Classification and detection of Plant Leaf Diseases using various Deep Learning techniques and Convolutional Neural Network

Partha P. Mazumder[1], Monuar Hossain[1] and Md Hasnat Riaz[1]

[1]Department of Computer Science and Telecommunication Engineering, Noakhali Science and Technology University, Noakhali 3814, Bangladesh

Abstract. In this paper, we developed a Convolutional Neural Network model for detecting and classifying simple leaves images of (mostly) diseased plants and healthy plants with the help of different types of deep learning methodologies. We used an open database from PlantVillage dataset of 54,306 images containing 14 different plants in a set of 38 distinct classes of (diseased plants and healthy plants) to train our model. Among different model architectures were trained, the best performance reaching a 99.22% success rate using 0.4%data as testing among the whole dataset in identifying the corresponding (diseased plant or healthy plant) combination. This significantly good amount of success rate ensures the model a very useful advisory or early warning tool, and also an approach that could be further extended to uphold an integrated plant disease identification system to operate in real cultivation conditions or a clear path toward smartphone-assisted crop disease diagnosis on a large amount of areas.

Keyword: Plant disease classification, Neural network, InceptionV3, Xception.

1 Introduction

Plant diseases have a longer lasting effect on agricultural products. There are estimated more than $30-50 billion annually monetary loss caused by plant diseases[1].Modern technologies have blessed human society the potential to produce ample rations to meet the request of more than 7 billion people. However the lack of side effects of the food always remains intimidated by a number of factors such as significant amount of change in climate(Tai et al., 2014)[2], the reduce in pollinators (from the reports of Plenary of the Intergovernmental Science-Policy Platform on Biodiversity Ecosystem and Services of its 4th session, 2016)[3], plant diseases (Strange and Scott, 2005)[4], and also many others. Also there are disease causing agents called pathogens. However we can lessen crop losses and also can take different types of measures to overpower specific micro-organisms if plant diseases are efficiently diagnosed and distinguished early. Thus, plant pathologists have shared their knowledge with farmers through farming communities. That's why machine learning comes into the picture. To improve the diagnostic results, several studies on machine learning-based automated plant diagnosis have been conducted. Convolutional neural networks (CNNs) are widely perceived as one of the most promising classification techniques among machine learning fields. The most attractive advantage of CNN is their ability to acquire requisite features for the classification from the images automatically during their learning processes. Recently, CNN have demonstrated excellent performance in large scale general image classification tasks [5], traffic sign recognition [6], leaf classification [7], and so on. Computer vision, and object recognition techniques in particular, has made immense advancements in the past few years. The PASCAL VOC Challenge (Everingham et al., 2010)[8], and more recently the Large Scale Visual Recognition Challenge (ILSVRC) (Russakovsky et al., 2015) [9]based on the ImageNet dataset (Deng et al., 2009)[10] have been widely used as yardstick for a quantity of visualization-related problems in computer vision, including object classification. In 2012, a large, deep convolutional neural network achieved a top-5 error of 16.4% for the classification of images into 1000 possible categories (Krizhevsky et al., 2012)[11]. In the next 3 years, different types of advancements in deep convolutional neural networks lessened the error rate up to 3.57% (Krizhevsky et al., 2012 [12]; Simonyan and Zisserman, 2014 [13]; Zeiler and Fergus, 2014 [14]; He et al., 2015[15]; Szegedy et al., 2015[16]).

2 Related Works

Research in agriculture section is focused towards improvement of the standards and the proportion of the product at less wasting with more take. The standard of the agricultural product may be debased due to different types of plant diseases. These diseases are caused by pathogens such as fungi, bacteria and viruses. With the help of different types of applications, many systems have been suggested to solve or at least to lower the problems faced by the farmers, by harnessing the use of image processing and different types of automatic classification tools.

Suhaili Kutty et al.[17]discussed the process to classify *Anthracnose* and *Downey Mildew*, watermelon leaf diseases using neural network analysis. They used a digital camera with specific calibration procedure under controlled environment. Their classification is based on color feature extraction from RGB color model where the RGB pixel color indices have been extracted from identified ROI(region of interest).To reduce noise from images and for segmentation median filter is used. And for classification of the image, neural network pattern recognition toolbox is utilized. Proposed method achieved 75.9% of accuracy based on its RGB mean color component.

