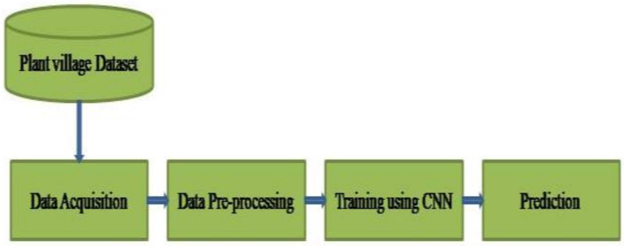

Fig 2: Proposed Method

The pre-processed images are trained using the deep convolutional neural network which consists of the convolutional, activation, pooling and fully connected layers.

B. DataAcquisition

The methodology was implemented using images from a publicly available dataset called the Plant Village dataset consisting of approximately 13787 images belonging to seven distinct classes. The dataset includes images of tomato leaf diseases that can affect the crop. The proposed work used two sets of injured images, namely bacteria spot and septoria spot. A subset of Plant Village, a repository containing 4930 images of tomato plants infected with two diseases, is the database used for evaluation.

The dataset is split into 80:20 proportions training dataset and test dataset. Each of the images downloaded belongs to the color space of the RGB and has been stored in uncompressed JPG format.

C. Data Pre-processing

The data set consists of images with minimal noise, so noise removal was not an essential pre-processing phase. The images in the dataset were resized to make the model training feasible computationally. The standardization technique of input or target factors tends to speed up the training process.

D. Training using Convolutional neural network (CNN)

Convolution Neural Network (CNN) is deep learning technique that plays a vital role in image processing applications such as object identification, recognition, detection, and classification. Deep Learning techniques are capable of learning feature representations from the data. A CNN model takes images as the input, applies convolution operation(s) and extracts the image features, resulting in dimensionality reduction of the input. The accuracy of image processing is directly influenced by those extracted features.

CNN model consists of layers such as Convolution, ReLU Layer, Pooling, Fully Connected, Flatten, and Normalization. Using CNN, the images would be compared piece by piece. Each piece is known as a feature or filter. From the input image, CNN uses the weight matrix and extract the specific features without misplacing the information about its spatial arrangement.

E. Convolution Layer

Convolution layer is the main and first layer of the CNN architecture. It is used to extract the characters from the input image. The convolutional layer maintains the association connects pixels by learning image features from input data.

The mathematical operation takes two input image matrices and 3×3 kernel filter. The 5×5 image pixel values and the convolution filter of 3×3.

The features after finishing the multiplication of the corresponding matrix. The CNN adds and divides the result by the total number of pixel and then create a map and put the value of the filter at that place. After that, it moves the feature to every other position of the image and obtains the output of the matrix and repeats the steps for the other filters. This layer moves the filter to every possible position on the image.

F. ReLU Layer

ReLU is an acronym for Rectified Linear Unit. It is the most widely used activation function implemented in *hidden layers* of neural network.In this layer, every negative value in the filtered image will be replaced with zeros as given in eqn.1

$$A(x) = max\ (0, x) \quad (1)$$

G. Pooling Layer

This layer compresses the image into a smaller size by taking the maximum value from the filtered image and obtains a matrix out of it. It also controls overfitting.

H. Fully Connected Layer

The neural network that is fully connected consists of a sequence of fully connected layers. Nodes are frequently referred to as "neurons" in fully connected networks. Consequently elsewhere in the literature, fully connected networks will commonly be referred to as "neural networks". The function of the fully connected layer from Rm to R n depends on the dimension of each input. When the networks get trained this feature vector is further used for classification, regression and other types of outputs. The FC layer holds composite and aggregated information from all the conv layers that matter. Let's dig a little deeper into what a fully connected network's mathematical form is

$$yi = \sigma(w1 \times 1 + \cdots + wm \times m) \quad (2)$$

Let x ∈ ℝ m represent the input to a layer that is completely attached. Let y(i) ∈ℝ be the i -the output from the layer that is fully attached.

