An Attention-Based Deep Network for Plant Disease Classification

Asish Bera[1,*], Debotosh Bhattacharjee[2,3] and Ondrej Krejcar[3,4,5]
[1] *Department of Computer Science and Information Systems,*
Birla Institute of Technology and Science, Pilani, Rajasthan, India
[2] *Department of Computer Science and Engineering,*
Jadavpur University, Kolkata, West Bengal, India
[3] *Center for Basic and Applied Science, Faculty of Informatics and Management,*
University of Hradec Králové, Czech Republic
[4] *Škoda Auto University, Mladá Boleslav, Czech Republic*
[5] *Malaysia Japan International Institute of Technology (MJIIT),*
Universiti Teknologi Malaysia, Kuala Lumpur, Malaysia
[*] *Corresponding author: Asish Bera (asish.bera@pilani.bits-pilani.ac.in)*

Abstract Plant disease classification using machine learning in a real agricultural field environment is a difficult task. Often, an automated plant disease diagnosis method might fail to capture and interpret discriminatory information due to small variations among leaf sub-categories. Yet, modern Convolutional Neural Networks (CNNs) have achieved decent success in discriminating various plant diseases using leave images. A few existing methods have applied additional pre-processing modules or sub-networks to tackle this challenge. Sometimes, the feature maps ignore partial information for holistic description by part-mining. A deep CNN that emphasizes integration of partial descriptiveness of leaf regions is proposed in this work. The efficacious attention mechanism is integrated with high-level feature map of a base CNN for enhancing feature representation. The proposed method focuses on important diseased areas in leaves, and employs an attention weighting scheme for utilizing useful neighborhood information. The proposed Attention-based network for Plant Disease Classification (APDC) method has achieved state-of-the-art performances on four public plant datasets containing visual/thermal images. The best top-1 accuracies attained by the proposed APDC are: PlantPathology 97.74%, PaddyCrop 99.62%, PaddyDoctor 99.65%, and PlantVillage 99.97%. These results justify the suitability of proposed method.

Keywords: agriculture, attention, Convolutional Neural Networks, CNNs, Deep Learning, plant disease classification.

1. Introduction

Modernization in agriculture is reckoned as an emerging research area. Decent growth has been achieved over conventional engineering and laborious farming technologies using artificial intelligence and machine learning [29, 32]. A myriad of diversified applications of computer vision, in conjunction with the plethora of machine learning (ML) techniques, are playing important roles in agricultural development and in supporting the sustainability. Still, agriculture needs to be improved further to meet growing global food demands as envisaged by scientists. Several key challenges are identified in allied areas of agriculture and related futuristic aspects, which seek more research attention, e.g., early disease prediction, crop yield estimation, crop health monitoring, and others [13, 34].

Automated plant disease prediction from leaf images using computer vision techniques is difficult due to wider variations in visual symptoms [40, 43]. In general, the images of various plants and crops are collected by the users/farmers and pre-processed with image processing techniques, such as noise removal, leaf-area detection, area of interest localization, edge map extraction, scaling, contrast adjustment, and others [27]. Several existing methods have applied pre-processing techniques for image segmentation, especially, segmented the region of interests (RoIs) representing infected regions/spots within the leaves, mask generation, and others [30]. Hence, these conventional pipelines essentially require a well-defined set of tasks to be accomplished before the feature extraction. To alleviate this, many deep learning methods have used actual images of plants and defined a deep network by integrating several sub-modules, such as generative adversarial networks (GAN) for augmentation [11] or U-Net for segmentation [41]. Some works have devised deep convolutional neural networks (CNNs) [10]. Also, lightweight CNNs have been studied for corn disease prediction and other applications due to lesser parametric complexities [13].

In recent years, attention mechanism plays as an indispensable component of modern deep architectures due its superior performance in solving diverse challenges in natural language processing, computer vision, and others [5, 7, 8, 46]. An attention method is effective for crop disease classification too [28]. Its aptness is witnessed for plant disease classification using self-attention [60]. Several prior works have used additional offline pre-processing, GAN-based augmentation, and additional sub-networks for localizing the infected leaf regions, as said above. Also, some methods are developed by transfer learning and ensemble techniques. Often, these existing techniques might overlook part and region based local information for subtle discrimination between infected similar types of leaves. Other than a global feature map, local descriptors are very useful for automated diagnosis and localizing finer details within a leaf. Because, various diseases can infect similar leaves of the same plant category [13, 47]. For example, the same tomato leaf can be infected by several diseases (e.g., mosaic, septoria, curl virus, etc.), and the differences among various plant leaves are naturally subtle. Thus, an efficient feature descriptor is crucial for discriminating and solving this problem.

The proposed Attention-based deep network for Plant Disease Classification (APDC) approach can be divided into three phases, shown in Fig. 1. A high-level feature map of an input leaf image is extracted using a backbone CNN in the first phase. The output feature vector is upsampled to a higher resolution for pooling the features from a set of fixed-size disjoint region proposals. These regions are spatially mapped with the upsampled base CNN's feature vector. Next, a bilinear pooling layer is applied to extract the upsampled convolutional features from each region [6]. The output dimension of these regions are kept the same as the output feature space of a base CNN. Overall, these

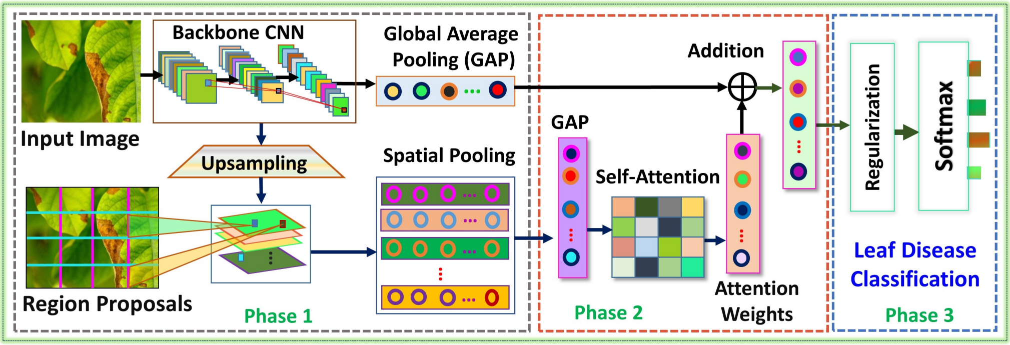

Fig. 1. The proposed APDC framework is divided into three phases: (1) Deep feature extraction from base CNNs and computing region proposals. (2) Attention-based weight computation for the candidate regions across the channel dimension with a residual connection. (3) Regularization of the learning task for plant disease classification using softmax activation.

region-based pooled feature maps are considered as the output of Phase 1. Then, intra-attention is computed for emphasizing the importance of various regions and assigning weights accordingly in Phase 2. The weighted attention score directs at accumulating a precise feature description relevant to classification. A residual path is added as a skip connection which is the output of a global average pooling layer applied to the base CNN's feature map. The added feature map combining the attention scores and skip path defines an efficient feature vector representing the output of Phase 2. A regularization technique is applied for handling the overfitting issues during the training of the proposed network, followed by a softmax layer for classification in the third phase. Experimentally, proposed APDC is found to be an effective and easy solution for leaf disease recognition.

The main contributions of this paper are summarized as:

- An attention-driven deep network integrating three key phases to emphasize the informativeness of complementary regions by weight assignment that represents an efficient feature vector for plant disease recognition.
- The proposed method is end-to-end trainable avoiding additional pre-processing module and bounding-box regression, implying a simple implementation.
- The proposed method has achieved state-of-the-art performance on four public datasets, representing visual and thermal leaf images of various plant classes.
- Rigorous experimental evaluation and ablation studies justify the importance of major components of the proposed deep network.

The rest of this paper is organized as follows: related works are summarized in Section 2. The proposed method is described in Section 3. The experimental results and ablation studies are discussed in Section 4. The conclusion is given in Section 5.