Sanjeev Sannaki et al[18] identify the disease with the help of image processing and AI techniques on images of grape plant leaf. In their proposed system , complex background with grape leaf image is taken as input. Noise is removed using anisotropic diffusion also the segmentation is done by k-means clustering. After segmentation, feature extraction is happened by computing Gray Level Co-occurrence Matrix. And finally classification takes place using Feed Forward Back Propagation Network classifier. Also Hue feature is used for more accurate result.

Akhtar et al[19]. have implemented the support vector machine(SVM) approach procedures for the classification and detection of rose leaf diseases as black spot and anthracnose. Authors have implied the threshold method for segmentation and Ostu's algorithm was mainly used to establish the threshold values. In this approach,different features of DWT, DCT and texture based eleven haralick features are extricated which are afterwards merged with SVM approach and predicts quite efficient accuracy value.

The study of Usama Mokhtar et al[20] incorporated with method that involves gabor wavelet transform technique to extract fitting features relevant to image of tomato leaf in coincidence with using Support Vector Machines (SVMs).They described technique of Tomato leaves diseases detection and diseases are: Powdery mildew and Early blight. Here gabor wavelet transformation is applied in feature extraction for feature vectors also in classification. Cauchy Kernel, Laplacian Kernel and Invmult Kernel methods are involved in SVM for output decision where tomato leaf infected with Powdery mildew or early blight. The proposed approach ensures excellent footnote with accuracy 99.5%.

Supriya et al[21] worked with the cotton leaves. They first captured the affected leaf and then pre-process converting into other color space. They also used Otsu's global thresholding method during segmentation. Also color-co-occurrence method is used for extracting different features such as color and texture. Multi SVM(Multi Support Vector Machine) classifier is used for detecting the diseases.

Ms. Kiran R. Gavhale et al. [22] presented number of image processing techniques to extract diseased part of leaf. For Pre-processing, Image enhancement is completed using DCT domain and thus color space conversion is done. After that segmentation is done with the help of k-means clustering. Feature extraction is done using GLCM Matrix. For classifying canker and anthracnose disease of citrus leaf, the use of SVM with radial basis kernel and polynomial kernel is done.

N.J. Janwe and Vinita Tajane [23] suggested for their medical plants disease identification using Canny Edge detection algorithm, Histogram Analysis and CBIR. The identification of medical plants according to its edge features. The leaf image converts to gray scale and calculate the edge histogram. The algorithm that purposed is canny edge detection.

3 Research Methodology

3.1 Dataset

We use PlantVillage Dataset for completing this classifier. We inspect total 54,306 images of plant leaves, which have a spread of 38 class labels alloted to them. Each class label is a crop-disease pair, and we ensure an attempt to estimate the crop-disease pair given just the picture of the plant leaf. In all the methods used in this paper, we reduce the sizes of the images upto 256x256 pixels, and we carry out both the model optimization and predictions on these downscaled images. Across all our experiments, we work with the colored version of the whole PlantVillage dataset

