Crossref
Similarity Check
Powered by iThenticate

Reason- A Technical Journal
ISSN 2277-1654
Volume XXI 2022

TOMATO LEAF DISEASE DETECTION USING CONVOLUTION NEURAL NETWORK

Arnab Nanda Goswami, Shubhajyoti Das and Satyendra Nath Mandal
Department of Information Technology, Kalyani Government Engineering College, Kalyani- 741235.
Corresponding email: satyen_kgec@rediffmail.com

Paper submitted on July 02 2022, Accepted after revision on June 05 2023
DOI: 10.21843/reas/2022/19-28/222960

Abstract: Tomato grains are an essential significant, abundant product in the Indian market with high commercial value. Diseases are detrimental to the plant's health, affecting its growth. It is crucial to monitor the condition of the crop for a sustainable farming system. There are many types of tomato diseases that affect the leaves of the crop at an alarming rate. This paper slightly modifies the evolutionary neural network model called InceptionV3 to detect and classify disease on tomato leaves. The main goal of the proposed work is to find solutions to the problem of tomato leaf disease detection using simple methods while using minimal computing resources to achieve results comparable to the latest technology. Neural network models employ automated feature extraction to classify the input image into the corresponding disease class. This proposed system has achieved an average accuracy of 94-95%, which indicates the feasibility of the neural network approach even in adverse conditions.

Keywords: leaf disease detection, neural network, convolution, inceptionV3

1. INTRODUCTION

India is a country where most of the population heavily depends on agriculture. Tomato is the most common vegetable used throughout India. Tomatoes contain three important antioxidants called Vitamin E, Vitamin C, and Beta-Carotene. They are also rich in potassium, an essential mineral for good health. The area under tomato cultivation in India is about 350,000 hectares, and the production volume is around 5,300,000 tons, making India the world's third largest tomato producer. Crop susceptibility to climatic conditions has made diseases common at all stages of tomato crop growth. Diseased plants make up 10-30%of the total crop loss [1]. Identifying such diseases in plants is very important to prevent massive damage in terms of yield and quantity of agricultural products. Manually monitoring plant diseases is difficult because of its complex nature and time-consuming process. Therefore, a manual effort for this work must be reduced while making accurate predictions and ensuring that farmer lives are hassle-free. Observable patterns are challenging to grasp, leading many farmers to make erroneous assumptions about the disease. As a result, preventive measures taken by farmers can be ineffective and sometimes harmful. Farmers usually gather and apply standard immunizations, as they need more expert advice on dealing with crop infestations. There are situations where due to inadequate knowledge or misinterpretation of the severity of the disease, excess or low doses of pesticides have caused crop damage. This is the implication of the proposed method, which aims to accurately identify and classify diseases of the tomato crop [2].

The methodology suggested in this paper depicts the most common diseases found in the tomato plant, like Bacterial spots, Black mold, Gray spots, Late blight, and Powdery mildew, among many others. Any leaf image given as input can be classified into one of the disease classes or deemed healthy. The database used includes around 3978 imagesof 5 different tomato leaf diseases. The methodology consists of three crucial steps: data acquisition, pre-processing, and classification. As mentioned earlier, the images used to implement the proposed methodology were acquired from a publicly available dataset called Plant Village. In the next step, the images were resized to a standard size before feeding them into the classification model. The final step is the classification of the input images using a slight variation of the deep learning convolution neural network (CNN) standard model called the InceptionV3, which consists of the convolution, activation, pooling, and fully connected layers.

The paper is organized as follows: Section II focuses on the prominent work done with the concerned field. Section III elucidates the proposed methodology and the model used, along with the steps taken to obtain the necessary results. Section IV pertains to the results and the analysis of the proposed methodology. Section V includes the conclusion of the paper and provides the scope for future work. The images used to implement the proposed method were obtained from a dataset having 3978 images. The images were resized to a standard size in the next step before feeding into the classification model. The final step is to classify the input images with a slight modification of the Deep Learning Convulsive Neural Network (CNN) [3] standard model, known as InceptionV3, which consists of the activation, pooling, and fully connected layers.

