Family-based Plant Disease Characterization using Deep Neural Networks

Sivasubramaniam Janarthan
Deakin University
Selvarajah Thuseethan (thuseethan@gmail.com)
Sabaragamuwa University of Sri Lanka
Suthharshan Rajasegarar
Deakin University
John Yearwood
Deakin University

Article

Keywords:

Posted Date: September 12th, 2022

DOI: https://doi.org/10.21203/rs.3.rs-2037645/v1

Family-based Plant Disease Characterization using Deep Neural Networks

Sivasubramaniam Janarthan[1,2], **Selvarajah Thuseethan**[1,3,*], **Sutharshan Rajasegarar**[1,2], **and John Yearwood**[1,2]

[1]School of Information Technology, Deakin University, Geelong, VIC 3220, Australia
[2]Deakin-SWU Joint Research Centre on Big Data, Faculty of Science, Engineering and Built Environment, Deakin University, Geelong, VIC 3125, Australia
[3]Department of Computing and Information Systems, Sabaragamuwa University of Sri Lanka, Belihuloya, 70140
[*]thuseethan@gmail.com

ABSTRACT

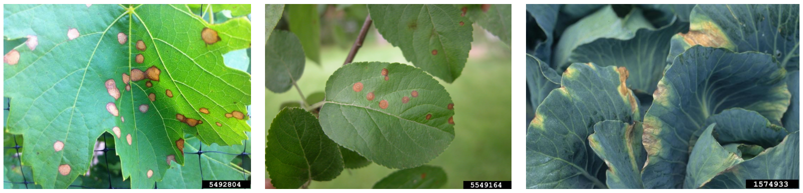
Over the years, researchers have applied various deep learning techniques to automatically recognise plant diseases from both raster and spectral images. The primary focus of the existing studies is developing individual species-specific or disease-specific models, where the former recognises diseases of single crop type and the latter recognises single diseases of single or multiple crop types. Building one global model to recognise diseases of multiple crops has also been widely explored, where a class is treated as a crop-disease combination. While training individual species-specific or disease-specific deep models is labour-intensive, embracing a vast number of crop species and inherent diseases present on this planet makes the model cumbersome. In order to address this problem, a more intuitive and feasible family-based plant disease characterisation approach with botanical reasoning is proposed in this study. This approach demonstrates the feasibility of six state-of-the-art deep neural networks through a set of extensive experiments incorporating six key strategies. The results on a newly built family-based plant disease dataset confirm that the proposed novel approach is convincing to be applied in a plant family-based disease recognition problem. Further, this study creates future opportunities for more intuitive plant disease data collection and benchmark classification model development.

Introduction

The world population is expected to reach 9.6 billion by 2050, which demands a 70% increase in the world food supply. Plant diseases have become the major threat to agriculture sustainability, and continuous food supply together with other dominating threats, such as plant pests, climate change and recently loomed environmental pollution[1,2]. For instance, even the less harmful target spot disease caused 10-42% of yield losses in America despite the fact that improved agricultural practices involving cultivar selection, crop succession, and intensive use of fungicides are in place[3]. Sudden yield losses cause not only financial difficulties but also put extra pressure on farmers and ag-allied professionals. Thus, it is necessary to provide automated solutions to accurately identify plant diseases resulting in increased productivity of farmers and allied professionals.

Researchers in the past explored various computer vision-based techniques for accurate plant disease recognition from both raster and spectral images. Deep learning-based techniques, such as convolutional neural networks (CNN), have achieved tremendous success in image classification tasks without requiring any explicit feature engineering. A systematically designed deep neural network trained with a task-based objective function using a gradient descent optimization algorithm results in a network that can automatically extract important features and classify them based on the trained objective. The first deep CNN with five convolutional layers and a fully connected layer, namely AlexNet[4], outperformed all the existing classical approaches on the ImageNet[5] dataset. Subsequently, many different architectures have been proposed for the purpose of improving accuracy[6–8] and enhancing computational efficiency[7,9,10]. As a pitfall, deep networks often require large data for parameter learning and generalization. This challenge is overcome by a technique called transfer learning, where the feature extraction ability of the pre-trained networks is fine-tuned to a new task with small training data[11]. Thus, several studies in the recent past have incorporated deep learning techniques to develop plant disease recognition systems from images[12–14].

The deep learning-based plant disease recognition models are being built in two different approaches: *individual models* and *global models*. The individual model approach has primarily focused on the species-specific or disease-specific scheme. In species-specific models, multiple disease categories of a particular crop type are considered for classification. For example, individual models are proposed for the recognition of diseases belonging to crop types, rice[15], potato[16], and maize[17]. In[15], the deep CNN model trained to identify ten disease types of the rice plant achieved 95.48% accuracy, which is much higher than the existing approaches. In a very recent study, solely the tomato diseases are taken into consideration for improved classification



Grape Apple Cabbage

Figure 1. Sample black rot images for grape, apple and cabbage crops (Grape black rot by Matthew Zidek, Texas A&M Agrilife Extension Service, Bugwood.org, apple black rot by Penn State Department of Plant Pathology & Environmental Microbiology Archives , Penn State University, Bugwood.org and cabbage black rot by Gerald Holmes, Strawberry Center, Cal Poly San Luis Obispo, Bugwood.org are licensed under CC BY 3.0).

accuracy[18]. The key objective of disease-specific models is to effectively recognize one disease type that can be observed in single or multiple crops. For instance, an improved deep network is used to recognize canker disease from citrus plant[19]. In another work, a MobileNetv2-based YOLOv3 model is proposed for early identification of tomato grey leaf spot disease[20]. Developing an individual model of both kinds is not only a time-consuming task but also requires a large storage capacity.

When it comes to the global model approach, one model is trained to classify multiple diseases of multiple crop types where each crop-disease combination is treated as a class[21–23]. In studies that proposed *global models*, the PlantVillage[24] is the most widely used dataset as it consists of multiple diseases of a range of crop types. Precisely, this dataset contains 54,309 images for 38 different diseases and healthy classes of 14 different crop types. By treating the crop-disease combinations of the PlantVillage dataset as individual classes, in[21], the GoogleNet[25] achieved 99.35% accuracy. In another work, the same approach is applied to another dataset comprising 87,848 diseased samples for 58 different diseases and healthy classes of 25 different crop types[22]. On this dataset, the VGG[26] network obtained the highest classification accuracy of 99.53%. Subsequently, there are numerous studies have been proposed using the global model strategy to handle the plant disease recognition task.

When the vast number of crops and their inherent diseases are considered, it is challenging to provide a workable global model with an expected level of performance. The performance of a global model reduces when there is an increase in the number of classes, which is a common hitch of any classification model. Further, due to the availability of the enormous varieties of crops, covering all of them that are native even to a region in a global mode is not practical. For instance, as per the Food and Agriculture Organization of the United Nations, around 6000 crop species are cultivated around the world for food consumption[27]. As an initial solution for this challenge, in[28], a novel strategy that focuses on possible non-optimal learning of the CNN models is proposed. The primary objective of the proposed model is to learn the disease symptoms instead of learning the shape of the leaf. In this work, diseased samples are regrouped based on the common names of the diseases rather than treating them as crop-disease combinations for individual classes. For instance, the Black rot disease samples obtained from the Apple and Grape plants are grouped into a single class. In the same way, the PlantVillage dataset is rearranged to form 21 classes, including 20 disease classes and the healthy class. This grouping is performed on the basis of the way the leaves are infected, not in terms of the kinds of pathogens that affect the diseases. The pre-trained VGG16 deep network obtained 98.98% classification accuracy on this data, which is a promising advancement compared to existing benchmarks.

Grouping the diseases according to the names is limited as they are caused by different pathogens and often show non-identical symptoms. Figure 1 shows the variations of Black rot disease samples diagnosed in three different crops such as grape, apple and cabbage. It can be observed that the Black rot disease symptoms are similar for grape and apple crops, while they are different from cabbage. They are also caused by different causal agents *Phyllosticta ampelicida* fungus, *Botryosphaeria obtusa* fungus and *Xanthomonas campestris pv. campestris* bacterium, in grape, apple and cabbage crops, respectively. Moreover, the significant variation that can be observed between the crop diseases with the same name can limit the classification performance. Hence, it is necessary to discover a more sophisticated and meaningful plant disease grouping scheme by considering causal agents responsible for the damages.