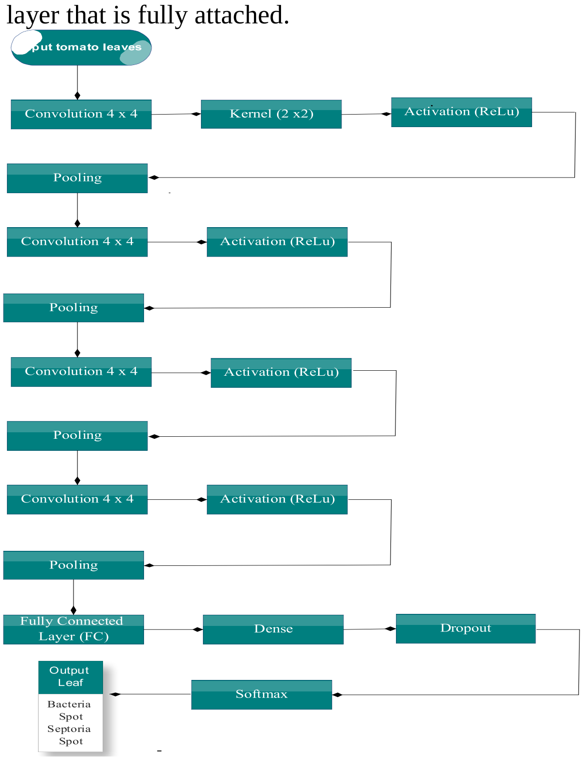

Fig. 3: Flowchart for Proposed model

IV. RESULTS AND DISCUSSION

On the Plant Village dataset, the application of the proposed methodology was carried out. It comprises of approximately 4930 pictures from 2 distinct types of tomato leaf diseases. Keras, a Python-based neural network API, was used to implement the model.

Each block consists of a layer of convolution, activation and maximum pooling layer. Four blocks are used in this architecture, followed by fully connected layers and softmax activation. For extraction of features, convolutional and pooling layers are used, while the fully connected layers are used to classify and predict. Activation layer are used to introduce non-linearity into the network.

The filter size is fixed as 5x5, while the amount of filters is gradually expanded as we move from block to block. To reduce the size of the feature maps, the max-pooling layer is used to speed up the training process. The max-pooling kernel size is 2x2.ReLU activation layer is used to introduce non-linearity in each of the blocks. Also, to prevent overfitting the training set, the dropout regularization method was used with a probability of 0.5.

Dropout regularization drops neurons randomly in the network during each training iteration to decrease the model's variance and simplify the network that helps to prevent overfitting. A softmax activation feature follows the second dense layer. 986 images were set aside for testing out of the 4930 pictures and 3944 images were used for training. To enhance the data set, Automatic data augmentation methods were used to rotate the pictures randomly by a tiny quantity of 20 degrees, horizontal flipping, vertical and horizontal shifting of images. The optimization was performed using Adam optimizer as the loss function with categorical cross-entropy. The batch size of 32 was used and for 40 epochs the model was trained.

Performance metrics

For quantitative evaluation and comparison, four metrics are used, including accuracy, positive predictive value (PPV) or precision, sensitivity or recall, and the harmonic mean of precision and sensitivity (f1-score). We denote TP, TN, FP, and FN as true positive, true negative, false positive and false negative, respectively. The evaluation metrics are defined as:

$$\text{Accuracy}= \frac{(\text{TP+TN})}{(TP+FP+FN+TN)} \quad (3)$$

$$\text{Precision} = \frac{\text{TP}}{(TP+FP)} \quad (4)$$

$$\text{Recall}= \frac{\text{TP}}{(TP+FN)} \quad (5)$$

$$\text{F1-score} = \frac{\text{2TP}}{(2TP+FP+FN)} \quad (6)$$

Table 1 displays the performance analysis of the proposed CNN classifier for bacterial spot disease. It is obvious from Table 1 that proposed classifier can achieve high performance in terms of precision, recall, f1-score. For testing, the test dataset contains 109 healthy leaf images and 222 unhealthy leaf images affected by bacterial spot disease. It could be observed from Table 1 that the proposed CNN classifier has achieved an accuracy of 0.92 in predicting healthy and bacteria spot disease affected leaves.

