Journal of Positive School Psychology
2022, Vol. 6, No. 3, 8623–8635 http://journalppw.com

Performance Evaluations of Convolutional Neural Network (CNN)-Based Models for Semantic Segmentation of Plant Leaf Diseases

Ali Abd Almisreb[1], Samir Z. Aliyev[2], Raseeda Hamzah[3], Nursuriati Jamil[4]

[1,2]Faculty of Engineering and Natural Sciences, International University of Sarajevo, Sarajevo, Bosnia, and Herzegovina
[1]aalmisreb@ius.edu.ba
[3,4] Faculty of Computer and Mathematical Sciences, Universiti Teknologi MARA, 40450 Shah Alam, Selangor, Malaysia
[3]raseeda@fskm.uitm.edu.my
[4]lizajamil@computer.org

Abstract

Plant disease identification is important to sustain food production. Automated plant disease identification using Convolutional Neural Network has shown highly potential to provide effective solution to high accuracy and real-time plant disease detection. This paper presented the evaluations of five CNN-based models, namely DeepLabV3+ network with Resnet18/Resnet50/Resnet101, modified Alexnet, and Segnet with VGG-16 for semantic segmentation and identification of plant leaf diseases. The leaf images were acquired from Leaf Disease on Kaggle comprising four types of leaf diseases: bacteria, fungi, nematodes and virus. A total of 196 images were labeled for ground-truth development and training dataset. Image augmentation was conducted to increase the training dataset followed by assigning class weightage to the imbalanced classes. A total of 1,918 labeled images were produced and these images were used to train the five CNN-based models. All the pre-trained CNN-based models were modified to cater to the new leaf disease dataset and to optimize the semantic segmentation. The results showed that DeepLabV3+ network with ResNet-18 outperformed other models achieving 95.8% global accuracy for segmentation of the leaf diseases. This is followed by Segnet with VGG-16, ResNet-50, ResNet-101 and modified AlexNet. However, upon closer study of the classes, the mean accuracy showed that AlexNet achieved better results compared to Segnet with VGG-16 and ResNet-50

Keywords: CNN-based network, semantic segmentation, intersection-over-union, leaf disease.

I. INTRODUCTION

The importance of agriculture was highlighted in the Sustainable Development Goals (SDG) Plan 2030, where SDG2 is specially dedicated to ending world hunger, achieve food security, improved nutrition and promote sustainable agriculture. Crop disease issues may lead to famine and food insecurity around the world. It is estimated that plant pathogens may account for annual crop yield losses of up to 16% globally (Oerke, 2006). Early and precise identification is an essential part of disease monitoring. The popular practice of disease identification was done by experienced farmers or plant pathologists using a visual examination of the plant. However, this manual practice may lead to inconsistencies and is also time-consuming. In isolated and poor regions, it will incur a further cost as plant pathologists are not readily available. It is also worth noting that many proper and thorough monitoring of agricultural areas are too expensive (Barbedo, 2018).

Automated plant disease identification using machine learning has shown tremendous potential to become a more effective solution providing high accuracy, real-time, low cost and simple plant disease detection operation. Conventional machine learning methods such as K-nearest neighbors (K-NN), logistic regression, decision tree, and support vector machine (SVM) (Annabel et al., 2019; Zhang et al., 2019) had made some progress in plant disease recognition. These approaches were combined with various image pre-processing methods to enhance feature extraction. However, the most recent work of plant disease identification by (Geetharamania and Pandian, 2019; Ferentinos, 2018; Barbedo; 2018; Too et. al, 2019) concluded that Convolutional Neural Networks (CNNs) were highly powerful and suitable for the automated detection and diagnosis of plant diseases through the analysis of simple leaves images. Since the introduction of the Deep Convolutional Neural Network (DCNN) of AlexNet (Krizhevsky et al., 2012), deep learning has made an unprecedented achievement in the development of computer vision.

There are several factors to be considered when using CNN models for plant disease identification (Barbedo, 2018). The factors that are extrinsic to the plant disease recognitions are limited annotated datasets, symptom representation, covariate shift, image background, and image-capture conditions. Another four factors that are intrinsically related to the plant disease are symptom segmentation, symptom variations, simultaneous disorder and disorder with similar symptoms. The use of transfer learning models and augmented data in plant disease recognition somewhat alleviated the need for huge annotated datasets (Ferentinos, 2018). However, a robust plant identification system should be able to recognize variations of diseases caused by pests, nutritional deficiencies, phytotoxicity and many more. At the time of writing, there are no such extensive datasets that represent these symptoms variety. Therefore, a model trained using several symptoms would not be able to identify other unseen symptoms. One popular dataset commonly used for plant disease identification is the PlantVillage dataset by (Hughes and Marcel, 2015). It contains 54,306 images of healthy and infected leaves images; comprising 26 diseases for 14 crop plants and divided into 38 different classes. However, training and testing a model using the same dataset usually resulted in non-realistic performance assessment (Barbedo, 2018). This is because the model has a high chance of failing when applied to other datasets (Mohanty et al., 2016). This problem is known as covariance shift.

Even though CNN models were able to detect diseases from a whole leaf image, it was shown that CNN model trained with individual lesions and localized symptom regions achieved 85% compared to the accuracy obtained using the original images at only 76% (Barbedo, 2018). Therefore, symptom segmentation is deemed to be an important factor for a more robust plant disease identification. This is the motivation of our work. This paper focused on comparing pre-defined CNN-based networks to perform symptom segmentation and solve plant disease detection tasks. Pixel-labeling on each leaf image was performed to classify the portions of the images into three (3) main classification areas: disease, healthy, and background parts. the diseases category includes four (4) disease types namely Bacteria, Fungi, Nematodes and Virus. Five CNN-based models that were DeepLabV3+ based on Resnet18, Resnet50, Resnet101, AlexNet, and Segnet based on VGG16 were selected and modified. The modifications were done based on the relevancy of the semantic segmentation tasks and to determine the most preferred CNN model.

