e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
Volume:03/Issue:06/June-2021 Impact Factor- 5.354 www.irjmets.com

LEAF DISEASE DETECTION USING CNN(CONVOLUTIONAL NEURAL NETWORK)

Suraj Sontakke[*1], Satyam Deshmukh[*2], Laxman Nemane[*3], Ashwini Jadhav[*4]

[*1,3,4]Dept. Of Computer Science, Genba Sopanrao Moze College Of Engineering, Pune-411033, Maharashtra, India.

[*2]Assistant Professor, Dept. Of Computer Science And Engineering, Genba Sopanrao Moze College Of Engineering, Pune-411033, Maharashtra, India.

ABSTRACT

In India, agriculture has become an important source of economic development. The farmer selects a suitable crop based on the type of soil, weather condition of the location, and economic value. The agriculture industries started searching for new methods to increase the production of food because of the increasing population, changes in weather and instant Deep learning with convolutional neural networks (CNN) has achieved great success in the classification of various plant diseases. In this study, a variety of neuron-wise and layer-wise visualization methods are applied using a CNN, trained with a publicly available plant disease image dataset. We showed that neural networks can capture the colors and textures of lesions specific to respective diseases upon diagnosis, which resembles human decision-making. The results provide an impetus for the CNN *black box* users in the field of plant science to better understand the diagnosis process and lead to further efficient use of deep learning for plant disease diagnosis.

Keywords: Leaf Disease Detection Using CNN(Convolutional Neural Network).

I. INTRODUCTION

In India, agriculture has become an important source of economic development. Farmer selects the suitable crop based on the type of soil, weather condition of the location, and economic value. The agriculture industries started searching for new methods to increase the production of food because of the increasing population, changes in weather, and instability in politics. This makes researchers search a new efficient and precise technologies for high productivity. Farmers can collect the information and data by use of precision agriculture in information technology to make the best decision on high output from the farm. Precision agriculture is a new technology, which provides advanced techniques to improve farm output. By utilizing these advanced technologies, it is possible to achieve economic growth in agriculture. Precision agriculture can be used for many applications like pest detection in plants, weed detection, yield production of crops and plant disease detection, etc. A farmer uses pesticides to control pests, prevent diseases, and increase crop yield. The diseases in the crop are creating the problem of low production and economic losses to farmers and agricultural industries. Therefore identification of disease and its severity based as become necessary. Disease identification in the plant is most important in a successful farming system. In general, a farmer recognizes the symptoms of disease in plants by using naked-eye observations and this requires continuous monitoring. However, this process is more expensive in large plantations and sometimes this may be less accurate. In some countries like India, farmers may have to show the specimen to experts, this makes it time-consuming and more expensive. The following sections in this paper contain general steps of plant diseases detection system and survey on machine learning classification techniques used to recognize and classify plant diseases.

II. METHODOLOGY

Pre-processing and Training the model (CNN): The database is Pre-processed such as Image reshaping, resizing, and conversion to an array form. Similar processing is also done on the test image. A database consisting of about 32000 different plant species is obtained, out of which any image can be used as a test image for the software. The training database is used to train the model (CNN) so that it can identify the test image and the disease it has. CNN has different layers that are Dense, Dropout, Activation, Flatten, Convolution2D, MaxPooling2D. After the model is trained successfully, the software can identify the disease if the plant species is contained in the database. After successful training and pre-processing, a comparison of the test image and trained model takes place to predict the disease. Database collection: Initial step for any image processing-based project is acquiring a proper database that is valid. Most of the time the standard database is preferred but in certain circumstances, we do not get a proper database. So in such conditions, we can collect the images and can form our database. The database is accessed from crowd AI which is a plant disease classification challenge. Data available here is not labeled. So the first task is to clean and label the database. There is a huge database so basically, the images with better resolution and angle are selected. After the selection of images, we should have deep knowledge about the different leaves and the disease they have. Huge research is done from the plant village organization repository. Different types of plant images are studied and corresponding. After a detailed study, labeling is done by segregating the images and with different diseases.