2. LITERATURE SURVEY

It is important to acknowledge previous research in this area in order to be able to move forward in the right direction. Plant leaf disease detection is a significant field of research where image processing and deep learning techniques have been widely used for proper classification. In this paper, authors discuss the strategies most popularly incorporated into the literature in the relevant field. Observing a large grain field is a tedious task if done manually. Plant care needs to reduce human effort. So it is a popular research domain that attracts many researchers. Several works related to plant diseases are observed in the literature. Deep learning methods [4] have been employed to solve problems that are easy for humans, such as playing games or identifying objects, but challenging to explain mathematically or comparatively prohibitively expensive. Image recognition, in particular, has experienced a paradigm shift, with progress sprouting everywhere. CNN is considered to be the most efficient architecture because of its ability to accelerate understanding of visual content and so help categorize. CNN can be classified into seven main categories depending on the architecture changes. Many methods have been built to improve CNN, including GoogleNet [5], DAG Net, and AlexNet [6]. Many previous studies have examined image recognition, and a specific classification has been used to classify images as healthy or unhealthy.

The primary source of identification is a disease of a tree is its leaves. Symptoms of the disease can be seen on the leaves. Over the past decades, K-Neighbor (KNN), Support Vector Machine (SVM), Fisher Linear Discriminant (FLD), Artificial Neural Network (ANN), Random Forest (RF) and other Classification algorithms have been widely used to identify crop diseases. Due to complications in the image of diseased leaves, photos of plant diseases are difficult to detect automatically; tasks include approaches to deep learning, especially the Convolutional Neural Network (CNN). It has recently risen to the top of the priority list by overcoming various obstacles. Both large and small-scale challenge classification is done by CNN, especially when image recognition is involved. It has demonstrated exceptional image processing and classification capabilities.

In addition, when deep learning algorithms are engaged in real-world situations, various factors affect their performance. As a result, the practical application of technology for automatic disease detection needs to be improved. Though there are several approaches to this problem, their accuracy or prediction level must give satisfactory results in the real-time scenario. Hence in this methodology, that problem is tried to solve in such a way that there will be no compromise in the level of prediction. Here with the help of this proposal, an accuracy level of approximately 96% has been achieved. The CNN model used here is InceptionV3, which gives a satisfactory result to identify the image of the tomato leaf disease accurately. In this study, authors first look at the general structure of the plant disease detection system, followed by a review of articles related to the study. In the end, authors give a comparative study of different papers. Here, performance analysis on both architectures is conducted by conducting model training in two ways [7]. This is done in the first case from scratch and in the second case using transfer learning. Transfer learning is matched to adapting pre-trained weights obtained through the training model in the InceptionV3 dataset. The model has been implemented and has given near about 96% accuracy. It illustrates the feasibility of this approach. However, when testing a trained model instead of a set of sample test images obtained from an online public data source that differs significantly from the train set, the model's accuracy falls to approximately 89%. This is a common problem encountered with outstanding neural networks for train and various distribution test sets. The authors have proposed a method where these conditions include illumination, complex background, and resolution of different images, size, and adaptation. They effectively demonstrate the accuracy of this method and the need for minimal computational effort.

3. PROPOSED METHODOLOGY

Deep learning is a technique that allows many classifications that use linear regression to work the base together, after which any activation function can be applied. The difference is the number of nodes between deep learning and traditional learning. Deep learning uses numerous nodes where a single node is present in the former. These nodes are collectively referred to as a neural network, and each node is named a Realization Deep learning can have any number of layers besides the input layer, the output layer [8]. Each layer contains neural units that can be hundreds or thousands. There is a layer called the number hidden layer, which contains nodes not part of the input or output layers. Another advantage of deep neural networks is that the classifiers build the network. The application of deep learning is widespread and is used effectively in image processing. The rest of the section details how deep learning is used to diagnose plant diseases.