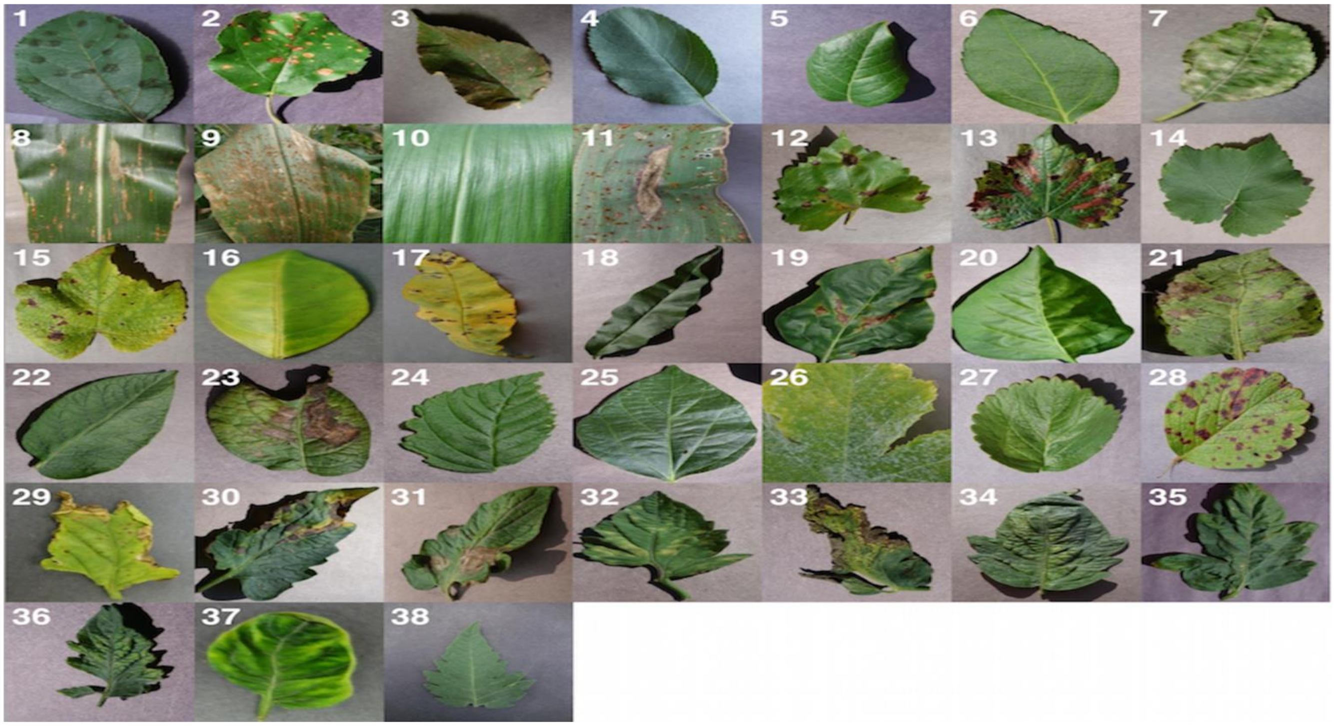

Fig.1. Example of leaf images from the Plant Village dataset, representing every crop-disease pair used .[24]

(1) Apple Scab, *Venturiain inaequalis* **(2) Apple Black Rot**, *Botryosphaeria obtuse* **(3) Cedar Apple Rust**, *Gymnosporangium juniperi-virginianae* **(4) Apple healthy,** *Malus* **(5) Blueberry healthy,** *Vaccinium sect.Cyanococcus* **(6) Cherry healthy,** *Prunus avium***(7) Cherry Powdery Mildew**, *Podoshaera clandestine* **(8) Corn Gray Leaf Spot**, *Cercospora zeae-maydis* **(9) Corn Common Rust**, *Puccinia sorghi* **(10) Corn healthy,** *Zea mays subsp.mays* **(11) Corn Northern Leaf Blight**, *Exserohilum turcicum* **(12) Grape Black Rot**, *Guignardia bidwellii*, **(13) Grape Black Measles(Esca)**, *Phaeomoniella aleophilum, Phaeomoniella chlamydospore* **(14) Grape Healthy,** *Vitis* **(15) Grape Leaf Blight**, *Pseudocercospora vitis* **(16) Orange Huanglongbing**(Citrus Greening), *Candidatus Liberibacter spp.* **(17) Peach Bacterial Spot**, *Xanthomonas campestris* **(18) Peach healthy,** *Prunus persica* **(19) Bell Pepper Bacterial Spot**, *Xanthomonas campestris* **(20) Bell Pepper healthy,** *Capsicum annuum Group* **(21) Potato Early Blight**, *Alternaria solani* **(22) Potato healthy,** *Solanum tuberosum* **(23) Potato Late Blight**, *Phytophthora infestans* **(24) Raspberry healthy,** *Rubus idaeus* **(25) Soy bean healthy,** *Glycine max* **(26) Squash Powdery Mildew**, *Erysiphe cichoracearum* **(27) Strawberry Healthy,** *Fragaria x ananassa* **(28) Strawberry Leaf Scorch**, *Diplocarpon earlianum* **(29)Tomato Bacterial Spot**, *Xanthomonas campestris pv.vesicatoria* **(30) Tomato Early Blight**,*Alternaria solani* **(31) Tomato Late Blight**, *Phytophthora infestans***(32)Tomato Leaf Mold**, *Passalora fulva* **(33)Tomato Septoria Leaf Spot**,*Septoria lycopersici* **(34) Tomato Two Spotted Spider Mite**, *Tetranychus urticae* **(35) Tomato Target Spot**, *Corynespora cassiicola* **(36) Tomato Mosaic (37) Tomato Yellow Leaf Curl (38) Tomato healthy,** *Solanum lycopersicum*

