Plant Disease Classification Using Hybrid Features

Vamsidhar Muthireddy, C.V. Jawahar

International Institute of Information Technology, Hyderabad

Abstract. Deep learning has shown remarkable performances in image classification, including those of plants and leaves. However, high-performing networks in terms of accuracy may not be using the salient regions for making the prediction and could be prone to biases. This work proposes a neural network architecture incorporating handcrafted features and fusing them with the learned features. Using hybrid features provides better control and understanding of the feature space while leveraging deep learning capabilities. Furthermore, a new IoU-based metric is introduced to assess the CNN-based classifier's performance based on the regions focused on making predictions. Experiments over multiple leaf disease datasets demonstrate the performance improvement with the model using hybrid features. Classification using hybrid features performed better than the baseline models in terms of P@1 and also on the IoU-based metric.

Keywords: Hybrid Network · Plant disease classification · PlantVillage.

1 Introduction

Agriculture has been a vital pillar in sustaining civilizations. It plays an irreplaceable role not just in feeding the populace but also in generating economic value. In recent times, the agricultural yield has declined [1]. There is a growing interest in Precision Agriculture (PA) to maintain sufficient conditions of the farm during the growth phase of the crop [36, 8] and the imaging tools for tracking the growth of the harvest and early detection of the crop diseases [28]. Effective designing of these tools leveraging Image processing techniques and machine learning algorithms for plant disease detection is an active research problem.

Traditional plant disease recognition algorithms used handcrafted features containing a combination of shape, colour, and texture-based features. A shallow classifier is subsequently used for the classification task [19]. Disease recognition systems built on such algorithms were difficult to scale and generalize due to the requirement of specific domain knowledge. Moreover, any change in imaging conditions would impact the efficiency of the systems. On the other hand, neural networks are capable of learning a data-specific feature space. Convolutional neural networks have been used for plant disease recognition [25]. However, to improve the baseline performance of the neural networks for a particular task, either a large amount of data or a change in architectural design is needed.

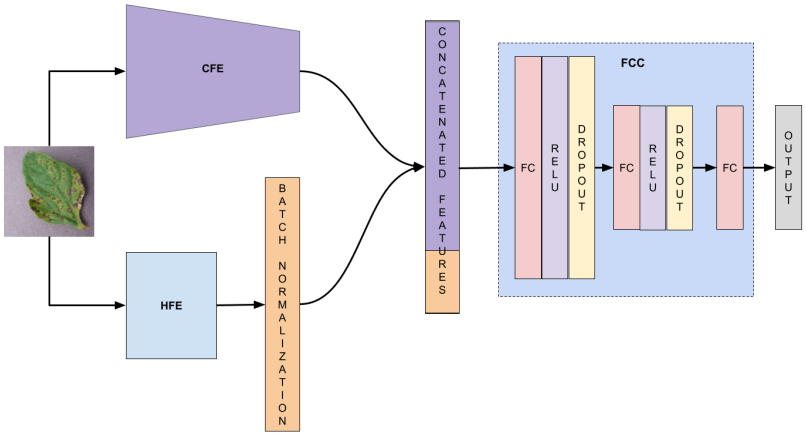


Fig. 1: This figure shows our proposed hybrid network. Convolutional block of Inception-V3 network is used as CFE. Features from the HFE block are concatenated and fed into FCNN for predicting output softmax scores.

Data fusion is a methodology where designed neural networks contain an architectural provision to inject additional features at different layers of the network [9]. We propose one such architecture to leverage handcrafted features in the training process of the CNN. We compare the efficacy of our designed network against other CNNs for plant disease classification. In addition, we compare the performance of our network against the baseline networks on similar datasets to show its efficacy on plant disease classification.

In a classification task, the algorithms are evaluated based on their ability to predict an accurate label for a given input [14]. This is in contrast to object detection tasks, where the algorithms are evaluated on their ability to localize the predicted label in the input image along with their ability to predict an accurate label. In this work, the necessity to integrate the localization capacity of the CNN while assessing its efficacy on the task of object classification is presented. In traditional classification algorithms, handcrafted features are designed to primarily encode the foreground object. This ensures that the shallow classifiers built using such features would be using encoded foreground objects in their task. However, CNN classifiers would select the feature space optimal for the classification task. With no restrictions in place to regulate or assess this behaviour, these classifiers might select a significant portion of the feature space from the background objects [31]. Coupled with the tendency to generate train, validation, and test sets to follow a similar data distribution would make it relatively complicated to diagnose this issue unless tested on real-world data. Therefore, CNNs trained on the classification task should be assessed on their ability to select the feature space from the foreground object. Employing visualization techniques could alleviate this issue by providing a gateway to understanding the feature space of the neural network with respect to an input image. We utilize one such approach to analyze our CNN-based classifiers trained on the Plant disease recognition task [35]. Further, we generate quantitative results to assert the importance of inte-

grating the localization capacity of the network while assessing its classification performance.

The contributions of this work can be summarized as follows:

- We propose a hybrid CNN architecture for plant disease recognition which utilizes the data fusion technique to inject handcrafted features into the first FC layer, allowing the CNN to learn the complimentary feature space to improve the classification accuracy.
- We provide the quantitative and qualitative experimental results that show the importance of assessing CNN classifiers on their ability to learn the feature space from the foreground object, along with the accuracy of predicting labels.

