# Few-shot Metric Domain Adaptation: Practical Learning Strategies for an Automated Plant Disease Diagnosis

**Shoma Kudo[1], Satoshi Kagiwada[2], Hitoshi Iyatomi[1]**

[1]Applied Informatics, Graduate School of Science and Engineering, Hosei University, Tokyo, Japan
[2]Clinical Plant Science, Faculty of Bioscience and Applied Chemistry, Hosei University, Tokyo, Japan
shoma.kudo.3n@stu.hosei.ac.jp, kagiwada@hosei.ac.jp, iyatomi@hosei.ac.jp

## Abstract

Numerous studies have explored image-based automated systems for plant disease diagnosis, demonstrating impressive diagnostic capabilities. However, recent large-scale analyses have revealed a critical limitation: that the diagnostic capability suffers significantly when validated on images captured in environments (domains) differing from those used during training. This shortfall stems from the inherently limited dataset size and the diverse manifestation of disease symptoms, combined with substantial variations in cultivation environments and imaging conditions, such as equipment and composition. These factors lead to insufficient variety in training data, ultimately constraining the system's robustness and generalization. To address these challenges, we propose Few-shot Metric Domain Adaptation (FMDA), a flexible and effective approach for enhancing diagnostic accuracy in practical systems, even when only limited target data is available. FMDA reduces domain discrepancies by introducing a constraint to the diagnostic model that minimizes the "distance" between feature spaces of source (training) data and target data with limited samples. FMDA is computationally efficient, requiring only basic feature distance calculations and backpropagation, and can be seamlessly integrated into any machine learning (ML) pipeline. In large-scale experiments, involving 223,015 leaf images across 20 fields and 3 crop species, FMDA achieved F1 score improvements of 11.1 to 29.3 points compared to cases without target data, using only 10 images per disease from the target domain. Moreover, FMDA consistently outperformed fine-tuning methods utilizing the same data, with an average improvement of 8.5 points.

## Introduction

Plant diseases cause significant damage to crops, posing a major challenge for farmers worldwide. According to the Food and Agriculture Organization (FAO), 20–40% of global food crops are lost annually to plant pests and diseases, resulting in economic losses estimated at approximately $290 billion per year (Food and A. O. of the United Nations 2019). Early detection of diseases in plants is critical to minimizing these losses. However, current diagnostic methods rely heavily on subjective visual inspections by experts, which may not always be accurate, especially in cases

of subtle or complex disease symptoms. This highlights the pressing need for a simple yet highly accurate automated plant disease diagnosis system.

Deep learning techniques, particularly convolutional neural networks (CNNs), have demonstrated remarkable performances in plant disease diagnosis, achieving high classification accuracy in numerous studies (Mohanty, Hughes, and Salathé 2016; Hiroki et al. 2018; Ferentinos 2018; Atila et al. 2021). However, subsequent large-scale and rigorous evaluations have revealed a significant drop in diagnostic performance when these models are validated on unseen datasets, captured under conditions different from those used during training (Mohanty, Hughes, and Salathé 2016; Boulent et al. 2019; Shibuya et al. 2021). In many earlier studies, exceptionally high accuracy was reported due to inappropriate train-test data splitting, since images within the same domain (environment) shared similar characteristics. When training and test data were probabilistically split without accounting for domain diversity, implicit data leakage occurred. Shibuya et al. addressed this issue by conducting a large-scale analysis using a dataset of 221,000 images of four crops collected from 24 prefectures, ensuring careful management of shooting fields to prevent data leakage. While they achieved accuracy levels comparable to prior studies (averaging 99%) on data from the same fields used for training, accuracy on data from different fields ranged from 64% to 88% (Shibuya et al. 2021). This discrepancy is attributed to the inherently small and diverse nature of disease symptoms, coupled with significant variations in visual appearance, cultivation methods, and photography conditions across fields (domains) and datasets. The limited diversity of training data, even with tens of thousands of images, fails to capture these variations, hindering the generalization of diagnostic performance to unseen domains. As such, the fundamental diagnostic capability of a system should be evaluated on data unseen to the model, from unknown domains. However, even state-of-the-art machine learning (ML) techniques struggle to achieve high accuracy on such scenarios, disadvantaged by limited diversity in training data.

