

Leaf Disease Classification Using Convolutional Neural Network

Ms. Mitali V. Shewale*, Dr. Rohin Daruwala

Veermata Jijabai Technological Institute (VJTI), Department of Electronics Engineering, Mumbai, Maharashtra

*Corresponding author

doi: <https://doi.org/10.21467/proceedings.114.4>

Abstract

Agriculture is a major domain that contributes a lot for building up the country's Economy; contributing to the GDP area synthesis of 17.9%. India stands second in production of agricultural products. Promising technologies such as Internet of Things, Machine Learning, Deep learning, Artificial neural networks contributes towards the most effective and reliable solutions by providing the most feasible solutions in making of different domain modernization through automation in monitoring and maintenance of agricultural fields with minimum human intervention. This paper presents a convolutional neural network based customized VGG framework and a lightweight architecture for the classification of tomato leaves affected with various diseases. Experimental analysis is performed on publically available PlantVillage dataset. After rigorous experiment we fined tuned the CNN model to obtain mAP of 83.33%.

Keywords: Agriculture, Convolutional neural network, Plant leaf classification

1 Introduction

Agriculture is the foundation of Indian Economy. As the future is heading towards an artificially smart world, machines are supplanting human specialists in each area. One such huge space is agribusiness, where the human specialists are searching for wise substitutes, which may make their chore simpler and far superior than human specialists. Such smart frameworks are extremely critical, and probably going to kill any odds of vagueness. A significant amount of work has been done through Internet of Things, Machine Learning, Deep learning, Artificial neural networks in agricultural area to develop smart farming solutions. State-of-the-art these technologies have brought a significant revolution in the agriculture domain by examining several challenges and complications in smart farming by detecting several issues faced in the agriculture providing solutions over them by increasing productivity, lowering the cost with reduced human interference [2]. Plant health detection through leaf texture and structure analysis has significant research importance in image processing and computer vision domain. Though several attempts have been made in these areas, lots of challenges are yet to be addressed. Most important amongst them are disease detection through plant leaf and their representation in order to characterize the plant species accurately and analysis of health condition and yield prediction could be made. However, in agro domain tremendous work has already been carried out and there is a requirement of in depth study on state-of-art technologies for the current research scenario to address various issues. This paper analyzes the plant leaf classification affected by several diseases and solutions in smart farming environment by transforming agricultural technologies in smart agriculture system through different innovations and utilization of different state-of-art methods.

Followed by the section I Introduction the further sections in the article consists of section II consisting of related work followed by section III the proposed workflow. Section IV consists of the experimentation results followed by section V consisting of conclusion and the future scope followed by the references.



2 Related Work

The system developed by Xiao- Zhi Gao *et al.* explained the use of morphological features and adaptive boosting methods for the recognition of plant species. The study demonstrated the use of Multilayer Perceptron with Adaboosting classifier for plant species for morphological feature extraction. k-NN, Decision Tree and Multilayer perceptron classifiers are used for calculating the performance of the implemented classifier. The study explored that AdaBoost methodology improved the precision accuracy by 95.42% that have used the FLAVIA plant public dataset. [1]

The system proposed by Xiaobo Yang *et al.* demonstrated the plant leaf images for instance segmentation and classification methods using ISC-MRCNN and APS-DCCNN. To enhance the network to learn more detailed features the lateral connection structure in ISC-MRCNN is utilized for feature mapping of plant leaf images from different depths followed by the suppression algorithm for improving the detection accuracy for overlapping objects in the acquired image followed by the implementation of pooling layer that reduced the precision loss during the feature mapping alignment between original image and its feature map. Finally to mask the complex backgrounds, a mask filter layer is implemented. In this process, to replace the softmax, the Support Vector Machine is used followed by an Particle Swarm Algorithm for optimizing it. Results showed that with different thresholds the accuracy is increased by 1.89% as compared with mask R-CNN and to 1.59% compared with traditional CNN. [2]