2 Related Works

Earlier works on plant disease detection performed the task by selecting the features by Molecular techniques [34, 32, 24]. These methods did not provide real-time feedback on the health of the plant. Features extracted from non-invasive methods like spectroscopy, and imaging techniques [6, 29, 33] have provided an opportunity to develop and deploy systems that can be used to assess the health of a plant in real-time. While able to identify plant diseases in real-time, these techniques have a heavy cost factor associated with the equipment needed to collect the data. A picture-based analysis has been proposed to identify plant diseases that have visible parts of the symptoms. These approaches usually collect image data using mobile phones [17]. Hand-crafted features are extracted from the images using classical image-processing techniques. These features in general are comprised of colour, shape, texture [27], a combination of the above [19, 41, 30, 38]. These features are combined with shallow discriminative classifiers to perform plant disease recognition [19, 18, 30, 38].

The classification pipelines built with hand-crafted features rely heavily on the prior knowledge of the task and dataset. Selecting a task-specific feature space or a classification algorithm takes tedious effort. To tackle these constraints, convolutional neural networks have been used for plant disease recognition [4, 5, 25, 39]. CNNs optimally select the feature space specific to the task, hence alleviating the task of feature selection [23]. In addition, when integrated with the training pipeline, the data augmentation framework increases the generalizability of the network [21, 13].

Although CNNs can learn optimal feature space, outperforming shallow classifiers built on hand-crafted features, they still lack the controllability of such learned space. Ensemble networks tackle this issue by combining the features extracted from the FC layer of the CNN with the hand-crafted features and feeding them into a shallow classifier [10]. However, in these types of architectures, CNNs do not learn a feature space that works along with the hand-crafted features since the network weights are not optimized after data fusion. To address this, data fusion techniques have been proposed that concatenate the hand-crafted features with the FC layers of a CNN at different levels of abstraction [9]. We

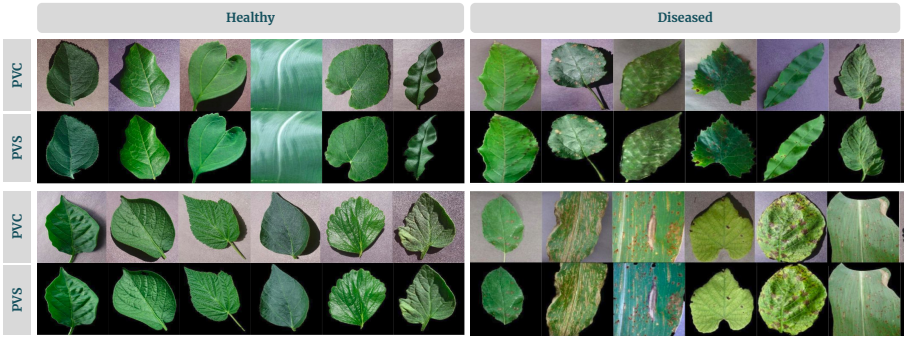


Fig. 2: This figure shows images of leaves belonging to 12 healthy plant species. Rows 1, 3 show the images from PlantVillage Colour(PVC) and rows 2, 4 show their counterparts from PlantVillage Segmented (PVS) dataset

propose a network along a similar approach to show the efficacy of the data fusion techniques for leaf disease recognition.

Accuracy, Precision@K, F score are some of the metrics used to obtain the efficacy of the models trained on the classification task [14]. Contrarily, models trained on object detection tasks include IoU to assess the localization behaviour of the model along with its classification capability. Enhancements have been made to the evaluation metrics of the object detection task [15, 42]. During inference, these methods return a confidence score by weighing classification accuracy with an IoU score. We propose using a similar metric for evaluating the models trained on the classification task. For a generation of the object mask to calculate IoU, we use the class-activated mapping obtained by Grad-CAM [35] to identify the salient regions in an image for a prediction. These salient regions are then thresholded at different values to obtain the object masks [3, 43]. In order to establish the necessity of modifying the evaluation methodology of classifiers, we provide qualitative and quantitative results demonstrating the utility of the proposed metrics.

3 Dataset

To validate the efficacy of the proposed hybrid pipeline, we compare its performance over datasets that are composed of images of diseased leaves. These datasets are described below.

PlantVillage dataset: The PlantVillage dataset is a corpus containing 54K images of healthy and unhealthy plant leaves. The dataset is divided into 38 different plant-disease pairs. Images in these classes belong to 16 plant species and contain 26 leaf diseases. Every image in the PlantVillage dataset contains a single leaf in three different configurations, *Color*, *Segmented*, and *Grayscale*.

Plant Pathology: Plant Pathology Dataset [7] contains leaf images of twelve plants in healthy and diseased conditions. It consists of 4503 images (2278 healthy; 2225 diseased). Overall, a total of 22 classes are available for a classification task. All the leaf images were collected from March to May in a controlled environment.

PlantDoc: PlantDoc [37] is a publically available dataset introduced in 2020. It covers 13 plant species through 2,598 images across 27 classes (10 healthy; 27 diseased) for image classification and object detection tasks. In contrast to the Plant Village dataset and Plant Pathology Dataset, where the images are taken in a controlled environment, the images in PlantDoc are curated from natural settings containing multiple leaves against different backgrounds and varying light conditions.