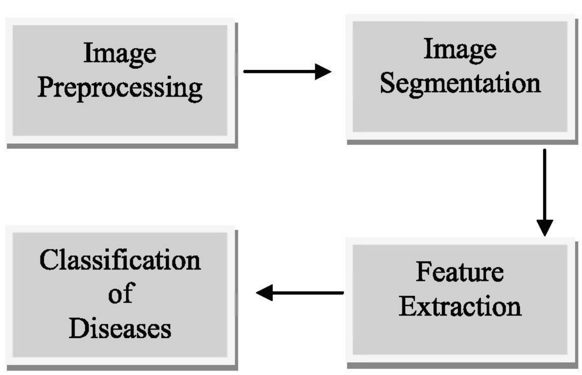

Fig. 1: Flow Diagram of Methodology

The step-by-step method of the plant disease detection system is illustrated in Fig.1 and is understood as follows:

The first step is to improve the captured image to remove unwanted materials and update the special features required in the following processing. Geometric image conversion (such as scale, translation, and rotation) is classified as a pre-processing technique. The second step is a process that breaks an image into many pieces. The main goal of this phase is to categorize artifacts or extract other useful information from digital images. The next step is to emphasize the critical elements that, when effectively demonstrated, provide the precise information needed for taxonomy. Feature extraction techniques can be applied to the object's geometry, shape, or color. Texture extraction is the most widely used technology in this area. The final stage determines whether the input can be classified as whether it is affected by the disease. Classification algorithms can be used to identify images based on the properties derived from them[9].

3.1. Data Acquisition

Pictures of tomato leaf disease were taken from the Plant Village store. Images (Fig. 2) for Diseases were downloaded using a Python script. The acquired dataset contains about 3978 images of six different categories consists of Bacterial Spot, Block Mod, Gray Spot, Healthy, Late Blight, Powdery Mildew. The dataset contains pictures of all major leaf diseases that may affect the tomato crop [10]. Each downloaded image is stored in RGB color space by default and in uncompressed JPG format.