3.2 Measurement of Performance

To have a proper sense of how our working will perform on newly unseen data, and also to remain a track of if any of our approaches are overfitting with the new data, we go through all our experiments across a whole range of train-test set splits, namely 80–20 (80% of the whole dataset used for training, and 20% for testing), 60–40 (60% of the whole dataset used for training, and 40% for testing), 40–60 (40% of the whole dataset used for training, and 60% for testing), 20–80 (20% of the whole dataset used for training, and 80% for testing).

3.3 Approach

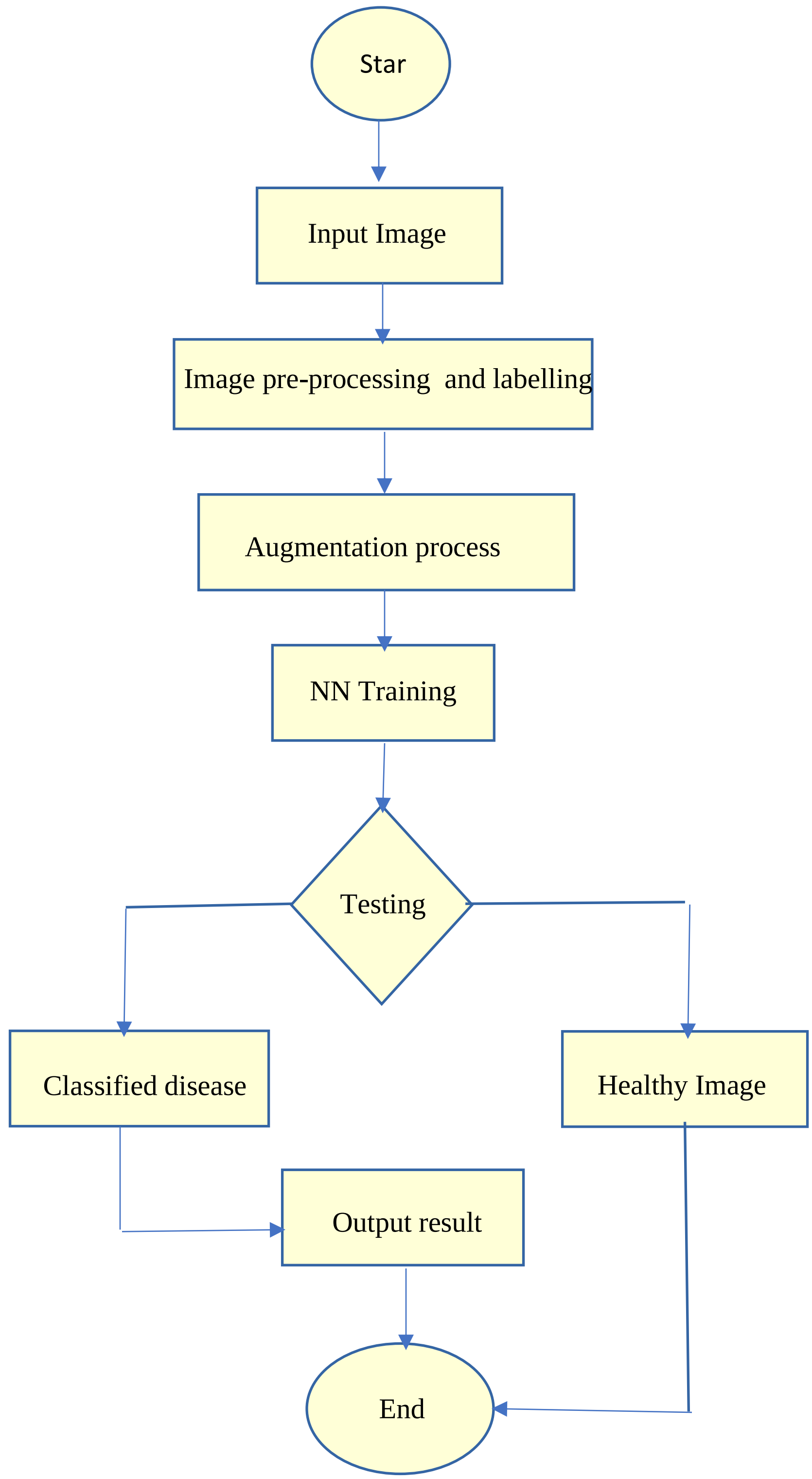

Fig.2. Flowchart of the entire work.

