# Embracing limited and imperfect training datasets: opportunities and challenges in plant disease recognition using deep learning

Mingle Xu[1†], Hyongsuk Kim[1†], Jucheng Yang[2*], Alvaro Fuentes[1], Yao Meng[1], Sook Yoon[3*], Taehyun Kim[4] and Dong Sun Park[1]

[1]Department of Electronic Engineering, Core Research Institute of Intelligent Robots, Jeonbuk National University, Jeonju, Republic of Korea, [2]College of Artificial Intelligence, Tianjin University of Science and Technology, Tianjin, China, [3]Department of Computer Engineering, Mokpo National University, Muan, Republic of Korea, [4]National Institute of Agricultural Sciences, Wanju, Republic of Korea

Recent advancements in deep learning have brought significant improvements to plant disease recognition. However, achieving satisfactory performance often requires high-quality training datasets, which are challenging and expensive to collect. Consequently, the practical application of current deep learning–based methods in real-world scenarios is hindered by the scarcity of high-quality datasets. In this paper, we argue that embracing poor datasets is viable and aims to explicitly define the challenges associated with using these datasets. To delve into this topic, we analyze the characteristics of high-quality datasets, namely, large-scale images and desired annotation, and contrast them with the *limited* and *imperfect* nature of poor datasets. Challenges arise when the training datasets deviate from these characteristics. To provide a comprehensive understanding, we propose a novel and informative taxonomy that categorizes these challenges. Furthermore, we offer a brief overview of existing studies and approaches that address these challenges. We point out that our paper sheds light on the importance of embracing poor datasets, enhances the understanding of the associated challenges, and contributes to the ambitious objective of deploying deep learning in real-world applications. To facilitate the progress, we finally describe several outstanding questions and point out potential future directions. Although our primary focus is on plant disease recognition, we emphasize that the principles of embracing and analyzing poor datasets are applicable to a wider range of domains, including agriculture. Our project is public available at https://github.com/xml94/EmbracingLimitedImperfectTrainingDatasets.

KEYWORDS

plant disease recognition, AI in Agriculture, deep learning in agriculture, smart agriculture, precision agriculture

# 1 Introduction

Plant diseases are responsible for significant yield losses (Savary et al., 2019), making their recognition a crucial task in crop cultivation. In the past decade, deep learning, characterized by two essential attributes inherited from classical machine learning methods (Kawasaki et al., 2015; Mohanty et al., 2016; Fuentes et al., 2017), has emerged as a promising approach for this purpose. First, deep learning possesses the remarkable ability to serve as a feature extractor (Singh et al., 2018; Bengio et al., 2021). This stands in contrast to traditional machine learning, which often necessitates human experts to manually design features, such as histograms of oriented gradients for Red-Green-Blue (RGB) images (Fan et al., 2022) and vegetation indices for hyperspectral and multispectral images (Abdulridha et al., 2020). However, designing effective features has proven challenging and often requires diversity for different tasks. Second, deep learning–based methods have demonstrated "decent performance" in numerous studies on plant disease recognition (Singh et al., 2018; Boulent et al., 2019; Abade et al., 2021; Liu and Wang, 2021; Ouhami et al., 2021; Singh et al., 2021; Thakur et al., 2022). Furthermore, the implementation of deep learning on farms offers the enticing advantage of liberating human labor and significantly reducing associated costs. This is particularly valuable in the present century, as the global population is expected to continue increasing while the number of agricultural workers has been steadily declining.

Although deep learning has demonstrated its potential, the requirement for high-quality datasets to achieve satisfactory performance remains a challenge. Unfortunately, collecting such datasets is often prohibitively expensive and extremely challenging in many real-world applications (Xu et al., 2022a; Xu et al., 2023). Conversely, poor datasets are prevalent, and current models may struggle when confronted with them. Recognizing this reality, we contend that embracing poor datasets presents new opportunities to advance plant disease recognition in real-world applications. To further enrich the relevant understanding of this embrace, we analyze the characteristics of the desired high-quality datasets: large-scale and desired annotation. Specifically, large-scale datasets provide a vast quantity of information within the images, whereas desired annotation ensures that the images are annotated in accordance with specific criteria and objectives. More details are discussed in Section 2. In contrast, poor-quality datasets are defined by their deviations from the characteristics. Specifically, a dataset not on a large scale is categorized as limited, whereas a dataset lacking the desired annotation is considered imperfect. Embracing poor datasets, therefore, entails embracing limited and imperfect dataset, each of which is further explored and analyzed in Sections 4 and 5, respectively. The challenges associated with embracing limited and imperfect datasets are explicitly defined within these sections. A novel taxonomy detailing these challenges is conceptually described in Section 3, and Table 1 provides a glimpse of the taxonomy.

This study distinguishes itself from existing survey papers on plant disease recognition using deep learning by adopting a "challenge-oriented" approach instead of a "technique-oriented" one. Whereas previous works such as those by Singh et al. (2018); Boulent et al. (2019); Abade et al. (2021); Liu and Wang (2021); Ouhami et al. (2021); Singh et al. (2021), and Thakur et al. (2022) have focused on summarizing existing techniques and relevant materials, we have identified the key challenges associated within this field. Specifically, we highlight the scarcity of large-scale annotated data and advocate for embracing the concept of limited and imperfect datasets when deploying deep learning in real-world applications.

To conclude, in pursuit of deploying deep learning for plant disease recognition in real-world applications with satisfactory performance, we offer a perspective that embraces limited and imperfect datasets, contrasting with high-quality data. Our main contributions are as follows.

- We *explicitly* argue embracing limited and imperfect datasets for plant disease recognition using deep learning,

TABLE 1  Taxonomy of challenges arising when embracing limited and imperfect datasets.

| Challenge | | | Definition |
|---|---|---|---|
| Limited dataset | Class-level | Few-shot | All classes have similar few annotated images, where trained models may get low performance for all classes. |
| | | Class imbalance | One class has many more annotated images than another class, where trained models may get high performance in the former class but suffer in the latter class. |
| | Dataset-level | Domain shift | The training and test datasets share the same label spaces but are in different distribution spaces, where trained models may get low test performance. |
| | | Unknown class | Unknown (new) classes exist in the test dataset, where trained models will consider the corresponding image into a known class and not distinguish the unknown from known classes. |
| Imperfect dataset | Incomplete annotation | | Training datasets have labeled and unlabeled images simultaneously, where utilizing the unlabeled images may contribute to the test performance. |
| | Inexact annotation | | Training datasets are given with only coarse-grained annotations, where utilizing these annotations is challenging to train models. |
| | Inaccurate annotation | | Some annotations may be inaccurate, where it is challenging to get decent test performance by utilizing these annotations to train models. |

A limited dataset suggests that the training dataset is not on a large scale, whereas an imperfect dataset implies that the annotations are not expected. In addition, the challenges of limited dataset encompass class-level challenges, which involve image variations among different classes within the training dataset, and dataset-level challenges, which pertain to the information gap between the training and test datasets.

motivated by the reality that collecting high-quality datasets is expensive and challenging.

- We analyze the underlying reasons behind the current necessity for high-quality datasets in Section 2.
- We present a *taxonomy* of challenges associated with the embrace in Section 3, with *formal definitions*. A concise overview of existing studies that tackle them is also given as discussed in Sections 4 and 5.
- We provide noteworthy questions and highlight potential directions for further exploration in Section 6.

# 2 Why is high-quality dataset desired?

In general, to achieve promising performance using deep learning models, the training datasets should have two characteristics: large-scale and annotated with desired strategies. This section aims to probe the underlying reasons behind the desired characteristics.

## 2.1 Large-scale dataset

### 2.1.1 Deep learning models require large-scale data

For two reasons, deep learning generally requires a large-scale training dataset to obtain a comparable test performance. First, there are enormous learnable parameters in deep learning–based models that require large-scale data (Krizhevsky et al., 2017; Sarker, 2021). This ensures that a better feature extractor could be learned; otherwise, the training data points could be remembered, resulting in a poor test performance (Xu et al., 2023). Second, the distribution of the training dataset is gradually approaching that of the test dataset when the training dataset becomes larger, supporting a better test performance. For example, a model trained with images captured in laboratories is not expected to be effective when tested with images captured on farms (Guth et al., 2023; Wu et al., 2023).

### 2.1.2 Huge image variation requires large-scale datasets

The requirement for a large-scale training dataset comes from not only deep learning but also the task of plant disease recognition, called image variation (Xu, 2023; Xu et al., 2023). Considering the types of plant diseases, it can be divided into intra-class, differences within the same plant disease, and inter-class, heterogeneity between the two plant diseases. The intra-class image variation, partially illustrated in Figure 1, originates from three main elements. The first one is "plant itself". For example, some plants may have different types, such as different types of tomatoes, with diverse leaf shapes and sizes, as shown in Figure 1A. The same type of plant and plant disease may also occur at various stages with individual visual patterns. For example, Figures 1B–D have shown the different stages of plant disease, flowers, and leaves. Second, the plants may be grown in different "environments", such as fields and greenhouses. Heterogeneous illuminations and backgrounds in the fields are shown in Figures 1E, F, respectively. Third, "imaging

processing" is another source of intra-class image variation. Arguably, optical sensors and platforms result in greater diversity, such as RGB and thermal images, phones, and satellites (Oerke et al., 2014; Mahlein, 2016). When the optical sensors and platforms are fixed, the distance between the plants and sensors results in a multiscale challenge, as shown in Figure 1G. In addition, these viewpoints also lead to variations, as shown in Figure 1H.