The crop species belong to the same plant family share intrinsic characteristics, such as growth and nutrient requirements, and are often affected by the same pathogens[29]. Further, a plant disease emerged in one of the crops is highly likely to spread across all other crops within the same family. For example, the pathogen *Alternaria solani* causes Early blight disease in most of the crops native to *Solanaceae* family, such as tomato, potato, eggplant, and bell pepper[30]. Farmers in a particular geographical area often cultivate the same family crops as they often require similar climate conditions for healthy growth. Interestingly, 80% of the food supply is produced from the crops primarily belonging to 17 plant families out of 452 already explored plant families[31]. The plant family list contributing to food production will be extended by another 20 plant families when all the other minor crops are considered. The plant family-based disease recognition approach is practically achievable as the entire crop production can be covered by building 37 (17+20) models. This approach can alleviate the challenges in designing plant disease recognition frameworks with individual models for each crop type or a global model covering all crop species. More importantly, plant family-based models are extensively flexible for adopting disease dynamics like emerging diseases.

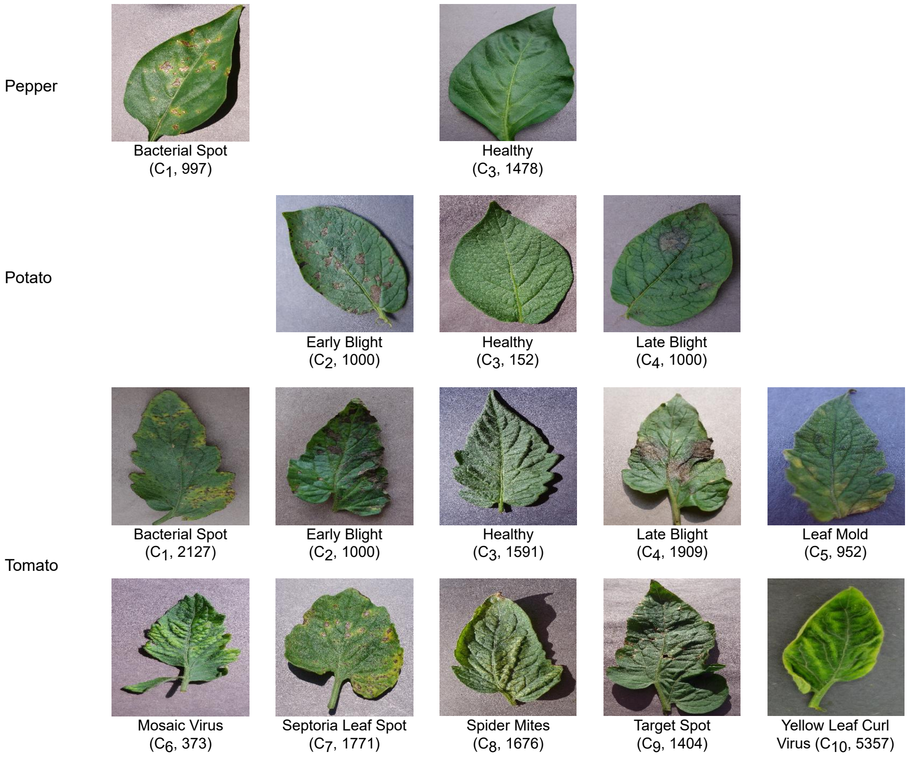


Figure 2. Sample diseased and healthy leaf images of three *Solanaceae* plant family crops: pepper, potato and tomato. Short names and the number of samples for all the diseases are given as tuples.

Table 1. Causal agents and the commonly impacted economically important plants for a set of selected disease types.

Disease	Causal agents	Commonly impacted plants
Bacterial Spot (C_1)	*Xanthomonas*[32]	Pepper and tomato
Early Blight (C_2)	*Alternaria tomatophila*[33]	Potato and tomato
Late Blight (C_4)	*Phytophthora*[34]	Potato, tomato
Leaf Mold (C_5)	*Cladosporium fulvum*[35]	Tomato
Mosaic Virus (C_6)	*Tobacco mosaic virus*[36]	Pepper, tomato and tobacco
Septoria Leaf Spot (C_7)	*Septoria lycopersici*[37]	Tomato
Spider Mites (C_8)	*Tetranychus evansi*[38]	Eggplant, potato and tomato
Target Spot (C_9)	*Thanatephorus cucumeris*[39]	Potato, tomato and tobacco
Yellow Leaf Curl (C_{10})	*Geminivirus*[40]	Pepper and tomato

In this study, using the samples taken from the PlantVillage dataset, a new plant family-based plant disease dataset is first constructed by extracting the diseased and healthy leaf samples for the plants belonging to the *Solanaceae* family, namely tomato, potato and bell pepper. Next, extensive experiments using six carefully designed experimental strategies are performed with six state-of-the-art deep neural networks (DNNs) that are built with various underlying neural network paradigms. The major contributions of this study can be summarized as 1) proposing a novel plant family-based plant disease recognition model development approach; 2) building the plant family-based dataset for *Solanaceae* family, using the PlantVillage dataset; 3) performing extensive experiments by systematically selecting the state-of-the-art DNNs with six carefully crafted experimental strategies, and 4) performing quantitative and qualitative assessments to select the best performing DNNs and appropriate experiment strategies.

Methods

Image dataset

The largest publicly available plant disease leaf images dataset, namely the Plant Village[24], is used in this study. The Plant Village dataset contains a total of 54,309 leaf images for 38 different diseased and healthy classes, spanning 14 different crop species. The image samples are collected from the experimental research stations associated with Land Grant Universities in the USA (Penn State, Florida State, Cornell, and others). In order to photograph the leaves, the infected leaves are first plucked from the plant and then placed against a grey or black coloured background. A standard point-and-shoot digital camera (Sony DSC - Rx100/13 20.2 megapixels) was used to shoot the images in a range of natural lighting conditions considering the real end-user applications. This can be considered an easy-to-follow semi-controlled environment setup. Therefore, the PlantVillage dataset is useful for building plant disease recognition models that can be deployed in in-field applications. To construct the plant family-based dataset, the diseased and healthy leaf images only belonging to the plants in *Solanaceae* plant family, such as pepper (2,475), potato (2,152) and tomato (18,160), are obtained.

In Figure 2, example images for ten disease types taken from the *Solanaceae* plant family crops are given. All the classes that include diseases and healthy are presented with short names (i.e., C_1-C_{10}) and sample counts. Hereinafter, throughout this text, the classes are referred to by their short names. Table 1 presents the causal agents that cause each disease type along with the corresponding commonly impacted economically important plants. The original PlantVillage dataset contains ten diseased and healthy leaf classes for tomato, while only three and two classes are available for potato and pepper, respectively. However, this data configuration is sufficient to evaluate the proposed plant family-based models and the novel experimental strategies used. The images of the newly constructed data are shuffled to create a snapshot of the order and maintained unchanged across all the experiments. This process is primarily conducted to simulate the randomness in sample selection. At last, the dataset is divided in the ratio of 60:20:20 and used for training, validation and testing.

Data augmentation

The image data augmentation is a process of enlarging the original dataset by generating additional instances of the images with higher variability in order to help the DNN model achieve better generalization. A set of randomly selected augmentation functions are employed in an online fashion while the batch of images is taken for training. This kind of online augmentation procedure is simple but effective, and it can be selectively applied during the training process[41].

Figure 3 shows the augmented images along with an original image obtained from the pepper bacterial spot disease class. In this online augmentation procedure, the applied augmentation functions are selected with a probability of 50% from the pool of available augmentation functions, namely colour jitter, Gaussian blur, resize corp, affine, rotation, horizontal flip and vertical flip. Additionally, each augmentation function is assigned with a random parameter configuration within an initialized range of values. The random augmentation boosts the learning and generalization capability of the network by exposing the network to both original and more challenging augmented samples[42].

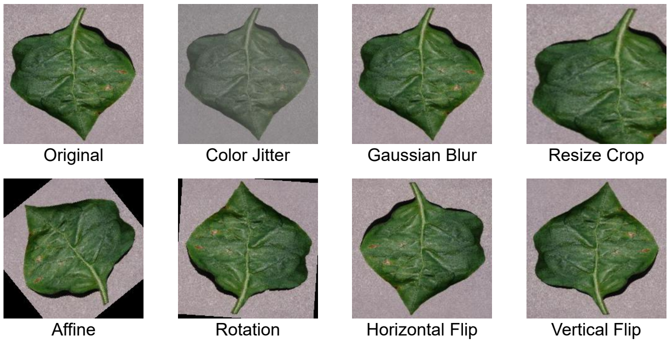


Figure 3. Visual representation of the augmented samples for a bacterial spot diseased pepper leaf image.

Deep learning architectures

Considering the performance, computational efficiency and underlying neural network paradigms, the ResNet50[6], MobileNetV2[9], EfficientNetB6[7], EfficientNetB0[7], ViT[8] and MobileViT[10] are selected in this study to evaluate the proposed plant family-based approach. While the ResNet50, EfficientNetB6 and ViT architectures focus on achieving high performance, the MobileNetV2, EfficientNetB0 and MobileViT target computational efficiency by trading off the performance. The computationally efficient DNNs are appropriate to be used in applications on resource-constrained devices, such as smartphones and tablets. Further, these networks are built based on different neural network paradigms, such as regular CNN, attention and transformer, which are discussed in details in the following subsections.