DiaMOS Plant: DiaMOS Plant [11] is a dataset of images of an entire growing season of pear trees (Cultivar: Septoria Piricola), from February to July, under realistic field conditions in Itlay. It comprises 3505 images, including 499 fruit images and 3006 leaves images.

4 Methods

Our proposed hybrid network utilizes data fusion techniques between two components - a convolutional neural network and a handcrafted feature extractor. In this section, we use the term Convolutional Feature Extractor (CFE) for the convolutional block of the neural network. The handcrafted feature block that extracts a set of pre-defined features from an input image is termed Handcrafted Feature Extracter (HFE). The fully connected block of the neural network is referred to as FCNN.

An overview of the hybrid network is given in Fig. 1. Features obtained from CFE and HFE are concatenated into a feature vector for an input image. This feature vector is fed into FCNN to obtain the classification result.

4.1 Handcrafted feature extractor (HFE)

For an input image, 258 features are extracted by the HFE. They contain 120 colour, 54 morphological, and 84 texture features. These features are concatenated to obtain a 258-dimensional feature vector. A batch normalization layer normalizes this feature vector.

Colour features are obtained by processing each channel of the RGB and HSI colour space. Each color channel's mean, standard deviation, minimum, and maximum intensity values are computed. Each channel is further processed to obtain a 16 binned histogram from 256 gray levels. All the obtained features are united to generate a 120-dimensional colour feature vector.

Morphological features include area, perimeter, major axis length, minor axis length, minimum radius, seven Hu moments, area of the convex hull, solidity, and Fourier descriptor composed of 20 harmonics. Fourier descriptor was obtained on the contour of the binary image. All the obtained features are united to generate a 54-dimensional morphological feature vector.

Texture features are extracted from the image's Red, Green, Blue, and Grayscale channels. They are composed of properties of gray level co-occurrence matrix(GLCM) and 13 Harlick features. Each of the channels is quantized to 32 intensity levels prior to the extraction of features. From the GLCM obtained for each channel, mean, standard deviation, contrast, correlation, homogeneity, dissimilarity, energy, and Angular Second Moment(ASM) are extracted. All these obtained features are united to generate an 84-dimensional texture feature.

4.2 Convolutional Feature Extractor (CFE)

We use the Inception-v3 [40] to build our CFE. We remove the final FC layer of the Inception-v3 to obtain the feature vector. This 2048 dimensional feature vector is extracted after the Average pooling layer. We use the PlantVillage dataset to explore plant disease recognition in this work. The disease parts in the images are not restricted to a particular size. Multi-scale convolutions in Inception-v3 will encode these saliency regions for better classification. Further, CFE composed of convolutional blocks of Inception-v3 will better establish the efficacy of our proposed hybrid model against the baseline.

4.3 Fully Convolutional Neural Network (FCNN)

Features extracted from the CFE and HFE are concatenated to form a 2316 dimensional feature vector. This feature is fed into FCNN to obtain the classification prediction. Our designed FCNN is composed of three Fully-Connected layers. The first and second FC layers are composed of 256 and 128 neurons, respectively, while the third FC layer is composed of c neurons, where c denotes the count of unique class labels in the classification task. Each of the first two FC layers is followed by a ReLU activation layer and a dropout layer for regularization. A softmax layer follows the third FC layer to perform classification.

4.4 Model validation

We validate the models trained for Plant disease classification using P@K and IoU. After the convolutional networks are trained, we use GradCAM [35] to obtain the heatmaps. These heatmaps are used as the predicted localization masks for the foreground objects. The localization mask is thresholded at k where $k \in (0, 1)$ to convert into the binary mask. The obtained binary mask is used to compute Intersection over Union (IoU) scores at a threshold of k. We evaluate the performance of the classification model using IoU based on the underlying assumption that the features used in a classification task should be obtained from the foreground objects in an image.

5 Experimental setup

Our experiments are composed of two tasks: finding an optimal network for the classification of plant diseases and evaluating the trained network using the gradient-activated heatmaps. To find the optimal network, first, we compare the classification performance of the proposed Inception-V3 hybrid network with the baseline Inception-V3 network. Further, we experimented with networks belonging to the ResNet family. All these networks are trained on the PlantVillage datasets (PVC, PVS) and pre-trained on the ImageNet dataset. Moreover, to further validate the proposed approach, multiple other datasets curated for plant diseases were used for experimentation. To explore the significance of in-domain dataset, the pretraining was done on the Imagenet or PVC dataset.

The datasets are split into train, validation, and test sets by keeping their ratio at 60:20:20. The batch size across all the experiments is set to 24. We use Stochastic Gradient Descent (SGD) as an optimizer with a momentum of 0.9.

Model	P@1	
	PVC	**PVS**
Inception-V3(baseline)	99.3400	99.2500
Resnet18	99.5672	99.3463
Resnet34	99.4383	99.2600
Resnet50	99.6133	99.3739
Resnet101	99.7145	99.4568
Resnet152	**99.7422**	99.3463
Inception-V3 hybrid	*99.7330*	**99.5580**

Table 1: The table shows the P@1 values of different models on the two configurations of PlantVillage dataset. Resnet152 outperforms the rest of the models in the PVC configuration with P@1 of 99.742. Inception-V3 Hybrid outperforms the rest of the models in the PVS configuration with P@1 of 99.558

Model	P@1 - PVC
AlexNet [12]	99.44
VGG [12]	99.53
MobileNet [20]	98.65
EfficientNet B5 [2]	98.42
CNN-RNN [22]	98.77

Table 2: P@1 values obtained by other works on PlantVillage dataset. [12] has reported the highest test accuracy of 99.53.