The inter-class variation assumes that one type of plant disease is visually different from another, such as tomato leaf mold and canker, as shown in Figures 2A, C. We emphasize that this assumption should be considered when formulating the application objectives. For example, early symptoms in RGB images of plant diseases really resemble healthy ones, and, consequently, finding related plant diseases at very early stages may not be reliable. In addition, inter-class image variation can be viewed as a relative challenge. Specifically, the visual differences between the two plant diseases could be larger than those of the counterparts of another pair. As shown in Figure 2, the visual deviations between tomato leaf mold and canker are larger than those between tomato magnesium deficiency and chlorosis virus. A strategy used for this scenario is to collect more data for the close pairs, in which the models required more evidence to make decisions.

## 2.2 Desired annotation strategy

Deep learning is first trained in a "training dataset" and then tested in a "test dataset". A "validation dataset" is usually utilized to select the best-trained model from different hyperparameters and other training settings. The training and validation datasets are accessible at both the training and test stages, whereas the test dataset is only accessible at the test stage. Furthermore, the training and validation datasets for most deep learning–based models are hypothesized to be annotated in the desired manner. A desired annotation strategy, called EEP, has three primary points: exclusion, extensiveness, and precision. This has suggested that every annotation included only one specific visual pattern of plant disease, whereas extensive indicated that every plant disease in the images should have been annotated. The last one requires a precise annotation. For example, in image classification, every image should have included a type of plant disease and be linked to a label, as illustrated in the first row of Figure 3. By contrast, as shown in the second row of Figure 1, object detection generalizes the idea that one image can cover multiple plant diseases. However, every region should be annotated with labels and locations (a bounding box with four values: two for the left point in the horizontal and vertical directions, and two for width and height).

Accordingly, segmentation is a task for recognizing plant disease, in which every pixel should be assigned a label, as suggested in the last row of Figure 3.

# 3 Challenge formulation with limited and imperfect datasets

As previously discussed, achieving satisfactory performance using deep learning often requires training datasets that are both annotated as desired and of large scale (Lu et al., 2022). However,
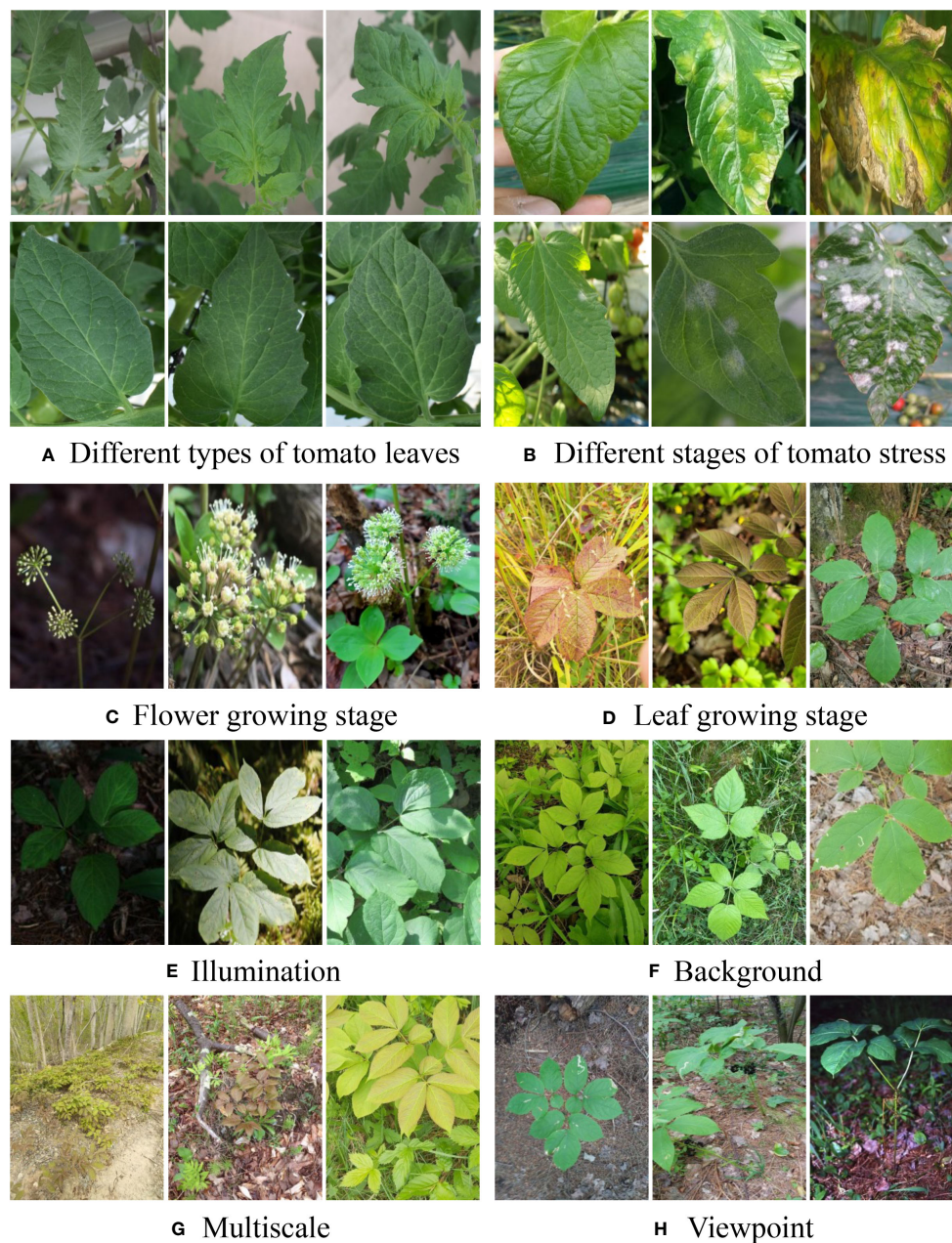
**FIGURE 1**

Illustration of the intra-class image variation. The images in **(A, B)** are tomato leaves. The images in **(C−H)** stem from a species, Aralia nudicaulis, in the PlantCLEF2022 dataset (Goëau et al., 2022). Every group suggests that the pictures belonging to the same plant disease may have visual diversities.

collecting such datasets is frequently challenging, time-consuming, and costly. Consequently, existing models may encounter limitations when applied to real-world scenarios without access to high-quality datasets. Therefore, a more convincing objective is to secure the satisfactory performance of models using limited and imperfect training datasets. However, this concept remains relatively unexplored within the context of plant disease recognition. The present study aims to shed light on this direction, with the ultimate goal of monitoring plant growth, thereby reducing human intervention and potentially mitigating the issues arising from plant diseases.