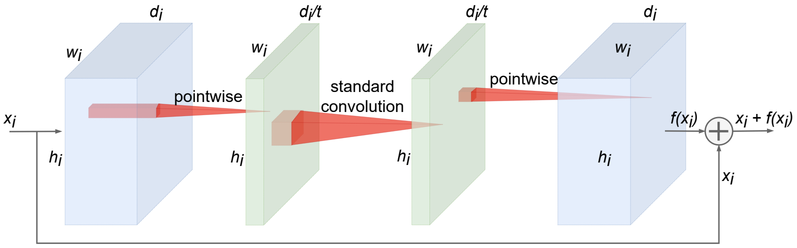


Figure 4. The structure of the residual unit proposed in the ResNet architectures.

ResNet50 (RN50)

Deeper CNN models have shown better performances in classification tasks in comparison to shallower networks. Deep networks with a larger number of convolution layers suffer from learning important features due to the gradient vanishing problem. The weight and biases of early layers in deeper networks are not updated effectively because of small gradients. The batch normalization technique solves this problem to a certain extent and enables the early layers to optimize effectively. However, the accuracy saturates at some point and continues to degrade while increasing the number of layers. In[6], a novel residual learning mechanism is proposed to alleviate the gradient vanishing challenge. As shown in Figure 4, this is achieved through shortcut connections employed between the subsequent layers of the DNN.

The architecture of the residual unit is also presented in Figure 4, which is the core building block of the ResNet. In the residual unit, the input x_i is added with the processed output $f(x_i)$ to produce the final output $y_i = x_i + f(x_i)$. The shortcut connections assist in realizing the potentiality of the residual learning and allow all the layers to learn optimally. In addition, it can be observed from Figure 4 that the feature dimensions are initially reduced using the pointwise convolution operation before a standard convolution operation is performed with 3×3 filters. At last, the feature dimensions are again increased using the pointwise convolution operation. This structure is known as the bottleneck design, which is employed to allow the standard convolution with 3×3 filters to operate on smaller feature dimensions for better computational complexity. Many variants of ResNet architectures have been built by stacking a various number of residual units and other layers. In this study, the best performing ResNet50 variant, which is a 50-layer network, is selected for our evaluation.

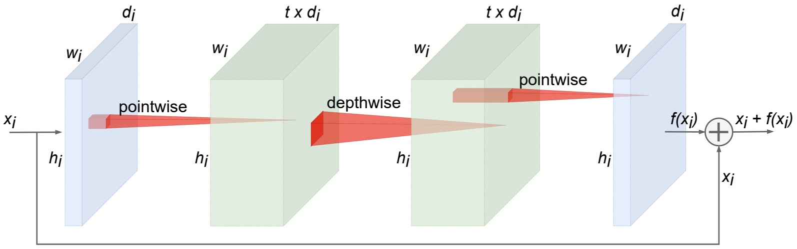


Figure 5. The structure of the inverted residual unit proposed in the MobileNetV2 architecture.

MobileNetv2 (MNV2)

The MobileNetv2 architecture proposed in[9] overcomes the larger computational requirements of the DNN models that hinder the applications on the resource constraint devices. The novel inverted residual unit, which is also known as MBConv[9,43] is introduced. In addition to the inverted residual unit, the shortcut connections similar to the ones used in the ResNet are also integrated with the architecture. The overall architecture of the inverted residual unit is shown in Figure 5. In this unit, as opposed to the residual unit in ResNet, the feature dimensions are first increased with the pointwise convolution operation. Next, the depthwise separable convolution is operated before the feature dimensions are finally reduced using the pointwise operation. As it can be noted, the depthwise convolution is performed in the higher dimensional space instead of the standard convolution. This approach drastically reduces the computational requirements of the DNN while maintaining higher accuracies that make it applicable on mobile devices[9].

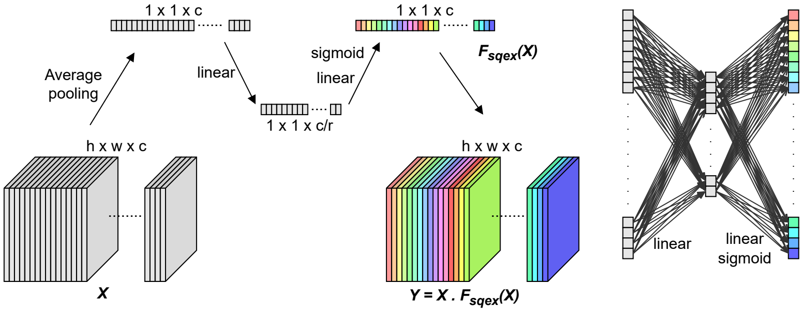


Figure 6. The structure squeeze-and-excitation block.

EfficientNets (ENB0, ENB6)

A family of CNN architecture, called EfficientNets, are introduced in[7] with a common aim of building computationally efficient models without abdicating the performance. A novel scaling scheme that uniformly scales the dimensions of the network, such as depth, width and resolution, with the help of a compound coefficient is proposed. A neural architecture search[43,44] is performed to obtain a family of neural network architectures with various computational complexities and accuracy performances. In addition, they use a modified MBConv[9,43] unit as the basic building block. They also add the squeeze-and-excitation block[45] next to the depthwise convolution, as shown in Figure 5, to enhance the features in the higher dimension. Figure 6 shows the structure of the squeeze-and-excitation block. In reality, it is a channel attention mechanism that learns the interactions between the individual channels of the input features and enhances the individual feature channels in the output[45]. The EfficientNets family of networks can thus be considered an attention network: another neural network paradigm. The mobile baseline model EfficientNetB0 (ENB0) and the EfficientNetB6 (ENB6) with the second-highest accuracy are selected for our experiments.

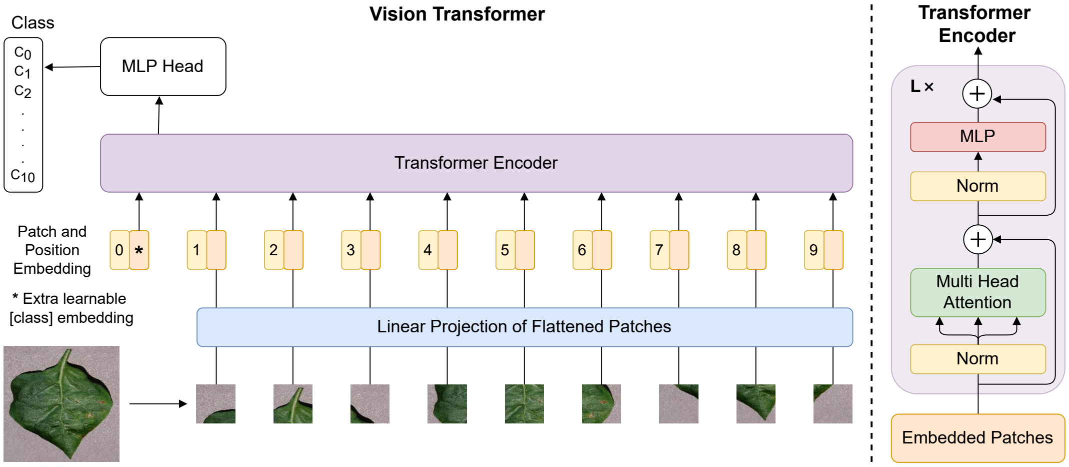


Figure 7. The block diagram of the ViT and the associated transformer encoder.

ViT

The vision transformer network (ViT) is a new paradigm of neural network introduced in[8], which is inspired by the high performance shown by the transformer networks in natural language processing tasks such as language translation. The block diagram of the ViT and the transformer encoder module integrated with it is presented in Figure 7. As can be seen in this figure, the input image is first divided into n number of patches. The divided patches are then flattened and presented to a linear layer to obtain the linear projections. As the next step, an additional learnable class embedding is also introduced. A position embedding is added for all linear projections, including the introduced class embedding. The embedded patches are then processed through the L number of transformer encoder modules. At last, the learned class embedding is presented to the Multilayer Perceptron (MLP) head to obtain the final classification output.

Unlike the CNN networks that use the convolution layers as the primary components, the ViT network uses only the linear layers. Especially, the self-attention mechanism is achieved by the multi-head attention modules[8], enabling a wider receptive field that helps to learn long-range or global dependencies. This is advantageous compared to the small receptive field of the convolution filters that can learn only the local dependencies in a short range. This capability of the ViT network helped to achieve very high accuracy in computer vision tasks. Considering this new paradigm and its capability to attend to the entire image, the ViT is picked as one of the networks in this study.