II. RELATED WORK

Machine learning methods were still actively being used in plant disease classification as demonstrated by the work done by (Singh et al., 2019) (Balakrishna et al., 2019) and (Rani and Paul, 2019). In (Singh et al., 2019), Support Vector Machine (SVM) was used to identify the early symptoms and classified fungal rust disease of pea at the microscopic level with an accuracy of 89.60%. On the other hand, K-NN and Probabilistic Neural Network (PPN) were utilized in a 2-stage classification of tomato leaves into healthy and unhealthy classes (Balakrishna; RAO, 2019, p. 63). In this work, PNN outperformed KNN when the leaf features were represented using Gray Level Co-occurrence Matrix (GLCM), Gabor filters and colours. Even though machine learning methods are still commonly used, Convolutional Neural Network (CNN) models for plant disease identifications are currently the adopted state-of-the-art methods (Barbedo, 2019).

As seen in Table 1, CNN-based networks have demonstrated high accuracies in detecting and classifying plant diseases using plant leaves. Even though several studies used their datasets such as Oppenheim et al. (2017) and Liu et al. (2018), the majority experimented using one or more crops in the PlantVillage dataset. However, Barbedo (2019) argued that PlantVillage images have limited diversity where a large proportion of the images contained homogeneous background particularly the earlier versions of 2015. Therefore, Barbedo (2019) developed his dataset comprising leaf images divided into individual lesions and spots captured under a diversity of conditions. More importantly, Table 1 also shows that CNN-based networks were tested and the accuracy proved that they are highly capable of identifying plant diseases.

Table 1 – Summary of CNN-based models in plant disease identification

Reference	CNN network	Best network	Dataset	Accuracy
(Too et al., 2019)	VGG 16, Inception V4, ResNet with 50, 101 and 152 layers DenseNets 121 layers	DenseNet	PlantVillage	99.75%
(KC et al., 2019)	MobileNet depthwise separable, Reduced MobileNet, Modified MobileNet	MobileNet depthwise separable	PlantVillage	98.65%
(Geetharamani and Pandian, 2019)	Deep CNN, SVM, Decision Tree , Logistic Regression, KNN	DeepCNN	PlantVillage	98.15%
(Ferentinos, 2018)	AlexNet, AlexNetOWTBn, GoogleNet, Overfeat, VGG	VGG	PlantVillage	99.48%
(Khamparia et al., 2019)	CNN+autoencoder		PlantVillage	97.50%
(Liang et al., 2019)	ResNet18-ResNet34,ResNet50,ResNet101, PD^2SE-Net50	PD^2SE-Net50	Synthetic dataset	98%
(Liu et al., 2018)	AlexNet		Own	98%
(Oppenheim et al., 2019)	VGG		Own	96%

The performance of CNN-based models that were VGG 16, Inception V4, ResNet with 50, 101, and 152 layers, and DenseNets with 121 layers were tested by (Too et al., 2019). The paper reported the DenseNets model to consistently improved accuracy and there was no indication of overfitting and deteriorated performance with the increasing number of epochs. DenseNets was also able to achieve a commendable accuracy of 99.75% with a considerably smaller number of parameters and reasonable computing time. KC et al. (2019) compared Modified MobileNet, Reduced MobileNet, and MobileNet depthwise separable using PlantVillage dataset and achieved an accuracy rate of 97.65%, 98.34% and 98.65%, respectively. In Geetharamani and Pandian (2019), a novel nine-layer deep CNN was proposed and compared with AlexNet, ResNet, VGG16, Inception-v3, SVM, Decision Tree, Logistic Regression and K-NN. The proposed deep CNN was trained using different training epochs, batch sizes, and dropouts and managed to achieve 96.46% accuracy. Another work by (Ferentinos, 2018) also used PlantVillage dataset to examine the performance of AlexNet, AlexNetOWTBn, GoogleNet, Overfeat, and VGG. The results showed VGG outperformed the other four networks achieving a 99.53% accuracy rate. Finally in Khamparia et al. (2019), a hybrid approach of combining CNN and autoencoders was done to build a network that is invariant to shadows, illumination, and skewed images. It was observed that the proposed architecture achieved variations in accuracy for the different number of epochs and different convolution filter size. They reached 97.50% accuracy for 2×2 convolution filter size in 100 epochs, while 100% accuracy for 3×3 filter size.

Plant disease identification using other datasets involving CNN-based networks has also shown promising results. Oppenheim et al. (2019) used the Visual Geometry Group (VGG) and achieved 96% accuracy in classifying the potato tuber diseases into five classes, namely, four disease classes and a healthy potato class. Meanwhile, (Liang et al., 2019) introduced a robust image-based Plant Disease Diagnosis and Severity Estimation Network (PD2SE-Net) comprising ResNet50 architecture as the basic model and shuffle units as the auxiliary structures. They used a synthetic dataset to conduct disease severity estimation, plant species recognition and plant disease classification, achieving overall accuracies of 0.91, 0.99 and 0.98, respectively.

III. METHODOLOGY

In this section, the image datasets, dataset labeling, data augmentation, training, and testing datasets, and architecture of the five CNN-based models are elaborated. Based on the literature findings, there is no work done to compare the accuracy of DeepLabV3+ (Chen et al., 2018) based on Resnet18, Resnet50, Resnet101 (He et al., 2016), AlexNet (Krizhevsky; Sutskever; Hinton, 2012), and Segnet based on VGG16 (Simonyan and Zisserman, 2015) for plant disease identification.

Image Dataset

The image dataset known as Leaf Disease was obtained from the Kaggle online (Sizlingdhairya, 2019) community of data scientists and machine learning engineers. The dataset contained 239 leaf images comprising four classes of leaf diseases namely Bacteria, Fungi, Nematodes, Virus, and one Healthy class. The quantity of images per class is shown in Table 2. Only Class 1, 2, 3 and 5 were considered in this paper as the main focus is to conduct semantic segmentation for plant disease identification. Therefore, 199 images were selected from the Leaf Disease dataset for this paper.