2. Related work

Various deep-learning techniques for plant disease detection and classification have been developed [10,27,38]. Common crop leaves such as the potato, rice, tomato, corn, wheat, etc., have been tested for solving disease identification [31]. A deep network consisting of object detector YoloX and siamese network is described for classifying rice diseases in RiceNet [33]. Multiple pest detection of orchard apples using improved faster R-CNN is presented [15]. A modified GoogLeNet is used for rice disease detection [50]. MobInc-Net is developed by combining MobileNet with the Inception module for disease recognition of 12 rice categories [12]. A dual-stream hierarchical bilinear pooling (DHBP) method is presented in [47]. Bacterial spot detection in the peach leaf images using Convolutional Autoencoders (CAE) and CNN is presented [4]. Six disease classes (e.g., anthracnose, etc.) of the maize crop is tested using NPNet-19 [31]. Pre-trained CNNs (e.g., Inception-v3, etc.) are used for transfer learning to detect 12 types of abnormalities, including huanglongbing of citrus [17].

A CNN is built with the Inception and residual architecture with a convolution block attention module (CBAM) is described in [56]. The method is tested on the epidemiological PlantVillage dataset [22], containing 54.3k images of 14 plant species. Fine-grained classification of infected tomato leaves of the PlantVillage dataset is tested [49]. A lightweight CNN for leaf disease identification is developed and tested on five datasets [45]. A multi-granular feature aggregation approach using self-attention is tested for crop disease classification [60]. A lightweight double fusion block with a coordinate attention network (DFCAnet) is developed [13]. A shuffle attention method and HardSwish function are introduced for recognizing tomato leaf diseases [52]. A cross-attention module, and bidirectional transposed feature pyramid Network is developed for apple disease detection [54]. A Multi-channel recurrent attention network is described for tomato leaf disease prediction [53]. The least important attention pruning algorithm selects the most important attention heads of multi-head self-attention module of each layer in the Transformer model for detecting Cassava leaf disease [43].

A convolutional vision transformer-based lightweight model (ConvViT) is proposed for apple leaf disease identification [26]. A Swin transformer is applied in the path aggregation Swin transformer network (PAST-Net) [48] for detecting and segmenting anthracnose-infected crops, e.g., apple, strawberry, pepper, etc. The Inception convolutional vision transformer (ViT) is developed [51]. The explainable ViT fuses vision transformers with CNN for plant disease identification [44]. A transformer-based with spatial convolutional self-attention transformer is developed for strawberry disease identification [25]. The GANs have been explored to enrich data diversity from small-scale various plant datasets [11]. GrapeGAN [23] follows a U-Net-like generator structure, and the discriminator is built with a convolution block and capsule structure. Four types of grape leaf images are generated by GrapeGAN. Fine grained-GAN method presents a local spot area data augmentation for grape-leaf disease classification [57]. Double GAN is applied for producing high-quality leaf images, representing five classes of PlantVillage dataset [55]. MergeModel identifies tea-leaf diseases [19]. It has applied the U-Net for segmentation and SinGAN for augmentation.

Thermal imaging is explored for crop yield estimation, disease detection, and classification [34]. Thermal images were tested for disease detection from tomato, wheat, and other leaves [18, 58]. The deep explainable artificial intelligence (PlantDXAI) classified plant diseases using CNN-16 in thermal images [3]. The PlantDXAI could be improved by adopting the class activation map and discriminator network during the training. Blight disease detection in rice plants using thermal images is tested [9]. A fusion of color information with thermal and depth information, could attain better accuracy for detecting diseases [35]. In this work, we have presented an attention-driven deep architecture tested on color and thermal leaf images for disease classification.

3. Proposed method

A global feature descriptor could be extracted from an input image using a backbone CNN. Sometimes, a global descriptor might overlook underlying detailed information and and might summarize an overall feature representation, which is relevant to a general classification problem. In contrast, the detailed and subtle information is essential for categorization of leaf sub-categories. An aggregation of partial feature descriptors extracted from complementary regions could effectively capture finer details. We aim to integrate subtle informativeness of several disjoint regions into a comprehensive feature descriptor. The proposed APDC method combines contextual information from complementary regions by aggregating their overall weighted attention scores, which improves holistic feature representation capability. The proposed APDC method is conceptualized in Fig. 1; it is divided into three phases for easier understanding. The extraction of base feature map, and region proposals are described in Phase 1. The attention module with weight computation from pooled regions and is performed Phase 2. The classification is discussed in Phase 3.

3.1. Disjoint region proposal

A region proposal generation method avoiding object detectors, segmentation modules, or bounding box annotations is devised to capture contextual descriptions from different locations of an input image. Let an input leaf image, $I_y \in \mathbb{R}^{h \times w \times 3}$, is to be fed into a backbone CNN with its class label y representing a leaf category. A backbone CNN extracts deep features $\mathbf{F} \in \mathbb{R}^{h \times w \times c}$, where h, w, and c denote the height, width, and channels, respectively. The feature vector $\mathbf{F}$ represents high-level information of input I_y. It could also be interpreted that a local region at low-level image representation is summarized within a small window of the high-level feature space $\mathbf{F}$. Thus, a correspondence between a local image-region with its feature map is necessary to correlate their significance holistically. We consider each uniform/regular region as a fixed rectangular dimension of $p \times p$ pixels. The window-size for spatial pooling from different uniform regions requires to be aligned because the spatial dimension of an I_y is squeezed to a lower size at the deeper levels through successive non-linear transformations in bottleneck layers of a standard CNN. Hence, $\mathbf{F}$ is upscaled to a higher spatial size $q \times q$ using a bilinear interpolation. The number of RoIs is $n = (q/p)^2$, generated without additional pixel-level adjustment during spatial pooling. The set of RoIs is denoted as $R = \{r_1, r_2, ..., r_n\}$, and feature map of r_i-th region is denoted as $\mathbf{F}_i$. The feature maps of all regions are $\mathbf{F}_R = \{\mathbf{F}_r\}_{r=1}^{r=n} \in \mathbb{R}^{n \times (h \times w \times c)}$. In addition to these key steps of Phase 1, a global average pooling (GAP) layer is added to optimize the output features of a base CNN across the channel dimension. A GAP layer squeezes the spatial dimension of a base CNN's output feature map. The pooled feature vector is $\mathbf{G}_R = \mathcal{GAP}(\mathbf{F}_R) \in \mathbb{R}^{n \times 1 \times c}$ maintaining the same channel dimension of $\mathbf{F}$.

3.2. Attention mechanism

The visual attention mechanism focuses on the most informative region(s) of an input image to improve the learning efficacy of a deep architecture by contriving long-range dependency of partial descriptors. Here, self-attention is applied across the channel dimension of feature maps for all regions [2, 46]. The self-attention captures channel-wise relationships among various regions. It correlates cross-channel feature interactions and explores essential parts, accordingly. The self-attention uses three similar feature vectors to compute attention scores: the *query* $\mathbf{Q}$, *key* $\mathbf{K}$, and *value* $\mathbf{V}$ which are derived from the same feature vector $\mathbf{G}_R$. The attention matrix is considered as a dot product of $\mathbf{Q}$ and $\mathbf{K}$, multiplied by $\mathbf{V}$ to produce a weighted feature vector. Here, intra-attention is applied to the r_n region and its neighbor r_m region such that $n \neq m$. The attention method generates feature vector $\mathbf{V}$ to focus on discriminative regions in I_y. The vectors $\mathbf{G}_n$ and $\mathbf{G}_m$ are computed from the r_n and r_m regions, respectively. The feature map is defined as

$$\phi_{n,m} = \tanh(\mathbf{W}_\phi \mathbf{G}_n + \mathbf{W}_{\phi'} \mathbf{G}_m + \mathbf{b}_\phi)\,, \tag{1}$$

$$\theta_{n,m} = \sigma\left(\mathbf{W}_\theta \phi_{n,m} + \mathbf{b}_\theta\right)\,, \tag{2}$$

where weight matrices $\mathbf{W}_\phi$ and $\mathbf{W}_{\phi'}$ compute attention vectors of r_n and r_m, respectively; and $\mathbf{W}_\theta$ is their nonlinear combination. The bias vectors are $\mathbf{b}_\phi$ and $\mathbf{b}_\theta$, and $\sigma(\cdot)$ is an element-wise activation function. The importance of each r_n is computed next using a weighted sum of the attention scores generated from all regions in R. The attention matrix $\hat{\mathbf{G}}_n$ indicates the importance to be given to a region conditioned on its neighborhood regions.