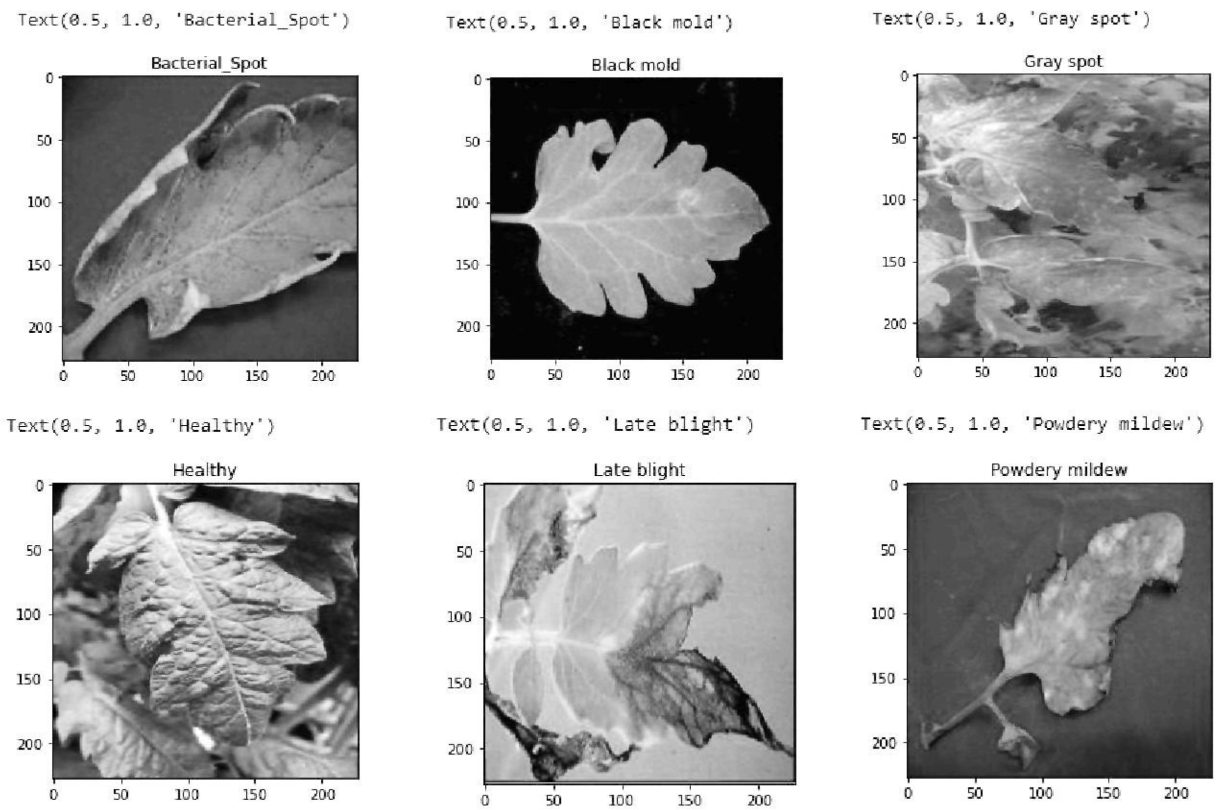

Fig. 2 : Training and Testing Classes for the Experiment.

3.2. Data pre-processing

Acquired datasets contain images with minimal noise, so word removal was not a necessary pre-processing step. Dataset images have been resized to 224 x 224 Resolution [11], making the training process faster and model training computationally possible. The process of standardizing input or target variables speeds up the training process. This is done by improving the numerical condition of the optimization problem. It also ensures that many default values associated with initialization and completion are appropriate. Using averages and standard deviations, one normalizes images to get all pixel values in the same range. In machine learning terminology, it is called Z-score.

2. CLASSIFICATION

The Convolution Neural Network (CNN) can be used to create a computational model that works on unorganized image inputs and converts them into corresponding classification output labels. These belong to the category of multi-layer neural networks that may be trained to learn the features required for classification purposes. They require less pre-processing than conventional methods and perform automatic feature extraction, which gives better performance, and the best results can be seen using variations of the InceptionV3. InceptionV3 is a typical CNN model consisting of convoluted, activation, pooling, and fully integrated layers. The architecture used to classify tomato leaf diseases is a variation of the InceptionV3 model. It consists of an extra block of convoluted, activation, and pooling layers compared to the original InceptionV3 architecture. Each block has a convoluted activation and maximum pooling layer. This architecture uses a fully integrated layer and three blocks followed by Softmax activation [12]. Convoluted and pooling layers are used for feature extraction, whereas fully connected layers are used for classification. Activation layers are used to introduce nonlinearity to the network. Authors have implemented a convoluted operation to remove the convoluted layer feature. As the depth increases, the complexity of the extracted properties increases. The filter size is fixed at 5 x 5, where the filters gradually increase as one moves from one block to another. This increase in the number of filters is necessary to reduce the size of the feature map due to the use of pooling layers in each block. Feature maps also have zero pads to save image size after applying the convolution operation. The maximum pooling level is used to reduce the size of the feature map, speed up the training process, and make the model less variable for minor changes in input. The kernel size for maximum pooling is 2 x 2. ReLu activation levels are used in each block to introduce nonlinearity [13]. Also, the dropout regulation technique has been used to avoid additional fitting of the train set. Dropout regularization randomly drops neurons in the network during each repetition of training to simplify the network, which helps to reduce and prevent model differentiation. The architecture of the model is described in Fig. 3.