The disease of different plants database:

- Apple black spot
- Apple broad leaf spot
- Apple needle leaf spot
- Apple normal
- Bell paper normal
- Cherry normal
- Cherry powder normal
- Corn blight
- Corn rust

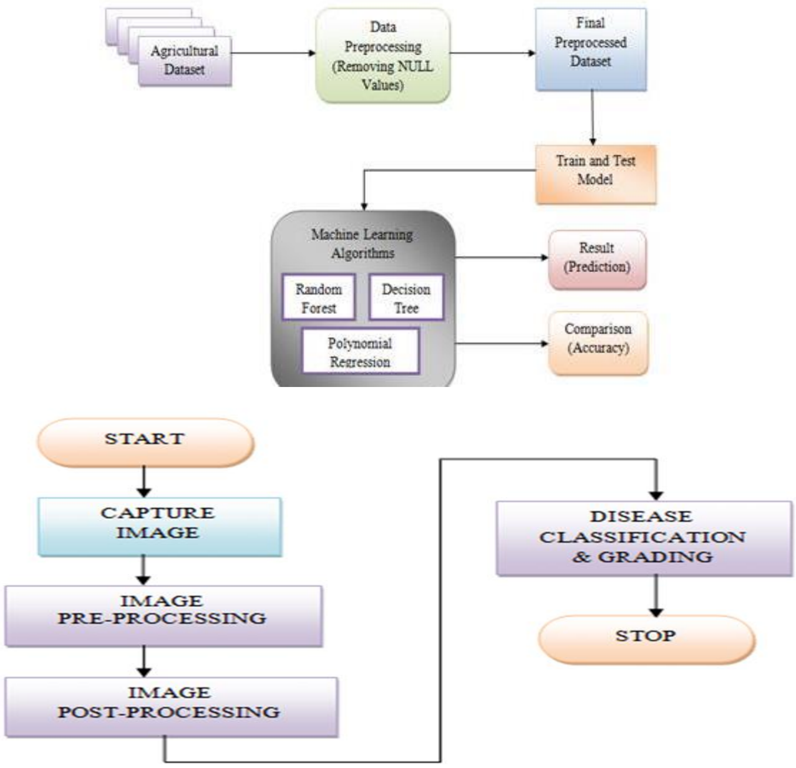

Figure.1: Flowchart for disease detection and grading in pomegranate

III. MODELING AND ANALYSIS

This section describes the steps involved in creating and deploying the classifier. Classification by CNN is divided into three phases that tackle separate tasks. All work involved in this research was completed on one machine, with specifications listed in Table 1.

A. Data Acquisition

All Potato and Tomato imagery derives from 'The PlantVillage Dataset', an open-access repository that contains in total 54,323 images. All Rice imagery originates from the "Rice Diseases Image Dataset" Kaggle dataset. For each species, a select number of classes are chosen, with details viewable in Table II. All images are captured in a controlled environment. Due to this, model bias is expected. To access this, a test dataset containing 50 images, sourced from Google is also established. These images contain additional plant anatomy, in-field background data and varying stages of the disease.

TABLE I. MACHINE SPECIFICATIONS

Hardware & Software	Characteristics
Memory	8.0GB
Processor	Intel(R) Core™ i5-9300H CPU @ 2.40GHz
Graphics	NVIDIA GeForce RTX 2060 6GB GDDR6
Operating system	Windows 10 Home 64

B. Data Pre-Processing

The dataset is divided into 80% for training and 20% for validation. First, augmentation settings are applied to the training data. These are generated 'on the fly, with each operation carrying a weighted probability of appearing in each epoch. The settings applied include flipping (random), padding mode (reflection) and zoom with crop (scale = (1.0,1.5)). 'Zoom with crop' was later omitted after discovering that it had inappropriately cropped areas of the infected leaf. Finally, all images are re-sized and normalized. Resizing is carried out using a compress function, to 150 x 150. As a pre-trained model is used, the RBG ImageNet statistics are used to normalize. A sample of the final pre-processed images is viewable in Fig.1.

