FOLIAGEN: FRAMEWORK FOR FOLIAGE IMAGE GENERATION FROM INDIVIDUAL CROP LEAF IMAGES

Anonymous authors
Paper under double-blind review

ABSTRACT

While machine learning (ML)-based crop disease classifiers mostly targeted individual leaf images, real-world applications call for disease classification on crop foliage images instead, because they usually rely on cameras mounted on unmanned aerial vehicles to capture foliage images across vast crop fields for automated disease identification. We found that known state-of-the-art (SOTA) classifiers on the only real-world soybean foliage image dataset all exhibited unsatisfactory performance, despite the dataset being modest-sized and including just two soybean disease categories (among many). Hence, it is desirable to make available large foliage image datasets with common crop disease categories for better evaluating and possibly improving SOTA crop disease classifiers on foliage images. This paper introduces a framework that generates crop foliage images utilizing available datasets of individual leaf images, termed Foliagen (short for foliage generation). A generated foliage image dataset can be arbitrarily sized, with each image emulating the natural distribution of diseased leaves with a specified disease rate. Being annotated by design, such generated datasets are valuable for (1) evaluating the SOTA classifiers when applied to practical use and (2) pre-training general SOTA classifiers, making it possible to effectively fine-tune them using any real-world foliage image dataset for improved classification performance. The Foliagen framework is exemplified by generating foliage image datasets for soybean and tomato. Our evaluation results indicate that five SOTA classifiers on generated datasets with nine disease categories achieve accuracy up to 87% for soybean and 86% for tomato under $\gamma = 5\%$, and that they all exhibit less than 92% in classifying the real soybean foliage image dataset (with just two disease categories). Foliagen makes it possible to generate crop foliage image datasets to evaluate future disease classifiers objectively, aiming at in-field applications.

1 INTRODUCTION

Crops have been indispensable to human civilization since its inception, serving as a fundamental source of food, medicine, clothing, shelter, and oxygen. Extensive pursuits in crop's structure, phyllotaxis phenomenon (8; 20; 25), life cycle, and disease have been undertaken, aiming at yield improvement to meet growing demands. According to the Food and Agriculture Organization (FAO), crop diseases cost approximately $220 billion annually in the world (1), with soybean alone accounting for a loss up to $3.9 billion USD in the USA (6). They usually show prominent symptoms, manifesting themselves as changes in the soybean's leaf foliar appearance and/or shape. For example, rust in soybean exhibits small, pale green to yellow spots on the upper surface of leaves (see Figure 1(f)). Since many crop disease categories possibly exist as illustrated in Figures 1 and 2, early disease identification makes it possible to apply proper measures at the onset of diseases for curbing damage they may cause, retaining crop yields as best as possible.

Instead of relying on experienced farmers for disease identification, machine learning (ML) has been adopted (2; 19; 26; 32; 33; 40; 42) recently to classify soybean and tomato diseased leaf images with success. ML models automate disease classification through training on large amounts of diseased and healthy leaf images. As exemplified in Figures 1 and 2, quality images of individual diseased soybean and tomato leaves exist in public datasets (4; 18; 31) for model training. A state-of-the-art (SOTA) ML model targeting soybean disease classification is shown to have an average accuracy above 85% on a non-public dataset of individual leaf images gathered from the soybean plantations

of the author's college (40). Meanwhile, two SOTA ML models for tomato disease classification are demonstrated to enjoy high accuracy rates, with one under two public datasets of individual leaf images (given in (13; 18)) to exceed 99% (33). Note that those SOTA classifiers often resort to specific augmentation strategies; for example, random masking on the individual soybean leaf images (40) or the Gaussian filter for enhancing and obscuring artifacts of individual tomato leaf images (33). Such augmentation strategies can be expensive and less effective when applied to images with large numbers of leaves, like crop foliage images (see Figure 5 and Figures 6 and 7).

Although SOTA ML models (e.g., (26; 33; 40; 42)) are high classification performers on the existing datasets of individual leaf images, they lack practicality, for in-field applications, where sequences of foliage images are usually captured over areas of interest by the cameras of unmanned aerial vehicles (UAVs) for disease identification. Therefore, accurate automated disease classification on foliage images is essential for practical applications. While one early work classifies soybean foliage images captured by UAVs via first employing a segmentation process to separate each foliage image into parts with diseased leaves and parts with solely healthy leaves (37), and then manually annotating all segmented parts individually before classification takes place. Although that early work dealt with a small collection of foliage images covering only two diseased categories, it required highly experienced agronomists with substantial time and effort to annotate segmented parts of foliage images for accurate classification, considered to be excessively expensive and impractical for in-field applications where large volumes of images are involved. So far, there is just one annotated soybean foliage image dataset available to the public, MH-SoyaHealthVision (28), and no publicly available dataset of tomato foliage images exists, to the best of our knowledge. Unfortunately, the MH-SoyaHealthVision dataset fails to include many predominant soybean disease categories (see Figure 5 and Figures 6 and 7 in Appendix) and is likely to be inadequate to train an effective classifier for identifying the diseases at their early stages, since its images were not captured at the disease onset and thus often had considerable diseased leaves each.

This paper introduces a framework to generate crop foliage images utilizing available datasets of individual leaf images, called Foliagen (short for foliage generation). A generated foliage image dataset can be arbitrarily sized, with each of its images having a specified rate of diseased leaves (denoted by γ) and the rest being healthy ones. Such generated datasets are annotated by design, tailored to in-field applications for foliage image classification. They are valuable for (1) evaluating the SOTA classifiers when applied to practical use and (2) pre-training general SOTA classifiers, making it possible to fine-tune them using any real-world foliage image dataset for improved classification performance. The Foliagen framework is exemplified by generating foliage image datasets for soybean and tomato, using the individual diseased leaf images from ASDID (4), PlantVillage (18), and Kaggle (31) datasets. Foliagen takes as its input (1) the disease category and (2) the rate (γ) of diseased leaves in a foliage image. Each generated foliage image dataset covers nine predominant disease categories for soybean and tomato (samples depicted in Figure 5 and Figures 6 and 7 in Appendix), respectively. Such datasets make it possible to train classifiers for early disease identification when foliage images are generated under a small γ (say, 5%).

Unlike an individual leaf image where background or noise takes up its considerable area, a foliage image synthesized by Foliagen is dominated by soybean or tomato leaves, with only a small fraction of its area being background (see Figure 5 and Figures 6 and 7 in Appendix). It is found from our evaluation results that SOTA models classify nine disease categories less accurately on the generated soybean foliage image datasets than on the original individual leaf images, at varying degrees. As listed in Table 1 for soybean disease classification, SOTA classifiers are subject to accuracy reduction by up to 3% (or 12%) on generated foliage image datasets with γ = 15% (or 5%), rendering the best performer (VGG19 (29)) for classifying individual soybean images to be less attractive for foliage image classification. Similar performance degradation is observed for tomato disease classification, up to 17% (or 22%) reduction in accuracy for Swin Transformer (22) under generated foliage image datasets with γ = 15% (or 5%), as shown in Table 3. The most effective classifier under foliage image datasets is DenseNet121 (17), instead of VGG19 (29) on the datasets of individual tomato leaf images. Hence, Foliagen establishes foliage image datasets useful for candidly assessing known and future classifiers to identify ones that are most effective for in-field applications.