$$\beta_{n,m} = \text{softmax}(\mathbf{W}_\beta \theta_{n,m} + \mathbf{b}_\beta) \, , \hat{\mathbf{G}}_n = \sum_{m=1}^{n} \beta_{n,m} \mathbf{G}_m \, , \tag{3}$$

where the weight matrix is $\mathbf{W}_\beta$, and b_β is the bias. Next, the feature map $\hat{\mathbf{G}}_n$ is undergone to produce a weighted attention map γ_m using a softmax activation over all regions. The output vector of attention importance scores is considered as attention weights representing a high level encoding of all regions and is denoted as $\mathbf{G}_A$. This overall attention map interprets underlying explanation of a given region by weighting its importance towards decision making, essential for plant disease recognition.

$$\mathbf{G}_A = \sum_{m=1}^{n} \gamma_m \hat{\mathbf{G}}_m \ , \quad \gamma_m = \text{softmax}(\mathbf{W}_\phi \hat{\mathbf{G}}_m + \mathbf{b}_\gamma) \, . \tag{4}$$

A residual path is connected by including a GAP layer to the feature maps of a base CNN. This residual path supports further refinement of attentional weighted feature description by improving the gradient flow from the output layers to early layers during the learning without any additional computational overhead. The GAP layer inherently selects the mean features by scaling down a high dimensional feature map precisely to $(1 \times 1 \times c)$, obtained from a base CNN by neglecting trivial information. Also, the GAP enriches the confidence scores for classification, and is robust to spatial translation. The rendered feature map is denoted as $\mathbf{H} = \mathcal{GAP}\,(\mathbf{F}) \in \mathbb{R}^{1 \times 1 \times c}$ where the feature mapping is $\mathbf{F} \rightarrow \mathbf{F}_{\text{gap}} : \mathbb{R}^{(1 \times 1 \times c)}$. Both $\mathbf{G}_A$ and $\mathbf{H}$ feature vectors are added to represent the final attentional feature vector $\mathbf{F}_A \in \mathbb{R}^{(1 \times c)}$.

$$\mathbf{F}_A = \text{Addition}\Big(\mathbf{G}_A, \mathbf{H}\Big) \ ; \ \mathbf{Y}_{\text{pred}} = \text{Softmax}(\mathbf{F}_A) \, . \tag{5}$$

3.3. Image classification

The dropout and batch normalization layers act as regularizers to ease overfitting issues, stabilizes and accelerate the speed of training. Thus, these two layers effective for enhancing the performance during the training. The final feature vector $\mathbf{F}_A$ is passed through a softmax layer to compute the output probability vector representing each class of leaf sub-categories. The categorical cross-entropy loss $\mathcal{L}_{\text{ce}}(Y_{\text{true}}, Y_{\text{pred}})$ optimizes the errors between the true class label (Y_{true}) and predicted class label (Y_{pred}). Overall, the attention technique strengthens the distinctness of feature vectors by capturing finer details without adhering to computational complexities, which is essentially required for leaf disease classification in the proposed APDC method.

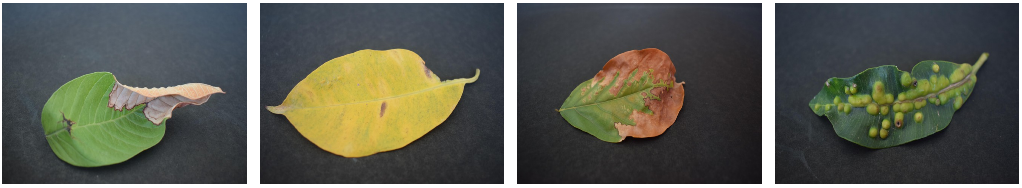

Fig. 2. Samples of leaf images of the PlantPathology-22 dataset.

3.4. Model implementation

The standard backbone CNNs are used for deep feature extraction in Phase 1 of the proposed APDC. The input image-size of 224×224 is fed into a deep CNN, e.g., MobileNet-v2 [39], NASNetMobile [59], DenseNet-169 [20], Inception-v3 [42], etc. During the image pre-processing stage, data augmentations of random rotation (±30 degrees), scaling (1±0.30), and random region erasing (within 0.2-0.7 scale) with a fixed RGB value $q = 127$, are applied for data diversity. Though the output feature dimension of various base CNNs are different, the feature maps are rescaled to a higher resolution using a bilinear interpolation for uniform spatial pooling in Phase 1. For example, a feature map of size 7×7 is upsampled to 40×40 and then the features of non-overlapping regions with a fixed size are computed. Three different sets comprised a total of 16 (4×4), 25 (5×5), and 36 (6×6) regions are generated for experiments. The upscaled resolution is 42×42 for 36 regions, and 40×40 for 25 and 16 regions. The purpose of using such resolutions is to maintain proper pixel alignment during spatial pooling with a fixed window size. However, no feature dimension is calibrated in Phase 2. The output dimension of attention and GAP layers are the same as the output channel dimension of a base CNN, e.g., $c = 1280$ for MobileNet-v2. Finally, a batch normalization and a dropout rate of 0.2 are applied for stabilization of input distributions and regularization for improving the training capacity prior to a softmax layer in Phase 3. Our model is trained with pre-trained ImageNet weights for initializing a base CNN, as well as trained from scratch, i.e., random initialization in different experiments to observe performance variation due to weight initialization not altering other parameters.

The Stochastic Gradient Descent (SGD) is used to optimize the categorical cross-entropy loss function ($\mathcal{L}_{\text{ce}}$) with an initial learning rate of 1×10^{-3}, and multiplied by 0.1 after every 75 epochs for smoother convergence of the learning parameters θ. The proposed APDC is trained for 200 epochs with a mini-batch size of 8 using a Tesla M10 GPU of 8 GB. The top-1 accuracy [%] is used for performance evaluation. The model parameter is estimated in million (M).

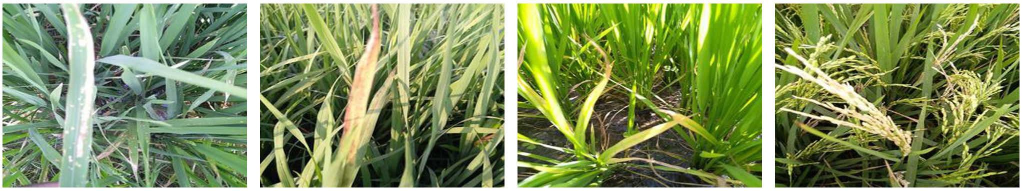

Fig. 3. Diseased leaves of the PaddyDoctor-13 dataset representing infected leaves of plants and crops collected in a natural field environment.

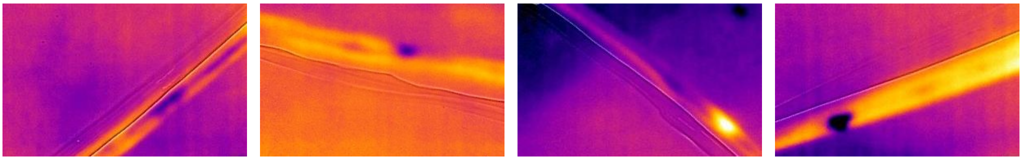

Fig. 4. Examples of diseased leaf images of PaddyCrop-6 thermal dataset.

4. Experimental results and discussions

First, a summary of various datasets tested in this work is briefed. Next, the experimental results, ablation studies, and visualizations are analysed.