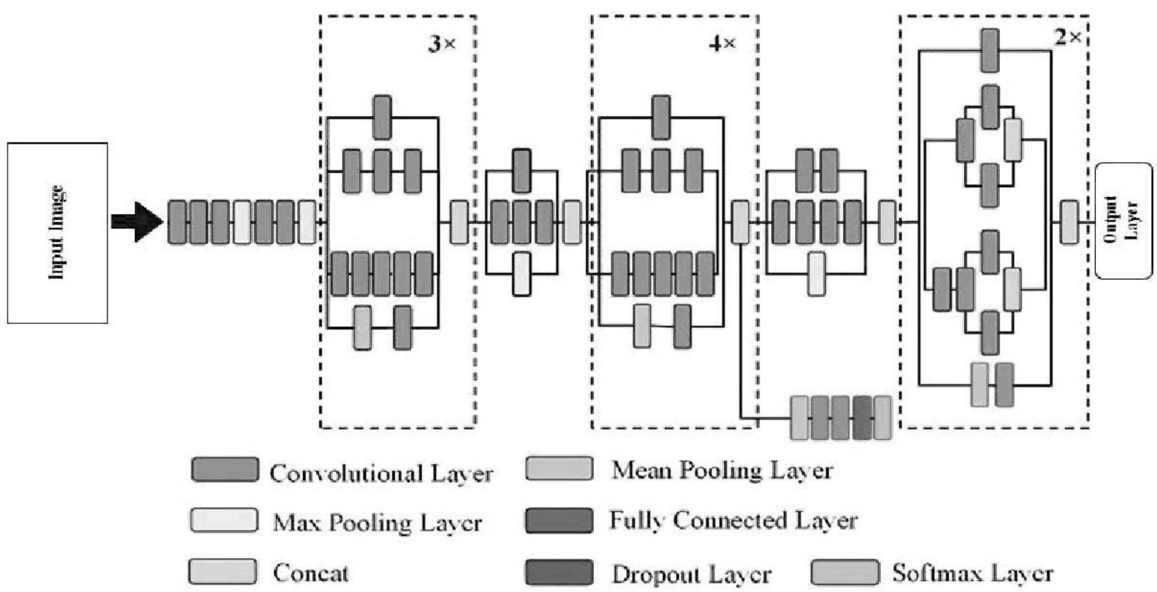

Fig. 3: Architecture of the Inception V3 model.

5. EXPERIMENTAL SETTINGS

In order to implement the proposed way on the Plant Village dataset, authors have followed a few steps. The dataset consists of around 4976 images belonging to 6 different classes of tomato leaf diseases and healthy leaves. Keras, a widespread neural network API widely used in Python, has been used for this model implementation. Of the 4976 images, 998 were set aside for testing, and 3978 were used for training. In order to increase the dataset, automatic data augmentation techniques have been used. Automatic rotational characteristics have been included with a fixed value for the rotational range 7. Side by side, the horizontal and vertical flip has also been set. Adam optimizer is used for optimization with categorical cross entropy as the loss function. A batch size 32 has been used, and the model has been trained for 200 epochs. The shear and zoom ranges are set to 0.2, and the class mode has been set to categorical.

6. RESULTS AND ANALYSIS

Fig. 4 and Fig. 5 describe the analysis of accuracy and loss respect to training and validation. It is observed that the model was not provided any over fitting and under fitting issues. Over fitting and under fitting issues can lead a model towards misclassified result. Over fitting issue can be defined as when a model works well on training data but it leads to incorrect result for unknown dataset. Under fitting provide incompetent result for train and test dataset.

The evaluation of the model is described in form of confusion matrix. Fig. 6 describes the confusion matrix of the model.

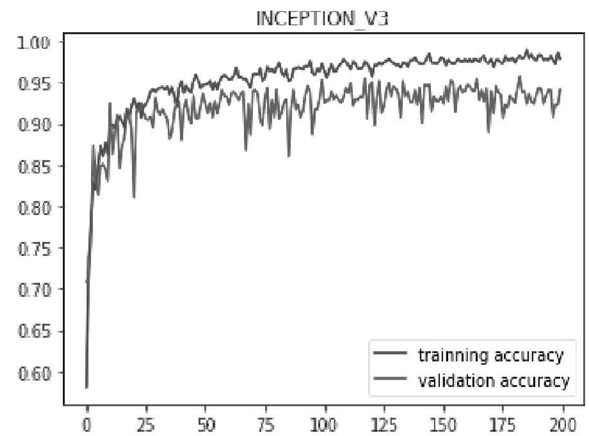

Fig. 4 : Training accuracy Vs Validation accuracy

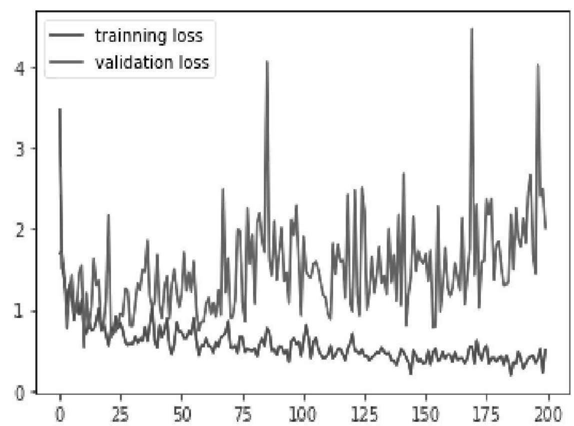

Fig. 5 : Training lossVs Validation loss

Predicted Class	Actual Class →	BP	BM	GP	LB	HLT	PM
	BP	107	002	001	006	000	004
	BM	002	094	012	008	000	011
	GP	001	000	118	003	000	010
	LB	004	002	009	138	000	011
	HLT	002	000	005	001	148	004
	PM	001	001	005	001	000	198