To comprehend the challenges that arise when dealing with limited and imperfect training datasets, we propose a novel taxonomy. Specifically, the term "limited dataset" refers to scenarios where the training dataset is not on a large scale, whereas "imperfect dataset" describes situations where the annotations of the training dataset deviate from the expected and desired. The limited dataset can be further divided into two subcategories: class-level, which examines deviations among different classes within the training datasets, and dataset-level, which analyzes the heterogeneity between the training and test datasets. On the other hand, imperfect dataset can be classified into
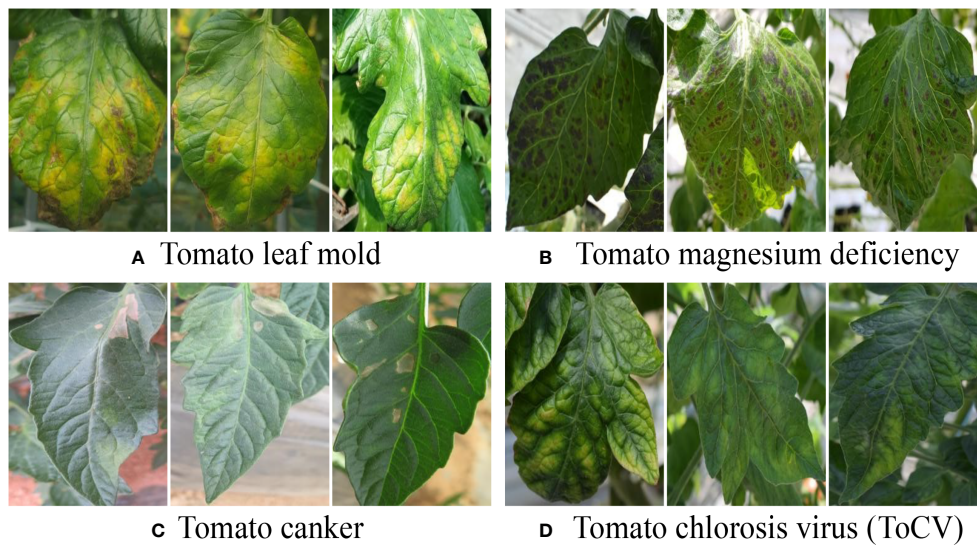
**FIGURE 2**
Illustration of inter-class image variation. It can be cast to a relative challenge where the visual deviations between **(A, C)** are larger than that between **(B, D)**. A corresponding strategy is to collect more data for **(B, D)** in that models need more evidence to make decisions for hard scenarios.
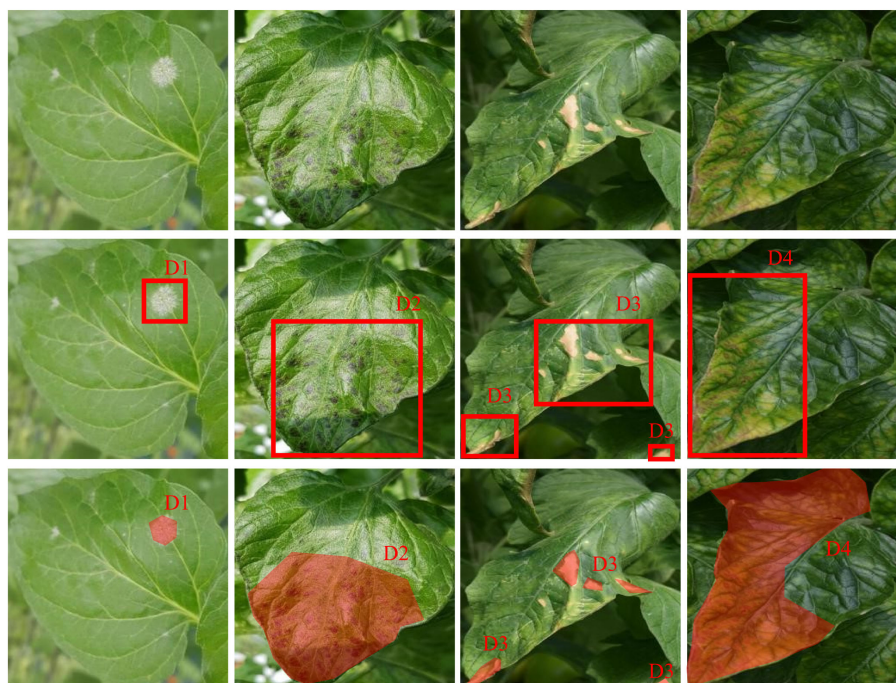


**FIGURE 3**
Annotation strategies in three primary tasks of plant disease recognition. From the first to the last rows are image classification, object detection, and segmentation, respectively. $Dm$ ($m$ = 1, 2, 3, 4) denote the types of plant disease, and every column suggests different plant diseases. In the simplest way, image classification refers to assigning a class to one image, whereas, in object detection, classes and their locations (bounding box) are entailed to predict. Segmentation requires class prediction at a pixel level. From the first to the last row, the annotation becomes more complicated and thus more time-consuming. The images in real-world applications tend to be more complex than these examples, such as multiple diseases existing in one leaf and one image including multiple leaves. Furthermore, a desired annotation strategy embraces three primary points: exclusive, extensive, and precise. The exclusive suggests that every annotation just includes one specific visual plant disease pattern, whereas the extensive denotes that every plant disease in the images should be annotated. The precise requires that the images should be annotated precisely. Violating the three points leads to the challenges of imperfect annotation.

three distinct types based on the nature of conflict: incomplete annotation, where a portion of images lacks annotations; inexact annotation, where some classes are annotated in a coarse-grained manner; and inaccurate annotation, where certain images are annotated with inaccuracies or even incorrect labels. Table 1 offers a glimpse into this comprehensive taxonomy.

# 4 Limited dataset

To make the following content easy to understand, several notations are first described. $D_X$ and $D_Y$ denote the training and test datasets, $X$ and $Y$ denoting the training and test domains. In general, the two datasets encompass $n$ plant disease classes: $c^1, c^2,…, c^N$. Let $n_{Xi}$ ($i = 1,2,…,N$) and $n_{Yj}$ ($j = 1,2,…,N$) denote the numbers of annotated images for class $c^i$ in $D_X$ and class $c^j$ in $D_Y$, respectively.

## 4.1 Class-level limited dataset

The class-level limited dataset, confined to the training dataset, is a case in which the number of annotated images for a class is small. Considering the differences across the different classes, they are divided into few-shot and class imbalances.

### 4.1.1 Few-shot

The few-shot challenge assumes that collecting and annotating images are expensive for *every* plant disease with the same number of annotated images. Formally, this challenge is *strictly* defined as $n_{Xi} = M$, where $M \in R^+$ is a small natural number. In general, $M$ may be equal to 5 or 10, which suggests that every plant disease has only 5 or 10 annotated images. Moreover, the few-shot challenge could be *generalized* as

$$n_{X_i} \approx M \qquad (1)$$

where every class contain *approximately M* annotated images. The essential issue is that a few annotated images could not provide sufficient evidence to train a deep learning–based model; thus, the trained model could not be generalized in the test dataset for every class. On the basis of this, we further extend the few-shot challenge from a small number of annotated images to a larger case, such as 100 and even 500, where *most deep learning–based models could not obtain good test performance for every plant disease*. This motivation is based on the observation that plant disease may have huge intra-class image variation and relatively low inter-class image variation.

To address this few-shot challenge, image manipulation, a set of traditional image processing methods, such as translation and flipping, is one of the simplest methods. It is hypothesized to retain class information and mimic image variations. For example, image translation changes the positions of objects in an image. This method is utilized to increase the number of training images from 350 to 39,010 for six plant diseases and healthy leaves (Gorad and Kotrappa, 2021). In particular, a new background is fused to the object of plant disease to create diverse backgrounds in the field rather than in the laboratory (Gui et al., 2021). Owing to its simple deployment, image manipulation is leveraged by default

with many other advanced methods in the general few-shot cases (Mohanty et al., 2016; Xu et al., 2022c). In addition, the image-generating models provide opportunities. Conditional generative adversarial networks (Mirza and Osindero, 2014) can generate new images, and, given a label, the generated images are assumed to be similar to the original images, but not the same (Abbas et al., 2021). Intuitively, image-generating models aim to learn image variations in the original training dataset and then create new image variations. However, learning an image-generating model requires data, which is often not satisfactory in the few-shot challenge.

Another effective and efficient method is transfer learning, which transfers knowledge for plant disease recognition from another task with a large-scale training dataset, assuming that learning one task with a large-scale training dataset is beneficial to plant disease recognition (Zhuang et al., 2020). In general, the plant disease recognition dataset is called the target dataset and the task is called the target task, whereas another one is termed as source dataset and task, respectively. With this strategy, the first issue is to choose a better model from many deep learning–based models (Mohanty et al., 2016; Kaya et al., 2019; Chen et al., 2020), such as ResNet (He et al., 2016) and DenseNet (Huang et al., 2017). The way to adopt a given pre-trained model is a question, such as freezing most of the models and training the remaining part (Sethy et al., 2020). In addition to models, choosing a better source dataset is another essential issue. A seminal work directly utilized a generic computer vision dataset, ImageNet (Deng et al., 2009). Although it has many variations, it is not that kindly related to plant disease recognition. Simultaneously, a source dataset related to the target dataset is more appealing (Neyshabur et al., 2020; Matsoukas et al., 2022). In this manner, plant-related datasets are considered, such as the AIChallenger2018 (Zhao et al., 2022) and PlantCLEF2017 (Kim et al., 2021) datasets. Furthermore, the image variations inside the source dataset should also be considered. For example, PlantCLEF2017 has more annotated images and higher intra-class variations than AIChallenge2018 where most images are collected in the laboratory. Considering the model and source dataset together is therefore encouraging. Embracing this idea, PlantCLEF2022 (Goëau et al., 2022; Xu et al., 2022b), a large-scale plant-relevant dataset with 2,885,052 annotated images for 80,000 classes, is leveraged with a ViT-based (Dosovitskiy et al., 2020) model rather than convolution neural network (CNN), to achieve versatile plant disease recognition (Xu et al., 2022c). With this strategy, a mean test accuracy of 86.29% over 12 datasets of plant disease recognition in a 20-shot case is achieved with a fast convergence speed, which is 12.76% higher than the current state-of-the-art accuracy of 73.53%.

Although transfer learning has been widely adopted, we argue that more opportunities still exist. First, transfer learning involves more segments beyond the models and source datasets, such as the loss function to pre-train and re-train the model. Considering these segments may provide motivation for improving the performance of plant disease recognition. We argue that the new methods in the general field of computer vision are useful for plant disease recognition. For example, a model could be trained using a source dataset with not only annotated images but also paired text to obtain improved semantic representations (Radford et al., 2021;

Wei et al., 2022). The utilization of transfer learning in practical applications is another issue. For example, current deep learning–based models have numerous parameters and thus should be compressed for embedding systems. In this case, transferring knowledge from a large to a small model is desirable (Abbasi Koohpayegani et al., 2020).