C. Classification by CNN

1) Phase One – Trialling of Image size

Phase one aims to investigate the effect that image size has on model performance. In total, five image sizes are tested ranging from 150 x 150 to 255 x 255. To begin, the Resnet34 pre-trained weights are downloaded. As a default of transfer learning, all layers except the final two layers are frozen. These contain new weights and are specific to the plant disease classification task. Freezing allows these layers to be disease separately trained, without back propagating the gradients. In exactly this way, the 1cycle policy is used to train the final layers. With this complete, the remaining layers are released. To aid the fine-tuning process, a plot displaying learning rate vs loss is generated and analyzed. From this, suitable learning is selected, and the model is run. With results recorded, the model is re-created to the additional four image sizes (Table III.). All steps remain consistent in each trial including the learning rate.

TABLE II. DATASET USED FOR CLASSIFICATION

Species	Class	No. of Images
Potato	Early blight	1000
Potato	Late blight	1000
Potato	Healthy	152
Tomato	Bacterial Spot	2119
Tomato	Leaf Mold	952
Tomato	Mosaic Virus	160
Tomato	Healthy	1000
Rice	Brown Spot	523
Rice	Leaf Blast	779
Rice	Healthy	1000

TABLE III. IMAGE SIZE TRIAL INFORMATION

Trial	Image Size	No. Epochs	Learner Rate
1	150 x 150	4	1e-05 and 1e-04
2	195 x 195	4	1e-05 and 1e-04
3	224 x 224	4	1e-05 and 1e-04
4	244 x 244	4	1e-05 and 1e-04
5	255 x 255	4	1e-05 and 1e-04

2) Phase Two – Model Optimisation

Using the most suitable image size, the ResNet34 model is optimized. To further improve the model's performance, additional augmentation settings are added (Fig. 2). Operations include brightness changes (0.4,0.7) and warp (0.5).

Next, the final two layers are isolated and trained at the default learning rate. With this complete, fine-tuning is performed, running multiple trials to test a series of learning rates and the number of epochs.

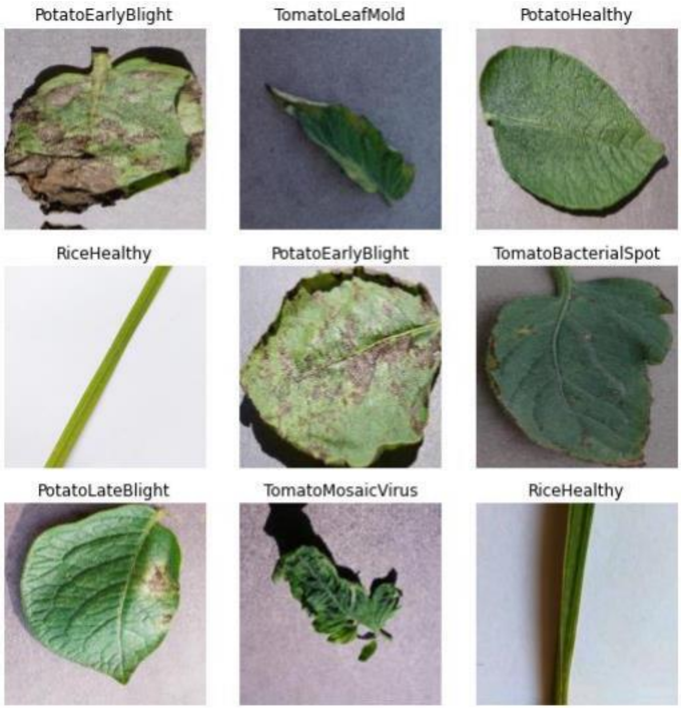

Fig.2: Pre-processed images - Phase One *augmentation settings = flipping (random), padding mode (reflection)*

IV. RESULTS AND DISCUSSION

1) Phase One – Trialling of Image Size: The results of Phase One prove that it is possible to achieve accuracy and F1 score of greater than 90% for image sizes 155 x 155 to 255 x 255. As expected, an increase in image size not only improves feature extraction but also increases running time (Table IV.). This initial analysis produced excellent results. As previously stated, the model would be accepted if it reached an accuracy of at least 80%. Even at this early stage, results far exceed the acceptance criteria. To achieve this result, each model was passed a range of learning rates from 1e-05 to 1e-04 and run for 4 epochs. Overall, image size 244 produced the best results including the highest accuracy and F1score. Although the literature suggests image size 224 x 224 to be suitable for plant disease classification tasks (10a), this model appears to marginally benefit from increased image size. For these reasons, image size 244 was chosen for the remainder of this research.