Table 2 – Leaf Disease dataset and class labels

Class number	Class name	Image Quantity	Label Class	Label No
1	Bacteria	50	Bacteria	1
2	Fungi	50	Fungi	2
3	Nematodes	49	Nematodes	3
4	Healthy	40	Healthy	4
5	Virus	50	Virus	5
	Total	**239**	Background	6

Image Labeling

The purpose of image labeling was to develop the ground-truth and the training dataset for the semantic segmentation. Every spot depicting the presence of disease was accurately segmented and the image region containing the healthy pixels and non-leaf pixels (i.e. background) was also segmented. Refer to Figure 1. Therefore, six labels were created for the labeled dataset as shown in the final two columns of Table 2 which were used for training of the five CNN-based models. During labeling, three images from the Virus class were eliminated because the image was blurry and noisy making it impossible to label each pixel in the image. Thus, a total of 196 images was labeled accordingly.

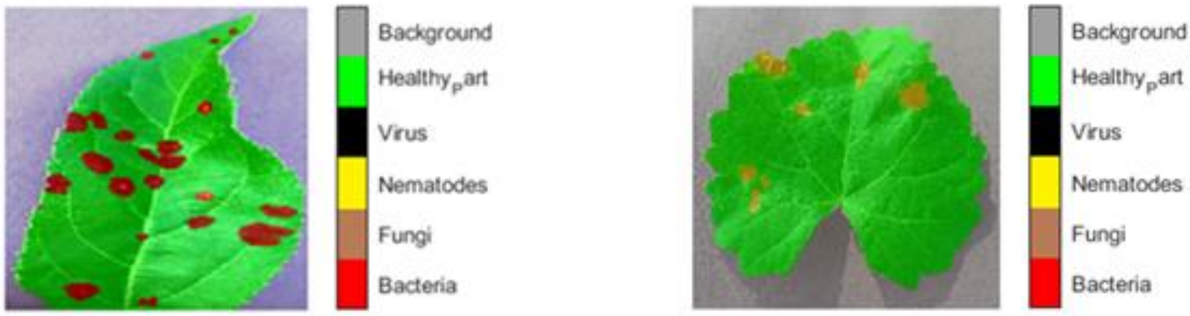
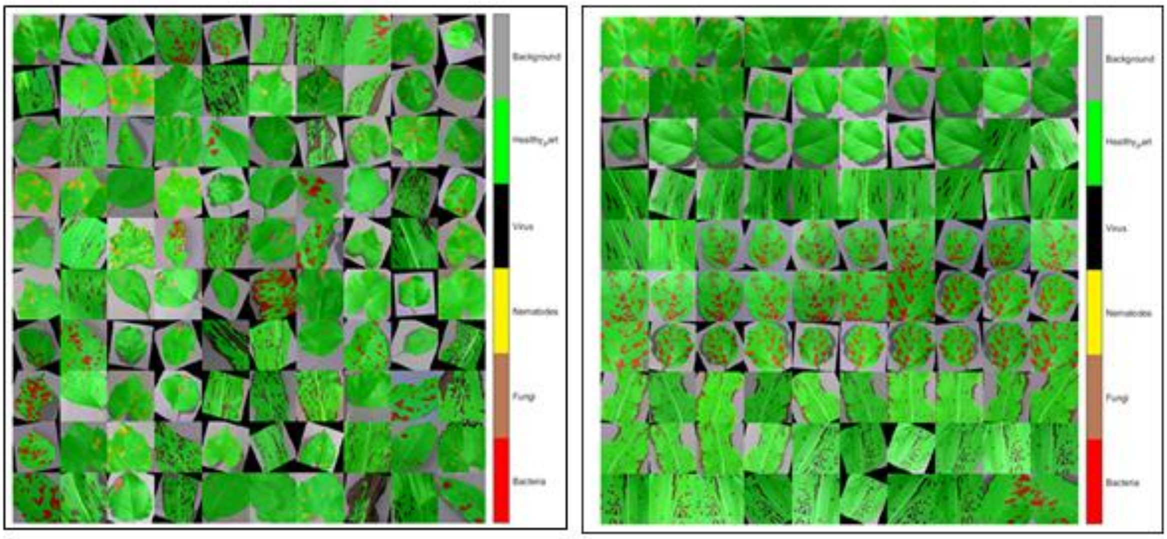


Figure 1 – Samples of labelled images

Image Augmentation

It is a known fact that a deep learning model requires a large dataset. Therefore, image augmentation was performed on the labelled dataset to increase the quantity of the training dataset without having to acquire more images. The training and testing dataset were distributed approximately according to the 70:30 ratio. Thus, 137 random images out of the 196 images were considered for the training dataset, while the remaining 59 images were used as the ground truth dataset for evaluation in the testing stage. The image augmentation process involved two stages of operations. In the first stage, the 137 labelled images were subjected to colour jittering, identical random scaling by a scale factor in the range [0.8 1.5], horizontal reflection and rotation in the range [-30, 30] degrees. See Figure 2a. The second stage of image augmentation creates a crop window centered on the image, then the image was cropped to 230 x 230 dimensions. See Figure 2b. The purpose of cropping is to reduce the background pixels produced during the first stage of image augmentation. Therefore, a more accurate representation of the leaf images was used for training the CNN-based models. A total of 1,918 augmented leaf images were produced.



(2a) (2b)

Figure 2 – Augmented images after colour jitter, scaling, horizontal reflection, rotation and cropping

Class Balancing

The number of images in each class was imbalanced, therefore class balancing using class inverse class frequency weighting was used to balance the classes. Each class frequency was calculated and the class weights were computer as the inverse class frequency over the labeled training dataset. The class weightage is shown in Table 3.