Other possible strategies, such as meta-learning (Huisman et al., 2021) and metric learning (Kaya and Bilge, 2019), have received attention over the last few years. Generic deep learning directly outputs the final results such as the type of plant disease and could be used to learn the relationship between samples and the corresponding ground truth. In contrast, meta-learning aims to learn how to learn, with which the output is rather parameters used to train a new task (Chen et al., 2021b; Li and Yang, 2021; Nuthalapati and Tunga, 2021). One of the primary issues is that current plant disease recognition datasets could not support an enormous number of source tasks, such as more than 110,000 tasks (Chen et al., 2021b). Furthermore, metric learning attempts to learn the differences between samples. In general, it pushes the samples closer within the same class and away from different classes (Afifi et al., 2020; Li and Yang, 2020; Egusquiza et al., 2022). Finally, the aforementioned strategies can be combined to further improve the performance, such as (Afifi et al., 2020) utilizing transfer learning and metric learning simultaneously.

## 4.1.2 Class imbalance

The few-shot challenge assumes that each type of plant disease occurs with a similar frequency, thus resulting in small image variations. However, some plant diseases occur at a higher frequency than others in the natural world. For instance, some plant diseases may appear more often than others, and, even for one specific type, different stages can be observed with diverse frequencies. In this case, one class may have a much higher number of annotated images than another class in the training dataset, termed a "class imbalance". Mathematically, the class imbalance challenge is formalized as $n_{Xi} \gg n_{Xj}$, where $\gg$ denotes much larger. A class with many more annotated images refers to the majority class; otherwise, it refers to the minority class (Xu et al., 2022a). The fundamental challenge is that the trained model tends to assign a high probability to the majority class during the test stage because it contributes more at the training stage (Xu et al., 2023). However, when the minority class also has many annotated images, the models may exhibit acceptable performance. Therefore, we propose a strict definition that is closer to the real applications:

$$\begin{cases} n_{X_i} \gg n_{X_j}, \\ n_{X_j} \leq M. \end{cases} \quad (2)$$

In the strict version, the number of annotated images of the minority class is not only much less than that of the majority class but should also be lower than a specific value. We argue that $M$ should not be fixed for all tasks. By contrast, this value depends on multiple factors, such as intra- and inter-class variations. Essentially, *deep learning–based models may not be able to learn robust features for the minority class in the class-imbalance datasets.* To mitigate this challenge, the primary idea is to increase the

performance of the minority class while maintaining that of the majority class.

Theoretically, the majority of strategies designed for few-shot could be utilized in that class imbalance becomes a few-shot challenge when the number of annotated images of the majority class reduces to a certain extent. This subsection introduces the methods aiming to specifically facilitate the minority classes. Compared with the majority class, the lower performance of the minority class is assumed to be due to the lower observation frequency of the models. This assumption inspires balancing the frequency for models by pushing the model to look at the images of the minority class more often (Nafi and Hsu, 2020). Similarly, models can also be punished more by using samples from the minority class (Nafi and Hsu, 2020; Oksuz et al., 2020). In addition, image augmentation aims to increase image variations to facilitate deep learning–based models. The methods belonging to image augmentation for the class imbalance differing from that of the few-shot is the basic motivation that the majority class can be leveraged for the minority. Conditional image-generating models implicitly utilize this insight by training a single model to learn from all classes (Mirza and Osindero, 2014; Abbas et al., 2021). By contrast, translating an image from one class to another directly utilizes the information among these classes (Cap et al., 2020; Nazki et al., 2020; Lu et al., 2022). To further consider the intra-class image variation from the majority class to the minority class, a specific loss is leveraged along with the image translation strategy (Xu et al., 2022a).

## 4.2 Dataset-level limited dataset

The limited dataset at the class level considers situations among the classes of the training dataset, whereas heterogeneity between the training and test datasets appears at the dataset level. It is further categorized into unknown classes and domain shifts. The former suggests that some classes in the test dataset, termed unknown classes, do not appear in the training dataset; whereas the latter emphasizes that the image variations in the test and training datasets are diverse. We emphasize that these two categories focus on specific essences and that their combination may exist at a higher frequency in real-world applications.

### 4.2.1 Unknown class

The class in the training dataset refers to the "known class", whereas a class existing in the test dataset but not in the training dataset is termed an "unknown class" (Geng et al., 2020). In the concept of plant disease, unknown classes may result in a large economic loss, and recognizing them is thus one of the fundamental demands. Simultaneously, collecting all the existing plant diseases is difficult and even impossible for real-world applications. Therefore, assuming the existence of unknown classes in the test dataset is encouraging. In this scenario, the task of plant disease recognition has two-fold; classifying known classes and rejecting unknown classes (Yang et al., 2021). This task refers to open set recognition or out-of-distribution and has witnessed significant developments in the field of computer vision (Geng et al., 2020; Salehi et al., 2021;

Yang et al., 2021). However, it has been rather underdeveloped for recognizing plant disease. In the following paragraphs, three key understandings from the computer vision field are first introduced, followed by a review of the literature on plant disease recognition.

First, thresholds are commonly employed to distinguish unknown classes, such as when an image larger than the threshold is categorized as known. In this case, two things are essential: a method to compute a value for a given image and a method to set a threshold. Currently, the probability (Liang et al., 2018), energy (Zhang et al., 2020), and reconstruction error (Sun et al., 2020) are the three main strategies for a given image. For example, known classes are assumed to have higher probabilities, lower energies, and smaller reconstruction errors than unknown classes. Simultaneously, a fixed threshold tuned in the training dataset is widely employed, such as the accuracy to maintain 95% of the images in the training dataset as known (Huang et al., 2022). This fixed one can be deemed at the dataset level and the class-level threshold has been recently considered in that different known classes probably behave diversely (Wang et al., 2022a).

Second, learning a good classifier with known classes is an effective and efficient strategy, such as utilizing strong image augmentation and longer training times (Vaze et al., 2022). A good classifier requires models to learn a robust feature space to distinguish one known class from another. In general, a robust feature space is tight for a specific class and the distances between the two classes are sufficiently large, with which unknown classes have more possibilities to be recognized. However, known classes with occlusions and unknown classes with features similar to those of known classes trigger problems in this strategy (Dietterich and Guyer, 2022). Finally, the exposure of potential novel classes, not the unknown class in the test dataset, is a convincing strategy because the primary challenge is that models are trained with only known classes and extra information about unknown classes can provide extra information (Zhou et al., 2021; Dietterich and Guyer, 2022). With this paradigm, the potential unknown classes and efficiency of sampling images from unknown classes are essential (Chen et al., 2021a).

In plant disease recognition, an existing work aims to learn a good classifier via metric learning (You et al., 2022), with the inspiration that the distances between two images from the same unknown class should be smaller than those from different known classes. In general, metric learning pushes models to learn robust feature spaces and thus implicitly contributes to the recognition of the unknown class. In addition, an extra probability branch is explicitly utilized to distinguish between known and unknown classes along with a generic classification branch for known classes (Jiang et al., 2022). Simultaneously, images belonging to unknown classes are utilized to train the models, where exposure to unknown classes is beneficial, although unknown classes in the training stage may also appear in the test stage. The ratio between the number of known and unknown classes is formally analyzed (Fuentes et al., 2021a). The experimental results suggest that the performance deteriorates with more unknown classes mainly because of the shortage of useful information in the training dataset.

### 4.2.2 Domain shift

Domain shift is a common problem in deep learning where the training and test data come from different domains. In such cases, the trained model may perform poorly on the test data, resulting in a phenomenon known as a "domain shift". Domain shifts can occur for several reasons such as differences in data distribution, scale, and quality. The general assumptions are as follows:

$$P(c\,|\,X) \neq P(c\,|\,Y) \qquad (3)$$

where $P(c|X)$ represents the probability distribution of one plant disease $c$ given the input images in the training dataset $X$, and $P(c|Y)$ represents the probability distribution of $c$ given the input images in test dataset $Y$. The inequality sign indicates that the two distributions are not equal, implying that the domain shift can lead to a significant decrease in the performance of the test data, making the model ineffective. The unknown class in the test dataset is a special form of the domain shift challenge, but, in this section, we aim to highlight the domain shift where the set of classes in the test dataset is a proper subset of that in the training dataset.

In the case of plant disease data, symptoms do not have well-defined boundaries and gradually change from healthy normal conditions to diseased regions, making it difficult to create homogeneity in the data (Barbedo, 2018). In addition, the inter- and intra-class image variations, as well as the explicit variations given by the domain used for the data collection, add complexity to the model. A frequent performance drop occurs when a model is trained on a dataset from a particular scenario but is further tested on data from a different scenario. A common scenario, for example, is that the training dataset is collected in one place by one person and the test dataset is collected in another place with different infrastructures and illumination by another person with their individual habit of taking pictures, such as training in the images collected in the laboratory and testing in the real world (Guth et al., 2023; Wu et al., 2023).

To address the domain shift problem, researchers have developed several techniques, such as domain adaptation (Wang and Deng, 2018) and domain generalization (Wang et al., 2022b). Domain adaptation aims to adapt a model to the test domain by modifying the training data or the model itself, whereas domain generalization aims to train a model that can perform well in unseen domains. The goal of generalization is to design a model that can operate efficiently in the same environment or across multiple environments. There are several approaches to domain adaptation for plant disease recognition using deep learning. One approach is to use transfer learning, which involves fine-tuning a pre-trained model on a new dataset. This approach can be effective when the new dataset is similar to the original dataset but may not work as well when there are significant differences between the domains. Another approach is to use domain adaptation techniques such as adversarial training or domain adaptation networks. These techniques involve training a model to recognize features specific to the target domain while also minimizing the differences between the source and target domains. This approach can be effective when there are significant differences between domains but may require more computational resources and training data.