2) Phase Two – Model Optimisation: Before fine-tuning, the model attained an accuracy of 0.9465 and an F1 score of 0.9359. (Fig. 3) To aid fine-tuning, a plot depicting learning rate (logarithmic scale) v loss was analyzed (Fig. 4). This demonstrates a relatively low loss between learning rates 1e-06 to 1e-04. As the learning rate increases past 1e-04 however, a dramatic increase in loss is experienced. These facts considered, several trials testing learning rate were carried out. A learning rate range of 1e-05 to 1e-04 produced the best results. By fine-tuning this hyperparameter, a slight increase in accuracy (1.5%) and F1-Score (1.3%) was accomplished. On the final epoch, however, the closing training and validation values indicate that the model may be slightly underfitting (Fig. 5). To correct this, the number of epochs was increased systematically. At Approximately the 10th epoch, there was an evident improvement to the fit of the model. A final reading presented an overall improvement of 2.8% in accuracy and 3.1 % in F1-score (Fig. 6). As stated earlier, the validation dataset consists of a very specific composition; one leaf and a plain background. For an accurate reading, akin to those stated in this section, the use of the classifier should mimic this image layout.

e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
Volume:03/Issue:06/June-2021 Impact Factor- 5.354 www.irjmets.com

Test	Image size	Train Loss	Valid Loss	Accuracy	F1 Score	Time (hours)
1	155	0.1660	0.1222	0.9557	0.9439	2:83
2	195	0.1588	0.1150	0.9585	0.9460	3.62
3	224	0.1778	0.1256	0.9522	0.9359	4.29
4	244	0.1310	0.1153	0.9603	0.9450	5.20
5	255	0.1607	0.1249	0.9562	0.944	5.42

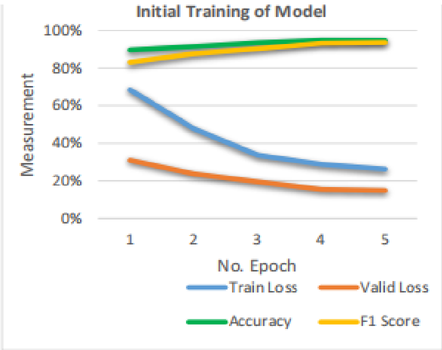

Fig. 3. Training the final layers (lr=1e-3)
Before fine-tuning the model attained an accuracy of 0.9465 and F1 score of 0.9359.

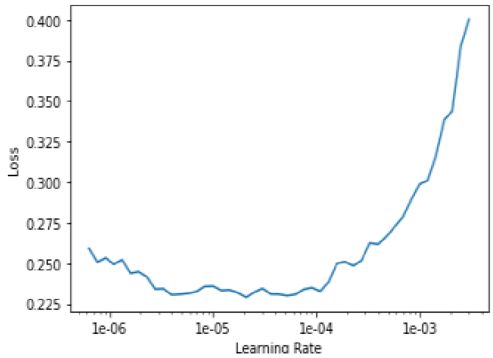

Fig. 6. Learning rate v loss
Used to guide the fine-tuning process. As the learning rate increases past 1e-04, a dramatic increase in loss is experienced.