In addition, classifiers pre-trained by a generated soybean foliage image dataset with γ = 15% is confirmed to perform equally well (with VGG19 to achieve 93^{+}% accuracy) under the only known real-world foliage image dataset (e.g., MH-SoyaHealthVision (28)) via employing a small fraction of the dataset (say, 10%) to fine-tune the pre-trained classifiers, as the result of transfer learning.

Classifiers so pre-trained learns abstract features of diseased soybean foliage images under various disease categories, making it possible to adapt soundly for classifying the real-world foliage images with high accuracy. The overall contributions of this paper are summarized as follows:

- A framework for generating annotated foliage image datasets (Foliagen) is introduced by utilizing public datasets of individual crop leaf images. Datasets so generated can be arbitrarily large, properly annotated, and aimed to cover various crop disease categories common in the field and to target early disease identification by setting a small γ (say, 5%).
- Foliagen is exemplified by generating foliage image datasets for soybean and tomato, with the generated datasets used for evaluating SOTA classifiers to determine the most effective ones for real-world applications.
- We have demonstrated that generated foliage image datasets can pre-train a general model for crop disease classification, so that the pre-trained model can then be fine-tuned by a small fraction of any real-world foliage image dataset for high classification accuracy under the dataset, as a result of transfer learning.

2 Related work

2.1 Plant leaf datasets

Many plant leaf datasets are available to the public, with some covering a variety of plant species each, such as PlantVillage (18) and PlantDoc (30), and others being plant-specific, including those for soybean and tomato given in (4; 13; 18; 28; 31). Existing plant leaf datasets are outlined briefly below, with more details about the PlantVillage dataset (18), ASDID (4), a Kaggle dataset (31), and the MH-SoyaHealthVision dataset (28) provided in Section 3.1.

PlantVillage dataset. PlantVillage (18) contains over 54,300 expertly curated healthy and diseased leaf images from various plants, including thirteen major crop species like soybean, tomato, etc. Its leaf classification has been attempted by GoogleNet (34) and AlexNet (21) to attain high accuracy.

PlantDoc dataset. The PlantDoc dataset (30) contains 2,598 single-leaf images of 17 disease categories across 13 plant species, including tomato and soybean. With its images gathered in a controlled laboratory environment, this dataset has limited applicability in real-world scenarios.

ASDID. Auburn Soybean Diseased Image Dataset (ASDID) (4) provides high-quality individual leaf images of the soybean plant, covering 9 disease categories, namely, bacterial blight, cercospora leaf blight, downy mildew, frogeye leaf spot, soybean rust, target spot, and potassium deficiency. The dataset was captured primarily at the EV Smith Agricultural Research Station in Tallassee, Alabama, and added with 80 images per disease category from the publicly available Image Database of Plant Disease Symptoms (PDDB) (3). Several ML-based classifiers (15; 17; 29) were employed to evaluate the dataset, with DenseNet201 (17) achieving the highest performance.

Kaggle dataset. Soybean Diseased Leaf Dataset contains individual leaf images from Kaggle (31), embracing 10 disease categories and having an artificially generated complex background added to each image. Its disease categories include brown spot, frogeye, mosaic virus, southern blight, sudden death syndrome, yellow mosaic, crestamento, ferrugen, powdery mildew, and septoria.

FieldPlant dataset. FieldPlant (24) is a dataset of individual crop leaf images annotated by pathologists, containing 8,629 images across 27 disease categories for three crops, including tomato. It aims at practical crop disease classification with every leaf image involving a complex background.

Tomato-Village dataset. Compensating for the negative effects due to the laboratory-controlled environment setup for gathering PlantVillage's leaf images, the Tomato-Village dataset (13) contains real-world images, which belongs to three groups, respectively for (1) multi-class tomato disease classification, (2) multi-label tomato disease classification, and (3) object detection-based tomato disease detection. This dataset covers seven disease categories, namely, early blight, late blight, leaf miner, magnesium deficiency, nitrogen deficiency, potassium deficiency, and spotted wilt virus, for multi-class classification applications.

MH-SoyaHealthVision dataset. As far as we know, MH-SoyaHealthVision (28) is the only publicly available and well-annotated dataset of diseased soybean foliage images. The dataset provides both ground-level leaf images and foliage leaf images, collected using a UAV, from the soybean

fields of Maharashtra, India. It comprises a total of 5,680 high-resolution images grouped into (1) single leaf images of 4 diseases categories (i.e., frogeye, mosaic virus, septoria brown spot, and rust), two types of pest attacks, and healthy leaves and (2) UAV-captured soybean foliage images, which belong to two disease categories of mosaic virus and rust, plus the healthy category.

All above datasets were reviewed and carefully examined for possible use by our Foliagen framework; however, only the PlantVillage, ASDID, and the Kaggle datasets were selected as single-leaf image sources due to their large numbers of well-annotated images, with high fidelity and clarity.

2.2 Crop leaf disease classification models

Automated disease classification studies have been conducted lately based on aforementioned publicly available datasets, plus certain privately collected datasets, to exhibit impressive classification outcomes (2; 5; 19; 26; 32; 33; 40; 42; 43). In particular, SoyNet (19) used a two-staged approach, with one stage to subtract complex background from leaf images and the other to employ a CNN-based technique to classify the single leaf images into 16 categories. Bevers *et al.* (4) in Auburn, AL collected a high-quality single leaf imagery dataset and employed standard CNN-based models, such as VGG19, DenseNet201, and ResNet50, etc., to classify their collected dataset, achieving high accuracy. Enhanced DenseNet121 (42) relied on transfer learning to tackle the intricate challenges of classifying individual soybean leaf images. Using a conditional generative adversarial network (C-GAN) to generate synthetic individual tomato diseased leaves, previous work (2) achieved very high performance employing DenseNet121 (17) as its classification model. Merging classical feature engineering with modern machine learning techniques, Bouni *et al.* (5) recently have employed a CNN pre-trained on ImageNet (9) under mutual information-based feature fusion to get a high performer for tomato disease classification. Previously, UAVs were deployed to capture soybean foliage images from plantation fields and applied a segmentation method to separate each foliage image into parts with diseased leaves and parts with only healthy leaves, before manually annotating all segmented parts, for high accuracy.