Table 3 – Class weights

Label name	Class weight
Bacteria	1.6714
Fungi	7.5047
Nematodes	4.2521
Virus	0.7134
Healthy_Part	0.2555
Background	0.2595

Construction of the CNN-Based Models

In this section, the comparison of five CNN-based models was described and the results presented. Figure 3 illustrates the process flow diagram of the research and each process is described in the subsections of Methodology. As depicted in the figure, the training dataset includes both original Leaf Disease images and the labelled images resized to 230x230x3 to accommodate the requirements of the CNN-based models. The performance of the semantic segmentation and plant disease identification was then measured using Intersection-over-Union (IoU) metrics and accuracy.

DeepLabV3+ network based on ResNet-18/50/101 models

The Residual Learning framework addressed accuracy saturation and degradation problems in deep networks. In this paper, we investigated three different layers of residual networks, namely ResNet-18, ResNet-50, and ResNet-101. We prepared the three ResNet models for the training stage by modifying them according to the DeepLabV3+, a type of CNN model. The modified parameters are presented in Table 4 and the detailed steps are as follows:

a) Firstly, Feature Extraction Layers (FEL) were selected from ResNet. For DeepLab v3+, the feature extraction layer was typically towards the end of the network, right before the classification layers.

b) After the FEL, all layers were removed in ResNet models.

c) The network image input size was amended appropriately to the training image size.

d) The network downsampling was reduced to 8 or 16 to preserve the spatial resolution required for accurate segmentation.

e) The convolution layer dilation factors were incremented to increase the receptive field size required to extract features from larger image regions.

f) The artrous spatial pyramid pooling module (ASPP) was added to the network.

g) Skip layer was selected to add a skip connection.

h) The decoder sub-network for DeepLab v3+ was added with skip layer and 6 is number of classes.

i) In the final step, Softmax and pixel classification layers were added to classify each pixel.

Table 4 – ResNet modifications for DeepLabV3+

Parameters	ResNet-18	ResNet-50	ResNet-101
FEL	'res5b_relu'	'activation_49_relu'	'res5c_relu'
Removed Layers	"pool5" "fc1000" "prob" "ClassificationLayer_predictions"	"avg_pool" "fc1000" "fc1000_softmax" "ClassificationLayer_fc1000"	"pool5" "fc1000" "prob" "ClassificationLayer_predictions"
Input Size	230 x 230 x 3		
Dilated Layers	"res5a_branch2b" "res5b_branch2a" "res5b_branch2b"	"res5a_branch2b" "res5b_branch2a" "res5b_branch2b" "res5b_branch2c" "res5c_branch2a" "res5c_branch2b" "res5c_branch2c"	"res5a_branch2b" "res5b_branch2a" "res5b_branch2b" "res5b_branch2c" "res5c_branch2a" "res5c_branch2b" "res5c_branch2c"
Skip Layer	'res2b_relu'	'activation_10_relu'	'res2b_relu'