4.1. Dataset description

One of the major challenges in agricultural disease diagnosis is the availability of a large realistic dataset of various crops and plants. Since the inception of the PlantVillage dataset, the largest crop dataset to date (to the best of our knowledge), several approaches have been tested for disease recognition and classification. However, this epidemiological dataset is curated in a controlled environment (Fig. 5) and not presented in a realistic manner (e.g., does not consider natural background, leaves are independent and isolated), which is considered as a restriction of this dataset while dealing with a real-world scenarios in agricultural fields. To alleviate this limitation, several other datasets representing various plants/crops are constructed (e.g., Fig. 3). However, most of these recent plant datasets are small-scale, which is further increased in size and image quality by leveraging GAN-based and other augmentations.

A summary of the datasets used in our study is listed in Table 1. Examples of diseased leaves from different datasets, namely PlantVillage-25 [22] (Fig. 5), PlantPathology-22 [14] (Fig. 2), PaddyDoctor-13 [24] (Fig. 3), and PaddyCrop-6 [3] (Fig. 4) are illustrated. The image examples imply that the PlantVillage and PlantPathology datasets are formulated in a simple and clear background condition. On the contrary, PlantDoc and PaddyDoctor represent realistic field environments and complex backgrounds.

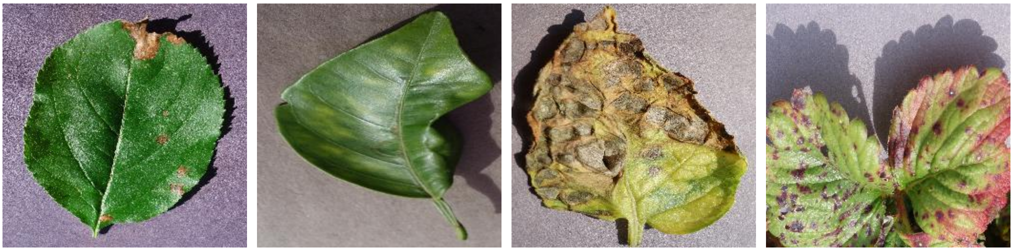

Fig. 5. Samples of diseased leaf images of the PlantVillage dataset.

Tab. 1. Summary of the datasets tested in this work.

Dataset Name	Train	Test	Class	Type
PlantVillage-25	24240	16053	25	RGB
PlantPathology-22	2695	1809	22	RGB
PaddyDoctor-13	12980	3245	13	RGB
PaddyCrop-6	397	240	6	Thermal

The PlantPathology-22 dataset represents healthy (2278) and diseased (2225) leaves from 12 different plants, containing a total of 4503 images and categorised into 22 fine-grained classes.

The thermal images of diseased and healthy leaves of paddy crops comprising a total of 636 samples representing 6 classes were collected using a high-resolution FLIR E8 Thermal camera. Details of this dataset are given in PlantDXAI [3].

4.2. Performance analysis

Firstly, the baseline performances on each dataset are evaluated using four base CNNs. Next, the performances of our method using 16 (4×4), 25 (5×5), and 36 (6×6) RoIs are evaluated in different sets of experiments. The results are given in Table 2. The results imply that the accuracy could be improved with a more number of regions. Because, the attention mechanism focuses on the most important regions of leaf images and emphasizes their inter-channel interactions for weighted feature aggregation. The attention scheme enhances overall prediction performances using four base CNNs. The experimental results, given in Table 2, are achieved by training with ImageNet weights for a fair comparison with existing works on diverse datasets. The model parameters (last column, Table 2) of various experiments remain almost the same for different RoIs and differ according to the backbone CNNs.

Next, the performances on these datasets are evaluated by training the networks from scratch, *i.e.*, initializing the APDC with random weights, and the results are reported

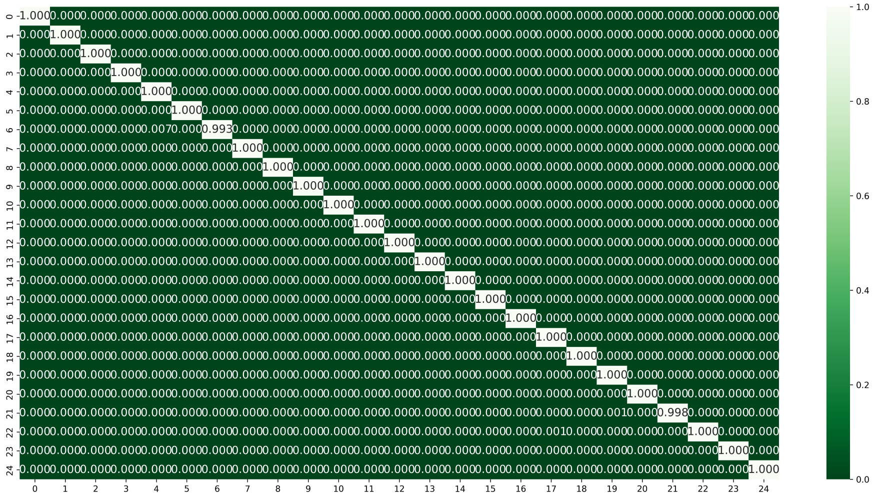

Fig. 6. Confusion Matrix of APDC (36 RoIs) on the PlantVillage-25 dataset.

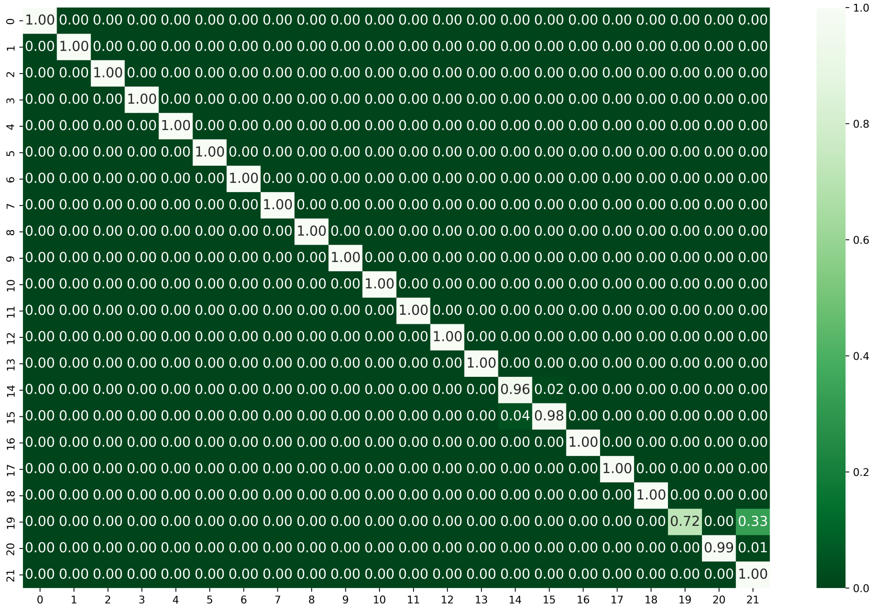

Fig. 7. Confusion Matrix of APDC (36 RoIs) using DenseNet169 on PlantPathology (left) dataset.

Tab. 2. Top-1 accuracy [%] of the proposed APDC using various CNNs backbones trained with ImageNet weights on five plant datasets. The accuracy of similar experiments attained by the CNNs trained from scratch is given in parenthesis. **Bold font** indicates the best performance(s) for each dataset.