Fig. 6 : Confusion Matrix

Table 1: Accuracy with respect to each class

Disease Name	Accuracy
Bacterial Spot Accuracy	89.17
Black Mold Accuracy	74.02
Gray Spot Accuracy	89.39
Late Blight Accuracy	84.15
Healthy Accuracy	92.50
Powdery Mildew Accuracy	96.12

So, the accuracy that is obtained from the confusion matrix is depicted in Table 1. A validation accuracy value of 95.82% was obtained when authors trained the model with 200 epochs, while a high 98.94% accuracy was obtained as training accuracy. An average validation accuracy of 95% accuracy has been obtained. It is an effective measure of classification created by the deep learning model. In contrast to the era, the accuracy and loss plots of trains and experiments indicate a way of visualization and the speed of model convergence. It can be seen that the model has stabilized in about 200 epochs. The results show that the model performs well on the dataset and can be used to classify six tomato leaf diseases with minimum resource requirements. The performance of the model is described in Table 2.

Table 2: Performance of the Proposed Model

Epochs	Accuracy	Precision	Recall	F1 Score
200	0.9612	0.9612	1.0	0.9802

The implementation process requires minimal hardware rather than large neural networks that typically require high computational resources or use graphics processing units. This is due to low training parameters for low-level presence with small filter sizes and small train size diagrams. The differences in the adopted InceptionV3 model are easy to understand and implement [14]. The model thus provides a simple and effective way to solve plant disease detection problems with comparative results, where the authors deal with plant diseases of multiple crops. With low resource constraints and minimal data, the model yields comparable results to the traditional state-of-the-art techniques. Along with the results, authors have also tried to obtain the class-wise precision and recall score to get the insight data and view of the obtained results. Precision is a metric defined as the ratio of corrected prediction of actual class and total samples. Precision is a metric defined as the ratio of corrected prediction of actual class and total samples. It is mathematically defined as

$$Precision = \frac{TruePositive\ (TP)}{(TruePositive\ (TP) + FalsePositive\ (FP))} \quad (1)$$

Recall is also a classification metric is defined as the ratio of correctly predicted class and the sum of correctly predicted class and incorrectly predicted class. The mathematically definition is

$$Recall = \frac{TruePositive\ (TP)}{(TruePositive\ (TP) + FalseNegative\ (FN))} \quad (2)$$

The Harmonic mean of Precision and Recall is used to get the F1 score. A high F1-score indicates high Precision and Recall. It can defined as

$$F1\ Score = 2 * \frac{(Precision * Recall)}{(Precision + Recall)} \quad (3)$$

Table 3 : described the model's classification report in terms of Precision, Recall and F1 Score.

Table 3: Classification Report of different class

Class	Bacterial Spot	Black Mold	Gray Spot	Late Blight	Healthy	Powdery Mildew
Precision	0.8917	0.7402	0.8939	0.8415	0.9250	0.9612
Recall	1.0	1.0	1.0	1.0	1.0	1.0
F1 Score	0.9427	0.8507	0.944	0.9139	0.9610	0.9802

The present model developed shows remarkable results and these can be comparable with the similar other works reported earlier [15, 16].

7. CONCLUSION AND FUTURE WORK

Agriculture is still one of the most critical sectors on which most of the Indian population depends. Diagnosis of these crops is vital for the growth of the economy. Tomato is one of the major crops that are produced in large quantities. Therefore, this paper aims to identify and classify six different disease stages of tomato crops. The proposed method for classifying tomato leaf diseases obtained from the Plant Village dataset uses a constitutional neural network model. The architecture used is a simple evolutionary neural network with a minimum number of layers so that tomato leaf diseases can be divided into six categories. Different learning rates and optimizers can also be used to test the proposed model in future work. It may also include testing with new architectures to improve model performance on train sets. Thus, the above model can be used as a decision tool to help and assist farmers in identifying the diseases that can be found in tomato plants. The proposed method, with 95-96% accuracy, can accurately detect leaf disease with little computational effort.