The system demonstrated by John Barron *et al.* discussed on plant species identification using a full plant leaf database consisting of occluded leaf images. Firstly, the two dimensional β -Spline curve is represented followed by extracting region of interest through Discrete Contour Evolution algorithm. Sub-graph matching is done by the graph nodes by using DCE points for closed leaf contour producing number of open curves and uses the Frechet distance metric by retaining the best η match. The energy functional is minimized using convex-concave relaxation framework that uses the local and global curvature. The best fit for the overall leaf is the one representing the best η curves amongst all the curves having minimum energy. Experiments carried out on three public leaf image dataset shows that the implemented method performance is outstanding with greater accuracy. The overall contour except the leaf area that is missing is measured in terms of percentage and is defined as the Occlusion. The implemented algorithm proved that, for plant leaves with a higher occlusions higher than 50% identifies the best full leaf match. [3]

The system implemented by Jiang Huixian *et al.* demonstrated the application of Deep Learning and Artificial Neural Network for the analysis of plants image recognition. The main contribution of this research is the identification and extraction of plant leaf features using image analysis technique. Dataset consists of 50 plant leaf compared and tested with KNN- classification, Kohonen network and support vector machine. Also it was observed that amongst all the 7 different plants from the leaf database the ginkgo leaves were easily identified. Also good recognition rate has been achieved for images with complex background. Results proved that the implemented method has the minimum recognition time of 617ms with the recognition accuracy of 85.94%. [4]

The system developed by Peng Jiang *et al.* explained the real-time detection and implementation using the Convolutional Neural Networks for apple leaf disease. Dataset consists of 26,377 apple leaf images. The research implemented the deep-CNNs GoogLeNet Inception and Rainbow concatenation structure. The results showed that through the implemented architecture the detection performance achieved is 78.80% with 23.13 FPS as the detection speed showing novelty of the algorithm for real-time detection performance with maximum accuracy and faster detection speed. [5]

The system proposed by S. Kaur *et al.* demonstrated the detection and classification of soybean culture using a semi-automatic leaf disease model. A rule based k-means automated system is designed and implemented for leaf disease detection that utilizes the PlantVillage dataset. Support vector machine classifier models are trained for detection purpose that utilizes features of leaf based on colour, texture and their combination and the average accuracy obtained with the implemented model is about 90% for all the combinations as compared to that of existing algorithms. Also, the implemented system computed the disease severity efficiently. [6]

The system demonstrated by M. Khan *et al.* discussed on correlation and genetic algorithm based segmentation and classification method for apple diseases. The implemented method utilizes the pipelining procedure. A hybrid method is used to enhance the apple leaf spots in the acquired image. After that, the infected regions are segmented and optimized using strong correlation method and genetic algorithm One-vs.-All M-SVM on Plant Village dataset. Research has been carried out for detection of healthy leaves, Blackrot, Rust, and Scab classes with accuracies 98.10%, 97.30%, 94.62 and 98% respectively. [7]

The system implemented by S. Salmpuria *et al.* demonstrated the application of CNN for the detection and recommendation system. In this research work the flow structure consists of preprocessing followed by feature extraction process of leaf images and the implementation of the network for disease classification and pesticide recommendation. The network consists of five, four & three layers for training the model and an android application for interaction. Results showed for 15 epochs in case of a 5-layer model the accuracy obtained is highest and is 95.05% and for 20 epochs for 5-layer model the validation accuracy achieved is 89.67% evaluated through tensor flow. [8]

The system developed by T. Mim *et al.* explained the image processing approach for disease detection on tomato leaf plant. This work utilizes AI and CNN method for detection mechanism. Six classifications including the healthy class of leaves disease have been detected. Accuracy achieved is 96.55% providing an efficient algorithm in correct classification and detection of Tomato leaf disease for increasing the productivity and providing more opportunities for research and in professional market. [9]

The system proposed by Xuan Nie *et al.* demonstrated the model based on Multi-Task Learning for strawberry wilt detection verticillium using Faster R-CNN and multi-task learning. The implemented system detects verticillium wilt through symptoms in plants. The dataset used for this research consists of 3531 images with 4 categories including healthy and diseased leaves along with the label for each input image indicating if it is suffering from wilt. Results demonstrated that the network achieved 77.54% mAP and a detection accuracy of 99.95%. [10]

The system demonstrated by U. Singh *et al.* discussed on mango leaves classification infected by anthracnose disease using a CNN. The research consists of real time dataset of 1070 images. The research aims at developing a MCNN network for effective, early and accurate diagnosis of fungal disease infected leaves and its symptoms providing a cost effective solution for plant health detection system. The implemented algorithm achieved the classification accuracy of 97%. [11]