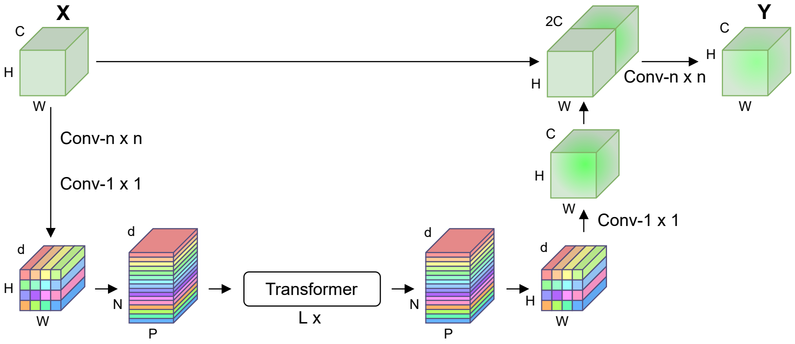


Figure 8. The overall architecture of the attention unit proposed in the MobileViT architecture.

MobileViT

Inspired by the ViT's ability to learn the global representations and the CNN's ability to learn representations with fewer parameters due to the image-specific spatial inductive biases, in[10], a low latency lightweight network called MobileViT is proposed for mobile vision tasks. The MobileViT network is constructed by stacking several MBConv modules defined in MobileNetv2 and a few MobileViT blocks.

Figure 8 illustrates the block diagram of the MobileViT module. The $H \times W \times C$ dimensional feature map is processed with a $n \times n$ standard convolution to encode local spatial information and then it is projected into a $H \times W \times d$ dimensional feature map by a pointwise convolution. The resulting feature map is unfolded into N non-overlapping flattened patches in the size of P to learn the global representations. Here, $N = H \times W/P$ and $P = w \times h$ with h and w are representing height and width of the patches. The patch size h and w are defined in such a way to satisfy the conditions, $h < n$ and $w < n$. The non-overlapping patches are processed through the L number of transformer modules to encode the inter-patch relationship, which is then folded back into $H \times W \times d$ dimensional feature map. The obtained feature map is projected to $H \times W \times C$ with the help of a pointwise convolution. The resultant feature map is then concatenated with the input feature map before presenting it to a $n \times n$ standard convolution layer that fuses the features to obtain the final output.

This architecture is unique and leverages both the ability of the ViT to learn global representation and CNN's ability to learn the representations with a few parameters. Hence, this network is chosen as the lightweight counterpart of the ViT in our evaluations.

Experimental Strategies

In this study, six novel experimental strategies are configured by considering different intuitions to understand the performances of deep learning models. A variety of underlying deep learning paradigms are taken into consideration while selecting the deep learning models. The proposed configurations are given below:

Strategy 1: This configuration follows the traditional approach, where the crop-disease combinations are treated as individual classes. There are a total of 15 classes defined in this configuration, including 2 classes of pepper, 3 classes of potato and 10 classes of tomato.

Strategy 2: A new perspective is presented in this strategy following the one proposed in[28], where the classes are defined only by the disease names. However, there are two significant improvements embraced in this approach. First, the samples are selected for the crops belonging to the same plant family, namely *Solanaceae*. Second, special consideration is given to ensure that the selected diseases are caused by the same or same family of pathogens, as listed in Table 1.

Strategy 3: In this strategy, the diseased and healthy samples obtained only from the tomato crop are used for training and validation. The trained models are then evaluated on the healthy and disease samples taken from the pepper and potato crops, in addition to the test set of the tomato data. The major objective of this setup is to understand the influence of underlying deep learning paradigms of the selected DNNs.

Strategy 4: Differing from Strategy 3, all ten disease classes of tomato and the healthy class from pepper and potato are chosen for training and validation. In addition to the test split of the selected data, the remaining 3 disease classes of pepper and potato are used to evaluate the models. The focus here is to evaluate whether the inclusion of healthy leaves into the training can positively contribute to the overall performance or not. This approach is more feasible as collecting healthy leaf samples is easier in reality, and any positive effect on the performance is beneficial.

Strategy 5: In contrast to Strategy 4, the bacterial spot, early blight and late blight samples of tomato crop are excluded in the training phase. The remaining seven classes of tomato and all the disease and healthy classes from both pepper and potato are included in training and validation. The test split of the selected data is then used for evaluation in recognising bacterial spot, early blight and late blight diseases. The primary intention is to evaluate whether combining tomato and other crops' healthy and disease samples can help to learn the disease symptoms effectively.

Strategy 6: This strategy is developed to evaluate the learning capability of the chosen architectures with small data. The plant disease data collection is challenging mainly because of their limited occurrences and varying disease dynamics on spatial and temporal scales. Hence, it is important for the architectures to perform effectively with small training data[46]. In order to evaluate that, three sets of models based on the following sub-strategies are built.

Sub-strategy 6.1: A set of models are trained with all the training and validation samples taken from the tomato crop. Further, 60 samples for training and 20 samples for validation are included from pepper and potato crops.

Sub-strategy 6.2: Another set of models is finetuned from the models created in Strategy 3 on a subset constructed by choosing 60 samples for training and 20 samples for validation from all the classes of three crops, as indicated in Strategy 2.

Sub-strategy 6.3: The final set of models is built by directly finetuning the pre-trained models with Imagenet weights on the same subset dataset created in Sub-strategy 6.2.

In order to carry out a fair evaluation, in all these three sub-strategies, the same images are selected for the training (60 images) and validation (20 images) phases.

Implementation and training

In this work, the Google Colab Pro[47] platform is utilized to perform the experiments. The PyTorch open source machine learning framework is used to implement, train, validate and test all the selected state-of-the-art DNNs. The ResNet50 and MobileNetV2 model implementations with Imagenet weights are taken from the PyTorch library. The EfficientNets, ViT and MobileViT implementations with Imagenet weights are taken from the public GitHub repositories, such as lukemelas[48], jeonsworld[49] and apple[50], respectively.

All the models are initialized with Imagenet weights and fine-tuned with a batch size of 32 and input image size of 224×224, using cross-entropy loss and Stochastic Gradient Descent (SGD) optimizer. In most cases, the training is performed for 20 epochs. However, the models are trained for 100 epochs when fine-tuning them with tomato diseases in Sub-strategy 6.2 and directly fine-tuning the pre-trained Imagenet models with small data in Sub-strategy 6.3. Further, the SGD optimizer is configured with an initial learning rate of 0.005, the momentum of 0.9 and weight_decay of $1e^{-4}$. The cosine decay function is scheduled for every step to reduce the learning rate reaching 0 in the last epoch.

Results

In this study, six main experiments are performed with different strategies, as described in Subsection *Experimental Strategies*, using six deep learning architectures that are developed with various underlying deep learning paradigms. The results and their interpretations for the experiments with first two strategies are discussed in Subsection *Strategies 1 and 2*, next three strategies are discussed in Subsection *Strategies 3, 4 and 5* and the last strategy is discussed in Subsection *Strategy 6*. Finally, a cumulative analysis and interpretations are presented in Subsection *Discussion*.

Strategies 1 and 2

The average disease recognition accuracies and individual class accuracies obtained for each crop type with experimental Strategy 1 and 2 is presented in Table 2 and Table 3, respectively. The results presented for Strategy 1 are for a 15-class classification problem as it considers the crop-disease combinations as individual classes. Strategy 2 is a 10-class classification problem as it considers each disease as an individual class by merging the same disease samples from multiple crop types.

Table 2. Average disease recognition accuracies of each crop type were obtained for each DNN model with the experimental Strategy 1 and 2.

Method	Accuracy - Strategy 1 (%)			Accuracy - Strategy 2 (%)		
	Tomato	Potato	Pepper	Tomato	Potato	Pepper
MNV2	99.61	99.53	100	99.7	100	99.8
RN50	99.72	100	99.80	99.67	99.77	99.8
ENB0	99.78	100	100	99.75	100	99.8
ENB6	99.56	100	99.59	99.56	99.77	99.8
MViT	99.48	100	100	99.67	100	99.59
ViT	99.75	100	100	99.64	100	99.8

However, the test results presented for Strategy 2 in Table 3 are for each disease of individual crop types. The weighted averages for the diseases grouped by crop types are presented in Table 2.