Method	PlantVill	PlantPath	Pad'Crop	Pad'Doc	Par
Mob'Net	97.98 (97.69)	94.96 (90.76)	96.66 (83.33)	98.24 (95.82)	2.3
16 RoI	99.32 (98.43)	97.34 (94.79)	97.91 (94.16)	99.02 (96.63)	2.4
25 RoI	99.58 (99.61)	97.45 (95.52)	98.75 (95.41)	99.47 (98.20)	2.4
36 RoI	**99.97** (99.90)	97.62 (97.12)	99.16 (98.25)	99.62 (98.85)	2.4
NasNet	98.49 (95.51)	95.13 (93.58)	95.00 (86.25)	98.14 (95.30)	4.3
16 RoI	99.73 (98.34)	97.46 (96.23)	97.50 (94.58)	99.04 (98.70)	4.4
25 RoI	99.85 (98.76)	97.61 (97.10)	98.33 (95.00)	99.25 (99.21)	4.4
36 RoI	99.93 (99.85)	97.65 (97.24)	99.52 (95.82)	99.60 (99.40)	4.4
DenseNet	99.31 (97.92)	96.73 (92.80)	95.83 (87.50)	98.40 (96.62)	12.7
16 RoI	99.52 (98.55)	97.56 (95.52)	99.16 (93.75)	99.26 (98.45)	12.9
25 RoI	99.67 (98.67)	97.61 (96.72)	99.50 (96.21)	99.58 (99.02)	12.9
36 RoI	99.94 (99.89)	**97.74** (**97.32**)	99.58 (**98.52**)	**99.65** (**99.43**)	12.9
Inception	99.37 (97.65)	96.23 (92.53)	97.00 (86.23)	98.00 (96.72)	21.9
16 RoI	99.91 (98.55)	97.51 (96.23)	98.75 (95.30)	99.41 (98.71)	22.0
25 RoI	99.92 (98.98)	97.60 (97.12)	99.28 (95.81)	99.56 (99.32)	22.1
36 RoI	**99.97** (**99.94**)	97.64 (97.21)	**99.62** (97.50)	99.63 (99.41)	22.1

Tab. 3. Performance Summary of APDC (36 RoI) using various metrics [%].

Dataset	Base CNN	Top-1	Top-5	Precision	Recall	F1-score
PlantPathology	DenseNet169	97.74	99.94	98.0	98.0	98.0
PaddyCrop	MobileNetV2	99.16	100.0	99.0	99.0	99.0
PaddyDoctor	MobileNetV2	99.62	99.97	100.0	100.0	100.0
PlantVillage	MobileNetV2	99.97	100.0	100.0	100.0	100.0

within parenthesis in Table 2. It signifies a clear distinction between the accuracy of APDC while trained with ImageNet weight vis-à-vis random weight initialization which requires more epochs to attain similar accuracy compared to the former. Our model is trained for 300 epochs from scratch in this test, while other hyper-parameters were unaltered. Whereas, 200 epochs are sufficient to attain satisfactory results with the ImageNet weights, which converged quickly. The influence of pre-trained ImageNet weights, compared to random weights, for plant disease prediction accuracy is notable. This accuracy gaps are small on the PlantVillage, and PlantPathology datasets. A reason could be the nature and characteristics of datasets. The samples of these two datasets (Fig. 5-2) were collected in a controlled manner with limited variations by following

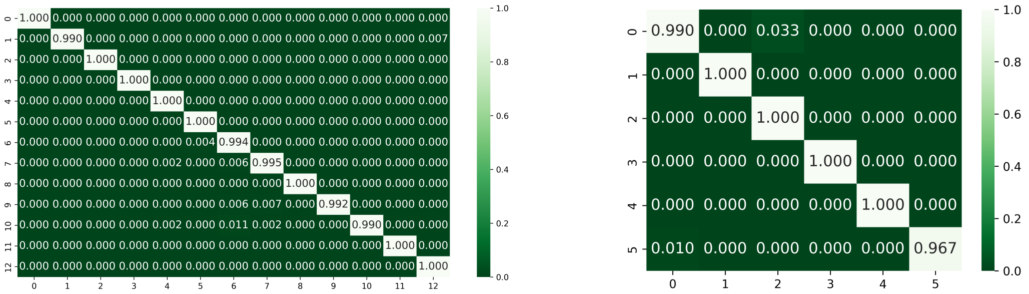

Fig. 8. Confusion Matrix of APDC (36 RoIs) using MoblieNetv2 on the PaddyDoctor (left) and PaddyCrop (right) datasets.

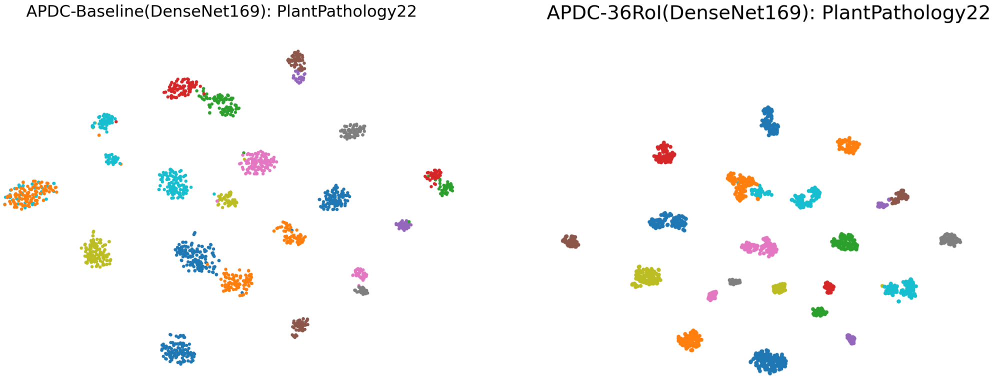

Fig. 9. t-SNE plots of baseline (left) and APDC (36 RoI) using DenseNet-169 (ImageNet) on the PlantPathology dataset.

simple image acquisition scenarios. A summary of the best performances (%) of APDC with 36 RoIs and ImageNet weights on five datasets using standard metrics, namely the top-1 accuracy, top-5 accuracy, precision, recall, and F1-score, are evaluated and reported in Table 3.

Also, one confusion matrix per dataset is shown in Fig. 6-8 for better clarity. In this assessment, MobileNetv2 (MN) is considered for the PlantVillage (Fig. 6), PaddyCrop, and PaddyDoctor datasets (Fig. 8). Whereas, DenseNet169 (DN) is used for generating the confusion matrices on the PlantPathology dataset (Fig. 7) for fair comparison.

Tab. 4. Performance comparison with SOTA on the PlantVillage dataset

Method [Ref]	Plant's Disease / #Class	Dataset Size	Accuracy [%]
GrapeGAN [23]	Grape leaf	4.1 K	96.13
Fine-grained-GAN [57]	Grape leaf-spot disease	1.5 K	96.27
ConvViT [26]	Apple disease	15.8 K	96.85
DenseNet-169 [1]	Corn Foliar disease, 4 cls.	9.1 K	99.50
PCA DeepNet [36]	Tomato, 10 classes	18.1 K	99.60
Double-GAN [55]	10 disease, 5 classes	31.3 K	99.70
PDD271 [27]	38 classes	50.3 K	99.78
FPDR (ResNet50) [16]	38 classes	50.3 K	99.84
APDC: MobileNet-v2	25 classes	40.3 K	**99.97**
DenseNet-169			**99.94**

4.3. Performance comparison

According to our study, many SOTA methods have achieved more than 99.50% accuracy on the PlantVillage dataset [27], and a few recent of them are listed in Table 4 for comparative study. The dataset was created in a controlled laboratory setup with a clear background. Hence, several deep-learning models achieved 99.50% accuracy. The gains in different successive works are competitively very small, e.g., 0.1% only between [36] and [55]. In this work, the average accuracy achieved by APDC with 36 RoIs is 99.95% with a standard deviation of ±0.02, considering four base CNNs trained with ImageNet weights (Table 2). The results on PlantVillage are computed with 25 classes of leaf categories. A brief description of existing disease prediction approaches and their accuracies are summarized in Table 4. The APCD (99.95%) has attained a competitive gain of 0.25% accuracy compared to Double-GAN (99.70%), whereas the accuracy gain over other methods is significant. The PCA DeepNet [36] reported 99.60% accuracy and 98.55% precision. Our APDC has gained 100% precision and F1-score (Table 3). In [27], 99.78% accuracy is obtained by ResNet-152, which is a heavier/deeper base model (≈60.4M) regarding the model parameters compared to lightweight backbones used here. The detailed results are given in Table 2. The IBSA-Net [52] has reported 99.40% accuracy, 98.90% precision, 99.30% and recall. Considering FPDR [16] as the previous best accuracy, 99.84% using ResNet-50 with ImageNet weights, the best 99.97% accuracy of APDC implies a 0.13% margin, with a lesser model parameters of MobileNet-v2. Nevertheless, to analyze the efficiency of our model, the gains on other datasets are significant. We have achieved SOTA performances on recently published public datasets. Rigours experiments have been conducted on the PlantPathology, PaddyDoctor, and PaddyCrop datasets. A fused multi-stream fusion (fsn) with learnable filters scheme [37] has reported 90.02% accuracy on the PlantPathology, curated with a clear background

Tab. 5. Comparison with SOTA on the PlantPathology, PaddyDoctor, and PaddyCrop Datasets [%]. The bottom row-set provides the accuracy of APDC with 36 RoIs using different base CNNs.