Domain adaptation techniques have shown promising results for tasks where training and target data exhibit domain differences (Ganin et al. 2016; Tzeng et al. 2017; Dinsdale, Jenkinson, and Namburete 2021; Li et al. 2023). These
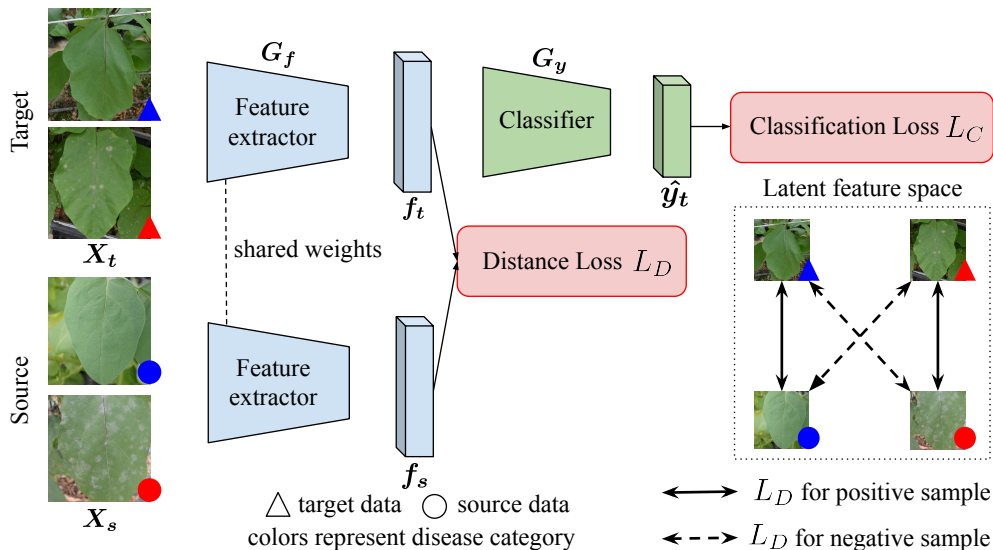
Figure 1: **Architecture of FMDA.** Triangles (△) represent target domain data, circles (◯) represent source domain data, and colors indicate the data labels (disease classes).

methods are particularly valuable for large-scale tasks involving data integration from multiple domain sources. Although domain adaptation has been applied to plant disease diagnosis with some success (Yan et al. 2021; Fuentes et al. 2021; Wu et al. 2023), existing approaches—often based on unsupervised domain adaptation (UDA)—require substantial amounts of unlabeled target data. Furthermore, their effectiveness diminishes when the distribution gap between source (training) and target (test) domains is too large (Wang et al. 2019; Zhao et al. 2019).

Efforts to address the limited diversity of training data include augmenting datasets using generative models. Techniques leveraging Generative Adversarial Networks (GANs) have achieved promising results by generating high-quality images (Nazki et al. 2020; Cap et al. 2022; Kanno et al. 2021). However, since GAN-generated images are derived from existing training data, they face limitations in scenarios with significant domain discrepancies. Recent advances in diffusion models (Ho, Jain, and Abbeel 2020), particularly latent diffusion models (Rombach et al. 2022), enable more flexible and diverse image generation by leveraging extensive pre-trained image datasets and textual information. These advancements hold potential for augmenting training datasets in plant disease diagnosis tasks.

Given the challenge of accurately diagnosing diseases in unseen domains with limited training data diversity, this study considers a practical scenario where a small amount of data from the target diagnostic domain is available. We aim to effectively utilize this limited data to improve diagnostic accuracy.

In this paper, we propose Few-shot Metric Domain Adaptation (FMDA), a supervised domain adaptation method that leverages a small amount of target data effectively. FMDA introduces a learning constraint to reduce the "distance" between feature spaces of the target and source domains,

thereby enhancing the model's ability to generalize. By extending conventional fine-tuning techniques with FMDA, we enable stable mitigation of domain shifts while utilizing existing knowledge from the source domain. FMDA is a simple and versatile approach applicable to various tasks with significant domain shifts. It is particularly effective for tasks like plant disease diagnosis, where intra-class variability is extremely high. In this paper, we demonstrate the application of FMDA to automated plant disease diagnosis, assuming scenarios where limited data from the target imaging environment is available. We validate its effectiveness through large-scale experiments.

The contributions of this study are as follows:

1. We propose FMDA, a supervised domain adaptation method that can be applied to various tasks with serious domain shifts using limited target data.

2. We develop a highly accurate automated plant disease diagnosis system capable of diagnosing target data with high precision.

## Related Works

### Improving Model Generalizability in Automated Plant Disease Diagnosis

**Region of Interest (ROI) detection** Saikawa et al. proposed a system that removes background features, which are considered significant sources of variability across fields, reporting an approximate accuracy improvement of 12% (Saikawa et al. 2019). However, subsequent experiments by Shibuya et al., using a larger dataset with higher-resolution images, revealed that the impact of background removal on diagnostic performance was limited (Shibuya et al. 2021). In contrast, methods utilizing object detection techniques have been reported to be more robust to variations in subject

distance while providing higher interpretability of results (Suwa et al. 2019; Chen et al. 2022). Specifically, Wayama et al. demonstrated through large-scale analysis that ROI detection is particularly effective for diagnosing pest infestations, which are smaller in size but exhibit less variability compared to plant diseases (Wayama et al. 2024).