In the literature, this problem has barely been investigated in plant disease recognition; however, it is an important issue for developing a more generalized model. Early work in this area (Fuentes et al., 2021b) shows the benefits of using control classes, such as background and healthy leaves, to lead the learning process toward classes of interest. It exhibited improved performance as an easier-to-adapt model across environments. However, data from different backgrounds and environments are required to achieve this goal. This issue was further investigated by (Fuentes et al., 2021a), where a bounding box detector was trained to obtain the regions of interest. Then, in the second stage, a domain adaption model obtained the features of data from a source farm with known diseases and transferred them to a target farm in which unlabeled data were used to assess the generalization capabilities of the model to recognize regions belonging to known classes or otherwise assigned them as unknown. This scenario showed how implementation could improve the recognition of target diseases and precisely estimate novel information by associating them with an unknown class.

Another important assumption is to address domain shifts across crops and environments. Shibuya et al. (Shibuya et al., 2021) utilized more than 221,000 labeled leaf images from different regions and crops to investigate the performance bias of evaluation within the same farm and the effect of Region-Of-Interest (ROI) detection. They found that even with many training images, the diagnostic performance for images in fields different from the training images is greatly degraded owing to covariate shifts. From this study, two important questions arise: first, what is the importance of data taken in different environments than the training data for evaluation, and, second, how do the pre-detection of regions of interest, including symptoms of diseases, affect the performance? Another essential aspect to investigate in the domain shift is the changes that occur in data collected in the laboratory compared with data collected under field conditions. The generalization capabilities of CNNs are investigated to learn the clear patterns from lab conditions that are to be detected again under new and more complex field conditions while avoiding overfitting (Guth et al., 2023). The important insight derived from this is that, in order to create useful tools for disease detection and classification using deep learning for image analysis, it is crucial to develop a final product that can handle a wide range of images from various crop conditions and locations, including inter-class and intra-class variations (Wu et al., 2023). This requires carefully designed datasets with a large number of image samples that can accommodate the significant variability in crop conditions in different areas.

In summary, domain shift is a critical challenge in plant disease recognition using deep learning. Developing models that can adapt to different domains is essential for building robust and accurate systems that can be used across a wide range of crops and regions. Although there are several approaches to the domain shift, researchers must continue to develop new techniques and datasets for ensuring that these models are effective in real-world scenarios.

# 5 Imperfect dataset

The limited dataset challenge considers annotated images within either a class or dataset in which all images in the training dataset are assumed to be annotated, whereas the imperfect dataset challenge instead hypothesizes that the annotation in the training dataset can be missing and not perfect. The imperfect dataset challenge is categorized into incomplete, inexact, and inaccurate annotations based on the violation of the EEP annotation strategy. Figure 4 provides a quick impression of the changes in performance when the annotations are different from the desired ones.

## 5.1 Incomplete annotation

Incomplete annotation indicates that the training dataset includes annotated and unannotated images, primarily because of economic issues and the annotation requirements of expert knowledge in plant science. The number of annotated images is much lower than that of unannotated images. A straightforward method is to discard unannotated images and just use the annotated images to train the models. By contrast, the use of unannotated images has become an active topic in recent years. One strategy to use unlabeled images is self-supervised learning (SSL), which aims to learn better representations, followed by a fine-tuning process within the annotated images. In SSL, a pretext task should first be defined without using annotation to train a deep learning–based model (Jing and Tian, 2020). Currently, there are many types of pretext tasks, but only a few are utilized to recognize plant disease. In particular, image augmentation does not change one image's type of plant disease and is deemed as a pretext task (Nagasubramanian et al., 2022). Furthermore, advanced image augmentation methods, such as Mixup (Zhang et al., 2018) changing the labels linearly, can also be utilized as a pretext task (Monowar et al., 2022) using the connections before and after image augmentation.

Semi-supervised learning, which is another active topic, attempts to leverage unlabeled and labeled images simultaneously. One branch directly adopts SSL methods along with a supervised loss function such as softmax. Another branch generates pseudo-labels for unlabeled images, where the labeled images can be leveraged to annotate the unlabeled images by first training a classifier (Li and Chao, 2021). In contrast, clustering methods find the similarity without label information and link the images in the test dataset to those in the training dataset (Fang et al., 2021).

Furthermore, active learning aims to select informative images labeled by humans later instead of machines, hoping to annotate fewer images yet obtain a better performance (Ren et al., 2021). Therefore, selecting informative images effectively and efficiently (Nagasubramanian et al., 2021) is essential. Moreover, the involvement of human experts in the training loop is difficult and inconvenient in real-world applications.
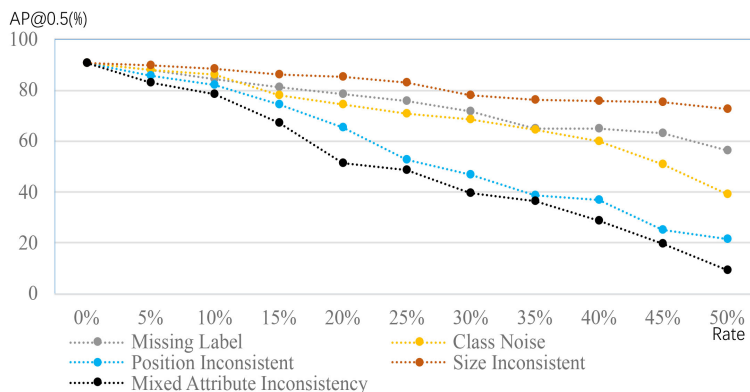
**FIGURE 4**
Performance comparison of object detection in a plant disease dataset, using different annotations; this figure adapted from Dong et al. (2022). The cases of missing labels and class noise suggest some patterns of plant disease have no labels and wrong labels. The inconsistencies of position and size suggest that the position and size are different from the desired. The mixed case is the combination of previous cases. Detection performance clearly degrades when the deviations from the desired ones are severe.

## 5.2 Inaccurate annotation

Inaccurate annotation, also called noisy annotation, denotes that some annotations in the training dataset may not be correct (Algan and Ulusoy, 2021; Dong et al., 2022). For example, plant disease is incorrectly annotated in the classification case, considering that experts may have conflicting decisions for a given image. Similarly, the bounding box used for object detection may be imprecise. Inaccurate annotation can be mitigated using multiple annotators, but this is expensive. Existing work (Dong et al., 2022) suggests that inaccurate annotation results in worse recognition performance, and different noise magnitudes have diverse impacts. Accordingly, facilitating the training process to reduce its impact is a straightforward approach (Li et al., 2019). Following this idea, new plant diseases are randomly generated for every image, and meta-learning is adopted to obtain consistent predictions (Zhai et al., 2022). In this manner, meta-learning aims to reduce the adverse impact of randomly generated labels. Although inaccurate annotations are facilitated, the corresponding images do not contribute to the trained models. Therefore, we highlight inaccurate annotations and employ relevant images by re-correcting the annotations (Liu et al., 2022; Wang et al., 2022c), although this idea has not yet been leveraged in plant disease recognition.

## 5.3 Inexact annotation

Inexact annotation refers to coarse-grained annotation, and the meaning varies for different tasks (Zhou, 2018; Zhang et al., 2021). For example, only image-level labels are accessible for object detection and segmentation, without bounding boxes and pixel-level classes, respectively. We emphasize that image classification also has a situation of inexact annotation, such as multiple diseases existing in one image but only one disease label. Simultaneously,

inexact annotations may appear along with precise annotations. A basic assumption in using inexact annotation is that a deep learning–based model may learn a significant area with coarse-grained annotations. Specifically, the activation value in every layer indicates that the pixels contribute to the final prediction diversely. Therefore, computing the most important pixels in an input image is one way of determining the exact annotation. This strategy has been employed for object detection of crop pests (Bollis et al., 2022) and segmentation of foliar diseases (Yi et al., 2021). However, this challenge currently receives less attention than incomplete and inaccurate annotations.

# 6 Concluding remarks and future perspectives

In this study, we advocated for embracing limited and imperfect training datasets for plant disease recognition using deep learning,
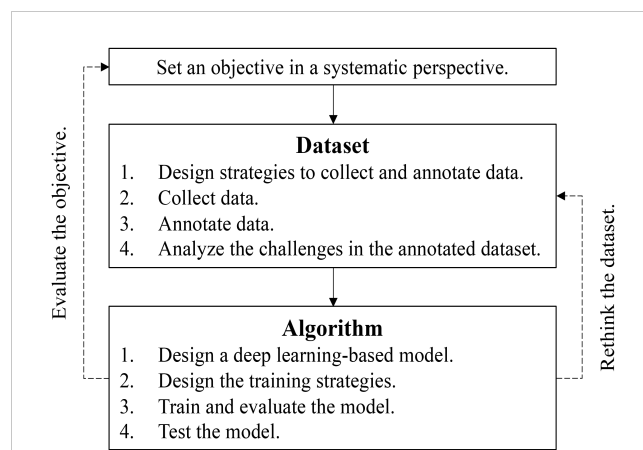


**FIGURE 5**
Flowchart to deploy deep learning in plant disease recognition. The evaluation of the project objectives and rethinking of the datasets are highlighted.