PlantPath'y	Acc	PaddyDoc	Acc	PaddyCrop	Acc
Multi-strm fsn [37]	90.02	MobileNet [24]	92.42	CNN16 [3]	88.63
DenseNet201 fsn [21]	96.14	ResNet34 [24]	97.50	PlantDXAI [3]	90.04
MobileNetv2	**97.62**	MobileNetv2	**99.62**	MobileNetv2	**99.16**
DenseNet169	**97.74**	DenseNet169	**99.65**	DenseNet169	**99.58**

like the PlantVillage. An ensemble of CNNs and statistical descriptors has reported an improved classification accuracy of 96.14% using DenseNet-201 [21]. Contrarily, our method has achieved at least 97.62% accuracy using MobileNet-v2 with 36 RoIs. The highest 97.74% accuracy is attained by DenseNet-169. The results are given in Table 5.

PaddyDoctor is a recent dataset on which transfer learning were tested [24]. The best 97.50% accuracy is achieved by ResNet-34, and MobileNet has attained 92.42% accuracy by training with ImageNet weights. The accuracy of APDC underlying on MobileNet-v2 (ImageNet weights) is 99.62%, and training from scratch achieves 98.85% accuracy. Also, APDC based on other CNNs has obtained SOTA results on PaddyDoctor (Table 5) irrespective of training scheme.

The PaddyCrop is a very small dataset containing thermal leaf images of infected rice crops. The PlantDXAI [3] is built with a CNN-16 and trained with class activation map and discriminator network. It has attained 90.04% accuracy on PaddyCrop. The accuracy achieved by our method underlying on DenseNet-169 is 99.58%, and Inception-v3 is 99.62%. Also, more than 99% accuracy is gained by APDC with 36 RoIs, while trained with ImageNet weights. The comparative results are given in Table 5. Overall result analysis evinces that our method outperforms existing works and achieves SOTA performances.

4.4. Ablation study

The necessity of major components of APDC is evaluated through two types of experiments. Firstly, various sets of regions avoiding the attention module are tested to understand their usefulness on different datasets using MobileNet-v2, NASNetMobile, and DenseNet-169 backbones. The results are given in Table 6. The contributions of various RoIs sets are notable using MobileNet-v2. However, in a few other cases, differences between the accuracies of 25 and 36 RoIs using various CNNs are small, e.g., PlantPathology. A reason could be the characteristics of dataset formulation which considered a simple background, such as the PlantPathology (Fig. 2). As a result, a

Tab. 6. Ablation Study I: Top-1 accuracy [%] in proposed APDC (ImageNet Weights) with RoIs Only, excluding attention mechanism.

Base CNN	RoIs	PlantPathology	PaddyCrop	PaddyDoctor
MobileNet-v2	16	96.18	94.58	98.85
	25	96.40	96.66	98.98
	36	96.79	98.75	99.10
NASNetMobile	16	95.06	95.46	98.80
	25	96.02	96.24	99.12
	36	97.01	96.66	99.44
DenseNet-169	16	96.90	98.33	99.16
	25	97.21	99.16	99.44
	36	97.34	99.50	99.56
Inception-v3	16	96.84	98.35	99.19
	25	97.06	99.16	99.41
	36	97.23	99.58	99.63

Tab. 7. Ablation Study II: Top-1 accuracy [%] of using attention on lightweight CNNs (random weight initialization) outputs, neglecting RoIs.

Base CNN	PlantPathology	PaddyyCrop	PddyDoctor
MobileNet	95.56	88.75	96.46
NASNet	95.44	87.25	96.23

few smaller regions may represent trivial information which directs the network to focus on central part of an image where crucial information about an infected leaf exists, neglecting other regions as insignificant.

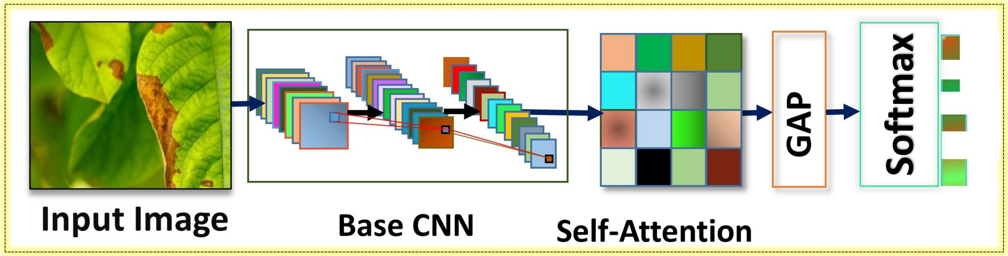

Fig. 10. A generalized CNN-based attention model excluding the regions from proposed APDC.

Next, lightweight MobileNet-v2 and NASNetMobile backbones are considered only and trained with random weight initialization. Here, the candidate regions are neglected from full model, and only intra-attention is applied to the base CNN's output features, followed by a GAP layer before a softmax layer. The deep network is shown in Fig. 10. The base CNN could be replaced by other backbones, e.g., ResNet, DenseNet, and other CNN families. The results are given in Table 7. In this test, the model parameters are reduced slightly, which also causes performance degradation in various datasets. The parameters of considering 36 RoIs for MobileNet-v2 based implementation are 2.46 M. Whereas, excluding the regions, 2.34 M parameters are required using the same MobileNet-v2. Similarly, the parameters for NASNetMobile based implementation are 4.34 M. These results (Table 7) are competitive on various datasets. This study justifies that complementary RoIs are effective to accomplish SOTA results on diverse plant datasets.

5. Conclusion

This paper proposes a deep architecture utilizing a visual attention mechanism, called APDC, for plant disease classification from visual/thermal images of leaves. Experiments were carried out using four plant datasets representing a wider variations in the plant categories, and background conditions. The proposed method follows an end-to-end trainable deep network and simple implementation using class labels only. It avoids extra pre-processing stage or sub-network for data pre-processing compared to existing techniques. The proposed APDC has achieved SOTA performances and emphasized lightweight CNN implementation balancing the accuracy with lower model parameters, unlike the existing ensemble of multiple CNNs-oriented techniques which are heavier models. The lightweight implementation of APDC requires lesser than 5M parameters only. We plan to develop a realistic approach for experimenting on larger and real-world datasets for plant disease classification in the future. A fusion with other sensory information such as soil data of agricultural fields will be another pertinent research direction for sustainable agricultural growth.

Acknowledgement

This work was supported by the New Faculty Seed Grant (NFSG/23-24) and necessary computational infrastructure at the Birla Institute of Technology and Science (BITS) Pilani, Pilani Campus, Rajasthan, 333031, India. This work was also supported in part by the project (2024/2204), Grant Agency of Excellence, University of Hradec Králové, Faculty of Informatics and Management, Czech Republic, and in part by Long-Term Conceptual Development of Research Organization (2024) at Škoda Auto University, Czech Republic.

References

[1] A. Ahmad, A. El Gamal, and D. Saraswat. Towards generalization of deep learning-based plant disease identification under controlled and field conditions. *IEEE Access*, 11, 2023. doi:10.1109/ACCESS.2023.3240100.

[2] D. Bahdanau, K. Cho, and Y. Bengio. Neural machine translation by jointly learning to align and translate. In: *Proc. 3rd International Conference on Learning Representations (ICLR)*. San Diego, CA, USA, 7-9 May 2015. doi:10.48550/arXiv.1409.0473.