Table 1 Performance Analysis of CNN Classifier for Bacteria Spot Disease

Prediction	Precision	Recall	F1-score	Support
Healthy Leaf prediction	0.97	0.80	0.87	109
Bacteria spot disease prediction	0.91	0.99	0.95	222
Accuracy			0.92	331
macro avg	0.94	0.89	0.91	331
Accuracy 0.92447 (92.44%)				

The pictorial representation of the performance of the classifier is shown in fig 4.

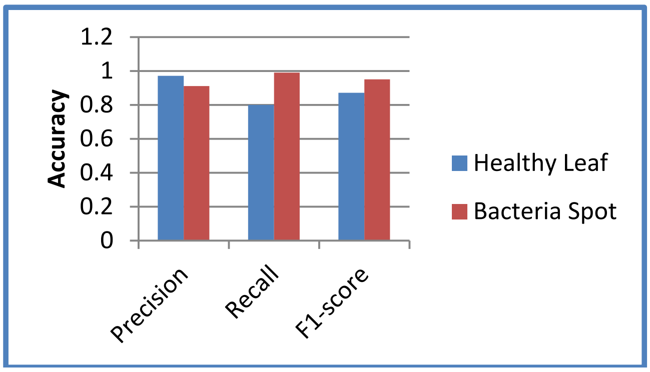

Fig 4 Performance Analysis of CNN Classifier

Table 2 displays the performance analysis of the proposed CNN classifier for septoria spot disease. Table 2 reveals that proposed classifier can achieve high performance in terms of precision, recall, f1-score. For testing, the test dataset contains 293 healthy leaf images and 361 unhealthy leaf images affected by septoria spot disease. It could be seen from Table 2 that the proposed CNN classifier has achieved an accuracy of 0.92 in predicting healthy and septoria spot disease affected leaves.

Table 2 Performance Analysis of CNN Classifier for Septoria Spot Disease

Prediction	Precision	Recall	F1-score	Support
Healthy Leaf prediction	0.96	0.90	0.90	293
Septoria spot disease prediction	0.92	0.99	0.94	361
Accuracy			0.91	654
macro avg	0.94	0.94	0.92	654
Accuracy 0.9466(94%)				

The pictorial representation of the performance of the classifier is shown in fig 5.

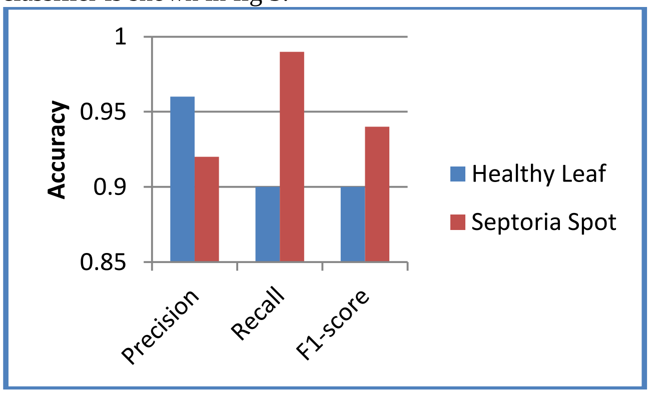

Fig 5 Performance Analysis of CNN Classifier

Table 3compares the accuracy of prediction for bacterial spot disease and septoria spot disease. The proposed CNN classifier has achieved good accuracy of above 92% in predicting healthy, Bacteria spot and Septoria spot diseases.

Table 3 Comparison analysis of Accuracy for Bacteria spot and septoria spot

CLASS	ACCURACY
Bacteria Spot	92.44
Septoria Spot	92.33

The pictorial representation of the performance of the classifier is shown in fig 6.

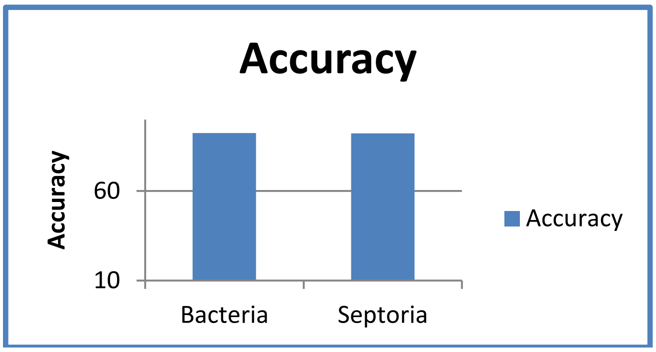

Fig 6 Comparison analysis of CNN classifier Accuracy

Table 4 displays the performance analysis of precision for bacteria spot disease and septoria spot disease and its graphical representation is shown in Fig. 7.