While CNN-based models are capable of capturing the prominent features of images, incorporating an attention mechanism to the models enables them to emphasize on the region that contributes most to performance improvement. Previous studies incorporating attention mechanisms (16; 22; 32; 33; 38; 39; 40) have been proven effective for computer vision tasks, such as image segmentation, image classification, and object detection–especially for leaf disease classification, where leaf colors and shapes are key features of interest (26; 40). Swin Transformer (22) introduces a hierarchical segmentation along with the vision transformer to implement shifted windows, which capture main features across different segmentation regions of an image. In contrast, SE-Net (16) focuses on channel-wise attention by modeling the inter-dependencies among channels to get the adaptive recalibration of channel-wise feature responses. Built upon SE-Net (16), ECA-Net (38) involves an efficient channel attention module realized by local cross-channel interaction without altering the dimension. TFANet (26) is a two-stage feature aggregation framework, which incorporates channel attention to exhibit high accuracy on the ASDID dataset. Meanwhile, CBAM-ConvNeXt (40) employs both channel attention and spatial attention along with ConvNeXt (23), to accurately classify the individual soybean leaf images, which are not publicly available yet. Merging EfficientNetV2 (35) and Swin Transformer (22), Eff-Swin (33) leverages the local features extraction capabilities of CNNs and global modeling ability of Transformers to achieve superior performance compared to individual network models on datasets sourced from PlantVillage (18) and TomatoVillage (13). While exhibiting impressive feats, all known studies (but (37)) only target classifying high-resolution images of individual crop leaves, as illustrated in Figures 1 and 2.

3 Methodology

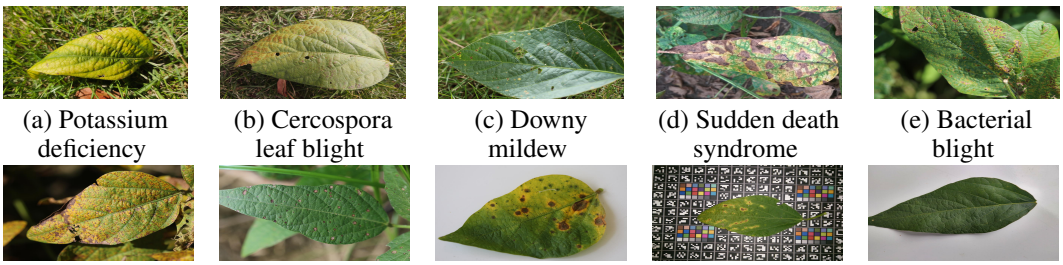
The Foliagen framework generates diseased foliage images out of available single-leaf diseased images for the classification of diseased foliage images for (1) objectively evaluating known and future crop disease classifiers when deployed for in-field applications and (2) pre-training general crop disease classifiers, making them tailored for specific fields with high classification performance after fine-tuned by a small number of annotated foliage images gathered in those fields, as the result of transfer learning. It is exemplified by generating foliage images of soybean and tomato, leveraging three publicly available individual-leaf image datasets, ASDID (4), a Kaggle dataset (31),

and the PlantVillage dataset (18), which are detailed in Section 3.1. Single soybean leaf images are preprocessed (as described in Section 3.1) before being utilized by Foliagen to generate diverse sets of foliage images for evaluating known classifiers objectively. Such a single image manipulation methods has been proven to be effective for image segmentation and object detection data augmentation (10; 14; 27). In addition, known classifiers after being pre-trained by generated foliage images that cover 9 disease categories are fine-tuned via 10% of images from the real-world soybean foliage image dataset, MH-SoyaHealthVision (28), and are found to classify the remaining 80% MH-SoyaHealthVision images with improved performance.

3.1 DATA COLLECTION

Soybean. A total of 10,722 high-resolution single leaf images covering 7 disease categories from ASDID (4) and 132 images (with 22 for Mosaic Virus and 110 for sudden Death Syndrome) from a Kaggle dataset (31) were chosen for foliage image generation. Images of those nine disease categories (as depicted in Figure 1) feature diverse and complex natural/artificial backgrounds, which are undesirable when generating foliage images. Therefore, we pre-process the images to remove their undesired backgrounds (i.e., to make them transparent) using an open-sourced and AI-enabled background remover, rembg (12), as shown in the first stage of Overall Foliagen depicted in Figure 3. Freely available soybean field soil images are then included in generated foliage images as their backgrounds, to emulate the natural habitat of soybean plants as best as possible.

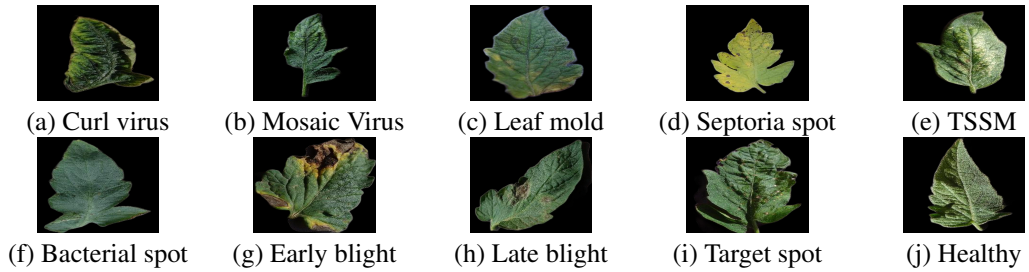
The only publicly available and properly annotated dataset of diseased foliage images, MH-SoyaHealthVision (28), comprises two disease categories: soybean rust and mosaic virus. The original images have a very high resolution of 3840×2160 and the dataset is imbalanced, with rust images outnumbering healthy images by a factor of four, leading to biased predictions favoring the majority class and consequently reducing the model's generalizability. To address these issues, we crop the high-resolution foliage images into ones with a lower resolution to ensure a more balanced data distribution and to obtain a total of 3210 images, comprising 1084 rust images, 1027 healthy images, and 1099 mosaic virus images, respectively.



(a) Potassium deficiency (b) Cercospora leaf blight (c) Downy mildew (d) Sudden death syndrome (e) Bacterial blight

(f) Rust (g) Frogeye (h) Target spot (i) Mosaic virus (j) Healthy

Figure 1: Soybean single-leaf images of 9 diseases, labeled by (a) to (i).



(a) Curl virus (b) Mosaic Virus (c) Leaf mold (d) Septoria spot (e) TSSM

(f) Bacterial spot (g) Early blight (h) Late blight (i) Target spot (j) Healthy

Figure 2: Tomato single leaf images of 9 diseases, labeled by (a) to (i).

Tomato. The single-leaf diseased images of tomato are from the PlantVillage dataset (18), with 9 primary tomato disease categories, and they are taken in a laboratory environment and with the dimension of 256×256, as shown in Figure 2.

3.2 FOLIAGEN FRAMWORK

Foliagen takes pre-processed individual leaf images as input to generate foliage images with a customizable rate of diseased leaves. After preprocessing single leaf images for background removal via rembg (12), it then involves 3 levels of generation, Leaf level, Plant level, and Foliage level, as depicted in Figure 3.