The system implemented by F. He *et al.* demonstrated tree species identification with higher leaves similarity based on CNN. The research focuses on the convolutional neural network, for distinguishing different features. The implemented methodology consists of addition of attention branch in all layers of network. Next to obtain a ROI a condensation process is adopted by the attention branch by designing the process for amplification of features difference. Next the attention branch and the normal branch is combined through the fusion process for improving the network performance. Experimentation is carried out on Leafsnap dataset and an accuracy of 91.43%. Furthermore, the implemented ABCNN network is also tested and achieved 98.27% classification accuracy, proving the robustness of the method. [12]

The system developed by J. Ta *et al.* explained the plant species classification using the vein structure. The work consists of a CNN method designed by extracting the features using AlexNet, fine-tuned AlexNet and D-Leaf networks. For feature classification SVM, ANN, k-NN, Naive-Bayes and CNN are used. The implemented model achieved 94.88% accuracy compared with AlexNet (93.26%) and fine-tuned AlexNet (95.54%) networks thus proving to be the effective and efficient automated model for plant species identification. [13]

The system proposed by D. Zhang *et al.* demonstrated the calculation of leaf area index of digital images of sugarcane. Image processing is used for estimating leaf area index. Dataset is collected from Guangxi province, China. The input sugarcane images are acquired using a digital video camera located on 6.0 m pole. The data is collected through portable handheld leaf area meter using a field-based LAI measurement and are processed through MATLAB environment extracting the different color vegetation indices through time series digital images. The highest correlation coefficients of the color vegetation indices are chosen and G-B is proved as the best fit to test the model for LAI values. [14]

The system demonstrated by S. Veni *et al.* discussed on cucumber leaf disease detection on using SVM. In this paper diseases such as Alternaria leaf blight, Bacterial wilt, Cucumber green mottle mosaic, Leaf Miner, Leaf spot, Cucumber Mosaic Virus are detected. A combination of K-means clustering and a SVM is used to address this problem. [15]

The system implemented by C. Cevallos *et al.* demonstrated the nutrient deficiency classification of tomato leaves using CNN network. The literature utilized a supervised vision Convolutional Neural Network based monitoring system for recognition and classification of different types of nutrients deficiency in plants. A dataset is created consisting of the leaf images having various symptoms. Results showed that the implemented model achieved 86.57% accuracy. [16]

The system proposed by P. Bhatt *et al.* demonstrated the application of comparative analysis of CNN Models and the participatory sensing for the purpose of crop health assessment. A smart-phone application is developed which is utilized by farmers for application involving crop health. Results shows that the classification accuracy obtained is 99.7% assuring the real-time crop diagnosis with minimum hardware capabilities. [17]

The system developed by A. Rau *et al.* explained the system for nutrient detection and disease analysis. In this work, an automated irrigation and fertigation cost-effective method developed using MATLAB based image processing for identification of magnesium and nitrogen deficiencies in rice plants enabling the farmers to monitoring weather conditions. [18]

3 Proposed Work Flow

In the proposed system all the images from the dataset are uniformly resized into 224*224. The proposed workflow consists of the acquisition of images from the dataset, followed by resizing and assignment of class labels to the images, followed by categorizing the images amongst training, testing and validation set selecting from the labels assigned to all the respective classes. Then the training and testing of the implemented CNN network from the training and testing images are done followed by the validation of implemented network model to check the performance.

Our CNN model is based on the VGG network, where we tried developing a lightweight architecture by fine-tuning and using dynamic programming. We increase the depth of the neural network to extract rich features for detecting the fine details. In the network, we used a large number of filters to minimize the error rate and to increase the accuracy. We used the ReLU activation function. Specifically, filters with the size 3*3 and 1*1 with the stride of one, different from that of LeNet-5 having large size filter and also from those of AlexNet consisting of smaller with a larger stride of four. Followed by most

of the convolutional layers are the max-pooling layer that is performed with the size of 2*2 and the same stride. The proposed system consists of multiple convolutional layers stacked together before the max-pooling layer. The restructuring is that the convolutional layers with smaller filters approximate the effect as that with a large-sized filter. To increase the accuracy of the model several data augmentation techniques such as resizing, transformation, scaling, etc. are performed on the dataset.

