Internatonal Journal of Science And Engineering
ISSN (Online): 2454 - 2016
Volume 11 Issue 01 May 2025

DOI: 10.53555/ephijse.v11i1.310

PROACTIVE DIAGNOSIS OF MUNG BEAN LEAF DISEASES IN A CONTROLLED ENVIRONMENT USING MACHINE LEARNING TECHNIQUES

Akruti Naik[1] , Nirav Desai[1]

[1] D-UIAS AND D-SIM&C,Valsad, INDIA

Corresponding author:
**Email: akrutinaikdesai@gmail.com*

Abstract.

This paper puts forth a model that classifies Mung (Vigna mungo L.) leaves to check if they are fit as a fiddle or have caught some disease, using a Machine Learning algorithm. The dataset was made in a proper controlled setup, where each data item (image) shows just one leaf against a white background, collected from the South Gujarat side of India. Support Vector Machine (SVM) models were trained for doing this work, focusing on spotting three types of Mung leaf diseases, as well as when a leaf is perfectly healthy. The model nicely pulls out the detailed features related to different diseases. Experiment results show that the SVM gets an identification accuracy of 86.4% on the Mung leaf image dataset. If diseases are caught early, it can help our kisaans in increasing their yield. The main aim was to automatically identify Mung leaf diseases using advanced machine learning techniques and image data.

Keywords: *Mung leaf, Machine Learning, SVM.*

1. INTRODUCTION

Mungbean, or green gram as it's commonly known, stands out as a very nourishing legume crop. 1 Its high protein content and how easily it digests make it a prized pulse. 2 India is the top producer of mungbean, placing it as the third most significant pulse crop in the country, grown across roughly 34.50 lakh hectares and yielding a total of 15.91 lakh tonnes (Vinod et al., 2018). To keep up with the world's needs, boosting the current average global yield is a must (Nair et al., 2012). To help our farmers spot and stop Mung leaf diseases early on, automated methods are key. Machine Learning, which is part of Artificial Intelligence, is all about making predictions using algorithms that get better on their own through learning from data. 3 These algorithms create a model based on the data they're trained on, so they can then make predictions on new data. 4 In this study, we used Support Vector Machines (SVM) to sort mung leaves into two groups: healthy or diseased. Our model showed encouraging outcomes, and from what we know, it gives useful knowledge for detecting diseases in Mung leaves. Furthermore, the non-destructive nature of image-based analysis offers a significant advantage for continuous monitoring in agricultural fields. The development of such systems can contribute to reduced pesticide usage and more sustainable farming practices. Considering the prevalence of smartphone usage in rural India, deploying these models on mobile platforms could provide accessible and timely disease diagnosis for farmers.

2. Literature review

Mungbean yields face significant threats from several diseases, notably Cercospora Leaf Spot, Powdery Mildew, and the particularly devastating Yellow Mosaic Virus (MYMV), which can cause yield losses of 85-90% and affects not only mungbean but also other important legumes across India and the subcontinent since its first report in 1940 (Nariani, 1960; Verma and Malathi, 2003). Cercospora Leaf Spot (CLS), the most widespread fungal disease, can lead to annual yield reductions of up to 58% (Lal et al., 2001), typically appearing a month or so after planting with characteristic reddish-brown spots (Munjal et al., 1960). Powdery Mildew, caused by Erysiphe polygoni, can also significantly impact yields, with losses ranging from 20-40% during reproductive stages (Fernandez and Shanmugasundaram, 1998) and potentially reaching 55% in epidemic conditions (Poehlman, 1991). The distinct visual symptoms of each of these diseases on leaves offer a pathway for classification using deep learning algorithms (Sladojevic et al., 2016; Huang et al., 2007; Price et al., 1993; Ciresan et al., 2012). While early deep learning approaches for plant disease detection relied heavily on leaf images (Fuentes et al., 2017; Ramcharan et al., 2017; Ferentinos et al., 2018; Godliver et al., 2018), their success hinges on the availability of large, manually labeled datasets, which can be a significant bottleneck (Valle et al., 2017; Galdran et al., 2017; Asperti et al., 2017; Wang et al., 2017), particularly in specialized domains like plant pathology. Data augmentation techniques offer a solution to this limitation by generating new training samples through transformations (Simard et al., 1992; Simard et al., 2003), although these may require specialized handling for tasks beyond simple classification (Abadi et al., 2019; Bloice et al., 2017; Jung, 2017). Training deep Convolutional Neural Networks (CNNs) traditionally demands extensive datasets (Lee et al., 2015), prompting researchers to explore transfer learning to leverage existing knowledge on smaller datasets (Pan et al., 2010). Studies have shown promising results using transfer learning with various CNN architectures (Srdjan et al., 2016), achieving high accuracy in classifying multiple plant diseases. Furthermore, research comparing different CNN models has indicated the effectiveness of architectures like GoogleNet and AlexNet (Mohanty et al., 2016), VGG16 (Wang et al., 2017), and DenseNets (Too et al., 2018) in plant disease identification, with DenseNets showing strong performance with fewer parameters and less computational cost. Approaches incorporating transfer learning (Selvaraj et al., 2019) and specialized techniques like regression and focus loss functions (Zhong et al., 2020) have also demonstrated high accuracy in identifying specific plant diseases. Liu et al. (2017) also highlighted the benefits of data augmentation in mitigating overfitting.