---

**BOX 1**  Outstanding questions.

- How can efficiently integrate plant science, including plant disease recognition, and artificial intelligence knowledge from collecting data to deploying a deep learning model?
- How to make a reliable dataset for the application-orientated challenges.
- Is there any other challenge to deploying deep learning in plant disease recognition, except the limited and imperfect dataset?
- What are the heterogeneities between plant disease recognition and generic computer vision tasks?
- How to design a preliminary automatic prototype to recognize plant disease as a real-world application?
- Considering the success of the large language models and foundation models, what can be done in plant disease recognition and the plant science field?

---

**BOX 2**  Future directions.

- Inspiration
  - Adopt the commonness between plant disease recognition and general computer vision tasks and then adapt the suitable concepts such as new problem formulations and methods.
  - Distinguish plant disease recognition from general computer vision tasks such as different plants having similar diseases and further leverage the difference.
- Dataset
  - Collect application-orientated datasets, such as for the domain shift.
  - Collect datasets from multiple sensors simultaneously, such as RGB and multispectral images.
  - Collect datasets from multiple observations such as spatial and temporary, inspired by the effectiveness of accumulated evidence.
  - Develop strategies to integrate datasets from the whole community.
- Model and algorithm
  - Develop strategies for specific challenges, such as for open set recognition.
  - Fine-tune large pre-trained models to achieve better performance in plant disease recognition tasks, and design strategies to achieve parameter-efficient fine-tuning (PEFT).
  - Employ small models for the embedding system.
  - Integrate large and small models to have decent performance yet for the embedding system.
- Analysis
  - Analyze the challenges in a dataset quantitatively.
  - Analyze the impacts of strategies to annotate datasets.
  - Analyze the relationship between performance, data quality and amount, computing resources, and model capacity.
- Application
  - Evaluate plant disease quantitatively, such as object detection and segmentation.
  - Deploy deep learning in real-world applications, such as robotic systems.
  - Design versatile plant disease recognition such as multi-plant and multi-dataset, rather than individual models for specific plants and datasets.
  - Consider plant disease recognition with other plant-related tasks, such as plant identification.

---

acknowledging the practical difficulties, expenses, and challenges associated with collecting high-quality datasets in real-world applications. Although this embrace is more convincing and practical, it also introduces new challenges. To enrich our understanding, we proposed a novel taxonomy of challenges with formal definitions. In addition, we provided a concise overview of strategies to address these challenges. One noteworthy finding is the limited research focused on dataset-level challenges related to limited datasets and imperfect annotation, whereas significant developments have been made concerning class-level limited datasets. Another essential point that we discovered is the severe shortage of benchmark datasets specifically tailored for real-world applications. By highlighting these insights, we aim to contribute to the advancement of deep learning techniques in real-world applications and foster progress in the domain of plant disease recognition. Although the primary focus of this study was on plant disease recognition, we emphasize that the concept of embracing limited and imperfect datasets is applicable to broader fields, such as deep learning in agriculture.

Building upon the challenges posed by limited and imperfect datasets, we propose a process tailored for real-world applications, depicted in Figure 5. Our objective is to emphasize the importance of evaluating and reevaluating the objectives and datasets. A fundamental assumption is that each class exhibits distinguishable visual patterns in the image space. However, certain classes may share remarkably similar patterns, making them challenging to distinguish. In such cases, the objectives of utilizing deep learning or the collected datasets should be polished, possibly by incorporating novel evidence. We also present several outstanding questions in Box 1 and outline potential future directions in Box 2, seeking to foster further research and advancements in the domain.

## Author contributions

MX: conceptualization, formal analysis, writing—original draft, and writing—review and editing. HK: project administration, funding acquisition, and writing—review and editing. JY: writing—review and editing. AF: writing—original draft, writing—review and editing. YM: writing—original draft and writing—review and editing. SY: supervision and writing—review and editing. TK: resources. DP:

supervision, project administration, funding acquisition, and writing—review and editing.

## Funding

## Acknowledgments

## Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

## Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

## Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2023.1225409/full#supplementary-material

## References

Abade, A., Ferreira, P. A., and de Barros Vidal, F. (2021). Plant diseases recognition on images using convolutional neural networks: A systematic review. *Comput. Electron. Agric.* 185, 106125. doi: 10.1016/j.compag.2021.106125

Abbas, A., Jain, S., Gour, M., and Vankudothu, S. (2021). Tomato plant disease detection using transfer learning with c-gan synthetic images. *Comput. Electron. Agric.* 187, 106279. doi: 10.1016/j.compag.2021.106279

Abbasi Koohpayegani, S., Tejankar, A., and Pirsiavash, H. (2020). Compress: Self-supervised learning by compressing representations. *Adv. Neural Inf. Process. Syst.* 33, 12980–12992.

Abdulridha, J., Ampatzidis, Y., Roberts, P., and Kakarla, S. C. (2020). Detecting powdery mildew disease in squash at different stages using uav-based hyperspectral imaging and artificial intelligence. *Biosyst. Eng.* 197, 135–148. doi: 10.1016/j.biosystemseng.2020.07.001

Afifi, A., Alhumam, A., and Abdelwahab, A. (2020). Convolutional neural network for automatic identification of plant diseases with limited data. *Plants* 10, 28. doi: 10.3390/plants10010028

Algan, G., and Ulusoy, I. (2021). Image classification with deep learning in the presence of noisy labels: A survey. *Knowledge-Based Syst.* 215, 106771. doi: 10.1016/j.knosys.2021.106771

Barbedo, J. G. (2018). Factors influencing the use of deep learning for plant disease recognition. *Biosyst. Eng.* 172, 84–91. doi: 10.1016/j.biosystemseng.2018.05.013

Bengio, Y., Lecun, Y., and Hinton, G. (2021). Deep learning for ai. *Commun. ACM* 64, 58–65. doi: 10.1145/3448250

Bollis, E., Maia, H., Pedrini, H., and Avila, S. (2022). Weakly supervised attention-based models using activation maps for citrus mite and insect pest classification. *Comput. Electron. Agric.* 195, 106839. doi: 10.1016/j.compag.2022.106839

Boulent, J., Foucher, S., Theau, J., and St-Charles, P.-L. (2019). Convolutional neural networks for the automatic identification of plant diseases. *Front. Plant Sci.* 10, 941. doi: 10.3389/fpls.2019.00941

Cap, Q. H., Uga, H., Kagiwada, S., and Iyatomi, H. (2020). Leafgan: An effective data augmentation method for practical plant disease diagnosis. *IEEE Trans. Automation Sci. Eng* 19 (2), 1258-1267. doi: 10.1109/TASE.2020.3041499

Chen, J., Chen, J., Zhang, D., Sun, Y., and Nanehkaran, Y. A. (2020). Using deep transfer learning for image-based plant disease identification. *Comput. Electron. Agric.* 173, 105393. doi: 10.1016/j.compag.2020.105393

Chen, L., Cui, X., and Li, W. (2021b). Meta-learning for few-shot plant disease detection. *Foods* 10, 2441. doi: 10.3390/foods10102441

Chen, G., Peng, P., Wang, X., and Tian, Y. (2021a). Adversarial reciprocal points learning for open set recognition. *IEEE Trans. Pattern Anal. Mach. Intell.* 44, 8065–8081. doi: 10.1109/TPAMI.2021.3106743

Deng, J., Dong, W., Socher, R., Li, L.-J., Li, K., and Fei-Fei, L. (2009). "Imagenet: A large-scale hierarchical image database," in *2009 IEEE conference on computer vision and pattern recognition.* (Miami, FL, USA: IEEE), 248–255.