4 Experimental Results

In the proposed work publically available PlantVillage dataset repository is used consisting of multiple images of plant leaves. Tomato plant leaves are considered for the experimentation consisting of total 16,012 images out of which 11212 is considered as training images, 3203 as the testing images and 1597 as the validation set images. This is achieved by random split of dataset into 70% as training set, 20% as testing set and 10% as validation set. The images categorization is done into 10 different classes of which 9 classes belong to disease class and 1 class belongs to the normal/healthy class. Disease classes include Target Spot (1404 images), Mosaic Virus (373 images), Yellow Leaf (3209 images), Bacterial Spot (2127 images), Early Blight (1000 images), Late Blight (1909 images), Leaf mold (952 images), Septoria (1771 images) and Spider Mites (1676 images). Some of the images from the dataset are as shown in the figure 1 below.

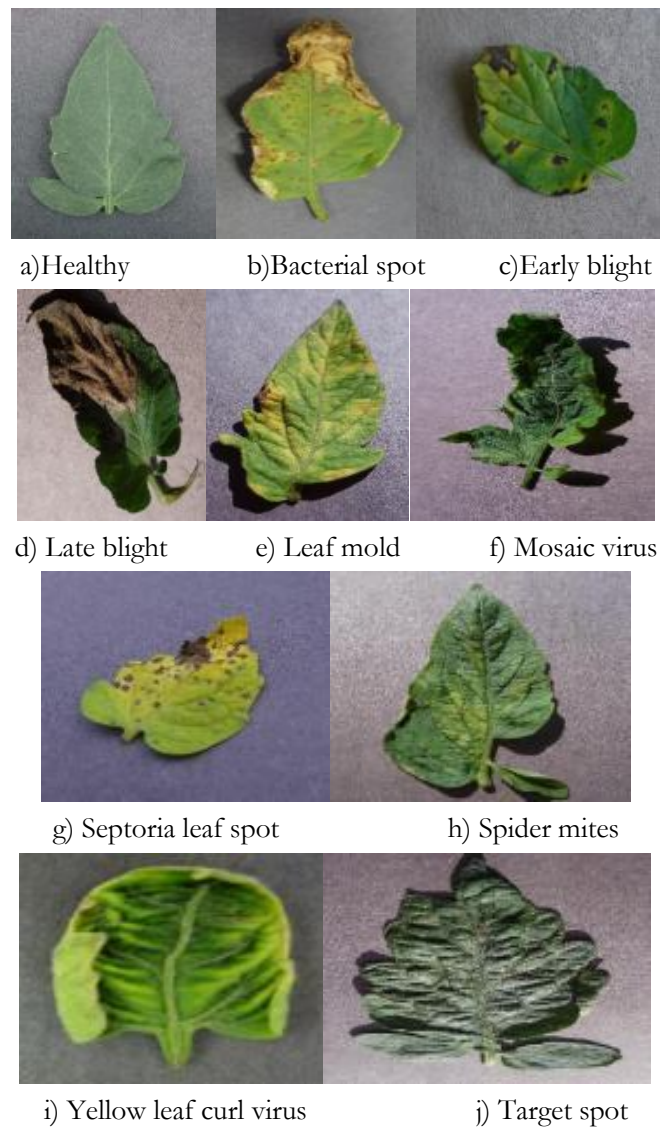


Fig. 1 Instances of images from dataset

The proposed system is implemented using a python language and an open source TensorFlow software framework. The input images (from a-j) and corresponding predicted (from i-x) series of images in case of 20 epochs for 0.01 learning rate is as shown in the figure 2. The first row corresponds to the input test images applied to the implemented network whereas the second row corresponds to their respective test images as predicted and classified by the network. Figure from a-j corresponds to bacterial spot, early blight, healthy, late blight, leaf mold, mosaic virus, septoria, spider mites, target spot and yellow leaf respectively. From the series of test images it is observed that the yellow leaf image and late blight image is misclassified as early blight image by the network.