Table 4 Comparison analysis of Precision for Bacteria spot and septoria spot

CLASS	PRECISION(0)	PRECISION(1)
Bacteria Spot	0.97	0.91
Septoria Spot	0.96	0.92

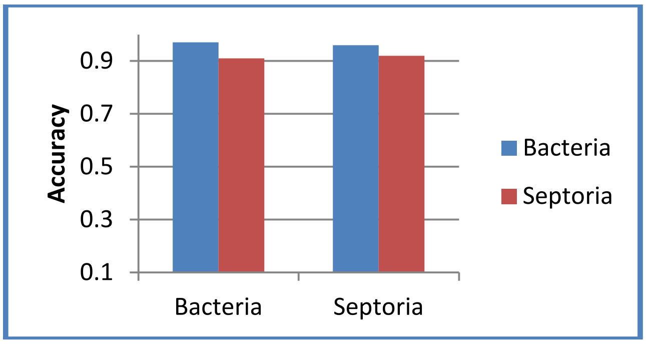

Fig 7 Comparison analysis of Precision for Bacteria spot and septoria spot

Table 5 displays the performance analysis of recall for bacteria spot disease and septoria spot disease

Table 5 Comparison analysis of recall for Bacteria spot and septoria spot

CLASS	RECALL(0)	RECALL(1)
Bacteria Spot	0.80	0.99
Septoria Spot	0.90	0.99

The pictorial representation of the performance of the classifier is shown in fig 8.

OPEN ACCESS
International Journal of Innovative Technology and Exploring Engineering (IJITEE)
ISSN: 2278-3075 (Online), Volume-9 Issue-1, November 2019

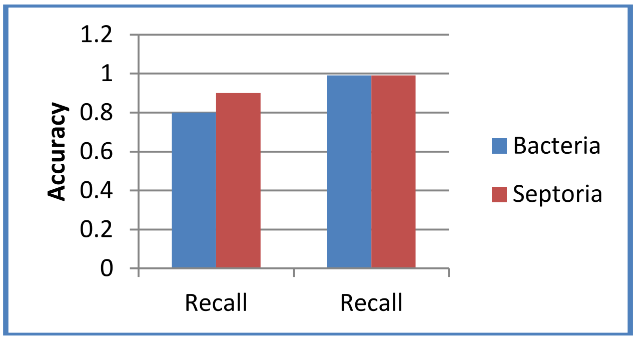

Fig. 8 Comparison analysis of recall for Bacteria spot and septoria spot

Table 6 compares the f1-score for bacteria spot disease and septoria spot disease.

Table 6 Comparison analysis of f1-score for Bacteria spot and septoria spot

CLASS	F1-Score(0)	F1-Score (1)
Bacteria spot	0.87	0.95
Septoria spot	0.90	0.94

The pictorial representation of the performance of the classifier is shown in Fig. 9.

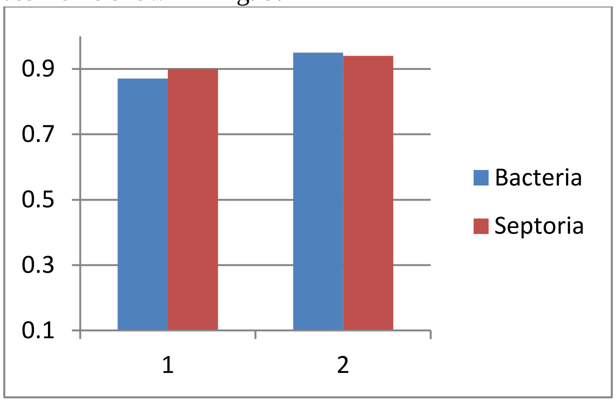

Fig 9 Comparison analysis of F1-score for Bacteria spot and septoria spot

V. CONCLUSION

Agriculture is still one of the largest sectors on which the majority of the Indian population depends. Disease detection in these plants is therefore critical to economic growth. Tomato is one of the main crops generated in big amounts. Therefore, in the tomato crop, this paper seeks to detect and predict two distinct diseases. While most of the previous study work focuses on binary classification, our model for multi-class classification provides significant improvement. Our proposed network can be very useful for the identification and prediction of tomato leaf disease using neural convolution network. Although the implemented model was only tested on the data set of Plant Village, we think it can be effectively used for other classification problems. In the future, we intend to assess the implemented model for the identification and classification of various datasets and plant leaf disease.