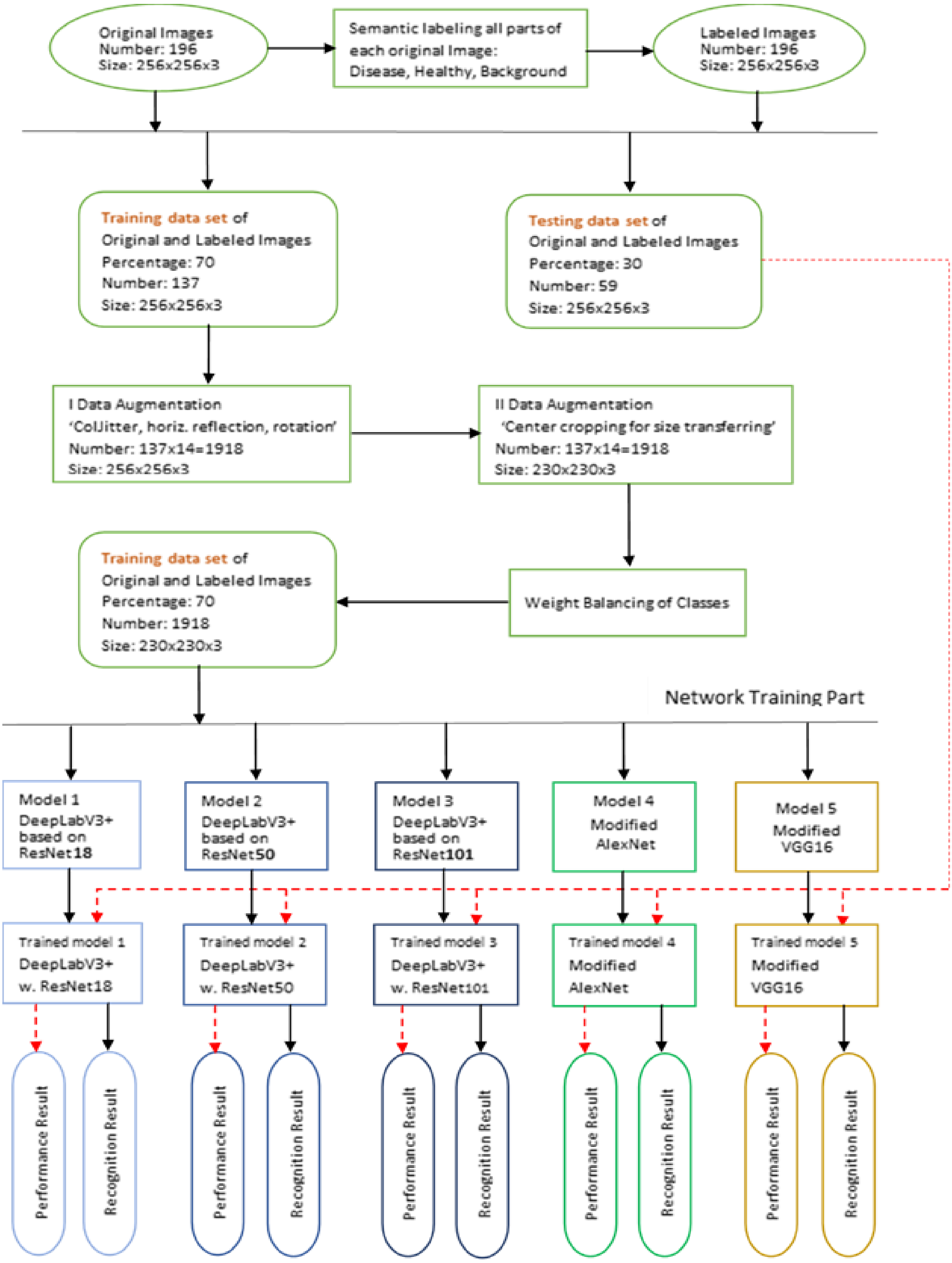


Figure 3 – Process flow of the research

Table 4 – ResNet modifications for DeepLabV3+

Parameters	ResNet-18	ResNet-50	ResNet-101
FEL	'res5b_relu'	'activation_49_relu'	'res5c_relu'
Removed Layers	"pool5" "fc1000" "prob" "ClassificationLayer_predictions"	"avg_pool" "fc1000" "fc1000_softmax" "ClassificationLayer_fc1000"	"pool5" "fc1000" "prob" "ClassificationLayer_predictions"
Input Size	230 x 230 x 3		
Dilated Layers	"res5a_branch2b" "res5b_branch2a" "res5b_branch2b"	"res5a_branch2b" "res5b_branch2a" "res5b_branch2b" "res5b_branch2c" "res5c_branch2a" "res5c_branch2b" "res5c_branch2c"	"res5a_branch2b" "res5b_branch2a" "res5b_branch2b" "res5b_branch2c" "res5c_branch2a" "res5c_branch2b" "res5c_branch2c"
Skip Layer	'res2b_relu'	'activation_10_relu'	'res2b_relu'

SegNet network based on VGG-16

VGG-16 network is made up of 13 convolutional layers (CL) and 3 fully connected (FC) layers. The SegNet is a modified form of VGG16 architecture and is an essential semantic segmentation tool (Badrinarayanan2017). A new pixel classification layer was created with the balanced classes stated in Table 2. The current pixel classification layer in the SegNet network was removed and replaced with the new updated pixel classification layer and then this layer was connected to the Softmax layer.

Modified AlexNet network

The pre-defined AlexNet network was modified in this paper to estimate the performance of the model for semantic segmentation. The image size of the input layer (the first layer) was resized to 230x230x3. Weights and bias parameters of the 17th and 20th layers (FC layers) were updated. The last two layers were removed from the predefined model. A new pixel classification layer was created with the balanced classes stated in Table 2. The current pixel classification layer in the SegNet network was removed and replaced with the new updated pixel classification layer and then this layer was connected to the Softmax layer.

Training of the CNN-based Models

After reconstructing the five CNN models, the training process commenced using stochastic gradient descent with momentum (SGDM) optimizer. The training was implemented in MATLAB 2019b platform utilizing deep learning and computer vision toolbox on a graphics processing unit NVIDIA GeForce GTX 1050 Ti, 4GB RAM, Intel Core i5 8th Gen CPU. Figure 4 shows the accuracy training performance up to 10,000th iterations, for Alexnet, Resnet-18, Resnet-50, Resnet-101, and VGG-16 models respectively. From the figure, AlexNet showed the lowest performance and ResNet-18 outperformed the other models based on the accuracy rate.

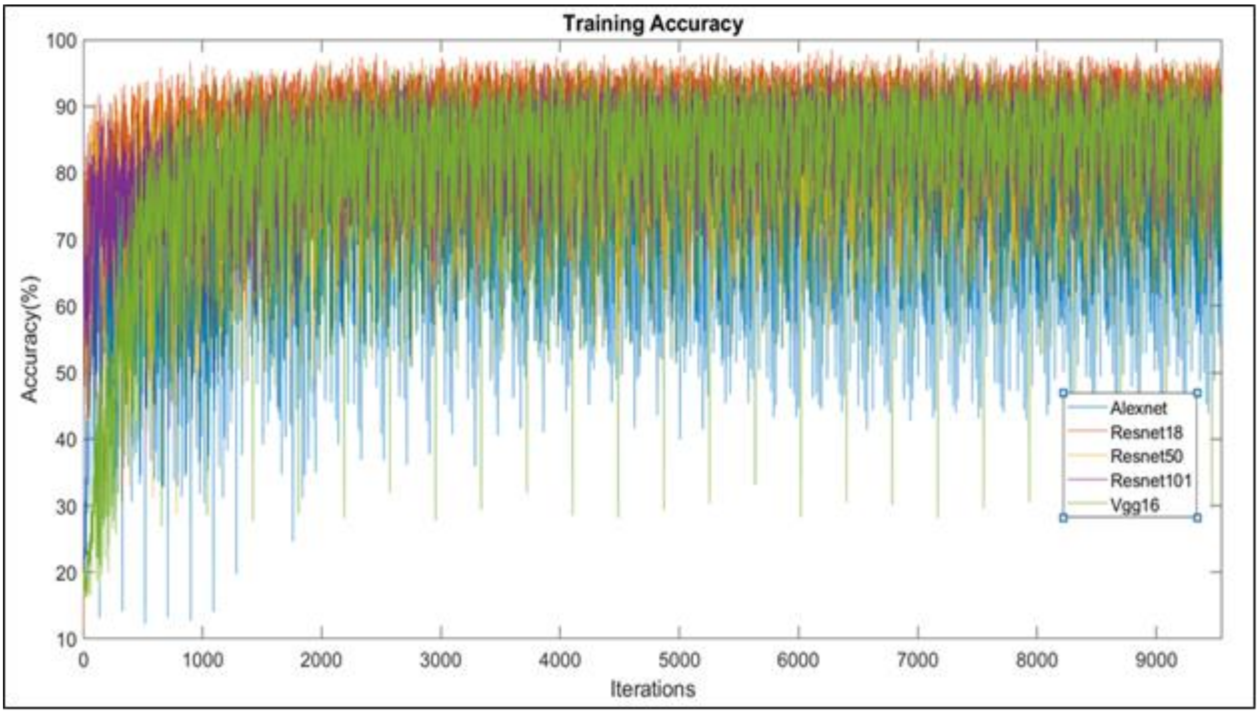


Figure 4 – Accuracy rate of training Alexnet, Resnet-18, Resnet-50, Resnet-101, and VGG-16 models