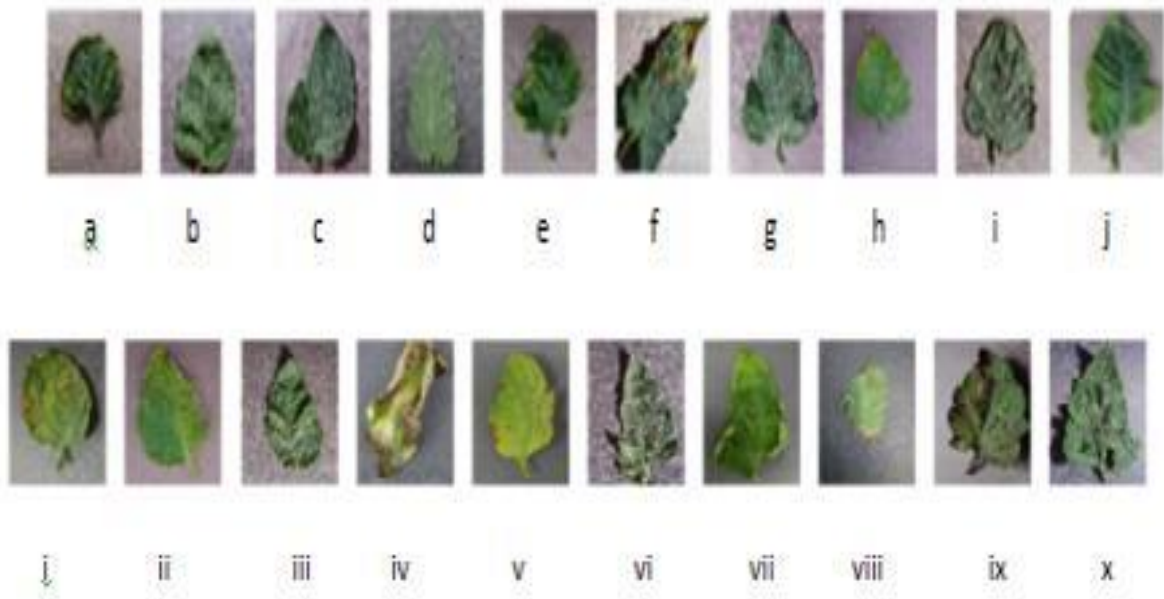


Fig. 2 Input and predicted series of images

The accuracy was evaluated by varying the epoch and learning rates of the implemented customized lightweight CNN. The accuracy score calculated by varying the parameters such as the learning rate and number of epochs is shown in the table 1 and its corresponding visualization in figure 3. It is observed that for 0.01 learning rate the maximum accuracy is achieved for the implemented network and further if the learning rate is reduced the accuracy decreases and from the figure it is observed that the maximum accuracy for the proposed system is achieved for 0.001 learning rate.

TABLE I. CLASSIFICATION ACCURACY

Epochs	Learning Rate		
	0.01	0.001	0.0001
20	0.8	1	0.7
50	0.9	1	0.7
100	0.9	0.8	0.7

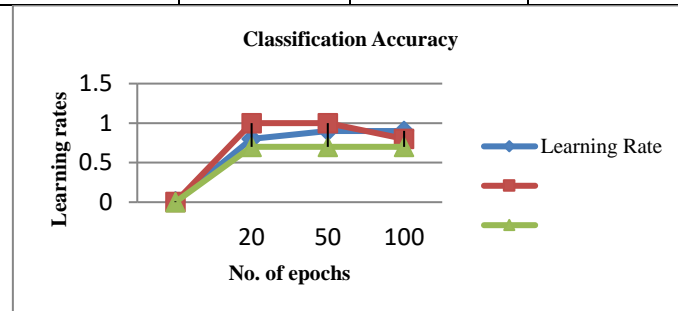


Fig. 3 Classification Accuracy

Figure 4 indicates the precision achieved by varying the no of epochs from 20 to 100 along with the learning rate from 0.01 to 0.0001. It is observed that initially the precision value for learning rate 0.01 is low as the learning rate is decreased to 0.001 for all the epoch maximum precision value of 1 is obtained after that as we decreased the learning rate further precision value also reduces.