REFERENCES

1. Dheeb Al Bashish, Malik Braik, and SuliemanBani- 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.
2. KeyvanAsefpourVakilian and JafarMassah, "An artificial neural network approach to identify fungal diseases of cucumber (Cucumissativus L.) Plants using digital image processing", Archives of Phytopathologyand Plant Protection, Vol. 46, Issue 13, Taylor & Francis 2013, pp.1580-1588.
3. MrunmayeeDhakate and Ingole A. B, "Diagnosis of Pomegranate Plant Diseases using Neural Network", IEEE 5th National Conference on Computer Vision, Pattern Recognition, Image Processing and Graphics (NCVPRIPG), Patna2015.
4. Ramakrishnan M. and SahayaAnselinNisha A., "Groundnut Leaf Disease Detection and Classification by using Back Propagation Algorithm". IEEE International Conference on Communications and Signal Processing (ICCSP), Melmaruvathur 2015, pp. 0964 –0968.
5. RashmiPawar and AmbajiJadhav, "Pomegranate Disease Detection and classification", IEEE International Conference on Power, Control, Signals, and Instrumentation Engineering (ICPCSI), Chennai 2017, pp.2475-2479.
6. SrdjanSladojevic, Marko Arsenovic, AndrasAnderla, DubravkoCulibrk and DarkoStefa-novice, "Deep Neural Networks Based Recognition of Plant Diseases by Leaf Image Classification", Computational Intelligence and Neuroscience, Article ID 3289801, 2016.
7. Sharada P. Mohanty, David P. Hughes, and Marcel Salathe, "Using Deep Learning for Image-Based Plant Disease Detection", Frontiers in Plant Science, Vol. 7, Article 1419,2016.
8. SeraworkWallelign, MihaiPolceanu, and Cedric Buche, "Soybean Plant Disease Identification Using Convolutional Neural Network", International Florida Artificial Intelligence Research Society Conference (FLAIRS-31), Melbourne, United States 2018, pp. 146-151.
9. Konstantinos P. Ferentinos, "Deep learning models for plant disease detection and Diagnosis", Computers and Electronics in Agriculture, Vol. 145, Elsevier 2018, pp. 311-318.

AUTHORS PROFILE

Dr. M. Mary Shanthi Rani holds Ph.D in Computer Science and has more than 14 years of teaching experience .She has great passion for teaching and is currently working as Assistant Professor in the Department of Computer Science and Applications, The Gandhigram Rural Institute (Deemed to be University), Gandhigram. She has nearly sixty two publications in International Journals and Conferences .Her research areas of interest are Image Compression, Information Security, Computational Biology and Machine Learning. She has served in various academic committees in designing the curriculum for B.Sc. and M.C.A courses as well. She has also served as reviewer of Peer-reviewed International Journals and Conferences. She is a Life member of Indian Society for Technical Education. She has the credit of being the Associate Project Director of UGC Indo-US 21st Knowledge Initiative Project. She is also the Lead Co-ordinator of GRIHAA (Gandhigram Research Incubator On Healthcare And Agriculture Applications)research group a part of LeadingIndia.ai project with a mission to promote Artificial Intelligence skilling and research. She has also organized three workshops on Deep Learning in Computer Vision using Python Programming.

Miss.R.Sangeetha is a Research Scholar (Full- time) in the Department of Computer Science and Applications, The Gandhigram Rural Institute – Deemed to be University, Dindigul, India. She received her Bachelor of Science (B.Sc) degree in Computer Science in the year 2014 from Madurai Kamaraj University and Master of Science (M.Sc) degree in the year 2016 from Loyola College and Master of Philosophy (M.Phil) in the year 2017 from St.Xavier's college. She is currently pursuing Ph.D. degree in Gandhigram Rural Institute-Deemed University. His research focuses on Data hiding in Medical Images.