Table 2 shows that all the models achieve very high accuracies regardless of the experimental strategies. All DNN models built with different underlying concepts, such as regular CNN, attention and transformer, can learn the task effectively according to the learning objective. The individual class accuracies obtained for the bacterial spot (C_1), early blight (C_2) and late blight (C_4) are approximately equal to the accuracies obtained when crop-disease combination-treated as individual classes and when they were combined as one class ignoring crop types. For instance, ENB0 achieved similar accuracies with Strategy 1 (100% [Tomato, C_1], 99% [Tomato, C_2], 98.95% [Tomato, C_4], 100% [Potato, C_2], 100% [Potato, C_4] and 100% [Pepper, C_1]) and Strategy 2 (99.76% [Tomato, C_1], 98% [Tomato, C_2], 99.48% [Tomato, C_4], 100% [Potato, C_2], 100% [Potato, C_4] and 100% [Pepper, C_1]).

The DNN models effectively handle the bacterial spot (C_1) of the different crops, such as tomato and pepper, as a single class when the learning objective treats them as single classes or as different classes when the objective treats them as separate classes. In a similar way, early blight (C_2) and late blight (C_4) diseases observed in tomato and potato crops are also effectively classified into a single class or multiple classes according to the objective function. Hence, *it is clear that combining the same diseases from multiple crops in any practical scenario does not technically degrade the performance.* This condition is *beneficial in identifying the same diseases that can be observed in unseen crop types from the same plant family,* which is further evaluated in the following Subsection *Strategies 3, 4 and 5*.

Table 3. Individual class accuracies of each disease of each crop type were obtained with all the DNN models for experimental strategies 1 and 2. Where, C_1: Bacterial Spot, C_2: Early Blight, C_3: Healthy, C_4: Late Blight, C_5: Leaf Mold, C_6: Mosaic Virus, C_7: Septoria Leaf Spot, C_8: Spider Mites, C_9: Target Spot and C_{10}: Yellow Leaf Curl.

	Method	Tomato										Potato			Pepper	
		C_1	C_2	C_3	C_4	C_5	C_6	C_7	C_8	C_9	C_{10}	C_2	C_3	C_4	C_1	C_3
Strategy 1	MNV2	99.76	97	99.69	99.21	100	100	99.72	100	99.64	99.91	100	100	99	100	100
	RN50	100	98.5	99.69	99.48	99.47	100	100	99.7	99.64	99.91	100	100	100	100	99.66
	ENB0	100	99	99.69	98.95	100	100	100	100	99.64	100	100	100	100	100	100
	ENB6	100	98	100	98.16	100	100	99.72	99.7	99.29	99.91	100	100	100	100	99.32
	MViT	99.76	98.5	99.37	97.9	100	100	99.72	99.7	98.93	100	100	100	100	100	100
	ViT	100	98.5	99.69	99.48	100	100	99.72	99.7	99.64	100	100	100	100	100	100
Strategy 2	MNV2	99.76	98	100	99.21	99.47	100	99.44	100	100	100	100	100	100	100	99.66
	RN50	100	98.5	99.37	99.48	100	100	100	99.7	98.57	100	100	100	99.5	100	99.66
	ENB0	99.76	98	100	99.48	100	100	99.44	100	100	100	100	100	100	100	99.66
	ENB6	100	98	99.69	99.48	99.47	98.65	99.44	100	98.57	99.91	100	100	99.5	100	99.66
	MViT	99.76	97.5	99.69	99.74	100	100	100	99.7	98.93	100	100	100	100	99.49	99.66
	ViT	100	98	99.69	99.48	100	100	100	99.4	98.57	100	100	100	100	100	99.66

Table 2 and Table 3 show that ENB0 network performed better in both Strategy 1 and 2. This network is based on a novel squeeze-and-excitation block[45], which is a channel attention block along with a network architecture search technique. Although ENB6 is also built with the same paradigm, it obtained lower accuracies in both strategies against ENB0. This behaviour is observed because ENB0 learns the task more effectively than ENB6. More importantly, ENB0 is lighter with 5.3M parameters in comparison to ENB6's 43M parameters[7]. In reality, state-of-the-art networks with large parameter sizes are built to perform more complex tasks, such as the 1000-class Imagenet classification problem. Classifying a maximum of 10 to 15 classes is a small task that can easily be achieved with smaller networks, such as ENB0. Moreover, the parameter constraint of ENB0 helped to better generalize in fewer class classification tasks.

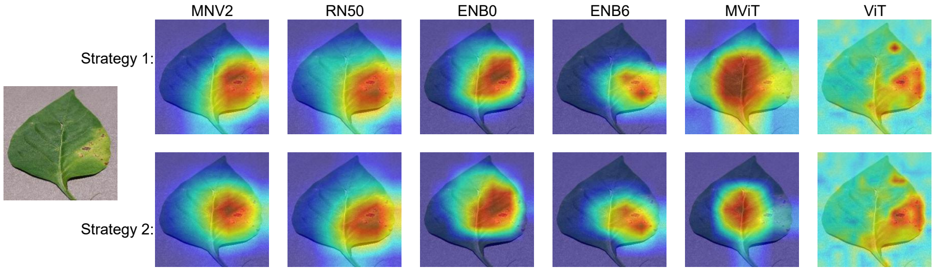


Figure 9. The attention map visualization of a pepper bacterial spot leaf image for all the six models in experimental strategies 1 and 2.

Figure 9 illustrates the visualization of the attention maps obtained by six models with experimental Strategies 1 and 2 for a pepper bacterial spot leaf image is given. The attention maps of the five models are generated using the GradCAM++[51], while the ViT attention maps are produced by the model as implemented in jeonsworld[49]. All the models except MViT attend the majority of the infected areas. The MViT mostly attends to the centre part of the leaf while partially covering the infected areas. Notably, the attention maps produced with Strategy 2 more precisely cover the diseased area compared to those made with Strategy 1. *The models learn the disease symptoms better when the same disease samples from multiple crop types are considered a single class, as specified in Strategy 2.*

Table 4. Individual class accuracies of each disease of each crop type were obtained with all the DNN models for the experimental strategies 3, 4 and 5. Where, C_1: Bacterial Spot, C_2: Early Blight, C_3: Healthy, C_4: Late Blight, C_5: Leaf Mold, C_6: Mosaic Virus, C_7: Septoria Leaf Spot, C_8: Spider Mites, C_9: Target Spot and C_{10}: Yellow Leaf Curl.

	Method	Tomato										Potato			Pepper	
		C_1	C_2	C_3	C_4	C_5	C_6	C_7	C_8	C_9	C_{10}	C_2	C_3	C_4	C_1	C_3
Strategy 3		✓	✓	✓	✓	✓	✓	✓	✓	✓	✓					
	MNV2	99.53	98	99.69	99.21	100	100	99.72	100	100	100	26	0	95.5	20.2	4.42
	RN50	100	98	99.37	99.48	100	100	99.72	99.7	99.29	100	46	0	94	24.75	5.1
	ENB0	100	98	99.69	99.21	100	100	100	100	100	100	51	20	93.5	5.05	11.9
	ENB6	100	98	99.69	99.21	100	100	99.72	99.7	99.29	99.91	46	63.33	92.5	7.07	36.39
	MViT	99.06	98	99.37	98.95	100	100	100	99.4	98.93	100	59.5	23.33	87.5	14.14	13.27
	ViT	100	99	99.69	99.48	99.47	100	99.72	99.4	100	100	25	16	95.5	36.36	9.18
Strategy 4		✓	✓	✓	✓	✓	✓	✓	✓	✓	✓		✓			✓
	MNV2	100	97.5	99.69	99.48	100	100	99.72	100	99.64	100	30	100	92.5	14.65	99.66
	RN50	100	99	99.69	98.95	100	100	99.72	99.4	98.93	100	68.5	100	89	20.71	100
	ENB0	100	97.5	100	99.48	100	100	100	100	99.29	100	53	100	90	13.63	99.66
	ENB6	100	98	100	98.69	99.47	97.3	99.72	99.7	98.93	100	57	100	94	4.04	100
	MViT	100	98	99.37	99.48	99.47	100	100	99.7	99.29	100	37.5	100	90.5	8.58	99.66
	ViT	100	98	100	99.21	100	100	100	99.7	98.21	100	26.5	100	96	19.19	99.66
Strategy 5				✓		✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
	MNV2	0.24	2	99.69	38.85	98.95	100	99.72	100	99.29	99.91	100	100	99.5	100	99.32
	RN50	0.94	6	99.69	43.57	100	100	99.72	99.7	99.64	100	100	100	99	100	99.66
	ENB0	0.94	0.5	99.37	43.31	100	100	100	99.4	99.64	100	100	100	100	100	99.66
	ENB6	0.23	3	100	46.19	100	100	99.72	99.7	98.93	100	100	100	99.5	100	99.32
	MViT	1.65	1	99.69	41.73	100	100	100	100	99.64	100	100	100	99.5	100	99.66
	ViT	2.36	3	99.69	48.82	100	100	99.72	99.4	98.57	100	100	100	100	100	100

Strategies 3, 4 and 5

The accuracies obtained for individual classes by MNV2, RN50, ENB0, ENB6, MViT and ViT networks with Strategies 3, 4 and 5 are presented in Table 4. The check-marks (✓) in Table 4 indicate that those diseased and healthy samples are used for training and validation.