We evaluate the applicability of deep convolutional neural networks for the classification problem described above. We focus on two popular architectures, namely *InceptionV3 [25]*, and *Xception [26]*.
To summarize-

1. Choice of deep learning architecture:
   I. *InceptionV3*
   II. *Xception*

2.Choice of training mechanism:
   i. *Transfer Learning.*
   ii. *Training from scratch*

3.Choice of dataset:
   i. *Color*

4.Choice of training-testing set distribution:
   i. *Train:80%, Test:20%*
   ii. *Train:60%, Test:40%*
   iii. *Train:40%, Test:60%*
   iv. *Train:20%, Test:80%*

To enable a fair comparison between the results of all the experimental configurations, we also tried to standardize the hyper-parameters across all the experiments, and we used the following hyper-parameters in all of the experiments:

- Base learning rate:0.001
- Batch size:32
- Default Image Size: tuple(256,256)
- Epoch:100
- Depth:3
- Optimizer: Adam

All the above experiments were conducted using Keras , which is a fast, open source framework for deep learning. The basic results, such as the overall accuracy can also be replicated using a standard instance of Keras.

3.4 RESULTS

The overall accuracy we obtained on the PlantVillage dataset varied from (*Train from scratch* **80.28%**(epoch=25,Optimizer=Adam) to **92.04%**(epoch=150,Optimizer=Adam)) and (*Train using InceptionV3* **85.53%**(epoch=25,Optimizer=Adam)).Also using Xception model the accuracy varies from **98.625%- 99.22%**.

Table 1. Different splits of test and train of dataset using Xception and accuracy at the end of 100 epochs

Model	Accuracy	Size
Xception(epoch-100) **0.2%Test & 0.8% Train**	99.13%	88mb
Xception(epoch-100) **0.4%Test & 0.6% Train**	99.22%	,,
Xception(epoch-100) **0.6%Test & 0.4% Train**	98.625%	,,
Xception(epoch-100) **0.8%Test & 0.2% Train**	98.32%	,,

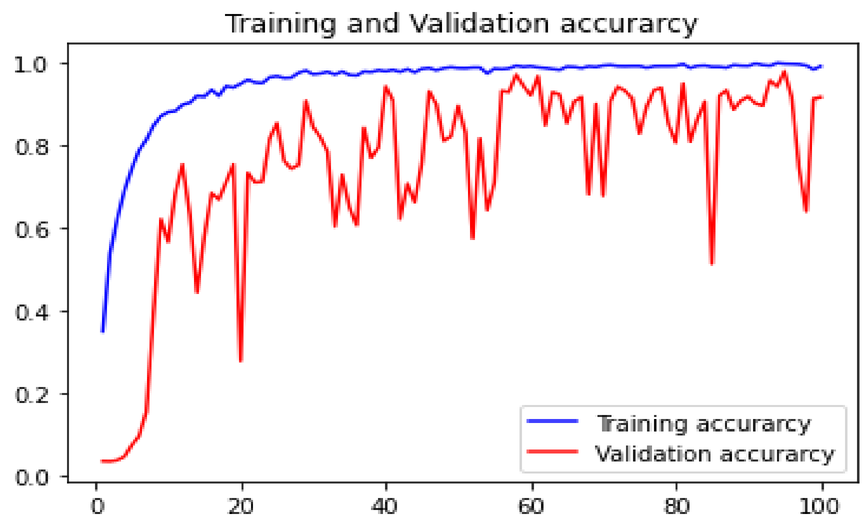

Fig.3. Training and Validation accuracy using training from Xception (0.6%Test&0.4%Train)