We keep the learning rate as 0.01 for networks belonging to ResNet architecture while we keep the learning rate of 0.005 for the Inception-V3 hybrid network. The learning rate is decayed by a factor of 0.1 for no improvement in validation loss for ten consecutive epochs.

We evaluate the trained networks under the fair assumption that a neural network needs to look at the foreground object of interest while classifying an image into a particular class. We generate the gradient-activated heatmaps utilizing the GradCAM network. These heatmaps represent the areas of interest in an input image identified by a network for a predicted class label. To obtain the binary masks, we threshold these gradient activations at bins ranging from 0 to 1. These binary maps are used to generate IoU scores across the multiple threshold values. We use this metric to evaluate the focus of the network. The results generated are presented in Table 4 and are further discussed in the following section.

6 Results

To show the efficacy of the proposed hybrid model, we compared it with different classification networks. Table 1 shows the obtained P@1 scores for the classification part of the experiments. All the tested models have a better P@1 compared to the baseline model in both configurations of the dataset. Resnet-152 has the best top P@1 accuracy in the PVC configuration of the dataset(P@1 = 99.74). Inception-V3 hybrid has the best accuracy in the PVS configuration of the dataset(P@1 = 99.56). Models trained on PVC configuration have higher P@1 when compared to their counterparts trained on PVS configuration of the dataset by a considerable margin.

Dataset	Pre-trained	Model	
		Inception-V3	Inception-V3 hybrid
Diamos	ImageNet	82.72 [11]	88.52
	ImageNet + PVC	89.02	**89.33**
plantDoc	ImageNet	41.10	51.69
	ImageNet + PVC	29.73	**53.81**
plantDocDetection	ImageNet	46.67 [37]	61.5
	ImageNet + PVC	62.06 [37]	**67.92**
Plant Leaves	ImageNet	94.55 [16]	98.56
	ImageNet + PVC	98.82	**99.16**

Table 3: The table shows the P@1 values obtained on multiple plant disease datasets using Inception-V3 (baseline) and Inception-V3 hybrid (proposed) networks. It can be noted that our proposed hybrid methodology outperforms the baseline reported. Pre-training on the domain specific PlantVillage dataset (PVC) improves the P@1 of the models

The comparison between the baseline Inception-V3 architecture with the proposed Inception-V3 hybrid approach across multiple plant disease datasets is presented in Table 3. It can be observed that our proposed hybrid model outperforms the baseline model for all the datasets. The difference in performance for the plantDoc and plantDocDetection detection datasets is most pronounced. Pretraining on the domain-specific dataset PVC yielded much better results than the pretraining on the ImageNet dataset.

Models trained on both dataset configurations are further interpreted using GradCAM architecture. Figure 5 shows the results obtained on healthy and unhealthy leaves. These heatmaps are generated from the last convolutional layer of each trained network in Table 1. This is done under the fair assumption that these layers feed directly into the softmax layer, which acts as a final classifier. The figure contains six rows of sub-figures; the odd positioned rows(1, 3, 5) have images from PVC configuration, while the even positioned rows (2, 4, 6) have their counterparts from the dataset's PVS configuration. These results were obtained to infer the model behaviour with respect to a classification; to find the parts of the leaf which the network is using to identify a particular plant-disease combination.

The heatmaps generated by Grad-CAM are used to convert the obtained qualitative results in the Figure 5 into quantitative results. These results are presented in Table 4. For a particular classification task, we assess the performance of the classification model based on the percentage of foreground regions chosen to classify an image. The Grad-CAM heatmaps are converted into binary masks at ten threshold ranges from 0 to 1. These are treated as the predicted binary masks. For the ground truth, we utilize the PVS configuration of the dataset to generate binary masks. We calculate IoU scores at each of these thresholds for different networks on both dataset configurations. Figure 3 shows the obtained results across different threshold values. We use the obtained IoU values to quantify the model performance in two ways. First, we measure the $IoU@k$ to show

Model	PVC		PVS	
	IoU@0.5	**meanIoU**	**IoU@0.5**	**meanIoU**
Resnet18	0.5125	0.3801	0.4747	0.3715
Resnet50	0.0666	0.1016	0.0919	0.1223
Resnet101	0.2174	0.2204	0.2905	0.2646
Resnet152	0.2572	0.2433	0.2821	0.2627
Inception-V3 hybrid	**0.5468**	**0.3986**	**0.5031**	**0.3771**