**Data Generation**   Nazki et al. proposed AR-GAN, an extension of CycleGAN (Zhu et al. 2017) that incorporates a loss function to maintain critical feature similarity between real and generated images. Experiments using a tomato leaf dataset showed that the inclusion of generated images during training enhanced the model's classification accuracy (Nazki et al. 2020). Quan et al. introduced LeafGAN, which ensures that disease symptoms are localized specifically within leaf regions, rather than appearing in the background, enabling the generation of high-quality images (Cap et al. 2022). Kanno et al. proposed Productive and Pathogenic Image Generation (PPIG), a two-stage generative process that first produces healthy leaf images and then overlays disease symptoms. This approach generates more diverse images, with disease symptoms being the only variable feature within otherwise identical compositions (Kanno et al. 2021). While these GAN-based methods have contributed to improving classification accuracy, the generated images are inherently constrained by the distribution of the original training data, reducing their effectiveness in addressing fundamental domain diversity issues. In contrast, latent diffusion models (Rombach et al. 2022) can generate more diverse images compared to GANs. Although still relatively new, these models hold great promise for advancing data augmentation techniques in plant disease diagnosis (Muhammad et al. 2023).

### Studies on Domain Adaptation in ML Tasks

**General Unsupervised Domain Adaptation (UDA)**   Domain adaptation aims to reduce the distributional gap (domain shift) between the "source domain" (training data) and the "target domain" (test data), thereby enhancing model performance on the target domain. In many real-world scenarios, test data cannot be used during training. A practical alternative is the transductive learning setup, where unlabeled test data is fully available for training, but the model is not expected to generalize beyond this scope. This setup has driven extensive research into unsupervised domain adaptation (UDA). Ganin et al. proposed the Domain Adversarial Neural Network (DANN), which simultaneously trains a domain discriminator and a primary classification task using source data. By reversing the gradient of the domain discriminator during backpropagation, DANN enforces domain-invariant feature extraction (Ganin et al. 2016). Tzeng et al. introduced Adversarial Discriminative Domain Adaptation (ADDA), which adversarially trains a target domain encoder while feeding its outputs into a domain discriminator. This approach constructs a domain-invariant encoder, leveraging a pre-trained source domain encoder (Tzeng et al. 2017). These adversarial approaches have been widely adopted across various ML application.

**UDA in Automated Plant Disease Diagnosis**   In the context of automated plant disease diagnosis, UDA has been applied in a limited number of studies. Fuentes et al. proposed a diagnostic framework that applies UDA to the bounding boxes of detected disease symptoms, mitigating domain shifts in a dataset of tomato leaf images captured across three fields (Fuentes et al. 2021). Wu et al. applied UDA to a training framework designed to learn diverse feature representations using a dataset comprising leaf images of various crops captured in both laboratory and outdoor environments (Wu et al. 2023). While these methods have demonstrated potential for improving diagnostic performance, they rely heavily on the availability of large amounts of unlabeled target data. Furthermore, when the data distributions of the source and target domains differ significantly, these methods often fail to bridge the domain gap effectively, thereby limiting diagnostic performance.

To address these challenges, this study considers a practical scenario where a small amount of labeled target data is available. We propose Few-shot Metric Domain Adaptation (FMDA), a domain adaptation method that effectively leverages this limited data to achieve superior diagnostic performance.

## FMDA: Few-shot Metric Domain Adaptation
### Problem Setting and Overview of FMDA

In this study, the labeled training dataset consists of $C$ classes of source data $\{X_s, y_s\}$ and a small number of labeled target data $\{X_t, y_t\}$ for the same $C$ classes. Here, $X$ represents the input data (e.g., images), and $y$ represents the corresponding class labels. While the source data is available in sufficient quantities, the target data is limited to $n$ samples per class, amounting to only $C \times n$ samples in total. The primary objective is to effectively leverage this limited labeled target data to train an ML model $G$ capable of achieving high diagnostic accuracy on unseen target data.

We propose Few-shot Metric Domain Adaptation (FMDA), a diagnostic framework designed to address challenging tasks, such as automated plant disease diagnosis, where domain discrepancies are significant. FMDA is both practical and capable of achieving excellent diagnostic performance. The architecture of FMDA is illustrated in Figure 1. The ML model $G$ in FMDA consists of a feature extractor $G_f$ and a classifier $G_y$, and the choice of models for $G_f$ and $G_y$ is flexible. FMDA training comprises two phases:

1. **Pre-training Phase:** The model $G$ is pre-trained on the source data $\{X_s, y_s\}$ using supervised learning.

2. **Adaptation Phase:** The pre-trained model $G$ is fine-tuned using the limited $C \times n$ target domain training samples $\{X_t, y_t\}$. This phase integrates metric learning into a standard fine-tuning strategy, enabling more effective utilization of the scarce but valuable target data to construct a robust model.

   Conventional fine-tuning applied in scenarios with extremely limited target data risks degrading the original model $G$ and causing overfitting. On the other hand, re-training the entire model using both source and target data is computationally expensive. Moreover, due to

the significantly smaller number of target samples compared to source data, the overall impact of such additional training is often limited. FMDA addresses these challenges through a simple yet effective combination of fine-tuning and metric learning. This approach ensures that the model not only retains the general knowledge learned from the source domain but also adapts effectively to the unique characteristics of the target domain, even with minimal data.