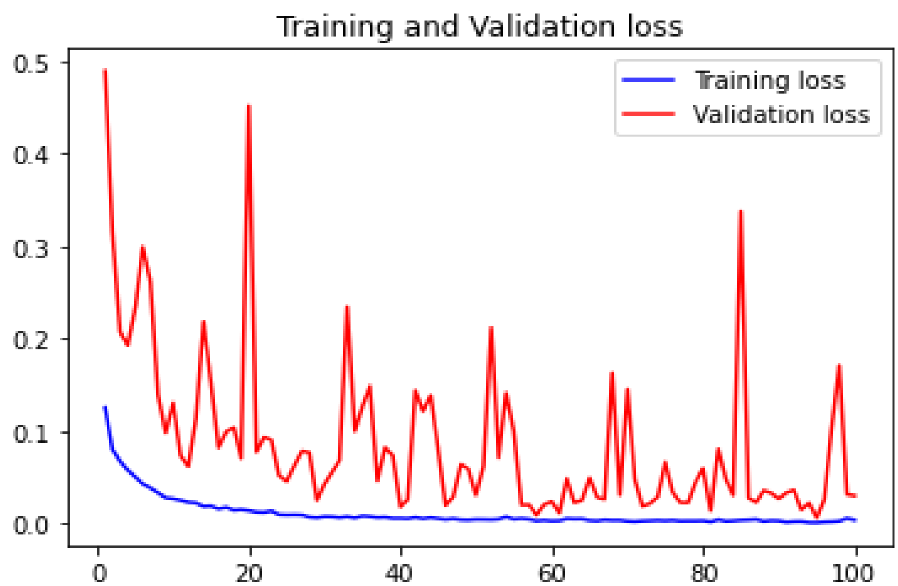

Fig.4. Training and Validation loss using training from Xception(0.6%Test&0.4%Train)

4 Conclusion

Pre-trained models have been greatly used in machine learning and computer vision applications also including plant disease identification. The achievements of convolutional neural networks in object recognition and image classification has made immense advancement in the past few years. The main purpose of this system is to improve the efficiency of the automatic plant disease detection. Our above mentioned results show that a large, deep convolutional neural network can achieve significant results on a highly challenging dataset with the help of purely supervised learning. Experimental results show that the proposed system can successfully detect and classify the plant disease with accuracy of 99.22%. In future work, we will extend our database for more plant disease identification and use large number of data as training data as training purpose in classification. As we increase the training data, the accuracy of the system will be high and then we can compare the accuracy rate and speed of system.