Dietterich, T. G., and Guyer, A. (2022). The familiarity hypothesis: Explaining the behavior of deep open set methods. *Pattern Recognition* 132, 108931. doi: 10.1016/j.patcog.2022.108931

Dong, J., Lee, J., Fuentes, A., Xu, M., Yoon, S., Lee, M. H., et al. (2022). Data-centric annotation analysis for plant disease detection: Strategy, consistency, and performance. *Front. Plant Sci.* 4937. doi: 10.3389/fpls.2022.1037655

Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., et al. (2020). An image is worth 16x16 words: Transformers for image recognition at scale. *arXiv preprint arXiv:2010.11929.* doi: 10.48550/arXiv.2010.11929

Egusquiza, I., Picon, A., Irusta, U., Bereciartua-Perez, A., Eggers, T., Klukas, C., et al. (2022). Analysis of few-shot techniques for fungal plant disease classification and evaluation of clustering capabilities over real datasets. *Front. Plant Sci.* 295. doi: 10.3389/fpls.2022.813237

Fan, X., Luo, P., Mu, Y., Zhou, R., Tjahjadi, T., and Ren, Y. (2022). Leaf image based plant disease identification using transfer learning and feature fusion. *Comput. Electron. Agric.* 196, 106892. doi: 10.1016/j.compag.2022.106892

Fang, U., Li, J., Lu, X., Gao, L., Ali, M., and Xiang, Y. (2021). Self-supervised cross-iterative clustering for unlabeled plant disease images. *Neurocomputing* 456, 36–48. doi: 10.1016/j.neucom.2021.05.066

Fuentes, A., Yoon, S., Kim, S. C., and Park, D. S. (2017). A robust deep-learning-based detector for real-time tomato plant diseases and pests recognition. *Sensors* 17, 2022. doi: 10.3390/s17092022

Fuentes, A., Yoon, S., Kim, T., and Park, D. S. (2021a). Open set self and across domain adaptation for tomato disease recognition with deep learning techniques. *Front. Plant Sci.* 12. doi: 10.3389/fpls.2021.758027

Fuentes, A., Yoon, S., Lee, M. H., and Park, D. S. (2021b). Improving accuracy of tomato plant disease diagnosis based on deep learning with explicit control of hidden classes. *Front. Plant Sci.* 12, 682230. doi: 10.3389/fpls.2021.682230

Geng, C., Huang, S.-j., and Chen, S. (2020). Recent advances in open set recognition: A survey. *IEEE Trans. Pattern Anal. Mach. Intell.* 43, 3614–3631. doi: 10.1109/TPAMI.2020.2981604

Goëau, H., Bonnet, P., and Joly, A. (2022). "Overview of plantclef 2022: Image-based plant identification at global scale," in *CLEF 2022-Conference and Labs of the Evaluation Forum*. (Bologna, Italy: Springer) 3180, 1916–1928.

Gorad, B., and Kotrappa, D. S. (2021). "Novel dataset generation for Indian brinjal plant using image data augmentation," in *IOP Conference Series: Materials Science and Engineering*. (IOP Publishing) 1065, 012041. doi: 10.1088/1757-899X/1065/1/012041

Gui, P., Dang, W., Zhu, F., and Zhao, Q. (2021). Towards automatic field plant disease recognition. *Comput. Electron. Agric.* 191, 106523. doi: 10.1016/j.compag.2021.106523

Guth, F. A., Ward, S., and McDonnell, K. (2023). From lab to field: An empirical study on the generalization of convolutional neural networks towards crop disease detection. *Eur. J. Eng. Technol. Res.* 8, 33–40. doi: 10.24018/ejeng.2023.8.2.2773

He, K., Zhang, X., Ren, S., and Sun, J. (2016). "Deep residual learning for image recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*. (Las Vegas, Nevada: IEEE), 770–778.

Huang, G., Liu, Z., van der Maaten, L., and Weinberger, K. Q. (2017). "Densely connected convolutional networks," in *Proceedings of the IEEE conference on computer vision and pattern recognition*. (Honolulu, Hawaii: IEEE), 4700–4708.

Huang, H., Wang, Y., Hu, Q., and Cheng, M.-M. (2022). Class-specific semantic reconstruction for open set recognition. *IEEE Trans. Pattern Anal. Mach. Intell* 45 (4), 4214-4228. doi: 10.1109/TPAMI.2022.3200384

Huisman, M., Van Rijn, J. N., and Plaat, A. (2021). A survey of deep meta-learning. *Artif. Intell. Rev.* 54, 4483–4541. doi: 10.1007/s10462-021-10004-4

Jiang, K., You, J., Dorj, U.-O., Kim, H., and Lee, J. (2022). Detection of unknown strawberry diseases based on openmatch and two-head network for continual learning. *Front. Plant Sci.* 13. doi: 10.3389/fpls.2022.989086

Jing, L., and Tian, Y. (2020). Self-supervised visual feature learning with deep neural networks: A survey. *IEEE Trans. Pattern Anal. Mach. Intell.* 43, 4037–4058. doi: 10.1109/TPAMI.2020.2992393

Kawasaki, Y., Uga, H., Kagiwada, S., and Iyatomi, H. (2015). "Basic study of automated diagnosis of viral plant diseases using convolutional neural networks," in Advances in Visual Computing: 11th International Symposium, ISVC 2015. (Las Vegas, NV, USA: Springer International Publishing). 638–645.

Kaya, M., and Bilge, H. S. (2019). Deep metric learning: A survey. *Symmetry* 11, 1066. doi: 10.3390/sym11091066

Kaya, A., Keceli, A. S., Catal, C., Yalic, H. Y., Temucin, H., and Tekinerdogan, B. (2019). Analysis of transfer learning for deep neural network based plant classification models. *Comput. Electron. Agric.* 158, 20–29. doi: 10.1016/j.compag.2019.01.041

Kim, B., Han, Y.-K., Park, J.-H., and Lee, J. (2021). Improved vision-based detection of strawberry diseases using a deep neural network. *Front. Plant Sci.* 11, 559172. doi: 10.3389/fpls.2020.559172

Krizhevsky, A., Sutskever, I., and Hinton, G. E. (2017). Imagenet classification with deep convolutional neural networks. *Commun. ACM* 60, 84–90. doi: 10.1145/3065386

Li, Y., and Chao, X. (2021). Semi-supervised few-shot learning approach for plant diseases recognition. *Plant Methods* 17, 1–10. doi: 10.1186/s13007-021-00770-1

Li, J., Wong, Y., Zhao, Q., and Kankanhalli, M. S. (2019). "Learning to learn from noisy labeled data," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. (Long Beach, CA, USA: IEEE), 5051–5059.

Li, Y., and Yang, J. (2020). Few-shot cotton pest recognition and terminal realization. *Comput. Electron. Agric.* 169, 105240. doi: 10.1016/j.compag.2020.105240

Li, Y., and Yang, J. (2021). Meta-learning baselines and database for few-shot classification in agriculture. *Comput. Electron. Agric.* 182, 106055. doi: 10.1016/j.compag.2021.106055

Liang, S., Li, Y., and Srikant, R. (2018). "Enhancing the reliability of out-of-distribution image detection in neural networks," in *6th International Conference on Learning Representations*. (Vancouver, BC, Canada: ICLR).

Liu, J., and Wang, X. (2021). Plant diseases and pests detection based on deep learning: a review. *Plant Methods* 17, 1–18. doi: 10.1186/s13007-021-00722-9

Liu, C., Wang, K., Lu, H., Cao, Z., and Zhang, Z. (2022). "Robust object detection with inaccurate bounding boxes," in *Computer Vision–ECCV 2022: 17th European Conference, Tel Aviv, Israel, October 23–27, 2022, Proceedings, Part X*. 53–69.

Lu, Y., Chen, D., Olaniyi, E., and Huang, Y. (2022). Generative adversarial networks (gans) for image augmentation in agriculture: A systematic review. *Comput. Electron. Agric.* 200, 107208. doi: 10.1016/j.compag.2022.107208

Mahlein, A.-K. (2016). Plant disease detection by imaging sensors–parallels and specific demands for precision agriculture and plant phenotyping. *Plant Dis.* 100, 241–251. doi: 10.1094/PDIS-03-15-0340-FE

Matsoukas, C., Haslum, J. F., Sorkhei, M., Soderberg, M., and Smith, K. (2022). "What makes transfer learning work for medical images: feature reuse & other factors," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. (Louisiana: IEEE), 9225–9234.

Mirza, M., and Osindero, S. (2014). Conditional generative adversarial nets. *arXiv preprint arXiv:1411.1784*.

Mohanty, S. P., Hughes, D. P., and Salathe, M. (2016). Using deep learning for image-based plant disease detection. *Front. Plant Sci.* 7, 1419. doi: 10.3389/fpls.2016.01419

Monowar, M. M., Hamid, M. A., Kateb, F. A., Ohi, A. Q., and Mridha, M. (2022). Self-supervised clustering for leaf disease identification. *Agriculture* 12, 814. doi: 10.3390/agriculture12060814

Nafi, N. M., and Hsu, W. H. (2020). "Addressing class imbalance in image-based plant disease detection: Deep generative vs. sampling-based approaches," in *2020 International Conference on Systems, Signals and Image Processing (IWSSIP)*. (Niteroi, Brazil: IEEE), 243–248.

Nagasubramanian, K., Jubery, T., Fotouhi Ardakani, F., Mirnezami, S. V., Singh, A. K., Singh, A., et al. (2021). How useful is active learning for image-based plant phenotyping? *Plant Phenome J.* 4, e20020.

Nagasubramanian, K., Singh, A., Singh, A., Sarkar, S., and Ganapathysubramanian, B. (2022). Plant phenotyping with limited annotation: Doing more with less. *Plant Phenome J.* 5, e20051. doi: 10.1002/ppj2.20051

Nazki, H., Yoon, S., Fuentes, A., and Park, D. S. (2020). Unsupervised image translation using adversarial networks for improved plant disease recognition. *Comput. Electron. Agric.* 168, 105117. doi: 10.1016/j.compag.2019.105117

Neyshabur, B., Sedghi, H., and Zhang, C. (2020). What is being transferred in transfer learning? *Adv. Neural Inf. Process. Syst.* 33, 512–523.