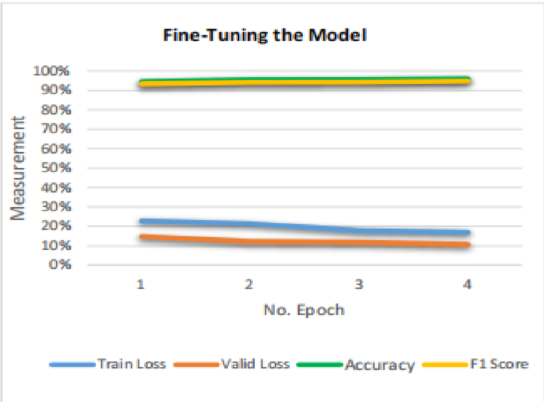

Fig. 4. Fine-tuning the model,
learning rate range = 1e-05, 1e,04, epochs = 4, Signs of underfitting apparent

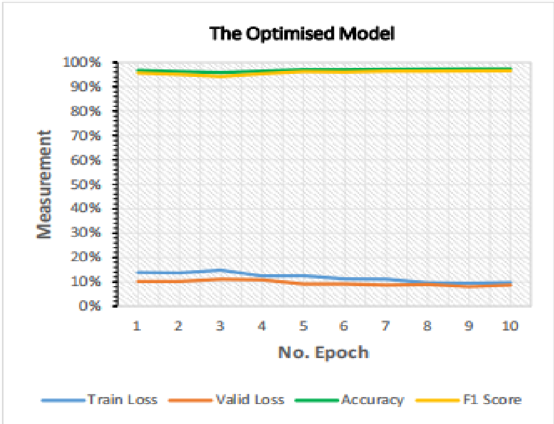

Fig. 5. The fianl optimized model,
learning rate range = 1e-05, 1e,04, epochs = 10

3) Phase Three– Visualisasations*:* An analysis of heat maps reveals the inner workings of the CNN. Colour, shape, and texture appear to be important factors in working to extract plant disease features (Fig.7, Fig.8). Colour appears to be especially crucial, helping to differentiate similar diseases, by adding an extra dimension of characterization. This explains the importance of RGB data to disease classification tasks, as was highlighted earlier [10, 20]. For all three species, CNN shows effectiveness in recognizing features. This is also true for rice disease classes, which contain smaller, and more difficult to distinguish symptoms. The confusion matrix presented in Fig. 10 lists the validation dataset results. Overall, no errors were recorded in any Potato or Tomato classes. Rice as a species, performed poorly, suggesting that there may be an underlying issue with the data. Rice Brown Spot was the highest misclassified class. 13.9% of these images were incorrectly classified as Healthy and a further 9.9% were misclassified as Rice Leaf Blast. A clear symptom of the brown spot is irregular dark spots. While this may be mistaken for similar lesions in leaf blasts, there should be overlapping characteristics with healthy samples. On average, 12.65% of each Rice class was misdiagnosed. To investigate this matter further, the misclassified images were plotted and sorted respectively to loss (Fig. 11). A closer inspection reveals that the quality of several images is questionable. Even to the skilled eye, an accurate diagnosis based on these images would be challenging. This data may have been mislabeled or is simply a poor class representation. As such data is not beneficial to the classifier, it should not be included in the training dataset. As expected, the model suffers a significant drop in accuracy when in-field imagery is tested. Out of 50 images, only 44% were accurately diagnosed (Fig. 12). This is due to a combination of factors; which augmentation could not overcome; including new plant anatomy and alternative background data. As the model was not trained on such data, adapting to such circumstances is extremely difficult. Diversifying the training data to include imagery that has been captured in this uncontrolled environment could stand to strengthen the model immensely. As highlighted earlier, there is a current lack of 'in-field plant disease imagery available. These results signify the importance of developing such resources. Finally, the model was deployed on Render to create a web application (Fig. 9). This provides the user with a live disease classification service and reflects the capabilities and limitations of both the validation and test dataset, which have been discussed in this section. The application is available to use on the following link:

Fig. 7. Heat map example 1

https://plants.onrender.com.