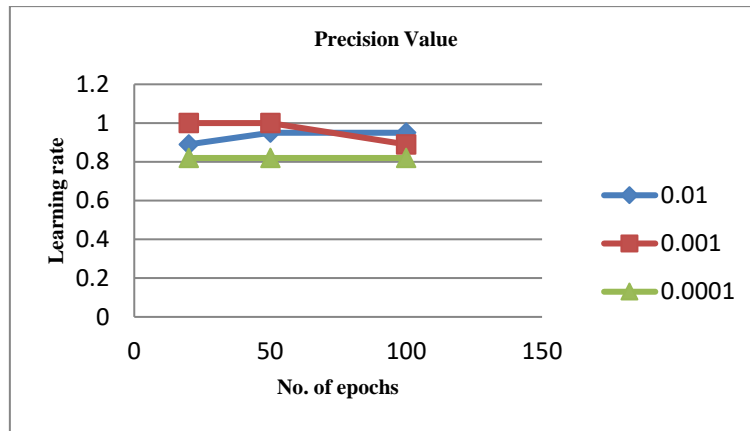


Fig. 4 Precision Chart

Figure 5 shows the recall chart for the varying number of epochs and the learning rate. As seen the maximum value of recall score of 1 for epoch 20 and 50 is obtained for learning rates 0.01 and 0.001 respectively. It gradually decreases with the further decrease in the learning rate.

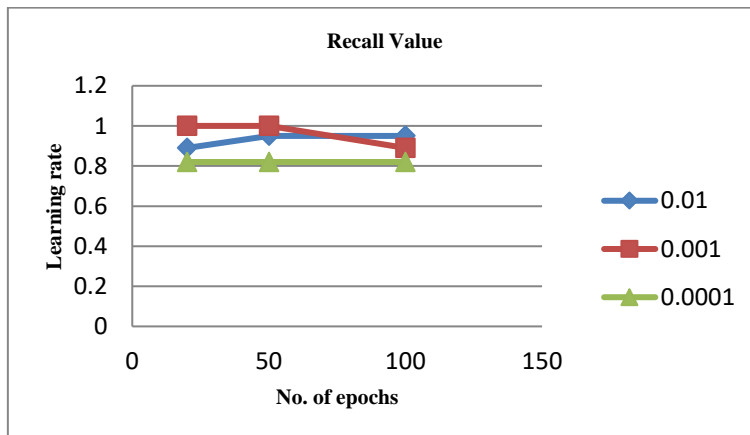


Fig. 5 Recall Chart

Figure 6 illustrates the F1-score values for the varying number of learning rate and epochs. From the graph it is observed that the F1-score values are highest that is 1 for 0.001 leaning rate, followed for the 0.01 rate which is on average about 0.95 and the lowest is observed in case of 0.0001 learning rate which is on an average about 0.82.

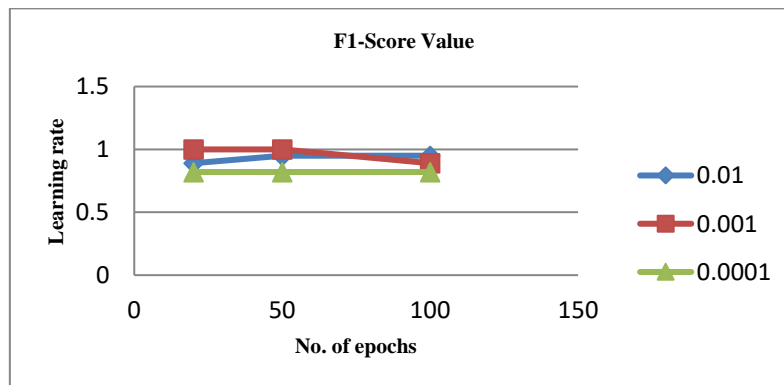


Fig. 6 F1-Score Chart

Table II represents the comparison of the implemented model with the other models used in the literature. It is observed that significant comparable accuracy is achieved by the network which can be further improved through extraction of more significant features from the images and by tuning the parameters.