[3] G. Batchuluun, S. H. Nam, and K. R. Park. Deep learning-based plant classification and crop disease classification by thermal camera. *Journal of King Saud University – Computer and Information Sciences*, 34(10):10474–10486, 2022. doi:10.1016/j.jksuci.2022.11.003.

[4] P. Bedi and P. Gole. Plant disease detection using hybrid model based on convolutional autoencoder and convolutional neural network. *Artificial Intelligence in Agriculture*, 5:90–101, 2021. doi:10.1016/j.aiia.2021.05.002.

[5] A. Bera, D. Bhattacharjee, and O. Krejcar. PND-Net: plant nutrition deficiency and disease classification using graph convolutional network. *Scientific Reports*, 14(1):15537, 2024. doi:10.1038/s41598-024-66543-7.

[6] A. Bera, O. Krejcar, and D. Bhattacharjee. Rafa-net: Region attention network for food items and agricultural stress recognition. *IEEE Transactions on AgriFood Electronics*, pp. 1–13, 2024. Early Access. doi:10.1109/TAFE.2024.3466561.

[7] A. Bera, M. Nasipuri, O. Krejcar, and D. Bhattacharjee. Fine-grained sports, yoga, and dance postures recognition: A benchmark analysis. *IEEE Transactions on Instrumentation and Measurement*, 72:5020613, 2023. doi:10.1109/TIM.2023.3293564.

[8] A. Bera, Z. Wharton, Y. Liu, N. Bessis, and A. Behera. SR-GNN: Spatial Relation-aware Graph Neural Network for fine-grained image categorization. *IEEE Transactions on Image Processing*, 31:6017–6031, 2022. doi:10.1109/TIP.2022.3205215.

[9] I. Bhakta, S. Phadikar, K. Majumder, H. Mukherjee, and A. Sau. A novel plant disease prediction model based on thermal images using modified deep convolutional neural network. *Precision Agriculture*, 24:23–39, 2022. doi:10.1007/s11119-022-09927-x.

[10] A. C. P. Calma, J. D. M. Guillermo, and C. C. Paglinawan. Cassava disease detection using MobileNetV3 algorithm through augmented stem and leaf images. In: *Proc. 17th Int. Conf. Ubiquitous Information Management and Communication (IMCOM)*, pp. 1–6. IEEE, Seoul, Republic of Korea, 3-5 Jan 2023. doi:10.1109/IMCOM56909.2023.10035648.

[11] Q. H. Cap, H. Uga, S. Kagiwada, and H. Iyatomi. LeafGAN: An effective data augmentation method for practical plant disease diagnosis. *IEEE Transactions on Automation Science and Engineering*, 19(2):1258–1267, 2020. doi:10.1109/TASE.2020.3041499.

[12] J. Chen, W. Chen, A. Zeb, S. Yang, and D. Zhang. Lightweight inception networks for the recognition and detection of rice plant diseases. *IEEE Sensors Journal*, 22(14):14628–14638, 2022. doi:10.1109/JSEN.2022.3182304.

[13] Y. Chen, X. Chen, J. Lin, R. Pan, T. Cao, et al. DFCANet: A novel lightweight convolutional neural network model for corn disease identification. *Agriculture*, 12(12):2047, 2022. doi:10.3390/agriculture12122047.

[14] S. S. Chouhan, U. P. Singh, A. Kaul, and S. Jain. A data repository of leaf images: Practice towards plant conservation with plant pathology. In: *Proc. 4th Int. Conf. Information Systems and Computer Networks*, pp. 700–707. IEEE, Mathura, India, 21-22 Nov 2019. doi:10.1109/ISCON47742.2019.9036158.

[15] F. Deng, W. Mao, Z. Zeng, H. Zeng, and B. Wei. Multiple diseases and pests detection based on federated learning and improved faster R-CNN. *IEEE Transactions on Instrumentation and Measurement*, 71:3523811, 2022. doi:10.1109/TIM.2022.3201937.

[16] P. Gui, W. Dang, F. Zhu, and Q. Zhao. Towards automatic field plant disease recognition. *Computers and Electronics in Agriculture*, 191:106523, 2021. doi:10.1016/j.compag.2021.106523.

[17] W. Gómez-Flores, J. J. Garza-Saldaña, and S. E. Varela-Fuentes. A huanglongbing detection method for orange trees based on deep neural networks and transfer learning. *IEEE Access*, 10:116686–116696, 2022. doi:10.1109/ACCESS.2022.3219481.

[18] I. C. Hashim, A. R. M. Shariff, S. K. Bejo, F. M. Muharam, K. Ahmad, et al. Application of thermal imaging for plant disease detection. In: *Proc. 10th IGRSM Int. Conf. and Exhibition on Geospatial & Remote Sensing*, vol. 540 of *IOP Conference Series: Earth and Environmental Science*, p. 012052. IOP Publishing, Kuala Lumpur, Malaysia, 20-21 Oct 2020. doi:10.1088/1755-1315/540/1/012052.

[19] G. Hu and M. Fang. Using a multi-convolutional neural network to automatically identify small-sample tea leaf diseases. *Sustainable Computing: Informatics and Systems*, 35:100696, 2022. doi:10.1016/j.suscom.2022.100696.

[20] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger. Densely connected convolutional networks. In: *Proc. IEEE Conf. Computer Vision and Pattern Recognition (CVPR)*, pp. 2261–2269. Honolulu, HI, USA, 21-26 Jul 2017. doi:10.1109/CVPR.2017.243.

[21] J. Huertas-Tato, A. Martín, J. Fierrez, and D. Camacho. Fusing CNNs and statistical indicators to improve image classification. *Information Fusion*, 79:174–187, 2022. doi:10.1016/j.inffus.2021.09.012.

[22] D. Hughes, M. Salathé, et al. An open access repository of images on plant health to enable the development of mobile disease diagnostics. *arXiv*, 2016. ArXiv:1511.08060v2. doi:10.48550/arXiv.1511.08060.

[23] H. Jin, Y. Li, J. Qi, J. Feng, D. Tian, et al. GrapeGAN: Unsupervised image enhancement for improved grape leaf disease recognition. *Computers and Electronics in Agriculture*, 198:107055, 2022. doi:10.1016/j.compag.2022.107055.

[24] B. Kiruba and P. Arjunan. Paddy Doctor: A visual image dataset for automated paddy disease classification and benchmarking. In: *Proc. 6th Joint Int. Conf. Data Science & Management of Data (10th ACM IKDD CODS and 28th COMAD)*, pp. 203–207. Mumbai, India, 4-7 Jan 2023. doi:10.1145/3570991.3570994.

[25] G. Li, L. Jiao, P. Chen, K. Liu, R. Wang, et al. Spatial convolutional self-attention-based transformer module for strawberry disease identification under complex background. *Computers and Electronics in Agriculture*, 212:108121, 2023. doi:10.1016/j.compag.2023.108121.

[26] X. Li and S. Li. Transformer help CNN see better: A lightweight hybrid apple disease identification model based on transformers. *Agriculture*, 12(6):884, 2022. doi:10.3390/agriculture12060884.

[27] J. Liu and X. Wang. Plant diseases and pests detection based on deep learning: A review. *Plant Methods*, 17(1):22, 2021. doi:10.1186/s13007-021-00722-9.

[28] Y. Liu, G. Gao, and Z. Zhang. Crop disease recognition based on modified light-weight CNN with attention mechanism. *IEEE Access*, 10:112066–112075, 2022. doi:10.1109/ACCESS.2022.3216285.

[29] J. Lu, L. Tan, and H. Jiang. Review on convolutional neural network (CNN) applied to plant leaf disease classification. *Agriculture*, 11(8):707, 2021. doi:10.3390/agriculture11080707.

[30] O. Mzoughi and I. Yahiaoui. Deep learning-based segmentation for disease identification. *Ecological Informatics*, p. 102000, 2023. doi:10.1016/j.ecoinf.2023.102000.

[31] M. Nagaraju and P. Chawla. Maize crop disease detection using NPNet-19 convolutional neural network. *Neural Computing and Applications*, 22:3075–3099, 2022. doi:10.1007/s00521-022-07722-3.