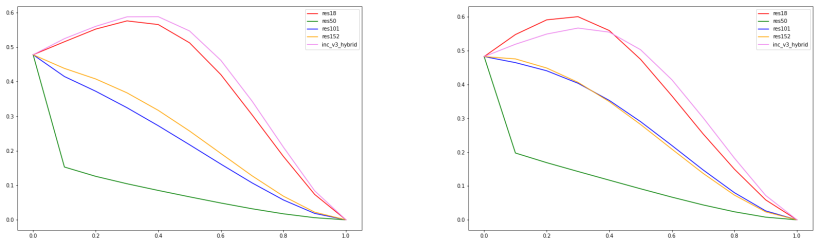
Table 4: The table shows the IoU values obtained between heatmaps generated by Grad-CAM with foreground binary mask. IoU@0.5 represents the scores when the Grad-CAM is thresholded at 0.5. Mean IoU scores are obtained by taking the mean of IoU across different threshold values. In both PVC, and PVS configuration of the data, Inception-V3 hybrid outperforms all other models in IoU@0.5 and mean IoU scores.

the network performance in being able to identify the foreground with higher values. Following the standard norms, we select $k = 0.5$. Second, we calculate the mean IoU values obtained by averaging the IoU at different threshold values to evaluate the model performance across the range of 0 to 1. From Table 4, it is evident that the Inception-V3 hybrid network outperforms all other networks across $IoU@k$ and $meanIou$ metrics. For PVC dataset configurations, Inception-V3 hybrid has $IoU@0.5$ of 0.5468 and $meanIoU$ of 0.3998. For PVS dataset configurations, Inception-V3 hybrid has $IoU@0.5$ of 0.5031 and $meanIoU$ of 0.3771. Both $IoU@0.5$ and $meanIoU$ values of the Inception-V3 hybrid network are higher for PVC when compared with the PVS configuration of the dataset.

We conducted the above analysis on the scan-like plant leaf datasets, Swedish-leaf, Flavia, Leafsnap, MEW-2012, and MEW-2014 datasets reported in [26]. We use the ResNet-101 trained on these datasets and Grad-CAM methodology to obtain the foreground heatmaps. By following the above-explained procedure, we generate $IoU@0.5$ and $meanIoU$ scores. The results obtained are reported in Table 5 and Figure 4. Swedish-leaf, Flavia, and Leafsnap have $IoU@0.5 \leq 0.06$. The IoU of the MEW-2014 dataset is 0.24, which is more significant in comparison to its subset, MEW-2012, having an IoU of 0.20. It should be noted that the ResNet-101 models used to generate the heatmaps are trained individually on each dataset. The accuracy of each of these models is above 97%.

6.1 Discussion

Different models are trained on both PVC and PVS configurations of the datasets. ResNet-152 in the PVC configuration and Inception-V3 hybrid in the PVS configuration outperform the rest of the models. We observe that all the models that we trained outperform the baseline model. It can also be noted that the Inception-V3 hybrid underperforms ResNet-152 in PVC configuration by a margin of 0.009%. The same network outperforms the second-best ResNet-101 on PVS by a margin of 0.1%, which hints at the efficacy of our hybrid model in the datasets having a similar structure to PVS.



(a) PVC (b) PVS

Fig. 3: This figure shows the IoU scores for ResNet-18, 50, 101, 152, Inception-V3 hybrid networks on threshold values ranging from 0-1. In PVS configuration, Inception-V3 hybrid has higher IoU scores across all threshold values. In PVS configuration, At lower threshold(≤ 0.5) ResNet-18 has higher IoU scores whereas at higher threshold(≥ 0.5), Inception-V3 hybrid has higher IoU scores.

To validate this, the Inception-V3 hybrid is tested on the plant disease datasets against the baseline Inception-V3 architecture as shown in Table 3. The hybrid model outperforms the baseline model in each of the datasets. Moreover, the improvement of P@1 when the models are pre-trained on PVC indicates that domain-specific initialization of the weights is a suitable way to obtain performance gains during training. The obtained results outline the efficacy of the proposed methodology over plant disease datasets.

It should also be noted that models trained on PVC outperform their counterparts trained on PVS, as shown in Table 1. Figure 2 shows the key difference between both configurations to be the presence of non-zero valued background in PVC and zero-valued background in PVS. Stating that the background is helping the network in making better classifications would be a fair conclusion in this scenario where the classifiers trained on the PVC outperform their counterparts trained on PVS. To verify this claim, Grad-CAM was used in further experiments. Using Grad-CAM, heatmaps were generated from the final convolutional layer of the networks trained on PVS and PVC configurations for each input image. Figure 5 shows the heatmaps overlapped with their respective input images for healthy and unhealthy leaves in the same order. A series of observations were made by analyzing these images. Firstly, each network identifies a specific part of an image for classifying an image. This region of interest varies from network to network. Secondly, some networks have chosen a significant part of the background as the region of interest for making a classification. This can be observed for outputs of ResNet-50, ResNet-101, and ResNet-152 on PVC, as shown in Figure 5. Thirdly, the regions of interest identified by some of the networks contain parts of the background in a significant majority. This could be interpreted as networks correctly classifying the input image by heavily relying on the background rather than the foreground, as shown in Figure 5 (a-g). Extreme cases such as Figure 5 (e) were observed where ResNet-50 generated heatmap has an uncanny resemblance to the shape of the leaf in the image;

Dataset	IoU@0.5	meanIoU
Swedish	0.0535	0.0784
Flavia	0.052	0.0743
Leafsnap	0.0481	0.0410
MEW-2012	0.2052	0.1790
MEW-2014	0.2413	0.2057

Table 5: IoU@0.5 and meanIoU values are obtained for Swedish-leaf, Flavia, Leafsnap, MEW-2012, and MEW-2014 datasets for the models reported in [26]. ResNet-101 is used too obtain the Grad-CAM heatmaps used to generate the foreground binary mask.