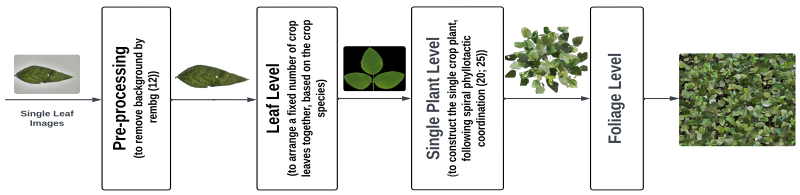


Figure 3: Overall Foliagen.

Leaf level. Soybean leaves arrange themselves into 3 leaflets in each petiole, with a slightly bigger central leaflet and two lateral leaflets on either side of the central leaflet. On the other hand, tomato leaves arrange themselves around a central axis, called the rachis. In practice, crop leaflets may or may not be diseased. Hence, a random number of healthy leaflets are included in each foliage image, governed by the disease rate γ.

Plant level. The number of leaves in an adult crop differs largely, based on the crop. An adult soybean plant might contain 30-40 trifoliates, while an adult tomato plant may have 20-40 leaves, assuming the determinate variety of tomato crop. Both crops exhibit spiral phyllotaxis (20; 25), in which leaves are arranged in a spiral arrangement that makes the golden angle, i.e, 137.5°, to maximize sunlight exposure and minimize leaf overlap. Most of the leaves in a crop are healthy in the early stages of a disease, so foliagen takes a small disease rate γ (e.g., 5% or 15%). A spiral phyllotactic coordinates generator is the vital part of this Plant level, following the formulae given below to determine the coordinates and the angles of leaves to maintain spiral phyllotaxis (25).

- Center coordinates: (x_0, y_0), Golden angle: $\theta_g = 137.5^\circ$, Scaling factor: $s = 35$
- Number of trifoliates: $N = X \sim \mathcal{U}\{30, 31, \ldots, 40\}$

For each trifoliate index $n \in \{1, 2, ..., N-1\}$:

$$\text{Angular displacement:} \quad \theta_n = n \cdot \theta_g^{\text{rad}} \qquad \text{Radial distance:} \quad r_n = s \cdot \sqrt{n}$$

$$\text{Cartesian coordinates:} \quad \begin{cases} x_n = x_0 + r_n \cos(\theta_n) \\ y_n = y_0 + r_n \sin(\theta_n) \end{cases}$$

The final discrete leaf positions are expressed by:

$$\mathbf{coords} = \{(\lfloor x_n \rfloor, \lfloor y_n \rfloor) \mid n = 1, 2, ..., N-1\},$$

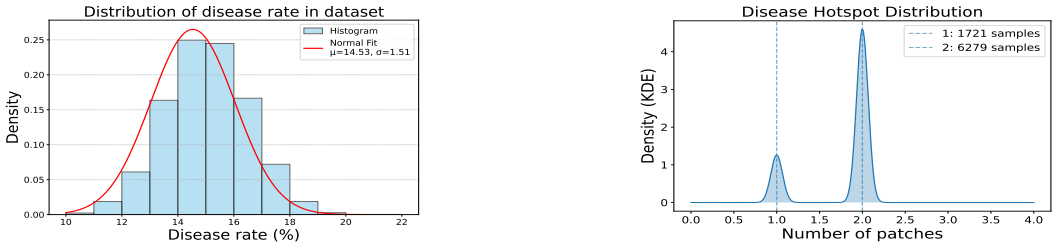
where $\lfloor \cdot \rfloor$ denotes the integer truncation.

As a tomato plant has branches, with a pair of leaves attached at a similar stem height and arranged in opposite directions, a sub-layer, called the branch layer, is added to create branches, each with 3-9 leaves. Such an emulated branch is then attached to the main stem in spiral phyllotactical order.

Foliage level. Foliage images usually consist of multiple rows of crops planted in a farm field. Observation from real-world images taken using UAVs (28) reveals that the major area of an image is covered by leaves, with only a small area being field soil (11). As a result, Foliagen generates images with three rows of crops in each image to emulate their natural appearance. Samples of generated foliage images are illustrated in Figure 5 and Figures 6 and 7 in Appendix.

3.3 DISEASE DISTRIBUTION

The distribution of disease in a plant leaf is shaped by multiple interacting factors, including insect vectors, wind-mediated spore dispersal, plantation age, and environmental conditions such as humidity, rainfall, and temperature. Although the spatial pattern of disease may vary considerably



(a) Disease rate distribution

(b) Hotspot distribution

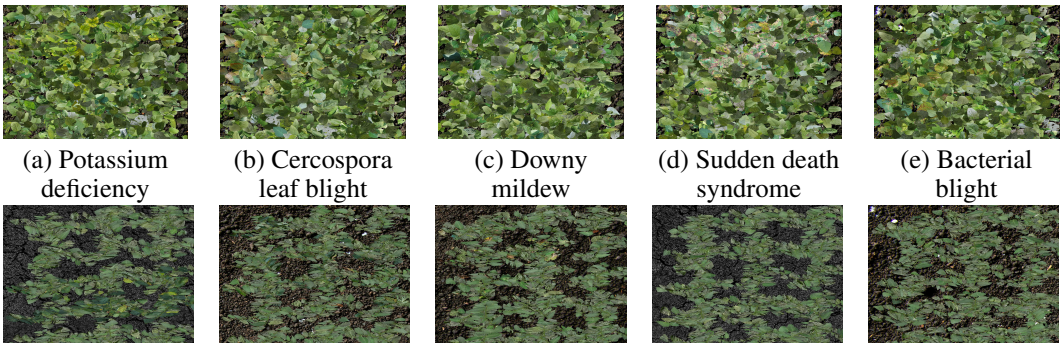
Figure 4: Disease distributions across diseased soybean foliage dataset (γ = 15%).

depending on these influences, a consistent phenomenon is that infections generally begin as localized hotspots on leaves or within a small plant patch and subsequently spread outward to one (or multiple) neighboring patch(es), ultimately forming larger area of diseased foliage (7; 36; 41). This diseased leaf distribution is confirmed by us through examining the foliage images of various disease categories in the available MH-SoyaHealthVision dataset (28), leading us to devise a three-level disease distribution pattern, as explained next.

Region level. The whole image is divided into a $\alpha \times \beta$ grid with a total of $\alpha \times \beta$ regions. e.g., $4 \times 3 = 12$ regions. Most regions are disease-free when the disease rate (γ) is small, say $\leq 20\%$. region is provided with a disease rate, so that all $\alpha \times \beta$ regions have the aggregate disease rate of γ. The diseased leaves in a diseased region are distributed normally across the region.

Patch level. A patch refers to a collection of regions with one hotspot and its neighboring regions. Based on the disease rate, the number of diseased patches in a single foliage image varies from 1 to 2 under our disease rate of interest to be less than 20% for early disease identification. More disease patches in a foliage image are expected for a higher disease rate. The distribution of the number of hotspots among the foliage images follows a skewed graph. Figure 4(b) shows the distribution of two hotspots for the mean disease rate of 15%, where the number of hotspot disease patches per image being 1 (or 2) equals 1,721 (or 6,279) out of 8,000 total generated foliage images.