Fig.3: Creation of the web application on Render

V. CONCLUSION

To prevent losses, smallholder farmers are dependent on a timely and accurate crop disease diagnosis. In this study, a pre-trained Convolutional Neural Network was fine-tuned, and the model was deployed online. The final result was a plant disease detection app. This service is free, easy to use, and requires just a smartphone and internet connection. Thus, the user's needs as defined in this paper have been fulfilled. A thorough investigation exposes the capabilities and limitations of the model. Overall, when validated in a controlled environment, an accuracy of 97.2% is presented. This achieved accuracy depends on several factors including the stage of disease, disease type, background data, and object composition. Due to this, a set of user guidelines would be required for commercial use, to ensure the stated accuracy is delivered. As the model was trained using a plain background and singular leaf, imitation of these features is best. Augmentation and transfer learning, in this case, proved beneficial to the model, helping the CNN to generalize more reliability. While this improved the model's ability to extract features, it was not enough when the model was presented with 'in field' imagery. In this case, the classifier ranked an accuracy of just 44% Above all, this highlights the importance of diversifying the training dataset to include alternative background data, additional plant anatomy, and varying stages of the disease. Overall, this study is conclusive in demonstrating how CNNs may be applied to empower small-holder farmers in their fight against plant disease. In the future, work should be focused on diversifying training datasets and also on testing similar web applications in real-life situations. Without such developments, the struggle against plant disease will continue.

VI. REFERENCES

[1] M. Hunter, R. Smith, M. Schipanski, L. Atwood and D. Mortensen, "Agriculture in 2050:Recalibrating Targets for Sustainable Intensification", *BioScience*, vol. 67, no. 4, pp. 386-391, 2017. Available: 10.1093/Biosci/bix010.

[2] "PlantVillage", Plantvillage.psu.edu, 2020. [Online]. Available: https://plantvillage.psu.edu/. [Accessed: 31- Jan- 2020].

[3] D. Klauser, "Challenges in monitoring and managing plant diseases in developing countries", Journal of Plant Diseases and Protection, vol. 125, no. 3, pp. 235-237, 2018. Available: 10.1007/s41348-018-0145- 9.

[4] A.Muimba-Kankolongo, Food crop production by smallholder farmers in Southern Africa. Elsevier, 2018, pp 23-27.

[5] J. Boulent, S. Foucher, J. Théau and P. St-Charles, "Convolutional Neural Networks for the Automatic Identification of Plant Diseases", *Frontiers in Plant Science*, vol. 10, 2019. Available: 10.3389/fpls.2019.00941."iNaturalist", iNaturalist, 2020. [Online]. Available: https://www.inaturalist.org/. [Accessed: 30- Jan- 2020].

[6] I. Ltd., "Success-stories - PlantSnap: Training the world's largest plant recognition classifier. | Imagga Technologies Ltd.", Imagga.com, 2020. [Online]. Available: https://imagga.com/success-stories/plantsnap case-study. [Accessed: 31- Jan- 2020].

[7] "Smartphone users worldwide 2020 | Statista", Statista, 2020. [Online]. Available: https://www.statista.com/statistics/330695/number-ofsmartphone-users-worldwide/. [Accessed: 22- Apr- 2020].

[8] K. Liakos, P. Busato, D. Moshou, S. Pearson and D. Bochtis, "Machine Learning in Agriculture: A Review", *Sensors*, vol. 18, no. 8, p. 2674, 2018. Available: 10.3390/s18082674.

[9] Y.Toda and F. Okura, "How Convolutional Neural Networks Diagnose Plant Disease", *Plant Phenomics*, vol. 2019, pp. 1-14, 2019. Available: 10.34133/2019/9237136.

[10] Atabay, H. A. 2016b. A convolutional neural network with a new architecture applied to leaf classification. IIOAB J 7(5):226–331

[11] A. Krizhevsky, I. Sutskever, and G. Hinton, "ImageNet classification with deep convolutional neural networks", *Communications of the ACM*, vol. 60, no. 6, pp. 84-90, 2017. Available: 10.1145/3065386.

[12] S. Jadhav, "Convolutional Neural Networks for Leaf Image-Based

[13] Plant Disease Classification", *IAES International Journal of Artificial Intelligence (IJ-AI)*, vol. 8, no. 4, p. 328, 2019. Available: 10.11591/Ajai.v8.i4.pp328-341.

[14] C. Szegedy and S. Ioffe, "Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift", *Arxiv.org*, 2015. [Online]. Available: https://arxiv.org/pdf/1502.03167v3.pdf. [Accessed: 10- Apr- 2020].