3. Proposed Method

In this study, we put forward the application of Support Vector Machine (SVM) algorithms for the purpose of detecting diseases affecting mung leaves by leveraging machine learning. Support Vector Machines (SVM) are versatile models suitable for both classification and regression problems. The fundamental principle of this algorithm involves the identification of a separating line, or hyperplane when dealing with more complex data structures, that effectively distinguishes between the various categories present in the dataset. The overall architecture of the proposed Mung Leaf disease detection system is visually represented in Figure 1.

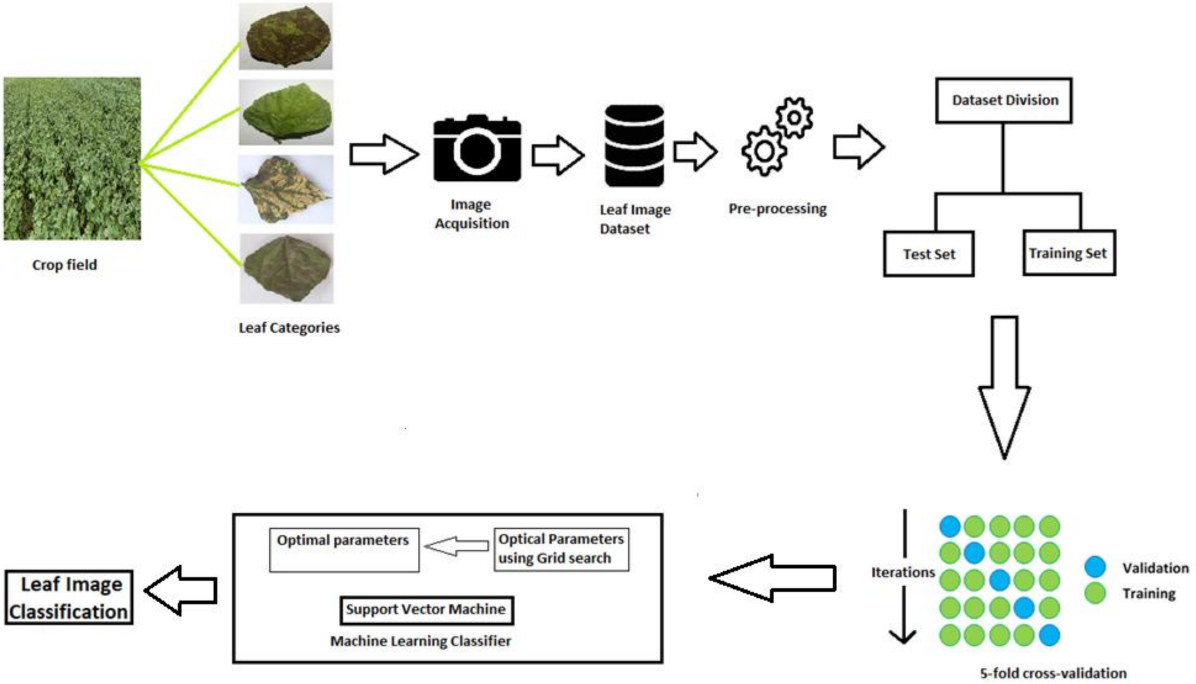

Figure. 1 The Structure of Mung Leaf Disease Detection

4. Data Collection and Feature Engineering

4.1 Data Collection

The dataset, for the most part, was created in a proper controlled setting, where every image shows just one mung leaf against a plain white background, so there's hardly any unwanted noise. Once we got the images, they were checked by hand to remove any that were the same and to make sure they were put in the right category in the dataset. The final dataset has 883 mung leaf images taken in this controlled way, and they are divided like this: Cercospora (224 images), Healthy (211 images), Powdery Mildew (225 images), and Yellow Mosaic (223 images). Each image was resized to 256 by 256 pixels. The leaves in the dataset are put into one of four types: Healthy, Cercospora, Yellow Mosaic, and Powdery Mildew.

The images of Mung leaves in 4 categories are shown in Figure 2.

(a) Yellow Mosaic Virus (b) Cercospora Leaf Spot (c) Powdery Mildew (d) Healthy

Figure. 2 All Categories (Controlled Environment)

4.2 Preprocessing

For this research, an SVM model has been developed to spot diseases. Initially, each image is loaded into a three-dimensional NumPy array and subsequently downsized to one-third of its initial dimensions specifically for training the SVM. Within the field of computer vision, the Histogram of Oriented Gradients (HOG) method is utilized for identifying objects. HOG functions as a feature descriptor by highlighting the form and arrangement of the subject within the image. A histogram is computed for every localized section of the image, recording the directions of the gradients and thereby supplying significant data for the classification task. By focusing on the shape and texture information encoded in the HOG features, the SVM model aims to accurately distinguish between healthy and diseased mung leaves. The initial step in our methodology involves reading the image data into a NumPy array, followed by a resizing operation to optimize the input for the SVM training process.