## Implementation Details of FMDA

Since the pre-training phase performs general training of the model using the source data $\{X_s, y_s\}$ as mentioned above, the adaptation phase of the FMDA is explained here. In the adaptation phase, the entire loss function $L$ the model to be minimized is defined as:

$$L = L_C(X_t, y_t) + \lambda L_D. \tag{1}$$

Here, $L_C$ is a classification loss based on the commonly used cross entropy, computed only by target data $\{X_t, y_t\}$:

$$L_C(X_t, y_t) = -\sum_{i=1}^{N} y_t^{(i)} \log(G_y(G_f(X_t^{(i)}))). \tag{2}$$

Here, $N$ denotes the batch size used during training.

The term $L_D$ corresponds to the distance-based loss, which is introduced to mitigate overfitting and domain shift. The hyperparameter $\lambda$ balances the contribution of classification loss and distance-based loss. Conventional fine-tuning often suffers from overfitting when only a small amount of target data is available. This makes the model difficult to generalize to unseen target data. By incorporating $L_D$, FMDA leverages information from both source and target domains, ensuring that the feature representations of target data are properly aligned with the source feature space. This alignment reduces domain discrepancies while preserving robust classification boundaries.

In this study, we define two types of $L_D$:

1. **Euclidean Distance Loss** ($L_{D,\text{L2}}$): This loss minimizes the Euclidean distance between the feature representations of target samples and positive samples ($X_{s_p}^{(i)}$: same labels from $X_t$).

$$L_{D,\text{L2}} = \frac{1}{N} \sum_{i=1}^{N} \left\| G_f(X_t^{(i)}) - G_f(X_{s_p}^{(i)}) \right\|_2. \tag{3}$$

2. **Improved Triplet Loss** ($L_{D,\text{Triplet+}}$): This loss enhances feature separation by maximizing the distance between target samples and negative samples ($X_{s_n}^{(i)}$: different labels from $X_t$) while minimizing the distance to positive samples.

$$\begin{aligned} L_{D,\text{Triplet+}} = \frac{1}{N} \sum_{i=1}^{N} \Big\{ & \left[ \left\| G_f(X_t^{(i)}) - G_f(X_{s_p}^{(i)}) \right\|_2 \right. \\ & \left. - \left\| G_f(X_t^{(i)}) - G_f(X_{s_n}^{(i)}) \right\|_2 + \alpha \right]_+ \\ & + \left[ \left\| G_f(X_t^{(i)}) - G_f(X_{s_p}^{(i)}) \right\|_2 - \beta \right]_+ \Big\}. \end{aligned} \tag{4}$$

$\alpha$ and $\beta$ are margin parameters, and $[z]_+ = \max(0, z)$.

The choice of distance function in FMDA has some influence on its performance, depending on the specific conditions. The Euclidean distance loss is computationally efficient and performs well in cases where intra-class variability is low, with features being tightly clustered. However, its performance may decrease in scenarios where intra-class variability is high, leading to greater diversity within the same class. In contrast, The Improved triplet loss offers additional benefits by minimizing the distance between positive pairs (target and source data with the same label) while increasing the separation between negative pairs (target and source data with different labels). This property makes it more effective in cases with high inter-class similarity or overlapping feature distributions. Practitioners using FMDA should weigh the computational cost against the complexity of feature distributions in the target domain. For applications requiring efficiency, Euclidean Distance Loss may be a practical choice, while Improved Triplet Loss could be advantageous for tasks demanding more discriminative feature spaces.

By incorporating $L_D$ into FMDA, the method effectively enhances generalization by balancing fine-tuning with domain adaptation. This integration enables FMDA to handle domain-shift-heavy tasks with improved flexibility and robustness.

# Experiments

## Experimental Setup

To evaluate the effectiveness of FMDA, we benchmarked it against the automated plant disease diagnosis task, a highly relevant application area, and conducted large-scale experiments using our proprietary dataset. The dataset comprises 223,015 images collected from 20 fields, covering 3 crops and a total of 30 disease classes. Details of the dataset are shown in Table 1. The dataset was split into source (Training) and target (Test) data for different farms taken or different periods of time. From the target data, only a very small number of samples ($n$ samples per disease class) were used for training. As shown in Figure 2, the domains differ significantly in features such as cultivation environments, backgrounds, and compositions.

In this study, EfficientNetV2-S (Tan and Le 2021) was employed as the feature extractor $G_f$, and the classifier $G_y$ was implemented as a simple fully connected layer with a softmax activation function throughout all phases of experiments. In the **pre-training phase**, the model was initialized with weights pre-trained on ImageNet1K (Russakovsky et al. 2015) dataset. Subsequently, pre-training was conducted using the source domain data $\{X_s, y_s\}$. The following conditions were applied during this phase:

- **Optimizer:** Adam (Kingma 2014) with a learning rate of 0.001.
- **Batch Size:** 128.
- **Data Augmentation:** Applied transformations included random cropping to $480 \times 480$ pixels, random 90-degree rotations, horizontal flips, and brightness adjustments.