[15] D. Wu, Y. Wang, S. Xia, J. Bailey, and X. Ma, "Skip Connections Matter: On the Transferability of Adversarial Examples Generated with ResNets", *arXiv.org*, 2020. [Online]. Available: https://arxiv.org/abs/2002.05990. [Accessed: 10- Apr- 2020].

[16] M. Ji, L. Zhang and Q. Wu, "Automatic grape leaf diseases identification via UnitedModel based on multiple convolutional neural networks", *Information Processing in Agriculture*, 2019. Available: 10.1016/j.inpa.2019.10.003.

[17] X. Zhang, Y. Qiao, F. Meng, C. Fan, and M. Zhang, "Identification of Maize Leaf Diseases Using Improved Deep Convolutional Neural Networks", *IEEE Access*, vol. 6, pp. 30370-30377, 2018. Available: 10.1109/access.2018.2844405.

[18] C. Buche, S. Wallelign and M. Polceanu, "Soybean Plant Disease Identification Using Convolutional Neural Network", 2018. [Accessed 22 April 2020].

[19] Isleib, "Signs and symptoms of plant disease: Is it fungal, viral or bacterial?", MSU Extension, 2012. [Online]. Available: https://www.canr.msu.edu/news/signs_and_symptoms_of_plant_disease_is_it_fungal_viral_or_bacterial. [Accessed: 11- Apr- 2020].

[20] P. Sharma, Y. Berwal and W. Ghai, "Performance analysis of deep learning CNN models for disease detection in plants using image segmentation", Information Processing in Agriculture, 2019. Available: 10.1016/j.inpa.2019.11.001.

[21] Padmavathi and K. Thangadurai, "Implementation of RGB and Grayscale Images in Plant Leave Disease Detection – Comparative Study", *Indian Journal of Science and Technology*, vol. 9, no. 6, 2016. Available: 10.17485/ijst/2016/v9i6/77739.

[22] S. Sankaran, A. Mishra, R. Ehsani and C. Davis, "A review of advanced techniques for detecting plant diseases", Computers and Electronics in Agriculture, vol. 72, no. 1, pp. 1-13, 2010. Available: 10.1016/j.compag.2010.02.007.

[23] 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).

[24] S. Mohanty, D. Hughes and M. Salathé, "Using Deep Learning for Image-Based Plant Disease Detection", Frontiers in Plant Science, vol. 7, 2016. Available: 10.3389/fpls.2016.01419.

[25] S. Lee, H. Goëau, P. Bonnet and A. Joly, "New perspectives on plant disease characterization based on deep learning", Computers and Electronics in Agriculture, vol. 170, p. 105220, 2020. Available: 10.1016/j.compag.2020.105220.

[26] P. Pawara, E. Okafor, L. Schomaker, and M. Wiering, "Data Augmentation for Plant Classification", 2017. Available:https://www.researchgate.net/publication/319990090_Data_Augmentaction_for_Plant_Classification. [Accessed 11 April 2020].

[27] B. Škaloudová, V. Křivan and R. Zemek, "Computer-assisted estimation of leaf damage caused by spider mites", *Computers and Electronics in Agriculture*, vol. 53, no. 2, pp. 81-91, 2006. Available: 10.1016/j.compag.2006.04.002.

[28] Y. Oo and N. Htun, "Plant Leaf Disease Detection and Classification using Image Processing", *International Journal of Research and Engineering*, vol. 5, no. 9, pp. 516-523, 2018. Available: 10.21276/ijre.2018.5.9.4

[29] K. Lin, L. Gong, Y. Huang, C. Liu, and J. Pan, "Deep Learning-Based Segmentation and Quantification of Cucumber Powdery Mildew Using Convolutional Neural Network", *Frontiers in Plant Science*, vol. 10, 2019. Available: 10.3389/fpls.2019.00155.

[30] C. Tucker and S. Chakraborty, "Quantitative Assessment of Lesion Characteristics and Disease Severity Using Digital Image Processing", *Journal of Phytopathology*, vol. 145, no. 7, pp. 273-278, 1997. Available: 10.1111/j.1439-0434.1997.tb00400.x.