4. Performance Evaluations

Testing was conducted using 59 leaf images to evaluate the performance of all five networks. An example of the segmentation results of each trained network, ResNet18/50/101, AlexNet, and VGG-16 can be seen in Figure 5. The original test image is shown in the first row, followed by the ground truths and the segmented results of all trained networks. Each pixel of the network's output was compared to the corresponding pixel in the ground truth labeled image. The overlapping pixels (IoU) are demonstrated in the last row of Figure 5. Particularly, the green and magenta regions highlighted areas where the segmentation outcomes differed from the ground truth.

Table 5 – Overall performance of the CNN-based models

CNN Model	GlobalAccuracy	MeanAccuracy	MeanIoU	WeightedIoU	MeanBFScore
ResNet-18	0.9580	0.8486	0.6994	0.9252	0.8399
ResNet-50	0.8769	0.5729	0.3956	0.8032	0.4376
ResNet-101	0.8760	0.7132	0.4634	0.8010	0.4609
VGG-16	0.9291	0.4192	0.3582	0.8895	0.7620
AlexNet	0.8483	0.6780	0.3825	0.7801	0.4206

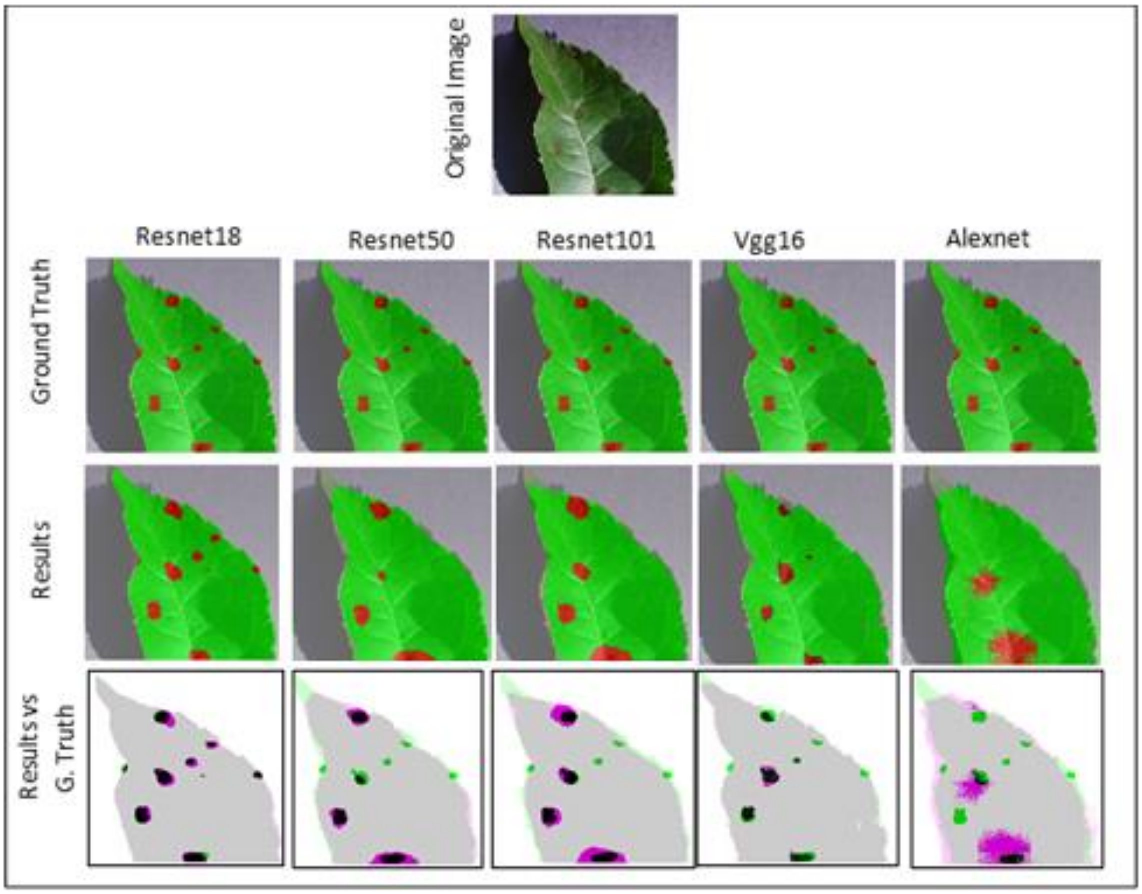


Figure 5 – An example of segmentation using ResNet18/50/101, AlexNet, and VGG-16 models.

The overall performance of the five models is shown in Table 5. Five metrics were used to evaluate the performance of the five CNN-based models. The global accuracy reflects the percentage of correctly classified pixels regardless of class, while the mean accuracy refers to the percentage of correctly identified pixels for each class. Again, ResNet-18 showed the highest percentage of correctly segmented pixels with 95.80% followed by VGG-16 at 92.91%. However, upon inspecting the mean accuracy, the highest correctly segmented pixels of VGG-16 scored the lowest average per class score of 41.92% indicating its performance for each class was not favourable. ResNet-18, on the other hand, remained as the top scorer of mean accuracy at 84.86%. Two metrics describing the Intersection-Over-Union were Mean IoU and Weighted IoU. In general, the Mean IOUs of all networks showed low percentages: ResNet-18 achieved a mere 69.94%, followed by ResNet-101 at 46.34 % and the other models achieving below 40%. The Mean IoU reflects that there are high false positives during the semantic segmentation of the leaf images. A fairer evaluation should utilize the Weighted IoU due to the disproportionately sized classes of the leaf images. The Weighted IoU shows a high score of 92.52% for ResNet-18 followed by 88.95% for the VGG-16 model. The final metric that was considered was MeanBFscore, which calculated how well the predicted boundary of each class aligned with the actual boundary. Based on Table 5, the ResNet-18 model consistently outperformed the other four models in all metric evaluations. Even though VGG-16 also showed promising results, its Mean Accuracy and Mean IoU were rather poor indicating its low robustness per class. Meanwhile, AlexNet achieved the lowest Global Accuracy, Weighted IoU and MeanBFScore indicating its low performance for semantic segmentation of leaf disease, in general.