Dataset level. The disease rate of the foliage images is normally distributed with the standard deviation (σ) of 1.5 and the variable mean ($\mu = \gamma$) of 15% (or 5%). Figure 4(a) depicts the disease rate distribution for γ = 15%, where the disease rate varies from 10% to 20%. For γ = 5%, the disease rate ranges from 1% to 9%, with a similar normal distribution as illustrated in Figure 4(a).



(a) Potassium deficiency

(b) Cercospora leaf blight

(c) Downy mildew

(d) Sudden death syndrome

(e) Bacterial blight

(f) Bacterial spot

(g) Early blight

(h) Late blight

(i) Target spot

(j) Healthy

Figure 5: Generated soybean foliage images (a)-(e) and tomato foliage images (f)-(j), for γ = 15% (All generated foliage image categories for soybean and tomato are shown respectively in Figures 6 and 7 in Appendix).

Crop Level Customization. Foliagen is a common framework, aiming to generate a foliage imagery dataset for various crops out of those crops' single leaf images. Given crops differ among one another in many factors, such as the leaf shapes, the leaf arrangements, phyllotaxis, number of leaves in single branch, numbers of leaves in single branches, etc., Foliagen is provisioned with a configuration file as its input to account for the crops' variability, with the file listing such crop-specific customization parameters as the disease rate, the size of individual leaf, the size of individual plant, the foliage size, disease categories, etc., to properly emulate crops' natural structures.

4 EVALUATION AND RESULT DISCUSSION

Extensive experiments to evaluate the performance of SOTA crop disease classifiers on generated foliage images are conducted on two workstations, with one housing 2 NVIDIA GeForce RTX 3090 GPUs (each with 24 GB of GDDR6X VRAM) and another housing 2 NVIDIA RTX 6000 Ada GPUs (each with 48 GB of GDDR6X VRAM). The foliage image datasets of soybean and tomato for γ = 5% and 15% have been generated for evaluation. Each generated foliage image dataset contains about 800 images for every disease category, plus a similar number of healthy foliage images. The Adam optimizer was used for model training, since it is known to converge faster with better performance by dynamically adjusting the learning rate for each parameter based on the first and second moments of gradients. Each model was trained for a maximum of 100 epochs, with an early stopping mechanism (with patience of 5 epochs) to avoid local minima. The batch size was set to 4, restricted by the GPU memory limitation, and the learning rate was initialized to 0.000001.

4.1 EVALUATION ON GENERATED SOYBEAN FOLIAGE IMAGE DATASETS

Various generated soybean foliage image datasets under different γ (the rate of diseased leaves in each foliage image) values have been produced by Foliagen to objectively evaluate the SOTA disease classification models of VGG19 (29), ResNet50 (15), DenseNet121 (17), Swin Transformer (22), and CBAM-ConvNeXt (40) under exactly the same set of foliage images without any classifier-specific data pre-processing or manipulation. Each produced foliage image dataset covers all crop disease categories that exist in the original datasets of individual leaf images, with a small γ (say, 5%) to indicate an early disease stage. The evaluation results shed light on choosing the best classifier among those SOTA models for real-world soybean applications, where disease identification is based on in-field images captured by cameras mounted on UAVs. The evaluation metric outcomes under a synthetic dataset with a larger γ are expected to be higher because more diseased leaves exist in each foliage image, making disease classification easier. The comparative performance evaluation results are obtained for γ ranging from 5% to 15% and beyond, and they are found to follow similar trends. For simplicity, only the results for γ = 5% to 15% are listed in Table 1. It is evident from the table that DenseNet121 prevails for both γ values, in terms of all the metrics.

Table 1: Comparative performance evaluation results (in %) under generated soybean foliage image dataset with γ = 15% (or 5%)

Models	**Accuracy**		**F1-score**		**Precision**		**Recall**	
VGG19 (29)	76.26	(65.45)	77.17	(65.67)	78.11	(66.19)	76.26	(65.15)
ResNet50 (15)	85.31	(81.91)	85.40	(83.58)	85.69	(85.37)	85.11	(81.87)
DenseNet121 (17)	**94.47**	(**87.56**)	**95.45**	(**90.81**)	**96.45**	(**94.45**)	**94.48**	(**87.44**)
Swin Transformer (22)	72.36	(65.38)	72.99	(64.50)	72.37	(63.62)	72.68	(65.41)
CBAM-ConvNeXt (40)	77.76	(66.33)	79.73	(69.59)	81.84	(73.15)	77.72	(66.36)

4.2 EVALUATION ON REAL SOYBEAN FOLIAGE IMAGE DATASET

Baseline. The MH-SoyaHealthVision dataset (28) was split into 80% for training, 10% for validation, and 10% for evaluation, enabling an objective evaluation of the same five SOTA classifiers under exactly the same set of real-world foliage images without any data preprocessing or manipulation. From the comparative performance results summarized under Baseline of Table 2, it is found that Swin Transformer achieves the highest performance, with accuracy exceeding 91% across all four metrics, whereas other models have the accuracy values ranging from $85^{+}\%$ to $89^{-}\%$. The baseline results indicate that Swin Transformer is the top performer for real-world applications.

Transfer learning. Generated foliage image datasets can pre-train crop disease classification models to get powerful disease classifiers suitable for general applications. After those five SOTA models are pre-trained by our generated foliage images to cover nine categories of predominant soybean diseases, they are expected to serve as general classifiers for effectively identifying any real-world dataset of soybean foliage images at hand by fine-tuning them using a small fraction of foliage images in the dataset, due to transfer learning. When classifiers are pre-trained by a generated foliage image dataset with a small γ (say, 5%), they are geared for identifying soybean diseases at an early stage, especially useful for real-world field applications. To this end, the trained models are evaluated under the real-world MH-SoyaHealthVision dataset (28), after being fine-tuned via 5% images

in the dataset, with the evaluation results listed under Pre-trained in Table 2. Note that the results are obtained when 10% and 85% dataset images are for validation and evaluation, respectively, after 5% images are employed for fine-tuning. Comparing the obtained evaluation results shown in Table 2, we find that the trained models with fine-tuning elevate performance metric values noticeably, to exceed 92% in accuracy for DenseNet121, ResNet50, CBAM-ConvNeXt, and Swin Transformer under γ = 15%. The performance results of pre-trained models are worse under γ = 5% than under γ = 15%, as expected, since the former aimed to detect diseases in an early stage, known to be harder but more useful. They also signify that Swin Transformer is the most desirable for in-field applications, when aiming at early disease detection (under γ = 5%).