REFERENCES

[1] Prajwala, T. M., Alla, P., Ashritha, K. S., Chittaragi, N. B., Koolagudi, S. G., Tomato Leaf Disease Detection using Convolutional Neural Networks, Proc. Eleventh International Conference on Contemporary Computing (IC3), 2-4 August, 2018, Noida, India.

[2] Bhakat, A., Nandakumar, N. and Rajkumar, S., An enhanced approach for Plant Leaf Disease Detection, Algorithms, Computing and Mathematics Conference, August 19 – 20, 2021, Chennai, India.

[3] Zhao, S., Peng, Y., Liu, J. and Wu, S., Tomato Leaf Disease Diagnosis Based on Improved Convolution Neural Network by Attention Module, Agriculture, Vol. 11, No. 7, p.651, 2021.

[4] Xie, S., Girshick, R., Dollar, P., Tu, Z., and He, K., Aggregated Residual Transformations for Deep Neural Networks, Computer Vision and Pattern Recognition, arXiv: 1611.05431.

[5] Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., Erhan, D., Vanhoucke, V., and Rabinovich, A., Going Deeper with Convolutions, in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 1–9, 2015.

[6] Krizhevsky, A., Sutskever, I. and Hinton, G. E., Imagenet Classification with Deep Convolutional Neural Networks, in Advances in Neural Information Processing Systems, 2012, pp. 1097–1105.

[7] Agarwal, M., Singh, A., Arjaria, S., Sinha, A. and Gupta, S., ToLeD: Tomato Leaf Disease Detection using Convolution Neural Network, International Conference on Computational Intelligence and Data Science (ICCIDS 2019).

[8] Wang, Q., Qi, F., Sun, M., Qu, J. and Xue, J., Identification of Tomato Disease Types and Detection of Infected Areas Based on Deep Convolutional Neural Networks and Object Detection Techniques, Computational Intelligence and Neuroscience, Vol. 2019, Article ID 9142753.

[9] Krizhevsky, A., Sutskever, I. and Hinton, G. E., Imagenet Classification with Deep Convolutional Neural Networks, Advances in neural information processing systems, 25 (NIPS 2012).

[10] Nagamani, H. S. and Sarojadevi, H., Tomato Leaf Disease Detection using Deep Learning Techniques, International Journal of Advanced Computer Science and Applications, Vol. 13, No.1, 2022.

[11] Gadade, H.D. and Kirange, D.K., Machine Learning Approach towards Tomato Leaf Disease Classification, International Journal of Advanced Trends in Computer Science and Engineering, Vol. 9, No. 1, pp. 490-495, 2020.

[12] Wu, Y., Xu, L. and Goodman, E. D., Tomato Leaf Disease Identification and Detection Based on Deep Convolutional Neural Network, Intelligent Automation & Soft Computing, Vol. 28, No. 2, 2021.

[13] Vinaya, P., Kumar, G. S. and Aravind, K. U., Classification of Tomato Plant Leaf Disease Using Neural Network, Proceedings of the International Conference on Innovative Computing & Communication (ICICC), 23 April 2021.

[14] Sreelatha, P., Udayakumar, S., Karthick, S., Chowdhary, S., Kavya, K. Ch. and Madiajagan, M., Managing the Tomato Leaf Disease Detection Accuracy Using Computer Vision Based Deep Neural Network, Journal of Contemporary Issues in Business and Government, Vol. 27, No.1, 2021.

[15] Elhassouny, A. and Smarandache, F., Smart Mobile Application to Recognize Tomato Leaf Diseases Using Convolutional Neural Networks, IEEE/ ICCSRE 2019, 22-24 July, 2019, Agadir, Morocco.

[16] Khurana, J., Sharma, A., Chhabra, H. S. and Nijhawan, R., An Integrated Deep Learning Framework of Tomato Leaf Disease Detection, International Journal of Innovative Technology and Exploring Engineering, Vol. 8, No.11S, 2019.