Table 1: Dataset details.

| Crop | ID_Name | Source | Target |
|---|---|---|---|
| | 00_HEALTHY | 16,016 | 5,583 |
| | 01_Powdery_Mildew | 7,757 | 1,905 |
| | 02_Gray_Mold | 636 | 174 |
| | 03_Anthracnose | 3,031 | 80 |
| | 08_Downy_Mildew | 6,946 | 2,586 |
| Cucumber | 09_Corynespora_Leaf_Spot | 7,557 | 1,820 |
| | 17_Gummy_Stem_Blight | 1,476 | 381 |
| | 20_Bacterial_Spot | 4,355 | 2,655 |
| | 22_CCYV | 5,961 | 186 |
| | 23_Mosaic_diseases | 26,854 | 1,633 |
| | 24_MYSV | 17,229 | 1,011 |
| | Total | 97,818 | 18,014 |
| | 00_HEALTHY | 12,431 | 1,122 |
| | 01_Powdery_Mildew | 7,936 | 938 |
| | 02_Gray_Mold | 1,024 | 166 |
| Eggplant | 06_Leaf_Mold | 3,188 | 732 |
| | 11_Leaf_Spot | 5,509 | 119 |
| | 18_Verticillium_Wilt | 3,176 | 354 |
| | 19_Bacterial_Wilt | 3,415 | 463 |
| | Total | 36,679 | 3,894 |
| | 00_HEALTHY | 8,120 | 2,994 |
| | 01_Powdery_Mildew | 4,490 | 4,250 |
| | 02_Gray_Mold | 9,327 | 571 |
| | 05_Cercospora_Leaf_Mold | 4,078 | 1,809 |
| | 06_Leaf_Mold | 2,761 | 151 |
| | 07_Late_Blight | 2,049 | 808 |
| Tomato | 10_Corynespora_Target_Spot | 1,732 | 1,350 |
| | 19_Bacterial_Wilt | 2,259 | 412 |
| | 21_Bacterial_Canker | 4,353 | 128 |
| | 27_ToMV | 3,453 | 49 |
| | 28_ToCV | 4,320 | 871 |
| | 29_Yellow_Leaf_Curl | 4,513 | 1,746 |
| | Total | 51,471 | 15,139 |

In the **adaptation phase**, the model weights initialized during the pre-training phase were further fine-tuned using the target data $\{X_t, y_t\}$. The experimental setup for this phase is as follows:

- **Optimizer:** Adam with a learning rate of 0.001.
- **Batch Size:** $2 \times C$, where $C$ represents the number of classes in the target data.
- **Data Augmentation:** Identical to the pre-training phase.

The primary focus was on evaluating the impact of domain adaptation learning. Therefore, only basic data augmentation were applied. Based on preliminary experiments, the hyperparameter $\lambda$, controlling the balance between the classification loss and the distance-based loss, was set to $\lambda = 1$.

### Evaluation

We evaluated the diagnostic performance of the classifier $G$ on test data (11 cucumber classes, 7 eggplant classes, and 12 tomato classes) using the F1 score. The following methods were compared:

1. **w/o Target data:** No information from target data is used.
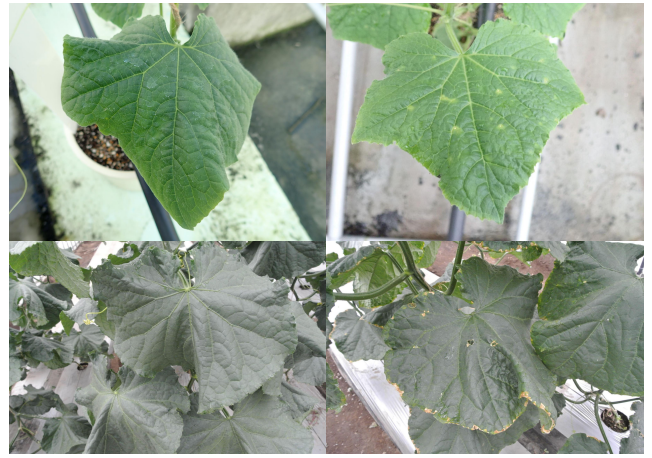


Figure 2: Examples of domain differences in the cucumber dataset. *HEALTHY_Source* (top-left), *Bacterial_Spot_Source* (top-right), *HEALTHY_Target* (bottom-left), *Bacterial_Spot_Target* (bottom-right).

2. **DANN:** The DANN framework (Ganin et al. 2016) is applied in a transductive setting where all target domain images are available as unlabeled data during training.
3. **finetune:** $n$ samples per disease from target data are used for fine-tuning.
4. **DANN-tune:** The representative UDA method DANN (Ganin et al. 2016) is adapted to fine-tuning with $n$ samples per disease.
5. **FMDA(L2):** The proposed method using Euclidean distance for $L_D$.
6. **FMDA(Triplet+):** The proposed method using Improved Triplet Loss (Cheng et al. 2016) for $L_D$.