Table 2: Comparative performance evaluation results (in %) under MH-SoyaHealthVision

	Baseline				**Pre-trained with γ = 15% (or 5%)**			
Models	**Accuracy**	**F1-score**	**Precision**	**Recall**	**Accuracy**	**F1-score**	**Precision**	**Recall**
VGG19	85.38	86.6	87.98	85.38	88.00 (84.03)	88.22 (84.06)	88.41 (84.08)	88.03 (84.03)
ResNet50	88.31	89.69	91.11	88.31	**95.83** (90.61)	**95.91** (90.86)	**95.98** (91.12)	**95.84** (90.60)
DenseNet121	87.5	87.95	88.37	87.5	94.40 (**94.47**)	94.48 (**94.19**)	94.55 (**94.29**)	94.40 (**94.08**)
Swin Transformer	**91.48**	**91.47**	**91.46**	**91.48**	92.32 (87.57)	91.33 (88.52)	90.39 (89.51)	92.29 (87.56)
CBAM-ConvNeXt	88.82	87.06	85.39	88.83	95.38 (86.76)	95.37 (88.06)	95.36 (89.42)	95.38 (86.74)

4.3 TOMATO

Foliagen synthesizes various disease datasets of tomato foliage images based on the PlantVillage dataset (18) (publicly available datasets of single leaf images with 9 primary tomato disease categories), under a range of γ for evaluating the SOTA classifiers. The evaluation metric outcomes under a generated dataset with a larger γ are expected to be higher because more diseased leaves exist in each foliage image, making disease classification easier. Given that the comparative evaluation results are obtained for γ ranging from 5% to 15% and beyond follow similar trends, Table 3 lists only the results for γ = 5% and 15%. As evident from the table, the considered models all perform better under 15% than under 5% with respect to the four performance metrics, underscoring the fact that they tend to struggle in early disease detection (under γ = 5%), especially for VGG19, ResNet50, and Swin Transformer. The evaluation results imply that DenseNet121 outperforms the rest consistently, making it the most desirable classifier for in-field applications for identifying tomato diseases according to foliage images captured in the field by UAVs.

Table 3: Comparative performance evaluation results (in %) under generated tomato foliage image dataset with γ = 15% (or 5%)

Models	**Accuracy**		**F1-score**		**Precision**		**Recall**	
VGG19 (29)	87.66	(80.53)	87.71	(81.01)	87.64	(81.51)	87.78	(80.50)
ResNet50 (15)	92.22	(78.84)	92.44	(81.39)	92.65	(84.11)	92.23	(78.83)
DenseNet121 (17)	**96.38**	(**86.93**)	**97.34**	(**89.47**)	**97.69**	(**92.16**)	**96.99**	(**86.93**)
Swin Transformer (22)	79.80	(66.53)	79.36	(66.71)	78.84	(66.81)	79.88	(66.60)
CBAM-ConvNeXt (40)	81.91	(64.85)	81.68	(65.56)	82.96	(66.24)	81.96	(64.89)

5 CONCLUSION

This article introduces a framework (called Foliagen) to generate rich and arbitrarily-sized datasets of crop foliage images to cover all disease categories that exist in publicly available datasets of individual leaf images, with a given rate of diseased leaves (γ) in each foliage image generated to emulate the real foliage images captured in farm fields when their crop diseases are at the stage corresponding to γ. The generated foliage datasets are employed to better and objectively evaluate state-of-the-art leaf disease classifiers without invoking classifier-specific data pre-processing or manipulation. The evaluation results make it possible to choose the most effective crop classifier among SOTA ones for in-field applications with UAV-captured images (rather than individual leaf images) for disease identification. Being a generated foliage dataset, its primary limitation lies in the lack of naturalness and limited real-world applicability; however, the strong performance of crop disease classifiers pre-trained on it suggests its potential viability for broader applications. With an available in-field foliage dataset, the pre-trained models can be fine-tuned using a small fraction of the dataset images to yield effective disease classifiers targeting the field where the foliage dataset is gathered. While Foliagen is exemplified for classifying soybean and tomato diseases via SOTA models in this paper, it is readily useful for other crops and for objectively evaluating future disease classifiers aiming at in-field applications.

REFERENCES

[1] Pathogens, precipitation and produce prices. *Nature Climate Change*, 11(8):635–635, 2021.

[2] Amreen Abbas, Sweta Jain, Mahesh Gour, and Swetha Vankudothu. Tomato plant disease detection using transfer learning with c-gan synthetic images. *Computers and Electronics in Agriculture*, 187:106279, 2021.

[3] Jayme Garcia Arnal Barbedo, Luciano Vieira Koenigkan, and Thiago Teixeira Santos. Identifying multiple plant diseases using digital image processing. *Biosystems Engineering*, 147:104–116, 2016.

[4] Noah Bevers, Edward J Sikora, and Nate B Hardy. Soybean disease identification using original field images and transfer learning with convolutional neural networks. *Computers and Electronics in Agriculture*, 203:107449, 2022.

[5] Mohamed Bouni, Badr Hssina, Khadija Douzi, and Samira Douzi. Synergistic use of handcrafted and deep learning features for tomato leaf disease classification. *Scientific Reports*, 14(1):26822, 2024.

[6] Carl A. Bradley, Tom W. Allen, Adam J. Sisson, Gary C. Bergstrom, Kaitlyn M. Bissonnette, Jason Bond, Emmanuel Byamukama, Martin I. Chilvers, Alyssa A. Collins, John P. Damicone, Anne E. Dorrance, Nicholas S. Dufault, Paul D. Esker, Travis R. Faske, Nicole M. Fiorellino, Loren J. Giesler, Glen L. Hartman, Clayton A. Hollier, Tom Isakeit, Tamra A. Jackson-Ziems, Douglas J. Jardine, Heather M. Kelly, Robert C. Kemerait, Nathan M. Kleczewski, Alyssa M. Koehler, Robert J. Kratochvil, James E. Kurle, Dean K. Malvick, Samuel G. Markell, Febina M. Mathew, Hillary L. Mehl, Kelsey M. Mehl, Daren S. Mueller, John D. Mueller, Berlin D. Nelson, Charles Overstreet, G. Boyd Padgett, Paul P. Price, Edward J. Sikora, Ian Small, Damon L. Smith, Terry N. Spurlock, Connie A. Tande, Darcy E.P. Telenko, Albert U. Tenuta, Lindsey D. Thiessen, Fred Warner, William J. Wiebold, and Kiersten A. Wise. Soybean yield loss estimates due to diseases in the united states and ontario, canada, from 2015 to 2019. *Plant Health Progress*, 22(4):483–495, 2021.

[7] Chen Chen, Harald Keunecke, Enzo Neu, Friedrich J Kopisch-Obuch, Bruce A McDonald, and Jessica Stapley. Molecular epidemiology of cercospora leaf spot on resistant and susceptible sugar beet hybrids. *Plant Pathology*, 74(1):69–83, 2025.

[8] Jonas Coussement, Michael Henke, Peter Lootens, Isabel Roldán-Ruiz, Kathy Steppe, and Tom De Swaef. Modelling leaf spectral properties in a soybean functional–structural plant model by integrating the prospect radiative transfer model. *Annals of Botany*, 122(4):669–676, 2018.