In Strategy 3, each model performed differently for each class when it was trained only on tomato samples and evaluated on potato and pepper samples. All models other than MViT achieved more than 90% accuracies in recognizing potato late blight (C_4) disease. Even the MViT achieved a competent accuracy of 87.5%. The potato early blight (C_2) disease is classified with around 50% accuracy by all the models. However, all the models except ENB6 suffered in accurately classifying the healthy potato leaf samples (C_3). A similar performance is observed for healthy pepper samples (C_3), where ENB6 obtained 36.39% accuracy while all other models showed very low performances. At the same time, ViT achieved the accuracy of 36.36% for the bacterial spot (C_1) and all the other models performed with low accuracies. The potato late blight (C_4) is recognized with an accuracy of 95.50% by MNV2 and ViT, while the potato early blight (C_3) is identified with 59.5% accuracy by MViT. The ENB6 identified the healthy potato (C_3) sample with 63.33% accuracy. Furthermore, ENB6 achieved better overall accuracy in identifying the diseases from unseen crop samples. *The results demonstrate that building plant family-based models can be useful for identifying unseen disease samples of the crops from the same family.*

Further, it can be observed that the recognition accuracies of pepper diseases are low compared to potato diseases. This behaviour can be due to the leaf structure of these three crop types. Figure 2 shows that the texture of tomato and potato leaves share common characteristics, even though the overall shape is different. Meanwhile, the texture of pepper leaves is not matching with the texture of tomato and potato leaves. Hence, the models struggled to classify the pepper disease and healthy samples more than the potato samples.

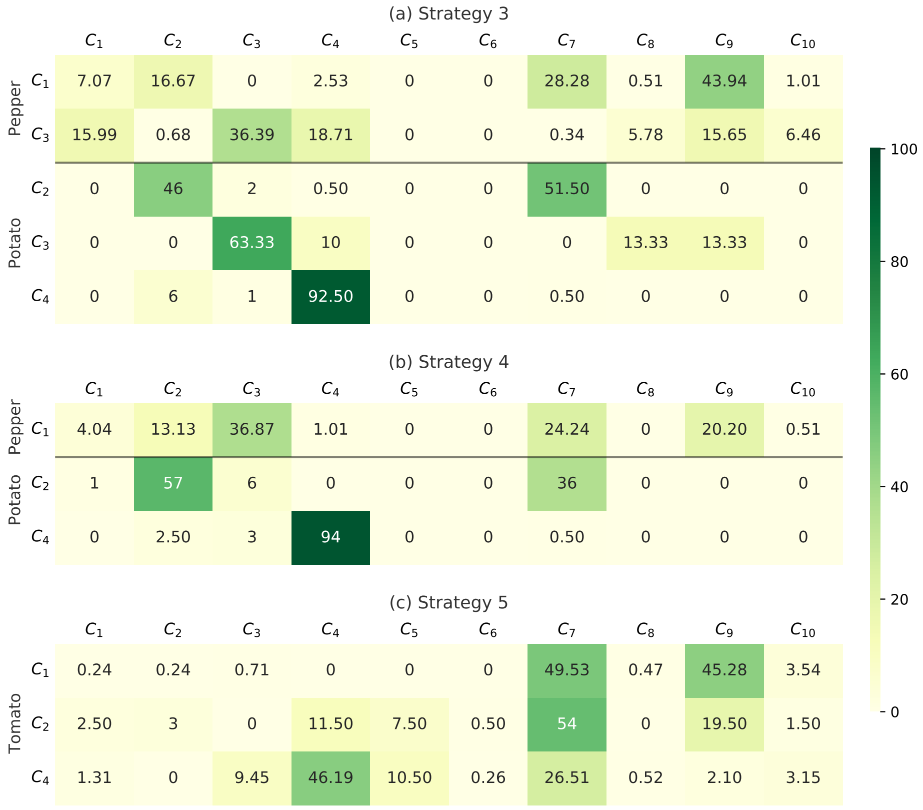


Figure 10. Confusion matrix visualizations for the results obtained with ENB6 for pepper and potato disease classes in strategies 3 and 4. Where, C_1: Bacterial Spot, C_2: Early Blight, C_3: Healthy, C_4: Late Blight, C_5: Leaf Mold, C_6: Mosaic Virus, C_7: Septoria Leaf Spot, C_8: Spider Mites, C_9: Target Spot and C_{10}: Yellow Leaf Curl.

A mixed behaviour is observed when healthy samples of the potato and pepper are included for training and validation, in addition to all tomato healthy and disease samples. While the accuracies for potato late blight (C_4) slightly decreases for MNV2, RN50 and ENB0 models, it increases for ENB6, MViT and ViT models. For potato early blight (C_2), the accuracy drastically dropped to 37.5% for MViT and increased in all the other cases. The accuracies obtained for pepper bacterial spot (C_1) are dropped for all models other than the ENB0. The overall accuracy significantly improved regardless of these tiny individual accuracy drops. The ViT achieved the highest accuracy of 96% for potato late blight (C_4) while the RN50 and ENB6 achieved 68.5% and 57% accuracies for potato early blight (C_2), which are significant increases. *Thus, it can be concluded that including healthy samples to cover all/most of the crops of a plant family can enhance the model performance.* This is further beneficial as collecting healthy leaf images is more straightforward than diseased leaf images. In this scenario, RN50 shows the best overall performance, followed by ENB6 and ENB0.

In strategy 5, the bacterial spot (C_1), early blight (C_2) and late blight (C_4) samples of tomato are held out of training, and the same class samples from pepper and potato are included. Samples from other tomato disease classes are considered in training the model. In this case, very low accuracies are recorded compared to Strategies 3 and 4. Notably, all the models showed low performances for the bacterial spot (C_1) and early blight (C_2) diseases of the tomato crop. The disease samples of tomato late blight (C_4) are reasonably recognized. However, they are significantly lower than potato late blight recognition accuracies obtained with experimental Strategies 3 and 4.

The confusion matrix for the disease classes that are not included in training with Strategies 3, 4 and 5 are presented in Figure 10. In Strategy 3, the pepper bacterial spot (C_1) samples are misclassified mostly with target spot (C_9) (43.94%) and septoria leaf spot (C_7) (28.28%). In Strategy 4, when the healthy pepper samples are included in the training, the pepper bacterial spot (C_1) samples become mostly confused with healthy (36.87%), septoria leaf spot (C_7) (24.24%) and target spot (C_9) (20.2%). This confirms that the shape and texture of the pepper leaves influenced the results. However, this behaviour is not observed for potato early blight (C_2) disease. In Strategy 3, 51.5% of the potato early blight samples are confused with the septoria leaf spot (C_7). It is reduced to 36% when the healthy potato samples are introduced into the training in Strategy 4. This might be because potato leaves' texture is similar to that of tomato leaves. Further, fewer samples of the potato late blight (C_4) are confused with other classes, as the disease symptoms of this class are distinct from other classes. *This demonstrates that the models can identify the unseen crop-disease samples more accurately if their inter-class variations are high.*

Further, the confusion tendency observed for tomato bacterial spot (C_1) and early blight (C_2) diseases with Strategy 5 is similar to the one observed with Strategy 3. A 49.53% and 45.28% of the tomato bacterial spot (C_1) samples are confused with septoria leaf spot (C_7) and target spot (C_9), respectively. The confusion with septoria leaf spot (C_7) happened due to the visual similarity of disease symptoms between tomato bacterial spot (C_1) and septoria leaf spot (C_7), which can be observed in Figure 2. In addition, the tomato early blight (C_2) samples confused with septoria leaf spot (C_7) (54%), target spot (C_9) (19.5%) and late blight (C_4) (11.5%). Despite the fact that 46.19% of the potato late blight samples are correctly classified, 26.51%, 10.5% and 9.45% samples are confused with septoria leaf spot (C_7), leaf mold (C_5) and healthy (C_3), respectively. Although some of these behaviours can be supported by the similarity between the samples of the respective classes, the major reason for this dilemma is the bias of the model. As the models are trained with more tomato (73.93%) leaf samples than the pepper (13.95%) and potato (12.12%) leaf samples, they tend to classify the unseen tomato disease samples into one of the disease classes representing more samples. Considering the training data that is evenly distributed among the crop types can mitigate this problem.