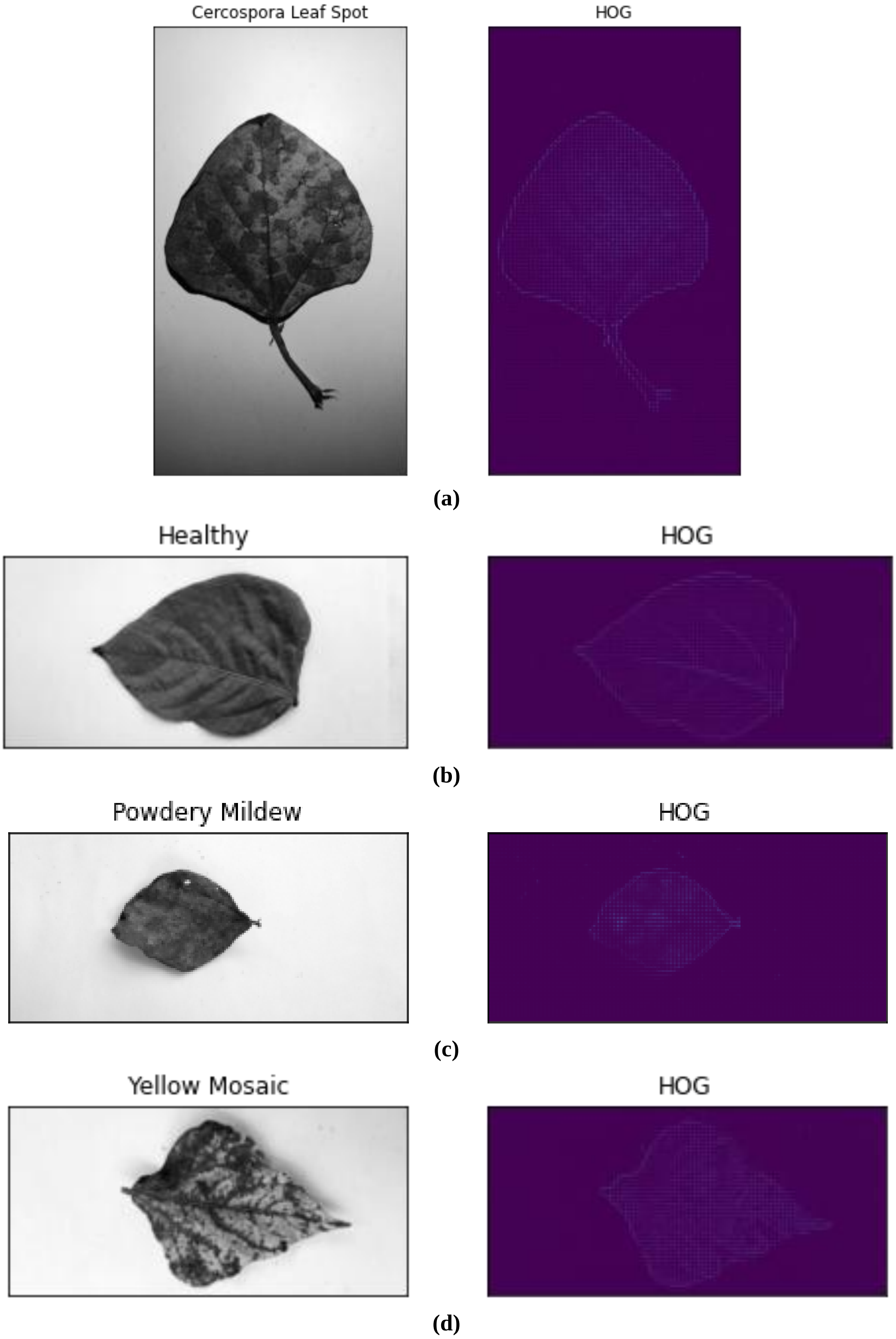

Figure. 3 A leaf image after applying HOG: (a) Cercospora Leaf Spot, (b) Healthy, (c) Powdery Mildew, (d) Yellow Mosaic Virus

Prior to employing the Histogram of Oriented Gradients (HOG) technique, each image undergoes a conversion to its grayscale equivalent. The corresponding disease labels are then transformed into integer values, specifically ranging from 1 to 4. Figure 3 provides a visual illustration of a converted grayscale leaf image alongside its computed HOG representation. Given that we are utilizing the SVM implementation from the Sklearn library, the application of one-hot encoding to the training labels is unnecessary. Instead, these labels are directly associated with integer values spanning from 1 to 4.

5. Algorithm Training

Support Vector Machines (SVMs) are recognized for their capacity to deliver superior accuracy compared to alternative machine learning models, including logistic regression. To effectively manage input spaces that exhibit non-linear characteristics, SVMs utilize a kernel trick. This technique operates by transforming the data into a higher-dimensional space, thereby facilitating the identification of a hyperplane capable of segregating the distinct classes. The sklearn.svm.SVC class in Python's scikit-learn library provides a robust Support Vector Classifier (SVC). In this particular study, an SVC was implemented employing a polynomial kernel and a defined regularization parameter (C), in conjunction with a "One-vs-One" strategy. This "One-vs-One" approach functions as a practical heuristic method wherein a binary classification task is executed for every unique pair of classes within the multiclass problem. Each individual binary classifier generates a prediction for a class label, and the label that receives the majority of predictions is ultimately chosen as the final classification output. Following the training phase, the SVM model achieved a test accuracy of 86.9%; however, it was observed to have overfitted the training dataset. Figure 4 provides a visual representation of the confusion matrix, illustrating the classification performance across the four Mung leaf categories, which encompass three diseased states and one healthy state. Complementarily, Table 1 presents a comprehensive classification report detailing the precision, recall, F1-score, and support for each of these categories within the controlled experimental environment using the SVM model. The "One-vs-One" strategy inherently increases the computational cost during training due to the creation of multiple binary classifiers but can often lead to improved classification accuracy in multiclass scenarios. Further optimization of the regularization parameter (C) and exploration of different kernel functions could potentially mitigate the observed overfitting and enhance the model's generalization capability on unseen data. The confusion matrix in Figure 4 provides a detailed breakdown of the model's performance, highlighting specific areas of confusion between different disease categories and the healthy class.