[9] Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei. Imagenet: A large-scale hierarchical image database. In *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 248–255. IEEE, 2009.

[10] Debidatta Dwibedi, Ishan Misra, and Martial Hebert. Cut, paste and learn: Surprisingly easy synthesis for instance detection. In *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, pages 1301–1310, 2017.

[11] Freepik. Freepik: Graphic resources for everyone. `https://www.freepik.com/`.

[12] Daniel Gatis. rembg: Remove image background. `https://github.com/danielgatis/rembg`, 2021. Accessed: 2025-05-08.

[13] Mamta Gehlot, Rakesh Kumar Saxena, and Geeta Chhabra Gandhi. Tomato-village: a dataset for end-to-end tomato disease detection in a real-world environment. *Multimedia Systems*, 29(6):3305–3328, 2023.

[14] Golnaz Ghiasi, Yin Cui, Aravind Srinivas, Rui Qian, Tsung-Yi Lin, Ekin D Cubuk, Quoc V Le, and Barret Zoph. Simple copy-paste is a strong data augmentation method for instance segmentation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 2918–2928, 2021.

[15] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition (CVPR)*, pages 770–778, 2016.

[16] Jie Hu, Li Shen, and Gang Sun. Squeeze-and-excitation networks. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 7132–7141, 2018.

[17] Gao Huang, Zhuang Liu, Laurens Van Der Maaten, and Kilian Q Weinberger. Densely connected convolutional networks. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 4700–4708, 2017.

[18] David Hughes, Marcel Salathé, 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.

[19] Aditya Karlekar and Ayan Seal. SoyNet: Soybean leaf diseases classification. *Computers and Electronics in Agriculture*, 172:105342, 2020.

[20] Kanahama Koki, Iwabuchi Kazunori, and Minoru Suda. Phyllotaxis on the main shoot of wild tomato plants calculated by the parastichy system. *Tohoku Journal of Agricultural Research*, 44:1–4, 1994.

[21] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. Imagenet classification with deep convolutional neural networks. In F. Pereira, C.J. Burges, L. Bottou, and K.Q. Weinberger, editors, *Proceedings of Advances in Neural Information Processing Systems*, volume 25. Curran Associates, Inc., 2012.

[22] Ze Liu, Yutong Lin, Yue Cao, Han Hu, Yixuan Wei, Zheng Zhang, Stephen Lin, and Baining Guo. Swin transformer: Hierarchical vision transformer using shifted windows. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (CVPR)*, pages 10012–10022, 2021.

[23] Zhuang Liu, Hanzi Mao, Chao-Yuan Wu, Christoph Feichtenhofer, Trevor Darrell, and Saining Xie. A convnet for the 2020s. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 11976–11986, 2022.

[24] Emmanuel Moupojou, Appolinaire Tagne, Florent Retraint, Anicet Tadonkemwa, Dongmo Wilfried, Hyppolite Tapamo, and Marcellin Nkenlifack. Fieldplant: A dataset of field plant images for plant disease detection and classification with deep learning. *IEEE Access*, 11:35398–35410, 2023.

[25] Karl J. Niklas. The role of phyllotactic pattern as a "development constraint" on the interception of light by leaf surfaces. *Evolution*, 42(1):1–16, 1988.

[26] Renyong Pan, Jianwu Lin, Jitong Cai, Licai Zhang, Jiaming Liu, Xingtian Wen, Xiaoyulong Chen, and Xin Zhang. A two-stage feature aggregation network for multi-category soybean leaf disease identification. *Journal of King Saud University-Computer and Information Sciences*, 35(8):101669, 2023.

[27] Tal Remez, Jonathan Huang, and Matthew Brown. Learning to segment via cut-and-paste. In *Proceedings of the European conference on computer vision (ECCV)*, pages 37–52, 2018.

[28] Sayali Shinde and Vahida Attar. MH-SoyaHealthVision: An indian UAV and leaf image dataset for integrated crop health assessment, 2024.

[29] Karen Simonyan and Andrew Zisserman. Very deep convolutional networks for large-scale image recognition. In *Proceedings of 3rd International Conference on Learning Representations (ICLR)*, pages 1–14, 2015.

[30] Davinder Singh, Naman Jain, Pranjali Jain, Pratik Kayal, Sudhakar Kumawat, and Nipun Batra. Plantdoc: A dataset for visual plant disease detection. In *Proceedings of the 7th ACM IKDD CoDS and 25th COMAD*, pages 249–253. 2020.

[31] Sivm205. Soybean diseased leaf dataset, 2023.

[32] Jianming Sun, Xuyao Hao, and Mengxin Zhao. Image recognition of soybean leaf disease based on bilateral branching network bbn. In *Proceedings of the 2024 4th International Conference on Bioinformatics and Intelligent Computing (BIC)*, pages 47–53, 2024.

[33] Yubing Sun, Lixin Ning, Bin Zhao, and Jun Yan. Tomato leaf disease classification by combining efficientnetv2 and a swin transformer. *Applied Sciences*, 14(17), 2024.

[34] Christian Szegedy, Wei Liu, Yangqing Jia, Pierre Sermanet, Scott Reed, Dragomir Anguelov, Dumitru Erhan, Vincent Vanhoucke, and Andrew Rabinovich. Going deeper with convolutions. In *Proceedings of the IEEE conference on computer vision and pattern recognition (CVPR)*, pages 1–9, 2015.

[35] Mingxing Tan and Quoc Le. Efficientnetv2: Smaller models and faster training. In *Proceedings of International Conference on Machine Learning*, pages 10096–10106. PMLR, 2021.

[36] Jiemeng Tao, Peijian Cao, Yansong Xiao, Zhenhua Wang, Zhihua Huang, Jingjing Jin, Yongjun Liu, Huaqun Yin, Tianbo Liu, and Zhicheng Zhou. Distribution of the potential pathogenic alternaria on plant leaves determines foliar fungal communities around the disease spot. *Environmental Research*, 200:111715, 2021.

[37] Everton Castelão Tetila, Bruno Brandoli Machado, Nícolas Alessandro Belete, David Augusto Guimarães, and Hemerson Pistori. Identification of soybean foliar diseases using unmanned aerial vehicle images. *IEEE Geoscience and Remote Sensing Letters*, 14(12):2190–2194, 2017.

[38] Qilong Wang, Banggu Wu, Pengfei Zhu, Peihua Li, Wangmeng Zuo, and Qinghua Hu. Eca-net: Efficient channel attention for deep convolutional neural networks. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 11534–11542, 2020.

[39] Sanghyun Woo, Jongchan Park, Joon-Young Lee, and In So Kweon. CBAM: Convolutional block attention module. In *Proceedings of the European Conference on Computer Vision (ECCV), pages=3–19*, 2018.