Table 5. Individual class accuracies of each disease of each crop type were obtained with all the DNN models for the sixth experimental strategy. Where, C_1: Bacterial Spot, C_2: Early Blight, C_3: Healthy, C_4: Late Blight, C_5: Leaf Mold, C_6: Mosaic Virus, C_7: Septoria Leaf Spot, C_8: Spider Mites, C_9: Target Spot and C_{10}: Yellow Leaf Curl.

	Method	Tomato										Potato			Pepper	
		C_1	C_2	C_3	C_4	C_5	C_6	C_7	C_8	C_9	C_{10}	C_2	C_3	C_4	C_1	C_3
Sub-strategy 6.1	MNV2	99.76	96.5	99.69	98.95	100	100	99.72	100	98.21	100	100	100	97	91.41	98.98
	RN50	100	98.5	99.69	99.48	100	100	100	99.4	98.57	100	99.5	100	97	93.43	99.32
	ENB0	99.76	98.5	99.69	99.48	100	100	100	100	100	100	99	100	97.5	88.38	98.3
	ENB6	100	98	99.69	98.69	98.95	100	100	100	98.93	100	98.5	100	94.5	91.41	98.64
	MViT	99.76	99	99.69	99.48	100	100	100	99.1	98.93	100	99.5	100	94.5	83.84	98.64
	ViT	100	98	99.69	99.74	98.95	100	100	99.4	99.29	100	99	100	97.5	90.4	99.66
Sub-strategy 6.2	MNV2	96.46	93	99.69	94.49	100	100	95.2	96.11	97.86	96.54	99.5	100	94	95.45	99.32
	RN50	93.87	94	99.69	93.96	100	100	98.31	98.5	98.21	97.2	100	100	95	94.95	99.66
	ENB0	97.88	93.5	99.69	97.9	100	100	97.46	99.4	98.57	97.94	100	100	97.5	97.98	99.32
	ENB6	98.82	97.5	99.69	97.11	98.42	100	99.15	99.4	99.29	98.41	99.5	100	97.5	96.46	97.96
	MViT	96.7	97.5	99.69	96.59	100	100	98.02	98.2	98.93	97.38	100	100	97.5	95.96	98.98
	ViT	99.29	95.5	99.69	97.9	99.47	100	99.44	99.4	98.21	99.72	99.5	100	99	97.47	98.98
Sub-strategy 6.3	MNV2	92.22	83.5	99.37	94.23	95.79	100	85.88	88.92	96.43	94.39	99.5	100	97	93.43	99.32
	RN50	95.28	92.5	99.69	91.6	99.47	100	94.35	95.21	96.43	95.42	100	100	94.5	95.45	99.66
	ENB0	97.64	83.5	99.69	91.6	97.37	100	87.85	96.71	83.21	95.42	99.5	100	94.5	96.97	98.3
	ENB6	93.87	87	96.86	89.24	96.84	98.65	94.07	95.81	94.29	95.42	99	100	96	94.44	98.3
	MViT	95.28	91	99.69	92.91	98.42	100	95.48	97.01	93.57	95.33	100	100	99	95.45	98.98
	ViT	96.23	87.5	99.37	93.7	97.37	100	94.07	98.5	94.64	95.7	98	100	98	97.47	98.98

Strategy 6

In Strategy 6, the models are evaluated with three sub-configurations, called Sub-strategy 6.1, 6.2 and 6.3. The aim of this strategy is to evaluate the learning capability of the chosen architectures with small data. Table 5 presents the results obtained with Strategy 6.

In Sub-strategy 6.1, very similar test accuracies are obtained with all the models for all the tomato classes and early blight (C_2) and healthy (C_3) classes of potato. However, the models showed slightly lower accuracies for potato late blight (C_2) and pepper healthy (C_3). Significantly lower accuracies are obtained for pepper bacterial spot (C_1) disease. In particular, MViT and ENB0 models showed the lowest accuracies ($<90\%$) for bacterial spot (C_1). As can be seen, the RN50 model showed better overall performance compared to other models.

In Sub-strategy 6.2, when the models trained with Strategy 2 are fine-tuned, all the models achieved better overall accuracies with $>93\%$ for all the diseases of all crops. However, the accuracies of tomato disease samples slightly dropped except for healthy (C_3), leaf mold (C_5), mosaic virus (C_6) classes of tomato crop and early blight (C_2) and healthy (C_3) classes of the potato crop. Overall, the ViT and ENB6 showed better resiliency when fine-tuning the pre-trained model with the same classes.

In Sub-strategy 6.3, lower accuracies are obtained for tomato disease samples compared to Sub-strategy 6.2. However, the accuracies achieved for potato and pepper disease classes are similar to Sub-strategy 6.2. The lowest accuracies are obtained for tomato early blight (C_2) with 83.5% for the ENB0 model. However, the MViT followed by ViT and RN50 models showed the best performances with small samples.

These experiments are conducted primarily to demonstrate the learning capability of the selected models with limited data conditions. The results show that different models show better performances in different scenarios. Hence, according to the availability and configuration of the data samples, the best-performing model and the appropriate training strategy need to be selected.

Discussion

Several suggestions and guidelines can be derived for future directions based on the obtained results and their interpretations. The observation that can be made with Strategies 1 and 2 is that all selected deep learning models perform well regardless of the method used for defining target classes. The models effectively learn with the objective set and perform well on the target task. Regardless of whether the individual classes are defined as crop-disease combinations or as diseases by ignoring the crop types, the models achieved similar accuracies. However, the qualitative assessments (i.e., attention map visualization) show that grouping the crop diseases from the same plant family together helps to learn the disease symptoms more effectively. Therefore, this is identified as a potential research direction in plant disease recognition with deep learning.

The experiments with strategies 3, 4 and 5 show that the trained models are effective in identifying the diseases on the other crops of the same family that are not used for training. In particular, higher accuracies can be achieved when the texture of the leaves is similar and inter-class variations are high. Hence, while training the models with diseased leaf samples, covering all the distinct texture types within the same family can help to improve the performance of those new or unseen crop-disease samples. Keeping a balanced sample size for each crop type is also important to avoid bias, which is observed in Strategy 5. These findings provides key guidelines for future data collection and plant family-based plant disease recognition model development.

The EfficientNets variants showed better performances in most of the strategies. The ENB0 showed better performance in Strategies 1 and 2, whereas ENB6 showed better performances in Strategies 3 and 4. More importantly, ENB6 effectively identified the unseen class samples. The transformer-based models, such as ViT and MobileViT, are selected considering their ability to attend to long-range dependencies and their recent benchmark performances on image classification tasks. However, they showed similar or slightly lower performances in the key strategies because of the nature of plant disease recognition tasks. The plant disease recognition task is primarily localizing the disease lesions or altered texture and classifying them based on the learned localized features. The recognition does not depend on any specific structure or disease symptoms from the other parts of the leaf. Hence, regardless of their smaller receptive field, CNN-based networks can effectively recognize plant diseases. The CNN's image-specific inductive bias helps to learn this task more effectively than the networks that can attend to global features. The SE channel attention layer, which is a part of the EfficientNets building block, still allows this family of networks to attend to the global features in the deeper layers[45]. In addition, the EfficientNets family of networks comprises variants targeting computational efficiency (ENB0) and high accuracy performance (ENB6 and ENB7). Therefore, EfficientNets-based models can be considered the first choice for the plant family-based disease recognition problem.

Conclusions

In this study, a novel plant family-based plant disease recognition approach is proposed for effectively identifying the plant diseases as opposed to the traditional approaches that build individual models for each crop type or build a single global model treating each crop-disease combination as individual classes or grouping all the diseases by their common names. The proposed approach reduces the data collection requirements and mitigates the challenges in building and maintaining individual crop-based models or a single global model. As 17 plant families involve in 80% of the agriculture crop production for food, building 17 disease recognition models for corresponding plant families help to cover a major part of the farming. More importantly, entire farming crops can be covered with another 20 models.

The experimental results show that the deep learning-based models can classify plant disease using the proposed grouping scheme with more than 99% accuracy. Further, these models are capable of recognizing the diseases from the other crops of the same family without their samples being used in training. The proposed deep learning models recognized the unseen diseases with 96% accuracy when the leaf texture is close to the one used in training, and the disease symptoms are more distinct from the other classes. Further, the results show that special attention should be given to plant family-based diseased leaves data collection and model development in future.