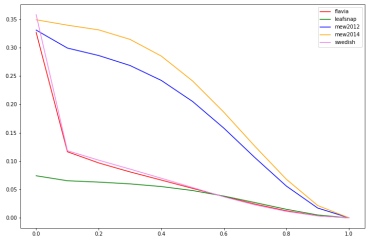


Fig. 4: This figure shows the IoU scores for ResNet-101 on threshold values ranging from 0-1 on Swedish-leaf, Flavia, Leafsnap, MEW-2012, and MEW-2014 datasets

only, in this case, it appears to have learned it from the background. ResNet 18 seems to work as expected but shows inferior performance in terms of accuracy. The inception-V3 hybrid network consistently showed better heatmaps where the major focus of the network was on the foreground. Fourthly, across all the networks, it is evident that the generated region of interest overlaps more with the foreground object in the case of PVS rather than PVC.

A series of experiments were conducted to verify the claims mentioned above and to identify the overlap between the generated heatmaps and foreground objects in each input image. The ground truth for these foreground objects was generated using the PVS version of the dataset. We binarize the obtained heatmaps by thresholding at $k \in (0, 1)$ and compute the IoU@k for the models in the Table 1. The results are presented in the Table 4. $k = 0.5$ is selected for a fair comparison of networks as the halfway point in its range. Inception-V3 outperforms other networks in both these metrics across PVC and PVS. ResNet-18 outperforms the rest of the deeper architectures with more convolutional layers among the ResNet family. It is rather surprising that deeper ResNets that outperform their shallow counterparts with respect to the accuracy underperform when it comes to IoU. It is to be noted that, even though these networks vary in terms of their depth, their architecture is similar with respect to ResNet blocks and their usage of skip connections.

Deeper networks with more trainable parameters have a higher capacity for selecting features from an image. They would identify and extract these features from the input image irrespective of whether they belong to the foreground or background object. When texture in the background is removed to make it a constant value, as in PVS, these same architectures tend to look more at the foreground object. In essence, ResNet-50, 101, and 152 have higher IoU in PVS when compared to PVC as seen in Table 4 and Figure 5. In contrast, ResNet-18 and Inception-V3 hybrid perform better in PVC when compared to PVS. Figure 3 shows IoU@k for different models for $k \in (0, 1)$. As noted, the Inception-V3 hybrid outperforms every network in PVC configuration across

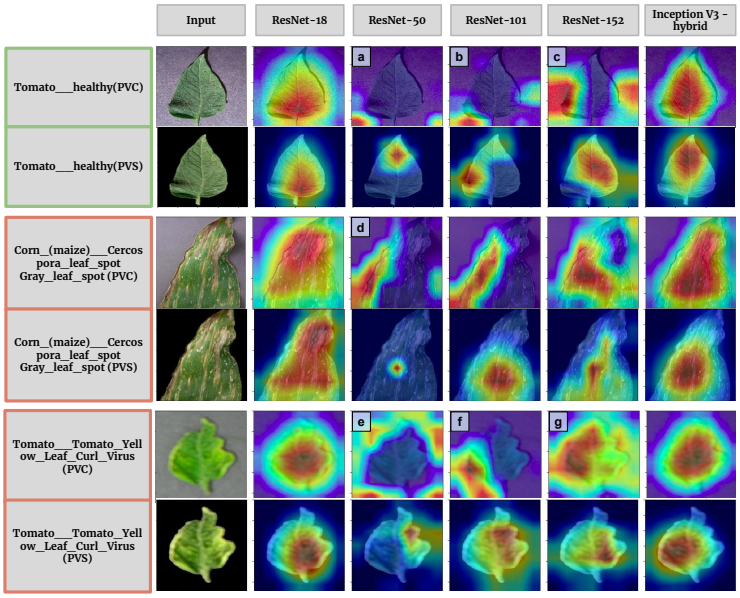


Fig. 5: This figure shows Grad-CAM heatmaps overlapped with their respective images. These heatmaps are obtained for healthy and unhealthy leaves from models specified in Table 1. Although all the images were correctly classified, the focussed regions by the different models vary.

all the values of k. In the PVS configuration, it outperforms other networks for $k \geq 0.5$. This indicates that the Inception-V3 hybrid has better success in identifying important areas of the foreground as regions of interest for making a classification decision when compared with other networks. The graph also shows that ResNet-50, 101, 152 have degrading IoU for an increase in threshold value. This also supports our claim that these networks have identified more parts of the background than the foreground as the region of interest for classifying an image.

To validate our findings, we conducted our recommended IoU based evaluation on the models trained on Swedish-leaf, Flavia, Leafsnap, MEW-2012, and MEW-2014 reported in [26]. Each of the ResNet-101 configurations used in the assessment has $P@1 \geq 0.97$. Swedish, Flavia, and Leafsnap have $IoU@0.5$ scores within a range of ~ 0.06 from each other. They have a difference in the number of classes and total images. The low IoU obtained on these datasets hints at the lack of direct correlation between dataset size and the focus of the neural network. However, there is an improvement in IoU from 0.20 in MEW-2012 to 0.24 in MEW-2014. MEW-2014 is an upgraded version of MEW-2012 containing more classes and images than its past subset. Introducing inter-class variability along with an increase in the size of the dataset has improved the focus of the neural network. The results obtained in Table 5 strengthen our claim of having an IoU based assessment that can also determine the focus of the network w.r.t the foreground object.