[40] Qinghai Wu, Xiao Ma, Haifeng Liu, Cunguang Bi, Helong Yu, Meijing Liang, Jicheng Zhang, Qi Li, You Tang, and Guanshi Ye. A classification method for soybean leaf diseases based on an improved convnext model. *Scientific Reports*, 13(1):19141, 2023.

[41] XB Yang, JP Snow, and GT Berggren. Patterns of rhizoctonia foliar blight on soybean and effect of aggregation on disease development. *Phytopathology*, 81(3):287–293, 1991.

[42] V Yogabalajee, Karthic Sundaram, and Kamaraj Kanagaraj. Soybean leaf disease classification using enhanced densenet121. In *Proceedings of the 10th International Conference on Advanced Computing and Communication Systems (ICACCS)*, volume 1, pages 1515–1520, 2024.

[43] Xiao Yu, Qi Gong, Cong Chen, and Lina Lu. Automatic diagnosis of soybean leaf disease by transfer learning. *American Journal of Biochemistry and Biotechnology*, 18(2):252–260, 2022.

A Appendix

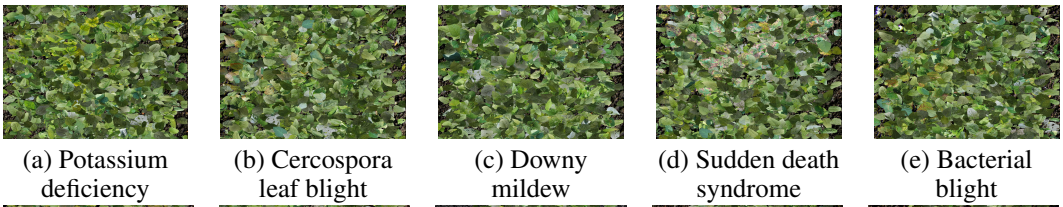
This section contains additional evidence to support our dataset and data generation method.

A.1 Generated Foliage Images

The generated diseased foliage images for both Soybean and Tomato are shown next.

A.2 Additional Experimental Results

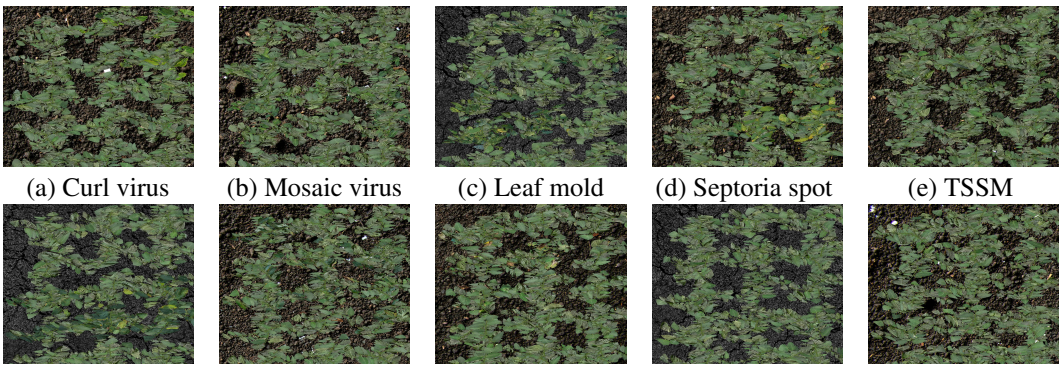
Different experiments have been performed to fortify our dataset's viability, which are presented next.



(a) Potassium deficiency (b) Cercospora leaf blight (c) Downy mildew (d) Sudden death syndrome (e) Bacterial blight

(f) Rust (g) Frogeye (h) Target spot (i) Mosaic virus (j) Healthy

Figure 6: Generated soybean foliage images (a)-(e) and tomato foliage images (f)-(j), for γ = 15%.



(a) Curl virus (b) Mosaic virus (c) Leaf mold (d) Septoria spot (e) TSSM

(f) Bacterial spot (g) Early blight (h) Late blight (i) Target spot (j) Healthy

Figure 7: Generated tomato foliage images of 9 diseases, labeled by (a) to (i), for γ = 15%.

A.2.1 Transfer Learning using ASDID

The models were pre-trained on the raw ASDID dataset and then fine-tuned with 5% of the MH-SoyaVisionHealth dataset. Table 4 illustrates the transfer learning performance with all the hyper-parameters configured as discussed in Section 4.

Table 4: Comparative performance evaluation results (in %) under MH-SoyaVisionHealth dataset, pre-trained on ASDID dataset.

Models	Accuracy	F1-score	Precision	Recall
VGG19 (29)	33.33	32.56	33.33	32.94
ResNet50 (15)	**74.84**	**75.31**	**75.80**	**74.82**
DenseNet121 (17)	35.37	34.43	33.55	35.36
Swin Transformer (22)	69.45	69.92	70.38	69.45
CBAM-ConvNeXt (40)	49.32	50.25	51.22	49.32

A.2.2 Varying size of leaves

The current version of the dataset reduces the size of the individual leaves to a similar size to the natural foliage leaves. This step gave a huge performance rise of the evaluated model. Many variations in the size of leaves, maintaining their aspect ratio, were used to create the dataset and were evaluated. One of the experimental results is shown in Table 5. As vivid from the table, upscaling the leaf size degrades the classifier's performance; this is also true when the leaves are downscaled.

Table 5: Comparative performance evaluation results (in %) under the generated tomato foliage image dataset with 2 $\times$ upscaled leaf size

Models	Accuracy	F1-score	Precision	Recall
VGG19 (29)	65.23	65.41	65.59	65.23
ResNet50 (15)	76.73	77.61	78.51	76.73
DenseNet121 (17)	**83.78**	**83.14**	**82.38**	**83.92**
Swin Transformer (22)	67.95	68.96	70.02	67.91
CBAM-ConvNeXt (40)	71.88	71.96	72.03	71.88

A.2.3 Comparison of Model metrics

In the current version of the dataset, the size of individual leaves was reduced to closely match the natural size of foliage leaves. This adjustment led to a significant improvement in the performance of the evaluated model. To construct the dataset, multiple variations in leaf size were generated while preserving their aspect ratios, and these variations were systematically evaluated. One such experimental result is presented in Table 5. As evident from the table, both upscaling and downscaling the leaf size result in a degradation of the classifier's performance, whereas maintaining leaf sizes closer to their natural scale yields the best outcomes.

Table 6: Comparative model metrics evaluation results under the generated dataset and the single-leaf image datasets of Soybean

For generated dataset (γ = 15%)			For single leaf images	
Models	**# of Parameters**	**Per epoch training time**	**of Parameters**	**Per epoch training time**
VGG19 (29)	139,611,210	459.17	139,611,210	331.03
ResNet50 (15)	23,581,642	433.38	23,581,642	244.10
DenseNet121 (17)	7,047,754	454.73	7,047,754	645.819
Swin Transformer (22)	27,527,044	419.10	27,527,044	290.10
CBAM-ConvNeXt (22)	29,727,934	437.34	29,727,934	331.03