References

1. Yadav, S. *et al.* Unravelling the emerging threats of microplastics to agroecosystems. *Rev. Environ. Sci. Bio/Technology* 1–28 (2022).
2. Lino-Neto, T. & Baptista, P. Distinguishing allies from enemies—a way for a new green revolution. *Microorganisms* **10**, 1048 (2022).
3. Edwards Molina, J., Navarro, B., Allen, T. & Godoy, C. Soybean target spot caused by corynespora cassiicola: a resurgent disease in the americas. *Trop. Plant Pathol.* 1–17 (2022).
4. Krizhevsky, A., Sutskever, I. & Hinton, G. E. Imagenet classification with deep convolutional neural networks. *Adv. neural information processing systems* **25** (2012).
5. Russakovsky, O. *et al.* Imagenet large scale visual recognition challenge. *Int. journal computer vision* **115**, 211–252 (2015).
6. 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 recognition*, 770–778 (2016).
7. Tan, M. & Le, Q. Efficientnet: Rethinking model scaling for convolutional neural networks. In *International conference on machine learning*, 6105–6114 (PMLR, 2019).
8. Dosovitskiy, A. *et al.* An image is worth 16x16 words: Transformers for image recognition at scale. *arXiv preprint arXiv:2010.11929* (2020).
9. Sandler, M., Howard, A., Zhu, M., Zhmoginov, A. & Chen, L.-C. Mobilenetv2: Inverted residuals and linear bottlenecks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, 4510–4520 (2018).
10. Mehta, S. & Rastegari, M. Mobilevit: light-weight, general-purpose, and mobile-friendly vision transformer. *arXiv preprint arXiv:2110.02178* (2021).
11. Tan, C. *et al.* A survey on deep transfer learning. In *International conference on artificial neural networks*, 270–279 (Springer, 2018).
12. Abade, A., Ferreira, P. A. & de Barros Vidal, F. Plant diseases recognition on images using convolutional neural networks: A systematic review. *Comput. Electron. Agric.* **185**, 106125 (2021).
13. Barbedo, J. G. A. Plant disease identification from individual lesions and spots using deep learning. *Biosyst. Eng.* **180**, 96–107 (2019).
14. Rangarajan, A. K., Purushothaman, R. & Pérez-Ruiz, M. Disease classification in aubergine with local symptomatic region using deep learning models. *Biosyst. Eng.* **209**, 139–153 (2021).
15. Lu, Y., Yi, S., Zeng, N., Liu, Y. & Zhang, Y. Identification of rice diseases using deep convolutional neural networks. *Neurocomputing* **267**, 378–384 (2017).
16. Oppenheim, D. & Shani, G. Potato disease classification using convolution neural networks. *Adv. Animal Biosci.* **8**, 244–249 (2017).
17. Haque, M. *et al.* Deep learning-based approach for identification of diseases of maize crop. *Sci. reports* **12**, 1–14 (2022).
18. Hong, H., Lin, J. & Huang, F. Tomato disease detection and classification by deep learning. In *2020 International Conference on Big Data, Artificial Intelligence and Internet of Things Engineering (ICBAIE)*, 25–29 (IEEE, 2020).
19. Wu, C.-M. & Chen, L.-I. Improved deep learning in citrus canker recognition system based on fpga. In *2021 IEEE International Conference on Consumer Electronics-Taiwan (ICCE-TW)*, 1–2 (IEEE, 2021).

20. Liu, J. & Wang, X. Early recognition of tomato gray leaf spot disease based on mobilenetv2-yolov3 model. *Plant Methods* **16**, 1–16 (2020).

21. Mohanty, S. P., Hughes, D. P. & Salathé, M. Using deep learning for image-based plant disease detection. *Front. plant science* **7**, 1419 (2016).

22. Ferentinos, K. P. Deep learning models for plant disease detection and diagnosis. *Comput. electronics agriculture* **145**, 311–318 (2018).

23. Borhani, Y., Khoramdel, J. & Najafi, E. A deep learning based approach for automated plant disease classification using vision transformer. *Sci. Reports* **12** (2022).

24. 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).

25. Szegedy, C. *et al.* Going deeper with convolutions. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, 1–9 (2015).

26. Simonyan, K. & Zisserman, A. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556* (2014).

27. Biodiversity report. https://www.fao.org/state-of-biodiversity-for-food-agriculture. (Accessed on 09/04/2022).

28. Lee, S. H., Goëau, H., Bonnet, P. & Joly, A. New perspectives on plant disease characterization based on deep learning. *Comput. Electron. Agric.* **170**, 105220 (2020).

29. Plant families. https://serc.carleton.edu/integrate/teaching_materials/food_supply/student_materials/805. (Accessed on 09/04/2022).

30. Marak, T. R., Ambesh, B. S. & Das, S. Cultural, morphological and biochemical variations of alternaria solani causing diseases on solanaceous crops. *The Bioscan* **9**, 1295–1300 (2014).

31. Kew, R. B. G. Naming and counting the world's plant families (2017).

32. Jones, J. *et al.* Systematic analysis of xanthomonads (xanthomonas spp.) associated with pepper and tomato lesions. *Int. J. Syst. Evol. Microbiol.* **50**, 1211–1219 (2000).

33. Rodrigues, T. *et al.* First report of alternaria tomatophila and a. grandis causing early blight on tomato and potato in brazil. *New Dis. Reports* **22**, 28–28 (2010).

34. Nowicki, M., Foolad, M. R., Nowakowska, M. & Kozik, E. U. Potato and tomato late blight caused by phytophthora infestans: an overview of pathology and resistance breeding. *Plant disease* **96**, 4–17 (2012).

35. Thomma, B. P., Van Esse, H. P., Crous, P. W. & De Wit, P. J. Cladosporium fulvum (syn. passalora fulva), a highly specialized plant pathogen as a model for functional studies on plant pathogenic mycosphaerellaceae. *Mol. plant pathology* **6**, 379–393 (2005).

36. Klug, A. The tobacco mosaic virus particle: structure and assembly. *Philos. Transactions Royal Soc. London. Ser. B: Biol. Sci.* **354**, 531–535 (1999).

37. Martin-Hernandez, A., Dufresne, M., Hugouvieux, V., Melton, R. & Osbourn, A. Effects of targeted replacement of the tomatinase gene on the interaction of septoria lycopersici with tomato plants. *Mol. plant-microbe interactions* **13**, 1301–1311 (2000).

38. Qureshi, A. H., Oatman, E. R. & Fleschner, C. Biology of the spider mite, tetranychus evansi. *Annals Entomol. Soc. Am.* **62**, 898–903 (1969).

39. Johnk, J. S., Jones, R., Shew, H. & Carling, D. Characterization of populations of rhizoctonia solani ag-3 from potato and tobacco. *Phytopathology* **83**, 854–858 (1993).

40. Moriones, E. & Navas-Castillo, J. Tomato yellow leaf curl virus, an emerging virus complex causing epidemics worldwide. *Virus research* **71**, 123–134 (2000).

41. Li, Y., Wang, H., Dang, L. M., Sadeghi-Niaraki, A. & Moon, H. Crop pest recognition in natural scenes using convolutional neural networks. *Comput. Electron. Agric.* **169**, 105174 (2020).

42. Shorten, C. & Khoshgoftaar, T. M. A survey on image data augmentation for deep learning. *J. Big Data* **6**, 1–48 (2019).

43. Tan, M. *et al.* Mnasnet: Platform-aware neural architecture search for mobile. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2820–2828 (2019).

44. Zoph, B. & Le, Q. V. Neural architecture search with reinforcement learning. *arXiv preprint arXiv:1611.01578* (2016).

45. Hu, J., Shen, L. & Sun, G. Squeeze-and-excitation networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, 7132–7141 (2018).

46. Janarthan, S. *et al.* Deep metric learning based citrus disease classification with sparse data. *IEEE Access* **8**, 162588–162600 (2020).

47. Welcome to colaboratory - colaboratory. https://colab.research.google.com/. (Accessed on 09/04/2022).

48. Github - lukemelas/efficientnet-pytorch: A pytorch implementation of efficientnet and efficientnetv2 (coming soon!). https://github.com/lukemelas/EfficientNet-PyTorch. (Accessed on 09/04/2022).

49. Github - jeonsworld/vit-pytorch: Pytorch reimplementation of the vision transformer (an image is worth 16x16 words: Transformers for image recognition at scale). https://github.com/jeonsworld/ViT-pytorch. (Accessed on 09/04/2022).

50. Github - apple/ml-cvnets: Cvnets: A library for training computer vision networks. https://github.com/apple/ml-cvnets. (Accessed on 09/04/2022).

51. Chattopadhay, A., Sarkar, A., Howlader, P. & Balasubramanian, V. N. Grad-cam++: Generalized gradient-based visual explanations for deep convolutional networks. In *2018 IEEE winter conference on applications of computer vision (WACV)*, 839–847 (IEEE, 2018).

Author contributions statement

S.J, S.T, S.R and J.Y conceived the study. S.J curated and processed image data, implemented the described models and conducted experiments. S.J and S.T analysed the results and wrote the manuscript. S.R and J.Y reviewed and edited the manuscript.

Data availability

The data used in this research is publicly available and can be accessed via: https://data.mendeley.com/datasets/tywbtsjrjv/1