Nuthalapati, S. V., and Tunga, A. (2021). "Multi-domain few-shot learning and dataset for agricultural applications," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition 2021* (Nashville, Tennessee: IEEE). 1399–1408.

Oerke, E.-C., Mahlein, A.-K., and Steiner, U. (2014). "Proximal sensing of plant diseases," in *Detection and diagnostics of plant pathogens* (Springer), 55–68.

Oksuz, K., Cam, B. C., Kalkan, S., and Akbas, E. (2020). Imbalance problems in object detection: A review. *IEEE Trans. Pattern Anal. Mach. Intell.* 43, 3388–3415. doi: 10.1109/TPAMI.2020.2981890

Ouhami, M., Hafiane, A., Es-Saady, Y., El Hajji, M., and Canals, R. (2021). Computer vision, iot and data fusion for crop disease detection using machine learning: A survey and ongoing research. *Remote Sens.* 13, 2486. doi: 10.3390/rs13132486

Radford, A., Kim, J. W., Hallacy, C., Ramesh, A., Goh, G., Agarwal, S., et al. (2021). "Learning transferable visual models from natural language supervision," in *International conference on machine learning* (PMLR), 8748–8763.

Ren, P., Xiao, Y., Chang, X., Huang, P.-Y., Li, Z., Gupta, B. B., et al. (2021). A survey of deep active learning. *ACM computing surveys (CSUR)* 54, 1–40.

Salehi, M., Mirzaei, H., Hendrycks, D., Li, Y., Rohban, M. H., and Sabokrou, M. (2021). A unified survey on anomaly, novelty, open-set, and out-of-distribution detection: Solutions and future challenges. *arXiv preprint arXiv:2110.14051*. doi: 10.48550/arXiv.2110.14051

Sarker, I. H. (2021). Deep learning: a comprehensive overview on techniques, taxonomy, applications and research directions. *SN Comput. Sci.* 2, 1–20. doi: 10.1007/s42979-021-00815-1

Savary, S., Willocquet, L., Pethybridge, S. J., Esker, P., McRoberts, N., and Nelson, A. (2019). The global burden of pathogens and pests on major food crops. *Nat. Ecol. Evol.* 3, 430–439. doi: 10.1038/s41559-018-0793-y

Sethy, P. K., Barpanda, N. K., Rath, A. K., and Behera, S. K. (2020). Deep feature based rice leaf disease identification using support vector machine. *Comput. Electron. Agric.* 175, 105527. doi: 10.1016/j.compag.2020.105527

Shibuya, S., Cap, Q. H., Nagasawa, S., Kagiwada, S., Uga, H., and Iyatomi, H. (2021). "Validation of prerequisites for correct performance evaluation of image-based plant disease diagnosis using reliable 221k images collected from actual fields," in *AI for Agriculture and Food Systems* (Arlington, Virginia: AAAI).

Singh, A. K., Ganapathysubramanian, B., Sarkar, S., and Singh, A. (2018). Deep learning for plant stress phenotyping: trends and future perspectives. *Trends Plant Sci.* 23, 883–898. doi: 10.1016/j.tplants.2018.07.004

Singh, A., Jones, S., Ganapathysubramanian, B., Sarkar, S., Mueller, D., Sandhu, K., et al. (2021). Challenges and opportunities in machine-augmented plant stress phenotyping. *Trends Plant Sci.* 26, 53–69. doi: 10.1016/j.tplants.2020.07.010

Sun, X., Yang, Z., Zhang, C., Ling, K.-V., and Peng, G. (2020). "Conditional gaussian distribution learning for open set recognition," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition 2020*. (Seattle, Washington: IEEE), 13480–13489.

Thakur, P. S., Khanna, P., Sheorey, T., and Ojha, A. (2022). Trends in vision-based machine learning techniques for plant disease identification: A systematic review. *Expert Syst. Appl.*, 118117. doi: 10.1016/j.eswa.2022.118117

Vaze, S., Han, K., Vedaldi, A., and Zisserman, A. (2022). "Open-set recognition: A good closed-set classifier is all you need," in *International Conference on Learning Representations*. (ICLR).

Wang, M., and Deng, W. (2018). Deep visual domain adaptation: A survey. *Neurocomputing* 312, 135–153. doi: 10.1016/j.neucom.2018.05.083

Wang, S., Gao, J., Li, B., and Hu, W. (2022c). Narrowing the gap: Improved detector training with noisy location annotations. *IEEE Trans. Image Process.* 31, 6369–6380. doi: 10.1109/TIP.2022.3211468

Wang, J., Lan, C., Liu, C., Ouyang, Y., Qin, T., Lu, W., et al. (2022b). *Generalizing to unseen domains: A survey on domain generalization* (IEEE Transactions on Knowledge and Data Engineering).

Wang, H., Li, Z., Feng, L., and Zhang, W. (2022a). "Vim: Out-of-distribution with virtual-logit matching," in *17th European Conference*. (Tel Aviv, Israel: Springer), 4921–4930.

Wei, L., Xie, L., Zhou, W., Li, H., and Tian, Q. (2022). "Mvp: Multimodality-guided visual pretraining," in *Computer Vision – ECCV 2022*. (Cham: Springer-Verlag), 337–353 (Springer-Verlag). doi: 10.1007/978-3-031-20056-420

Wu, X., Fan, X., Luo, P., Choudhury, S. D., Tjahjadi, T., and Hu, C. (2023). From laboratory to field: Unsupervised domain adaptation for plant disease recognition in the wild. *Plant Phenomics* 5, 0038. doi: 10.34133/plantphenomics.0038

Xu, M. (2023). Enhanced plant disease recognition with limited training dataset using image translation and two-step transfer learning (Jeonbuk National University). doi: 10.13140/RG.2.2.27298.91846

Xu, M., Yoon, S., Fuentes, A., and Park, D. S. (2023). A comprehensive survey of image augmentation techniques for deep learning. *Pattern Recognition* 137, 109347. doi: 10.1016/j.patcog.2023.109347

Xu, M., Yoon, S., Fuentes, A., Yang, J., and Park, D. S. (2022a). Style-consistent image translation: a novel data augmentation paradigm to improve plant disease recognition. *Front. Plant Sci.* (Bolonga, Italy: Springer), 12, 3361. doi: 10.3389/fpls.2021.773142

Xu, M., Yoon, S., Jeong, Y., Lee, J., and Park, D. S. (2022b). "Transfer learning with self-supervised vision transformer for large-scale plant identification," in *International conference of the cross-language evaluation forum for European languages*. 2253–2261 (Springer).

Xu, M., Yoon, S., Jeong, Y., and Park, D. S. (2022c). Transfer learning for versatile plant disease recognition with limited data. *Front. Plant Sci.* 13, 4506. doi: 10.3389/fpls.2022.1010981

Yang, J., Zhou, K., Li, Y., and Liu, Z. (2021). Generalized out-of-distribution detection: A survey. *arXiv preprint arXiv:2110.11334*. doi: 10.48550/arXiv.2110.11334

Yi, R., Weng, Y., Yu, M., Lai, Y.-K., and Liu, Y.-J. (2021). Lesion region segmentation via weakly supervised learning. *Quantitative Biol* 10 (3), 239. doi: 10.15302/J-QB-021-0272

You, J., Jiang, K., and Lee, J. (2022). Deep metric learning-based strawberry disease detection with unknowns. *Front. Plant Sci.* 13. doi: 10.3389/fpls.2022.891785

Zhai, D., Shi, R., Jiang, J., and Liu, X. (2022). Rectified meta-learning from noisy labels for robust image-based plant disease classification. *ACM Trans. Multimedia Computing Communications Appl. (TOMM)* 18, 1–17.

Zhang, H., Cisse, M., Dauphin, Y. N., and Lopez-Paz, D. (2018). "mixup: Beyond empirical risk minimization," in *International Conference on Learning Representations*. (Convention center in Vancouver, Canada: ICLR).

Zhang, D., Han, J., Cheng, G., and Yang, M.-H. (2021). Weakly supervised object localization and detection: A survey. *IEEE Trans. Pattern Anal. Mach. Intell.* 44, 5866–5885.

Zhang, H., Li, A., Guo, J., and Guo, Y. (2020). "Hybrid models for open set recognition," in *Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part III 16*. 102–117 (Springer).

Zhao, X., Li, K., Li, Y., Ma, J., and Zhang, L. (2022). Identification method of vegetable diseases based on transfer learning and attention mechanism. *Comput. Electron. Agric.* 193, 106703. doi: 10.1016/j.compag.2022.106703

Zhou, Z.-H. (2018). A brief introduction to weakly supervised learning. *Natl. Sci. Rev.* 5, 44–53. doi: 10.1093/nsr/nwx106

Zhou, D.-W., Ye, H.-J., and Zhan, D.-C. (2021). "Learning placeholders for open-set recognition," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. (Nashville, Tennessee: IEEE), 4401–4410.

Zhuang, F., Qi, Z., Duan, K., Xi, D., Zhu, Y., Zhu, H., et al. (2020). A comprehensive survey on transfer learning. *Proc. IEEE* 109, 43–76. doi: 10.1109/JPROC.2020.3004555