For these experiments, the number of target data samples was set to $n = 3$ and $n = 10$ per disease in each crop. Each experiment was repeated five times with different target data, and results were evaluated diagnosis performance. The number of training iterations for the pre-training and adaptation phases was set to 100 and 1,000, respectively, based on preliminary experiments. In addition, since the diagnostic performance of some methods varies with the number of training iterations, the performance at the point of maximum test performance was reported along with the corresponding number of iterations. To qualitatively evaluate the extent to which each method reduces domain differences, the feature distribution $f$ of both domain data obtained from the feature extractor $G_f$ was visualized using t-SNE (Van der Maaten and Hinton 2008).

### Results

Table 2 summarizes the diagnostic performance of the models after 1,000 training iterations. Although the dependence of the diagnostic performance on the number of training iterations varied by method, the observed trends were generally consistent across all crops. As a representative example, Figure 3 shows the relationship between the number of training

Table 2: Comparison of plant disease diagnosis performance at the 1,000th epoch in macro F1 (%: mean ± SD).

| Crop | Cucumber | | Eggplant | | Tomato | |
|---|---|---|---|---|---|---|
| # added samples ($n$ / class) | 3 | 10 | 3 | 10 | 3 | 10 |
| 1. w/o Target data | 48.0 | | 72.7 | | 48.0 | |
| 2. DANN | 47.9 | | 67.9 | | 44.5 | |
| 3. finetune (baseline) | 69.73±2.69 | 70.05±2.79 | 78.81±3.56 | 78.39±2.15 | 62.54±5.39 | 73.72±2.42 |
| 4. DANN-tune | 61.45±4.04 | 69.08±1.99 | 73.99±2.75 | 78.89±2.14 | 62.61±3.78 | 67.30±2.91 |
| 5. FMDA(L2) (ours) | 72.11±2.62 | 74.57±1.53 | **81.61±1.03** | 82.73±3.03 | 65.75±1.53 | 74.19±2.78 |
| 6. FMDA(Triplet+) (ours) | **72.91±1.45** | **77.29±1.84** | 80.62±2.90 | **83.83±1.97** | **65.91±2.76** | **74.37±1.42** |

Scores in bold indicate best results in the category.

Table 3: Reference results: Comparison of plant disease diagnosis performance at the best-performing epoch (epoch count) in macro F1 (%: mean ± SD).

| Crop | Cucumber | | | | Eggplant | | | | Tomato | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| # added samples ($n$ / class) | 3 | | 10 | | 3 | | 10 | | 3 | | 10 | |
| 3. finetune (baseline) | 69.85±1.42 | (200) | 76.32±1.50 | (120) | 81.04±1.44 | (600) | 87.17±1.70 | (50) | 65.24±2.40 | (400) | 74.85±2.41 | (180) |
| 4. DANN-tune | 68.82±1.67 | (100) | 75.64±0.81 | (20) | 77.71±3.12 | (500) | 86.53±1.17 | (10) | 62.86±3.75 | (900) | 74.85±2.41 | (180) |
| 5. FMDA(L2) (ours) | 72.38±1.73 | (700) | 79.39±1.57 | (400) | **85.36±1.39** | (400) | 89.08±1.63 | (50) | 70.74±2.02 | (200) | 79.26±0.75 | (90) |
| 6. FMDA(Triplet+) (ours) | **73.79±2.30** | (400) | **79.90±1.31** | (300) | 85.20±2.04 | (500) | **89.15±1.11** | (30) | **70.79±2.58** | (200) | **80.38±1.51** | (60) |

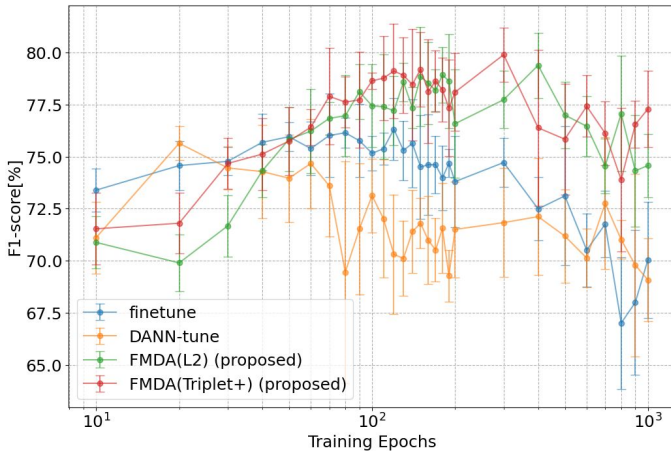Scores in bold indicate best results in the category.



Figure 3: Performance trends for the cucumber dataset.

iterations and the diagnosis performance for the cucumber dataset when $n = 10$.