[32] A. Pal and V. Kumar. AgriDet: Plant leaf disease severity classification using agriculture detection framework. *Engineering Applications of Artificial Intelligence*, 119:105754, 2023. doi:10.1016/j.engappai.2022.105754.

[33] J. Pan, T. Wang, and Q. Wu. RiceNet: A two stage machine learning method for rice disease identification. *Biosystems Engineering*, 225:54–68, 2023. doi:10.1016/j.biosystemseng.2022.11.007.

[34] M. Pineda, M. Barón, and M.-L. Pérez-Bueno. Thermal imaging for plant stress detection and phenotyping. *Remote Sensing*, 13(1):68, 2021. doi:10.3390/rs13010068.

[35] S.-e.-A. Raza, G. Prince, J. P. Clarkson, and N. M. Rajpoot. Automatic detection of diseased tomato plants using thermal and stereo visible light images. *PloS ONE*, 10(4):e0123262, 2015. doi:10.1371/journal.pone.0123262.

[36] K. Roy, S. S. Chaudhuri, J. Frnda, S. Bandopadhyay, I. J. Ray, et al. Detection of tomato leaf diseases for agro-based industries using novel PCA DeepNet. *IEEE Access*, 11:14983–15001, 2023. doi:10.1109/ACCESS.2023.3244499.

[37] N. S. Russel and A. Selvaraj. Leaf species and disease classification using multiscale parallel deep CNN architecture. *Neural Computing and Applications*, 34(21):19217–19237, 2022. doi:10.1007/s00521-022-07521-w.

[38] M. H. Saleem, J. Potgieter, and K. M. Arif. Plant disease detection and classification by deep learning. *Plants*, 8(11):468, 2019. doi:10.3390/plants8110468.

[39] M. Sandler, A. Howard, M. Zhu, A. Zhmoginov, and L.-C. Chen. MobileNetV2: Inverted residuals and linear bottlenecks. In: *Proc. IEEE/CVF Conf. Computer Vision and Pattern Recognition (CVPR)*, pp. 4510–4520. Salt Lake City, UT, USA, 18-23 Jun 2018. doi:10.1109/CVPR.2018.00474.

[40] D. Singh, N. Jain, P. Jain, P. Kayal, S. Kumawat, et al. PlantDoc: A dataset for visual plant disease detection. In: *CoDS COMAD 2020: Proc. 7th ACM IKDD CoDS and 25th COMAD*, pp. 249–253. Hyderabad, India, 5-7 Jan 2020. doi:10.1145/3371158.3371196.

[41] C. K. Sunil, C. D. Jaidhar, and N. Patil. Cardamom plant disease detection approach using EfficientNetV2. *IEEE Access*, 10:789–804, 2021. doi:10.1109/ACCESS.2021.3138920.

[42] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna. Rethinking the inception architecture for computer vision. In: *Proc. IEEE Conf. Computer Vision and Pattern Recognition (CVPR)*, pp. 2818–2826. Las Vegas, NV, USA, 27-30 Jun 2016. doi:10.1109/CVPR.2016.308.

[43] H.-T. Thai, K.-H. Le, and N. L.-T. Nguyen. Formerleaf: An efficient vision transformer for Cassava Leaf Disease detection. *Computers and Electronics in Agriculture*, 204:107518, 2023. doi:10.1016/j.compag.2022.107518.

[44] P. S. Thakur, P. Khanna, T. Sheorey, and A. Ojha. Explainable vision transformer enabled convolutional neural network for plant disease identification: PlantXViT. *arXiv*, 2022. ArXiv:2207.07919. doi:10.48550/arXiv.2207.07919.

[45] P. S. Thakur, T. Sheorey, and A. Ojha. VGG-ICNN: A lightweight CNN model for crop disease identification. *Multimedia Tools and Applications*, 82(1):497–520, 2022. doi:10.1007/s11042-022-13144-z.

[46] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, et al. Attention is all you need. In: *Advances in Neural Information Processing Systems: Proc. NIPS 2017*, vol. 30. Curran Associates, Inc., 2017. `https://papers.neurips.cc/paper_files/paper/2017/hash/3f5ee243547dee91fbd053c1c4a845aa-Abstract.html`.

[47] D. Wang, J. Wang, Z. Ren, and W. Li. DHBP: A dual-stream hierarchical bilinear pooling model for plant disease multi-task classification. *Computers and Electronics in Agriculture*, 195:106788, 2022. doi:10.1016/j.compag.2022.106788.

[48] Y. Wang, S. Wang, W. Ni, and Q. Zeng. PAST-net: a swin transformer and path aggregation model for anthracnose instance segmentation. *Multimedia Systems*, 29(3):1011–1023, 2022. doi:10.1007/s00530-022-01033-2.

[49] G. Yang, G. Chen, Y. He, Z. Yan, Y. Guo, et al. Self-supervised collaborative multi-network for fine-grained visual categorization of tomato diseases. *IEEE Access*, 8:211912–211923, 2020. doi:10.1109/ACCESS.2020.3039345.

[50] L. Yang, X. Yu, S. Zhang, H. Long, H. Zhang, et al. GoogLeNet based on residual network and attention mechanism identification of rice leaf diseases. *Computers and Electronics in Agriculture*, 204:107543, 2023. doi:10.1016/j.compag.2022.107543.

[51] S. Yu, L. Xie, and Q. Huang. Inception convolutional vision transformers for plant disease identification. *Internet of Things*, 21:100650, 2023. doi:10.1016/j.iot.2022.100650.

[52] R. Zhang, Y. Wang, P. Jiang, J. Peng, and H. Chen. IBSA_Net: A network for tomato leaf disease identification based on transfer learning with small samples. *Applied Sciences*, 13(7):4348, 2023. doi:10.3390/app13074348.

[53] Y. Zhang, S. Huang, G. Zhou, Y. Hu, and L. Li. Identification of tomato leaf diseases based on multi-channel automatic orientation recurrent attention network. *Computers and Electronics in Agriculture*, 205:107605, 2023. doi:10.1016/j.compag.2022.107605.

[54] Y. Zhang, G. Zhou, A. Chen, M. He, J. Li, et al. A precise apple leaf diseases detection using bctnet under unconstrained environments. *Computers and Electronics in Agriculture*, 212:108132, 2023. doi:10.1016/j.compag.2023.108132.

[55] Y. Zhao, Z. Chen, X. Gao, W. Song, Q. Xiong, et al. Plant disease detection using generated leaves based on DoubleGAN. *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, 19(3):1817–1826, 2021. doi:10.1109/TCBB.2021.3056683.

[56] Y. Zhao, C. Sun, X. Xu, and J. Chen. RIC-Net: A plant disease classification model based on the fusion of Inception and residual structure and embedded attention mechanism. *Computers and Electronics in Agriculture*, 193:106644, 2022. doi:10.1016/j.compag.2021.106644.

[57] C. Zhou, Z. Zhang, S. Zhou, J. Xing, Q. Wu, et al. Grape leaf spot identification under limited samples by fine-grained GAN. *IEEE Access*, 9:100480–100489, 2021. doi:10.1109/ACCESS.2021.3097050.

[58] W. Zhu, H. Chen, I. Ciechanowska, and D. Spaner. Application of infrared thermal imaging for the rapid diagnosis of crop disease. *IFAC-PapersOnLine*, 51(17):424–430, 2018. doi:10.1016/j.ifacol.2018.08.184.

[59] B. Zoph, V. Vasudevan, J. Shlens, and Q. V. Le. Learning transferable architectures for scalable image recognition. In: *Proc. IEEE/CVF Conf. Computer Vision and Pattern Recognition (CVPR)*, pp. 8697–8710. Salt Lake City, UT, USA, 18-23 Jun 2018. doi:10.1109/CVPR.2018.00907.

[60] X. Zuo, J. Chu, J. Shen, and J. Sun. Multi-granularity feature aggregation with self-attention and spatial reasoning for fine-grained crop disease classification. *Agriculture*, 12(9):1499, 2022. doi:10.3390/agriculture12091499.