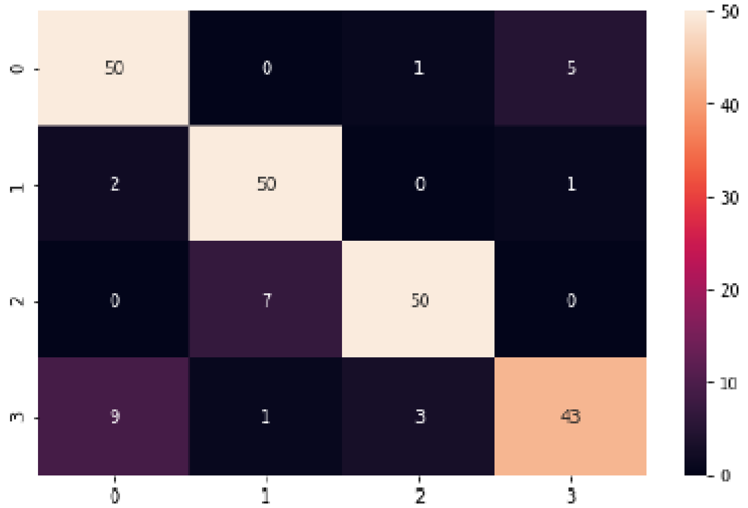

Figure. 4 Confusion Matrix (Controlled); 0-Cercospora, 1-Healthy, 2-Powdery Mildew, 3-Yellow Mosaic

	precision	**recall**	**f1-score**	**support**
0	0.82	0.89	0.85	56
1	0.86	0.94	0.90	53
2	0.93	0.88	0.90	57
3	0.88	0.77	0.82	56
accuracy			0.87	222
macro avg	0.87	0.87	0.87	222
weighted avg	0.87	0.87	0.87	222
0 – Cercospora		1 – Healthy		
2 – Powdery Mildew		3 – Yellow Mosaic Virus		

Table 1 Classification Report

To tackle the problem of overfitting, a range of values for the regularization parameter, along with other hyperparameters, were evaluated. The Grid Search technique was employed to pinpoint the hyperparameter settings that yield the highest accuracy while effectively avoiding overfitting. Several hyperparameters, namely C, Gamma, Kernel, Degree, and Strategy, were investigated across a defined set of potential values, utilizing a 5-fold cross-validation approach within the Grid Search framework. Following the implementation of the optimal hyperparameter values identified through the grid search process, the training accuracy of the model reached a perfect 100%. However, this adjustment resulted in a slight decrease in the test accuracy, which settled at 86.4%.

Conclusion

Within the supervised confines of the controlled environment, the relevant features were effectively extracted from the mung leaf images and subsequently classified using the Support Vector Machine (SVM) model, culminating in a test accuracy of 86.4%. The carefully managed setting, characterized by distinct, noise-absent images, facilitated precise feature detection, thereby enabling the model to reliably differentiate between healthy and afflicted foliage. Notwithstanding the elevated training accuracy achieved, the model exhibited a degree of overfitting, evidenced by the marginal reduction in test accuracy. This observation implies that while the model demonstrates strong performance on the data it was trained on, further refinement and adjustments are warranted to enhance its capacity to generalize to novel, previously unseen data.