Table 3 the F1 scores achieved at the best-performing number of training iterations for each method. This provides additional insight into the peak performance potential of each approach.

Figure 4 visualizes the feature distribution of the tomato dataset's source and target data using t-SNE, at $n = 10$. The visualization highlights how well each method aligns the target domain features with the source domain, offering a qualitative measure of domain adaptation effectiveness.

## Discussion

When target data is not utilized, the diagnostic performance remains extremely low, even when state-of-the-art ML classifiers are trained on 30,000–90,000 images per crop. This reaffirms that automated plant disease diagnosis is inherently a domain-shift-heavy problem, making it challenging to achieve practical diagnostic accuracy without addressing domain discrepancies.

Among the methods leveraging label information from the test data, DANN-tune exhibited the lowest diagnostic performance. DANN reduces domain discrepancy by training the feature extractor $G_f$ to produce domain-invariant features. However, this approach does not guarantee that target data with the same labels as source data will be mapped to nearby locations in the feature space. As illustrated in the feature distribution (Figure 4), DANN-tune incorrectly maps target data into inappropriate class distributions, particularly in datasets with significant domain shifts. Furthermore, the DANN (Transductive Learning) setting, which utilized all unlabeled target domain images, also performed worse than the baseline w/o Target. This counterintuitive result suggests that using unlabeled target data without proper alignment of features can exacerbate domain shifts, leading to degraded diagnostic performance. These findings highlight that UDA approaches, such as DANN, are ineffective for tasks with substantial domain discrepancies, such as the current task of automated plant disease diagnosis.

In contrast, FMDA consistently achieved high diagnostic performance regardless of the distance loss function used. This was evident not only in its practical performance after a fixed number of training iterations but also in its best-performing potential, where FMDA demonstrated a higher upper limit of achievable performance compared to existing methods. Figure 4 demonstrates that FMDA effectively leverages target label information to map target features appropriately onto the source feature space. This indicates that FMDA can successfully learn from limited target data, reducing domain discrepancies while preserving diagnostic accuracy.

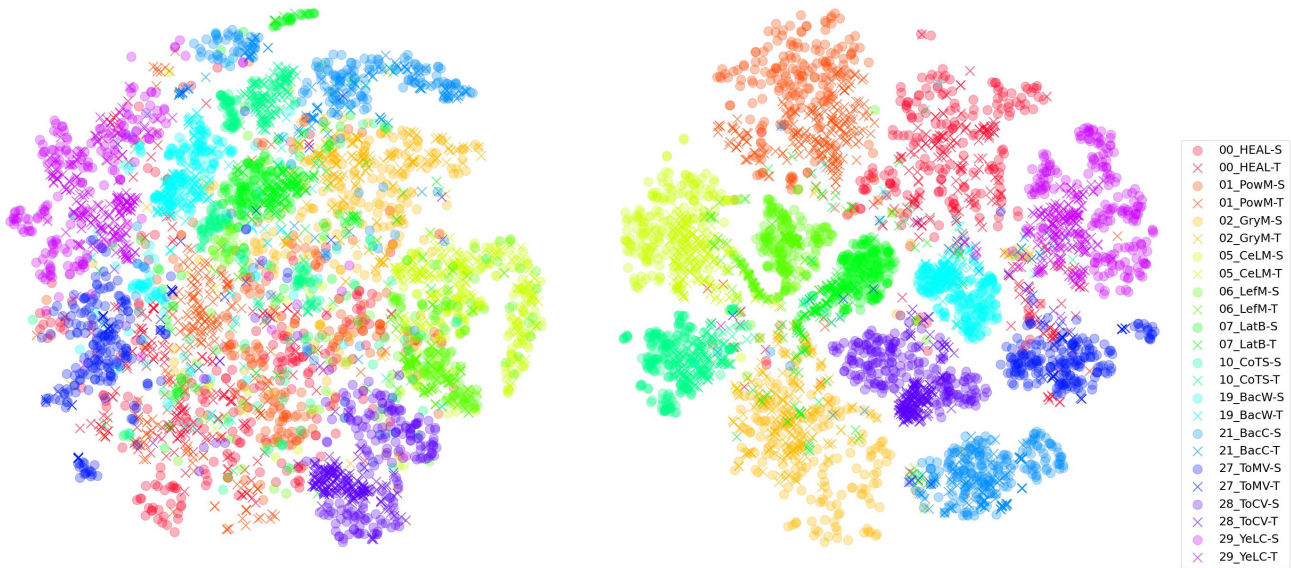Incorporating as few as three labeled target data sam-

Figure 4: Feature distribution for the tomato dataset. DANN-tune (left), FMDA(Triplet+) (right). Circles (◯) represent source data, and crosses (×) represent target data.

ples per disease resulted in an average F1 score improvement of 9.8–16.9 points across methods, while using ten target samples yielded an average improvement of 15.5–22.3 points. Although performance improves as more target data becomes available, even minimal amounts — such as three samples per disease — significantly enhance diagnostic performance. This underscores the importance of utilizing even limited labeled target data in domain-shift-heavy tasks like automated plant disease diagnosis.