6.2 Conclusion

This paper presents an approach for image classification by including handcrafted features in deep learning. We propose a hybrid network that fuses the handcrafted features with the learned features from the neural network. This architecture utilizes the generalization capacity of the neural networks and its ability to select an optimal feature space while giving more control over the learned feature space. We demonstrate the superior results from our approach on different datasets and the improvement with pretraining on domain-specific datasets.

On evaluating the classification model, we have highlighted an unexplored problem in neural network based architectures while training on datasets like PVC. It is a reasonable expectation that the high-performing neural networks for classification learn the feature space from the foreground object. Our analysis demonstrates that this is not always the case. Therefore, it is necessary to further analyze a trained classification model for the regions it focuses on for predicting the label. A properly trained model is the one that not only gives high classification performance in terms of conventional metrics but also has significantly identified the salient regions of the image for making the prediction. We proposed an evaluation methodology based on IoU for quantifying the network's focus on the salient regions. Our proposed hybrid network performed the best in terms of accuracy and the IoU based evaluation. Though the experiments have been done on leaf datasets, the methods are general and can be applied to other domains.

References

1. Alexandratos, N., Bruinsma, J.: World agriculture towards 2030/2050: the 2012 revision (2012)
2. Atila, Ü., Uçar, M., Akyol, K., Uçar, E.: Plant leaf disease classification using efficientnet deep learning model. Ecological Informatics **61**, 101182 (2021)
3. Bae, W., Noh, J., Kim, G.: Rethinking class activation mapping for weakly supervised object localization. In: European Conference on Computer Vision. pp. 618–634. Springer (2020)
4. Barbedo, J.G.: Factors influencing the use of deep learning for plant disease recognition. Biosystems engineering **172**, 84–91 (2018)
5. Brahimi, M., Boukhalfa, K., Moussaoui, A.: Deep learning for tomato diseases: classification and symptoms visualization. Applied Artificial Intelligence **31**(4), 299–315 (2017)
6. Chaerle, L., Lenk, S., Hagenbeek, D., Buschmann, C., Van Der Straeten, D.: Multicolor fluorescence imaging for early detection of the hypersensitive reaction to tobacco mosaic virus. Journal of plant physiology **164**(3), 253–262 (2007)
7. Chouhan, S.S., Singh, U.P., Kaul, A., Jain, S.: A data repository of leaf images: Practice towards plant conservation with plant pathology. In: 2019 4th International Conference on Information Systems and Computer Networks (ISCON). pp. 700–707. IEEE (2019)
8. Cisternas, I., Velásquez, I., Caro, A., Rodríguez, A.: Systematic literature review of implementations of precision agriculture. Computers and Electronics in Agriculture **176**, 105626 (2020)

9. Cruz, A.C., Luvisi, A., De Bellis, L., Ampatzidis, Y.: X-fido: An effective application for detecting olive quick decline syndrome with deep learning and data fusion. Frontiers in plant science **8**, 1741 (2017)
10. Çuğu, İ., Şener, E., Erciyes, Ç., Balcı, B., Akın, E., Önal, I., Akyüz, A.O.: Treelogy: A novel tree classifier utilizing deep and hand-crafted representations. arXiv preprint arXiv:1701.08291 (2017)
11. Fenu, G., Malloci, F.M.: Diamos plant: A dataset for diagnosis and monitoring plant disease. Agronomy **11**(11), 2107 (2021)
12. Ferentinos, K.P.: Deep learning models for plant disease detection and diagnosis. Computers and Electronics in Agriculture **145**, 311–318 (2018)
13. Fujita, E., Kawasaki, Y., Uga, H., Kagiwada, S., Iyatomi, H.: Basic investigation on a robust and practical plant diagnostic system. In: 2016 15th IEEE international conference on machine learning and applications (ICMLA). pp. 989–992. IEEE (2016)
14. Hossin, M., Sulaiman, M.N.: A review on evaluation metrics for data classification evaluations. International journal of data mining & knowledge management process **5**(2), 1 (2015)
15. Huang, Z., Huang, L., Gong, Y., Huang, C., Wang, X.: Mask scoring r-cnn. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 6409–6418 (2019)
16. Huertas-Tato, J., Martín, A., Fierrez, J., Camacho, D.: Fusing cnns and statistical indicators to improve image classification. Information Fusion **79**, 174–187 (2022)
17. Hughes, D., Salathé, M., et al.: An open access repository of images on plant health to enable the development of mobile disease diagnostics. arXiv preprint arXiv:1511.08060 (2015)
18. Islam, M., Dinh, A., Wahid, K., Bhowmik, P.: Detection of potato diseases using image segmentation and multiclass support vector machine. In: 2017 IEEE 30th canadian conference on electrical and computer engineering (CCECE). pp. 1–4. IEEE (2017)
19. Johannes, A., Picon, A., Alvarez-Gila, A., Echazarra, J., Rodriguez-Vaamonde, S., Navajas, A.D., Ortiz-Barredo, A.: Automatic plant disease diagnosis using mobile capture devices, applied on a wheat use case. Computers and electronics in agriculture **138**, 200–209 (2017)
20. Kamal, K., Yin, Z., Wu, M., Wu, Z.: Depthwise separable convolution architectures for plant disease classification. Computers and Electronics in Agriculture **165**, 104948 (2019)
21. Kawasaki, Y., Uga, H., Kagiwada, S., Iyatomi, H.: Basic study of automated diagnosis of viral plant diseases using convolutional neural networks. In: International symposium on visual computing. pp. 638–645. Springer (2015)
22. Kaya, A., Keceli, A.S., Catal, C., Yalic, H.Y., Temucin, H., Tekinerdogan, B.: Analysis of transfer learning for deep neural network based plant classification models. Computers and electronics in agriculture **158**, 20–29 (2019)
23. LeCun, Y., Bengio, Y., Hinton, G.: Deep learning. nature **521**(7553), 436–444 (2015)
24. Li, W., Abad, J.A., French-Monar, R.D., Rascoe, J., Wen, A., Gudmestad, N.C., Secor, G.A., Lee, M., Duan, Y., Levy, L.: Multiplex real-time pcr for detection, identification and quantification of 'candidatus liberibacter solanacearum'in potato plants with zebra chip. Journal of Microbiological Methods **78**(1), 59–65 (2009)
25. Mohanty, S.P., Hughes, D.P., Salathé, M.: Using deep learning for image-based plant disease detection. Frontiers in plant science **7**, 1419 (2016)