References

1. Sastry, K.S.: Plant Virus and Viroid Diseases in the Tropics, vol. II. Springer, Heidelberg (2013)
2. Tai, A. P., Martin, M. V., and Heald, C. L. (2014). Threat to future global food security from climate change and ozone air pollution. *Nat. Clim. Chang* 4, 817–821. doi: 10.1038/nclimate2317
3. Report of the Plenary of the Intergovernmental Science-PolicyPlatform on Biodiversity Ecosystem Services on the work of its fourth session (2016). *Plenary of the Intergovernmental Science-Policy Platform on Biodiversity and Ecosystem Services Fourth session*. Kuala Lumpur. Available online at: http://www.ipbes.net/sites/default/files/downloads/pdf/IPBES-4-4-19-Amended-Advance.pdf, last accessed 2021/01/04.
4. Strange, R. N., and Scott, P. R. (2005). Plant disease: a threat to global food security. *Phytopathology* 43, 83–116. doi: 10.1146/annurev.phyto.43.113004.133839
5. Krizhevsky, A., Sutskever, I., Hinton, G.E.: ImageNet classification with deep convolutional neural networks. In: Advances In Neural Information Processing Systems,pp. 1–9 (2012)
6. Hall, D., McCool, C., Dayoub, F., Sunderhauf, N., Upcroft, B.: Evaluation of features for leaf classification in challenging conditions. In: 2015 IEEE Winter Conference on Applications of Computer Vision, pp. 797–804 (2015)
7. Jin, J., Fu, K., Zhang, C.: Traffic sign recognition with hinge loss trained convolutional neural networks. IEEE Trans. Intell. Transp. Syst. **15**, 1991–2000 (2014)
8. Everingham, M., Van Gool, L., Williams, C. K., Winn, J., and Zisserman, A. (2010). The pascal visual object classes (voc) challenge. *Int. J. Comput. Vis.* 88, 303–338. doi: 10.1007/s11263-009-0275-4
9. Everingham, M., Van Gool, L., Williams, C. K., Winn, J., and Zisserman, A. (2010). The pascal visual object classes (voc) challenge. *Int. J. Comput. Vis.* 88, 303–338. doi: 10.1007/s11263-009-0275-4
10. Deng, J., Dong, W., Socher, R., Li, L.-J., Li, K., and Fei-Fei L. (2009). “Imagenet: A large-scale hierarchical image database,” in *Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on. (IEEE).*
11. Krizhevsky, A., Sutskever, I., and Hinton, G. E. (2012). “Imagenet classification with deep convolutional neural networks,” in *Advances in Neural Information Processing Systems*, eds F. Pereira, C. J. C. Burges, L. Bottou, and K. Q. Weinberger (Curran Associates, Inc.), 1097–1105.
12. Krizhevsky, A., Sutskever, I., and Hinton, G. E. (2012). “Imagenet classification with deep convolutional neural networks,” in *Advances in Neural Information Processing Systems*, eds F. Pereira, C. J. C. Burges, L. Bottou, and K. Q. Weinberger (Curran Associates, Inc.), 1097–1105.
13. Simonyan, K., and Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition. arXiv:1409.1556.
14. Zeiler, M. D., and Fergus, R. (2014). “Visualizing and understanding convolutional networks,” in *Computer Vision–ECCV 2014*, eds D. Fleet, T. Pajdla, B. Schiele, and T. Tuytelaars (Springer), 818–833.
15. He, K., Zhang, X., Ren, S., and Sun, J. (2015). Deep residual learning for image recognition. arXiv:1512.03385.
16. Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., et al. (2015). “Going deeper with convolutions,” in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*.
17. Kutty, S. B., Abdullah, N. E., Hashim, H., Kusim, A. S., Yaakub, T. N. T., Yunus, P. N. A. M., & Abd Rahman, M. F. (2013, April). Classification of watermelon leaf diseases using neural network analysis. In *2013 IEEE Business Engineering and Industrial Applications Colloquium (BEIAC)* (pp. 459-464). IEEE.

18. Sannakki, S. S., Rajpurohit, V. S., Nargund, V. B., & Kulkarni, P. (2013, July). Diagnosis and classification of grape leaf diseases using neural networks. In *2013 Fourth International Conference on Computing, Communications and Networking Technologies (ICCCNT)* (pp. 1-5). IEEE.
19. Akhtar, A., Khanum, A., Khan, S. A., & Shaukat, A. (2013, December). Automated plant disease analysis (APDA): performance comparison of machine learning techniques. In *2013 11th International Conference on Frontiers of Information Technology* (pp. 60-65). IEEE.
20. Mokhtar, U., Ali, M. A., Hassenian, A. E., & Hefny, H. (2015, December). Tomato leaves diseases detection approach based on support vector machines. In *2015 11th International Computer Engineering Conference (ICENCO)* (pp. 246-250). IEEE.
21. Patki, S. S., & Sable, G. S. (2016). Cotton leaf disease detection & classification using multi SVM. *International Journal of Advanced Research in Computer and Communication Engineering*, *5*(10), 165-168.
22. Gavhale, K. R., & Gawande, U. (2014). An overview of the research on plant leaves disease detection using image processing techniques. *IOSR Journal of Computer Engineering (IOSR-JCE)*, *16*(1), 10-16.
23. Tajane, Vinita, and N. J. Janwe. 2014. Medicinal Plants Disease Identification Using Canny Edge Detection Algorithm, Analysis and CBIR. *International Journal of Advance Research in Computer Science and Software Engineering*. 4(6): 530-536.
24. Mohanty, S. P., Hughes, D. P., & Salathé, M. (2016). Using deep learning for image-based plant disease detection. Frontiers in plant science, 7, 1419. (pg-3)
25. Szegedy, C., Ioffe, S., Vanhoucke, V., & Alemi, A. (2017, February). Inception-v4, inception-resnet and the impact of residual connections on learning. In *Proceedings of the AAAI Conference on Artificial Intelligence* (Vol. 31, No. 1).
26. Chollet, F. (2017). Xception: Deep learning with depthwise separable convolutions. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 1251-1258).