References

1. Abadi, M. Tensorflow: Large-scale machine learning on heterogeneous systems. Accessed 8.
2. Asperti, A. and Mastronardo, C. The effectiveness of data augmentation for detection of gastrointestinal diseases from endoscopical images. CoRR, 03689:1â7.
3. Bloice, M., Stocker, C., and Holzinger, A. Augmentor: An image augmentation library for machine learning. CoRR, 04680:1â5.
4. Ciresan, D., Meier, U., and Schmidhuber, J. Multicolumn deep neural networks for image classification. CVPR, In.
5. Ferentinos, K. Deep learning models for plant disease detection and diagnosis. Comput Electron Agric, 145:311â8.
6. Fernandez, G. and Shanmugasundaram, S. The avrdc mungbean improvement program: The past, present and future. In Mungbean. Eds. Shanmugasundaram, S. and McLean, B., editors, Proceedings of the Second International Symposium held at Bangkok, page 58â70, Thailand.
7. Fuentes, A., Yoon, S., Kim, S., and Park, D. A robust deep-learning-based detector for real-time tomato plant diseases and pestâs recognition. Sensors, 2017;17.
8. Galdran, A. Data-driven color augmentation techniques for deep skin image analysis. CoRR, 03702:1â4.
9. Godliver, O., Friedrich, M., Mwebaze, E., Quinn John, A., and Biehl, M. Machine learning for diagnosis of disease in plants using spectral data. Intâl Conf. Artificial Intelligence, ICAIâ18.
10. Huang, K. Application of artificial neural network for detecting phalaenopsis seedling diseases using color and texture features. Comput Electron Agric, 57:3â11.
11. Jung, A. imgaug: a library for image augmentation in machine learning experiments. Accessed
12. Lal, G., Kim, D., Shanmugasundaram, S., and T, K. Mungbean production. AVRDC, page 6.
13. Lee, S., Chan, C., and Wilkin, P. Deep-plant: Plant identification with convolutional neural networks. In IEEE International Conference on Image Processing, volume 2015, page 452â456.
14. Liu, B., Zhang, Y., and He, D. Identification of apple leaf diseases based on deep convolutional neural networks. Symmetry, 10(1).
15. M., P., G., P. Y., S., D., M., T., K., G., G., R. R., and B, V. Hyperspectral remote sensing of yellow mosaic severity and associated pigment losses in vigna mungo using multinomial logistic regression modelsâ. Elsevier Crop Protection, 45:132â140.
16. Mohanty, S., Hughes, D., and SalathÃ c , M. Using deep learning for image-based plant disease detection. Front Plant Sci, 7(1419).
17. Munjal, R., Lall, G., and Chona, B. Some cercospora species from india-iv. Indian, (ytopathololology,13):144â149.
18. Nariani, T. Yellow mosaic of mungbean (phaseolus aureus. Indian Phytopathol, 13:24 29.
19. Pan, S. Yang qa survey on transfer learning. IEEE Transactions on Knowledge Data Engineering, 2010,22(10):1345-1359.
20. Poehlman, J. The mungbean. Westview Press, Boulder, Colo.
21. Price, T., Gross, R., Wey, J., and Osborne, C. A comparison of visual and digital image-processing methods in quantifying the severity of coffee leaf rust (hemileia vastatrix. Aust J Exp Agric, 33:97â101.
22. Ramcharan, A., Baranowski, K., McCloskey, P., Ahmed, B., Legg, J., and Hughes, D. Deep learning for image-based cassava disease detection. Front Plant Sci, 8(1852).
23. R.M., N., Schafleitner, R., Kenyon, L., Srinivasan, R., Easdown, W., Ebert, A., and Hanson, P. Genetic improvement of mungbean productivity. In Proc. Of the 12th SABRAO Congress on Plant Breeding towards 2025: Challenges in a Rapidly changing World, page 27â28, Chiang Mai, Thailand.
24. Selvaraj, M., Vergara, A., and Ruiz, H. Aipowered banana diseases and pest detection. Plant Methods, 15(92):7â019â0475â.
25. Simard, P., Steinkraus, D., and Platt, J. Best practices for convolutional neural networks applied to visual document analysis. In editor., editor, Society IC, volume ICDARâ03), vol. 2, page 958â64, Edinburgh. IEEE Computer Society.
26. Simard, P., Victorri, B., LeCun, Y., and Denker, J. Tangent prop â a formalism for specifying selected invariances in an adaptive network. In Proceedings of the 4th International Conference on Neural Information Processing Systems (NIPSâ91). Advances in Neural Information Processing Systems, volume 4, page 895â903, Denver. MIT Press.
27. Sladojevic, S., Arsenovic, M., Anderla, A., Culibrk, D., and Stefanovic, D. Deep neural networks based recognition of plant diseases by leaf image classification. Comput Intellig Neurosci, 2016(11).
28. Srdjan, S., Marko, A., and Andras, A. Deep neural networks based recognition of plant diseases by leaf image classification. Comput Intell Neurosci, 6:1â11.
29. Too, E., Li, Y., and Njuki, S. A comparative study of fine-tuning deep learning models for plant disease identification. Comput Electron Agric, 161:272â9.

30. Valle, E. Data, depth, and design: Learning reliable models for melanoma screening. CoRR, 00441:1â10.
31. Verma, A., Dhar, A., and Malathi, V. Cloning and restriction analysis of mungbean yellow mosaic virus. In International Conf. Virology in the Tropics Lucknow, India, page â14.
32. Vinod, K. and Pandey, S. Current status of mungbean in madhya pradesh - a review. International Journal of Current Microbiology and Applied Sciences, ISSN, 7706:1062â1072.
33. Wang, X. Chestx-ray8: Hospital-scale chest x-ray database and benchmarks on weakly-supervised classification and localization of common thorax diseases. In Proceedings of the 2017 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPRâ17). CVPR â17, Hawai. IEEE Computer Society.
34. Yong, Z. and Ming, Z. Research on deep learning in apple leaf disease recognition. Comput Electron Agric, 168(105146):2019 10514 6.