26. Muthireddy, V., Jawahar, C.: Computer vision for capturing flora. In: Digital Techniques for Heritage Presentation and Preservation, pp. 245–272. Springer (2021)
27. Ojala, T., Pietikainen, M., Maenpaa, T.: Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. IEEE Transactions on pattern analysis and machine intelligence **24**(7), 971–987 (2002)
28. Patrício, D.I., Rieder, R.: Computer vision and artificial intelligence in precision agriculture for grain crops: A systematic review. Computers and electronics in agriculture **153**, 69–81 (2018)
29. Purcell, D.E., O'Shea, M.G., Johnson, R.A., Kokot, S.: Near-infrared spectroscopy for the prediction of disease ratings for fiji leaf gall in sugarcane clones. Applied Spectroscopy **63**(4), 450–457 (2009)
30. Qin, F., Liu, D., Sun, B., Ruan, L., Ma, Z., Wang, H.: Identification of alfalfa leaf diseases using image recognition technology. PLoS One **11**(12), e0168274 (2016)
31. Ribeiro, M.T., Singh, S., Guestrin, C.: " why should i trust you?" explaining the predictions of any classifier. In: Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining. pp. 1135–1144 (2016)
32. Ruiz-Ruiz, S., Ambrós, S., del Carmen Vives, M., Navarro, L., Moreno, P., Guerri, J.: Detection and quantitation of citrus leaf blotch virus by taqman real-time rt-pcr. Journal of virological methods **160**(1-2), 57–62 (2009)
33. Sankaran, S., Mishra, A., Maja, J.M., Ehsani, R.: Visible-near infrared spectroscopy for detection of huanglongbing in citrus orchards. Computers and electronics in agriculture (2011)
34. Saponari, M., Manjunath, K., Yokomi, R.K.: Quantitative detection of citrus tristeza virus in citrus and aphids by real-time reverse transcription-pcr (taqman®). Journal of Virological Methods **147**(1), 43–53 (2008)
35. Selvaraju, R.R., Cogswell, M., Das, A., Vedantam, R., Parikh, D., Batra, D.: Grad-cam: Visual explanations from deep networks via gradient-based localization. In: Proceedings of the IEEE International Conference on Computer Vision (2017)
36. Shafi, U., Mumtaz, R., García-Nieto, J., Hassan, S.A., Zaidi, S.A.R., Iqbal, N.: Precision agriculture techniques and practices: From considerations to applications. Sensors **19**(17), 3796 (2019)
37. Singh, D., Jain, N., Jain, P., Kayal, P., Kumawat, S., Batra, N.: Plantdoc: a dataset for visual plant disease detection. In: Proceedings of the 7th ACM IKDD CoDS and 25th COMAD, pp. 249–253 (2020)
38. Singh, V., Misra, A.K.: Detection of plant leaf diseases using image segmentation and soft computing techniques. Information processing in Agriculture **4**(1), 41–49 (2017)
39. Sladojevic, S., Arsenovic, M., Anderla, A., Culibrk, D., Stefanovic, D.: Deep neural networks based recognition of plant diseases by leaf image classification. Computational intelligence and neuroscience **2016** (2016)
40. Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J., Wojna, Z.: Rethinking the inception architecture for computer vision. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 2818–2826 (2016)
41. Wang, H., Li, G., Ma, Z., Li, X.: Image recognition of plant diseases based on backpropagation networks. In: 2012 5th International Congress on Image and Signal Processing. pp. 894–900. IEEE (2012)
42. Wu, S., Li, X., Wang, X.: Iou-aware single-stage object detector for accurate localization. Image and Vision Computing **97**, 103911 (2020)
43. Ye, W., Yao, J., Xue, H., Li, Y.: Weakly supervised lesion localization with probabilistic-cam pooling. arXiv preprint arXiv:2005.14480 (2020)