Further investigations were done to study the impact of each class contributing to the overall performance of each network. The average Accuracy, IoU, and MeanBFScore results of each class are presented in Table 6. Overall, the class Background which contains pixels belonging to non-leaf has the highest accuracy rate, IoU and MeanBFScore. Healthy_part class which represented the non-infected pixels of the leaf was also another class that achieved good results indicating that the pixels were generally correctly segmented. The other four classes which represented the leaf diseases, however, were not segmented well. The Fungi class achieved the highest accuracy of 67.53% among the disease classes, followed by the Bacteria class at 49.29%, Nematodes at 48.47% and Virus at 37.62%. A closer inspection was then done to evaluate the semantic segmentation according to the CNN-based models.

Table 6 – Overall performance of the class measured by average accuracy, IoU and MeanBFScore

Class	Accuracy	IoU	MeanBFScore
Bacteria	0.4929	0.2581	0.40
Fungi	0.6753	0.2736	0.3164
Nematodes	0.4847	0.2187	0.3222
Virus	0.3762	0.2584	0.3914
Healthy_Part	0.9038	0.8376	0.5448
Background	0.9451	0.9123	0.7774

Table 7 listed the accuracy, IoU, and MeanBFScore of each class. The ResNet-18 model achieved the highest accuracy rates for all four disease classes compared to the other CNN-based models. Interestingly, even though VGG-16 achieved the second-highest global accuracy rate as stated in Table 5, its MeanAccuracy was low because all the accuracies of the 4 disease classes were very poor. The Bacteria class scored an accuracy rate of 39.11%, followed by the Virus class at 18.40% and both the Nematodes and Fungi classes failed miserably. Also, even though AlexNet was ranked lowest based on the GlobalAccuracy rate in Table 5, the average accuracy of all the disease classes segmented by AlexNet achieved better accuracy than VGG-16 and ResNet-50. Therefore, based on per class accuracy ResNet-18 is the best CNN-based model in segmenting plant disease images followed by ResNet-101, AlexNet, Res-50 and VGG-16 model.

Table 7 – Overall performance of the CNN-based models

	ResNet-18			**ResNet-50**		
Classes	**Accuracy**	**IoU**	**MeanBFScore**	**Accuracy**	**IoU**	**MeanBFScore**
Bacteria	0.6802	0.5444	0.6664	0.4614	0.1588	0.2840
Fungi	0.9165	0.6507	0.7879	0.7964	0.2448	0.3176
Nematodes	0.8856	0.5946	0.8438	0.2043	0.1543	0.2099
Virus	0.6644	0.5102	0.6901	0.1562	0.1322	0.1294
Healthy_Part	0.9643	0.9264	0.8387	0.9070	0.8079	0.3787
Background	0.9803	0.9699	0.9379	0.9117	0.8758	0.7023
	ResNet-101			**VGG-16**		
Classes	**Accuracy**	**IoU**	**MeanBFScore**	**Accuracy**	**IoU**	**MeanBFScore**
Bacteria	0.5063	0.2797	0.2929	0.3911	0.1392	0.5432
Fungi	0.7997	0.3016	0.3087	0.0001	0.0001	0.0164
Nematodes	0.6636	0.2369	0.3680	0.0001	0.0001	0.0133
Virus	0.5207	0.2979	0.2901	0.1840	0.1416	0.5618
Healthy_Part	0.8849	0.7973	0.3753	0.9594	0.9103	0.7623
Background	0.9041	0.8668	0.6965	0.9805	0.9578	0.8603
AlexNet						

Classes	Accuracy	IoU	MeanBFScore
Bacteria	0.4257	0.1688	0.2135
Fungi	0.8638	0.1706	0.1515
Nematodes	0.6702	0.1084	0.1761
Virus	0.3561	0.2102	0.2856
Healthy_Part	0.8034	0.7459	0.3692
Background	0.9488	0.8911	0.6898

5. Conclusion

This paper compared the performance evaluation of five commonly used CNN-based models for the semantic segmentation of plant leaf diseases. The experiments were conducted using the Leaf Disease image dataset to segment four common leaf diseases, namely bacteria, fungi, nematodes and virus. The initial overall results revealed that the ResNet-18 model has the highest correctly classified pixels with the highest global accuracy and VGG-16 achieved the second-highest accuracy. AlexNet, however, recorded the lowest global accuracy among all the CNN-based models. Upon closer review of the per class evaluations, the mean accuracy was retained by the ResNet-18 model indicating that this model consistently classified the disease pixels correctly both at the image and per class levels. Among the disease classes, the Fungi class achieved the highest accuracy followed by the Bacteria class, Nematodes and Virus class. For unknown reasons, the VGG-16 model performed worst at segmenting the disease pixels. The findings showed that when performing semantic segmentation, it is important to analyze accuracy at global level and class level. Even though the CNN-based network achieved good accuracy at global level, the same was not true for class level. Therefore, if class identification is crucial, a CNN model with high mean accuracy is recommended. On the other hand, if only general detection is to be done, a CCN model with satisfactory global accuray suffices. For future research, the class imbalanced issues need to be resolved for better semantic segmentation. Also, deeper insights into the segmentation at the class level should be investigated to better understand how the CNN-based models worked.