TABLE II COMPARISON WITH DIFFERENT MODELS

Dataset	Number of images	Model	Accuracy	Weakness
Flavia	Total: 1907; Training: 80%, Testing: 20%	Multilayer perceptron with AdaBoost [1]	95.42%	Simple architecture, Disregard spatial information
Folio, CSUFT- 20	9200 (Folio)+99413 (CSUFT-20)=108613 (Total); Training: 60%, Verification: 20%; Testing: 20%	AlexNet,Vgg16, DCCNN [2]	93.08%, 93.13%, 94.94% with AlexNet,Vgg16, DCCNN respectively	More computational time with failure to capture fine details in the object
Swedish, Flavia, Leafsnap	1125 (Swedish), 50 (Flavia), 184 (Leafsnap)	Beta-Spline [3]	74.8% on Swedish dataset, 74.4% on Flavia dataset, 71.9% on Leafsnap dataset	More complexity with significant reduction in accuracy
Self-generated	200; Training:180; Testing: 20	KNN adjacent classification, Kohonen net, SVM, BP Neural network [4]	85.94% (KNN), 86.78% (Kohonen net), 86.45% (SVM), 92.47% (BP Neural network)	After a particular threshold as the complexity increases algorithm fails to learn and capture the detailing in training process
Self-generated	2029; Training: 75%, Testing: 25%	Deep-CNN [5]	78.80%	Moderate CNN network
Plant-Village	4775; Models: 1: (50% training-50% testing); 2: (60% training -40% testing); 3: (70% training-30% testing)	SVM [6]	Model: 1: 83%; Model 2: 84.25%; Model 3: 85.65%	Moderate accuracy and slow performance in real time application
Plant-Village	630; Training:315, Testing: 315	M-SVM [7]	94.62%	Moderate accuracy and slow performance in real time application
Plant-Village	Training:18917, Testing: 3000	CNN (3 Layer, 4 Layer, 5 Layer) [8]	88.24% (with 3 layer network), 92.19% (with 4 layer network), 91.01% (with 5 layer network)	CNN is trained on smaller dataset
Self-generated and GitHub	6000; Training: 80%, Testing: 20%	CNN [9]	96.55%	Slow learning model with good accuracy

Self-generated	3531; Training: 56%, Validation: 14% , Testing: 30%	R-CNN [10]	99.25%	Uses selective search algorithm which is slow and time consuming
Self-generated and Plant-Village	1070 self-generated + 1130 PlantVillage = 2200 (Total) Training: 80%, Testing: 20%	CNN [11]	97.13%	-
Leafsnap	7719 Training: 6109 Testing: 1610	ABCNN [12]	91.43%	-
Self-generated	1290; Training: 80%, Testing: 20%	CNN [13]	94.88%	Training on smaller dataset
Self-generated	-	K means clustering along with SVM [15]	89%	Simple classifier
Plant-Village	596; Training: 70%, Testing: 30%	CNN [16]	86.57%	Simple network with small dataset
Plant-Village	3750	CNN [17]	99.7%	Slow learning rate
Plant-Village	16,012	CNN (Proposed work)	83.33% (Average)	Scope to increase the accuracy of the network

5 Conclusion and Future Scope

Across the globe technologists and agriculturalists are exploring state-of-art methods and solutions to enhance the agricultural productivity by focusing on smart, better and more efficient crop growing methods along with identification of plant health detection. This paper presented a customized light weight CNN model for the leaf classification of healthy tomato plant from multiple disease types and the rigorous analysis from several papers were done to propose their insights and to increase the performance for the classification and achieved the overall accuracy of almost 100% for 0.001 learning rate. Also the effects of varying the parameters such as epochs and learning rate influencing the performance evaluation parameters such as precision, learning arte and accuracy is analyzed. The implemented system thus proved to be computationally simple and efficient in classifying the healthy and unhealthy leaves. In future instead of using the Softmax as an activation function and the publically available dataset, experimentation can be carried out on the real-time dataset and evaluating for different activation function. Also the focus can be made on the disease severity computation and yield prediction considering other plant parts. Also the IoT technology can be incorporated to monitor the classification and the real time behavior of the system.