BIBLIOGRAPHY

1. ANNABEL, L.S.P.; ANNAPOORANI, T.; DEEPALAKSHMI, P. Machine learning for plant leaf disease detection and classification - A review. Proc. 2019 IEEE Int. Conf. Commun. Signal Process. ICCSP 2019, pp. 538–542, 2019.
2. ZHANG S.; ZHANG C.; WANG, S.; SHI, Y. Cucumber leaf disease identification with global pooling dilated convolutional neural network. Comput. Electron. Agric., v. 162, p. 422–430, December 2018.
3. BADRINARAYANAN, V.; KENDALL, A.; CIPOLLA, R. SegNet: A Deep Convolutional Encoder-Decoder Architecture for Image Segmentation, IEEE Trans. Pattern Anal. Mach. Intell., v. 39, n. 12, p. 2481–2495, 2017.
4. BARBEDO, J.G.A. Factors influencing the use of deep learning for plant disease recognition. Biosyst. Eng., v. 172, pp. 84–91, 2018.
5. BARBEDO, J.G.A. Plant disease identification from individual lesions and spots using deep learning. Biosyst. Eng., v. 180, p. 96–107, 2019.
6. BALAKRISHNA K.; RAO, M. Tomato plant leaves disease classification using KNN and PNN, Int. J. Comput. Vis. Image Process., v. 9, n. 1, pp. 51–63, 2019.
7. CHEN, J.; LIU, Q.; GAO, L. Visual tea leaf disease recognition using a convolutional neural network model. Symmetry (Basel)., v. 11, n. 3, 2019.
8. CHEN, L.C.; ZHU, Y.; PAPANDREOU, G.; SCHROFF, F.; ADAM, H. Encoder-decoder with atrous separable convolution for semantic image segmentation. In Proceedings of the European

Conference on Computer Vision (ECCV), p. 801-818, December 2018.
9. FERENTINOS, K.P. Deep learning models for plant disease detection and diagnosis. Comput. Electron. Agric., v. 145, p. 311–318, January 2018.
10. GEETHARAMANI G.; PANDIAN, A. Identification of plant leaf diseases using a nine-layer deep convolutional neural network. Computers & Electrical Engineering, v. 76, p. 323-338, 2019.
11. HE, K.; ZHANG, X.; REN, S.; SUN, J. Deep residual learning for image recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognitionp. p. 770-778, 2016.
12. HUGHES, D.; SALATHÉ, M. An Open Access Repository of Images on Plant Health to Enable the Development of Mobile Disease Diagnostics. arXiv preprint arXiv:1511.08060, 2015.
13. KAGGLE, “Image Dataset on Kaggle,” Kaggle, 2019. [Online]. Available: https://www.kaggle.com/sizlingdhairya1/leaf-disease.
14. KC, K.; YIN, Z.; WU, M.; WU, Z. Depthwise separable convolution architectures for plant disease classification, Comput. Electron. Agric., v. 165, p. 104948, 2019.
15. KHAMPARIA, A.; SAINI, G.; GUPTA, D.; KHANNA, A.; TIWARI, S.; DE ALBUQUERQUE, V.H.C. Seasonal crops disease prediction and classification using Deep Convolutional Encoder Network, Circuits, Systems, and Signal Processing, v. 39, n. 2, p. 818-836, 2019.
16. KRIZHEVSKY, A.; SUTSKEVER, I.; HINTON, G.E. Imagenet classification with deep convolutional neural networks. Advances in Neural Information Processing Systems, v. 25, p. 1097-1105, 2012.
17. LIANG, Q.; XIANG, S.; HU, Y.; COPPOLA, G.; ZHANG, D.; SUN, W. PD 2 SE-Net: Computer-assisted plant disease diagnosis and severity estimation network, Comput. Electron. Agric., v. 157, p. 518–529, 2019.
18. LIU, B.; ZHANG, Y.; HE, D.; LI, Y. Identification of apple leaf diseases based on deep convolutional neural networks. Symmetry, v. 10, no. 1, p. 11, 2018.
19. MOHANTY, S.P.; HUGHES, D.P.; SALATHÉ, M. Using deep learning for image-based plant disease detection. Frontiers in Plant Science, v. 7, p. 1419, 2016.
20. OERKE, E.C., Crop losses to pests. The Journal of Agricultural Science, v. 144 no. 1, pp.31-43, 2006.
21. OPPENHEIM, D.; SHANI, G.; ERLICH, O.; TSROR, L. Using deep learning for image-based potato tuber disease detection. Phytopathology, v. 109, n. 6, p. 1083–1087, 2019
22. PETRELLIS, N. Plant disease diagnosis with color normalization. 2019 8th Int. Conf. Mod. Circuits Syst. Technol. MOCAST 2019, pp. 1–4, 2019.
23. RAMCHARAN, A.; MCCLOSKEY, P.; BARANOWSKI, K.; MBILINYI, N.; MRISHO, L.; NDALAHWA, M.; LEGG, J.; HUGHES, D.P. A mobile-based deep learning model for cassava disease diagnosis. Frontiers in plant science, v. 10, p. 272, 2019.
24. SIMONYAN K.; ZISSERMAN, A. Very Deep Convolutional Networks for Large-Scale Image Recognition, 3rd Int. Conf. Learn. Represent. ICLR 2015 - Conf. Track Proc., p. 1–14, 2015.
25. SINGH, K.; KUMAR, S.; KAUR, P. Support vector machine classifier based detection of fungal rust disease in Pea Plant (Pisam sativam), Int. J. Inf. Technol., v. 11, n. 3, pp. 485–492, 2019.
26. TOO, E.C.; YUJIAN, L.; NJUKI, S.; YINGCHUN, L. A comparative study of fine-tuning deep learning models for plant disease identification, Comput. Electron. Agric., v. 161, p. 272–279, 2019.