References

- [1] X. Yang, A. Chen, G. Zhou, J. Wang, W. Chen, Y. Gao, R. Jiang, "Instance segmentation and classification method for plant leaf images based on ISC-MRCNN and DCCNN", *IEEE Access*, vol.8, pp. 151555-151573, Aug. 2020, doi: 10.1109/ ACCESS.2020.3017560
- [2] M. Kumar, S. Gupta, X. Gao, A. Singh, "Plant species recognition using morphological features and adaptive boosting methodology", 2020 *IEEE Access*
- [3] A. Chaudhury, J. Barron, "Plant species identification from occluded leaf images", *IEEE/ACM transactions on computational biology and bioinformatics*, Aug. 2018, doi: 10.1109/TCBB.2018.2873611
- [4] J. Huixian, "The analysis of plants image recognition based on deep learning and artificial neural network", *IEEE Access special section on Data Mining for Internet of Things*, vol. 8, pp. 68828-68841, April 2020, doi: 10.1109/ACCESS.2020.2986946

-
- [5] P. Jiang, Y. Chen, B. Liu, D. He, C. Liang, "Real-time detection of apple leaf diseases using deep learning approach based on improved convolutional neural networks", *IEEE Access special section on Advanced optical imaging for extreme environments*, vol. 7, pp. 59069- 59080, May 2019, doi: 10.1109/ACCESS.2019.2914929
- [6] S. Kaur, S. Pandey, S. Goel, "Semi-automatic leaf disease detection and classification system for soybean culture", *The Institution of Engineering and Technology (IET) Journal on Image Processing*, vol. 12, Issue 6, pp. 1038-1048, Feb 2018, doi: 10.1049/iet-ipr.2017.082
- [7] M. Khan, M. Lali, M. Sharif, K. Javed, K. Aurangzeb, S. Haider, A. Saud, T. Akram, "An optimized method for segmentation and classification of apple diseases based on strong correlation and genetic algorithm based feature selection", *IEEE Access Transaction on Data Mining*, vol. 1, pp. 1-18, 2019, doi: 10.1109/ ACCESS.2019. 2908040
- [8] P. Kosankar, V. Kulkarni, K. Mantri, "Leaf disease detection and recommendation of pesticides using convolution neural network", 2018 Fourth International Conference on Computing Communication Control and Automation (ICCUBEA)
- [9] T. Mim, M. Sheikh, R. Shampa, M. Reza, M. Islam, "Leaves diseases detection of tomato using image processing", *Proceedings of the 8th SMART-2019 IEEE International Conference on System Modeling and Advancement in Research Trends*, pp. 244-249, Nov. 2019
- [10] X. Nie, L. Wang, H. Ding, M. Xu, "Strawberry verticillium wilt detection network based on multi-task learning and attention", *IEEE Access*, vol. 7, pp. 170003-170011, Nov 2019, doi:10.1109/ACCESS.201902954845
- [11] U. Singh, S. Chouhan, S. Jain, "Multilayer convolution neural network for the classification of mango leaves infected by anthracnose disease", *IEEE Access*, pp. 1-10, 2019, doi: 10.1109/ACCESS.2019.2907383
- [12] Y. Song, F. He, X. Zhang, "To identify tree species with highly similar leaves based on a novel attention mechanism for CNN", *IEEE Access*, vol. 7, pp. 163277-163286, Nov. 2019, doi: 10.1109/ACCESS.2019.2951607
- [13] J. Tan, S. Chang, S. Kareem, H. Yap, K. Yong, "Deep learning for plant species classification using leaf vein morphometric", *IEEE/ACM transactions on computational biology and bioinformatics*, 2018, doi: 10.1109/ACCESS.2018.2848653
- [14] D. Zhang, X. Song, R. Lamin, X. Wu, H. Sun, L. Li, Z. Kuang, H. Jin, L. Xu, "Estimating leaf area index of sugarcane based on multi-temporal digital images", *IEEE Access*, pp. 1-5
- [15] P. Krithika, S. Veni, "Leaf disease detection on cucumber leaves using multiclass support vector machine", *IEEE WiSPNET 2017 Conference*, pp. 1276-1281, 2017
- [16] C. Cevallos, H. Ponce, E. Albor, J. Brieva, "Vision-based analysis on leaves of tomato crops for classifying nutrient deficiency using convolutional neural networks", *IEEE Access*, Oct. 2020
- [17] P. Bhatt, S. Sarangi, S. Pappula, "Comparison of CNN models for application in crop health assessment with participatory sensing", *IEEE 2017*
- [18] A. Rau, J. Sankar, A. Mohan, D. Krishna, J. Mathew, "IoT based smart irrigation system and nutrient detection with disease analysis", *2017 IEEE Region 10 Symposium (TENSYP)*, 2017