Fine-tuning with a small amount of target data experienced a performance drop of 6.3 points due to overfitting, whereas FMDA exhibited only a modest decline of approximately 2.5 points. This suggests that fine-tuning is highly susceptible to overfitting in low-data scenarios, while FMDA effectively mitigates overfitting and facilitates stable learning.

While FMDA demonstrates significant improvements in diagnostic performance, certain limitations remain. For example, its reliance on labeled target data, albeit in small quantities, may not be feasible in real-world scenarios where labeling is costly or impractical. Future research could investigate integrating FMDA with semi-supervised learning techniques to harness a larger pool of unlabeled data. Additionally, while this study focuses on plant disease diagnosis, FMDA's framework could be extended to other domain-shift-heavy tasks, such as medical imaging or remote sensing. Validating FMDA in these domains would further assess its generalizability and robustness under varied conditions.

## Conclusion

In this study, we proposed Few-shot Metric Domain Adaptation (FMDA), a supervised domain adaptation method designed to address tasks with significant domain shifts. FMDA demonstrated its ability to effectively and reliably

learn from limited target domain information, even in scenarios where UDA-based approaches fail due to large domain discrepancies.

## Acknowledgement

## References

Atila, Ü.; Uçar, M.; Akyol, K.; and Uçar, E. 2021. Plant leaf disease classification using EfficientNet deep learning model. *Ecological Informatics*, 61: 101182.

Boulent, J.; Foucher, S.; Théau, J.; and St-Charles, P.-L. 2019. Convolutional neural networks for the automatic identification of plant diseases. *Frontiers in plant science*, 10: 941.

Cap, Q. H.; Uga, H.; Kagiwada, S.; and Iyatomi, H. 2022. LeafGAN: An Effective Data Augmentation Method for Practical Plant Disease Diagnosis. *IEEE Transactions on Automation Science and Engineering*, 19(2): 1258–1267.

Chen, Z.; Wu, R.; Lin, Y.; Li, C.; Chen, S.; Yuan, Z.; Chen, S.; and Zou, X. 2022. Plant disease recognition model based on improved YOLOv5. *Agronomy*, 12(2): 365.

Cheng, D.; Gong, Y.; Zhou, S.; Wang, J.; and Zheng, N. 2016. Person re-identification by multi-channel parts-based cnn with improved triplet loss function. In *Proceedings of the iEEE conference on computer vision and pattern recognition*, 1335–1344.

Dinsdale, N. K.; Jenkinson, M.; and Namburete, A. I. 2021. Deep learning-based unlearning of dataset bias for MRI harmonisation and confound removal. *NeuroImage*, 228: 117689.

Ferentinos, K. P. 2018. Deep learning models for plant disease detection and diagnosis. *Computers and electronics in agriculture*, 145: 311–318.

Food and A. O. of the United Nations. 2019. New standards to curb the global spread of plant pests and diseases. https://www.fao.org/newsroom/detail/New-standards-to-curb-the-global-spread-of-plant-pests-and-diseases/en. Accessed: 2024-11-12.

Fuentes, A.; Yoon, S.; Kim, T.; and Park, D. S. 2021. Open set self and across domain adaptation for tomato disease recognition with deep learning techniques. *Frontiers in Plant Science*, 12: 758027.

Ganin, Y.; Ustinova, E.; Ajakan, H.; Germain, P.; Larochelle, H.; Laviolette, F.; March, M.; and Lempitsky, V. 2016. Domain-adversarial training of neural networks. *Journal of machine learning research*, 17(59): 1–35.

Hiroki, T.; Kotani, R.; Kagiwada, S.; Hiroyuki, U.; and Iyatomi, H. 2018. Diagnosis of multiple cucumber infections with convolutional neural networks. In *2018 IEEE Applied Imagery Pattern Recognition Workshop (AIPR)*, 1–4. IEEE.

Ho, J.; Jain, A.; and Abbeel, P. 2020. Denoising diffusion probabilistic models. *Advances in neural information processing systems*, 33: 6840–6851.

Kanno, S.; Nagasawa, S.; Cap, Q. H.; Shibuya, S.; Uga, H.; Kagiwada, S.; and Iyatomi, H. 2021. PPIG: Productive and pathogenic image generation for plant disease diagnosis. In *2020 IEEE-EMBS Conference on Biomedical Engineering and Sciences (IECBES)*, 554–559. IEEE.

Kingma, D. P. 2014. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*.

Li, J.; Xu, R.; Ma, J.; Zou, Q.; Ma, J.; and Yu, H. 2023. Domain adaptive object detection for autonomous driving under foggy weather. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, 612–622.

Mohanty, S. P.; Hughes, D. P.; and Salathé, M. 2016. Using deep learning for image-based plant disease detection. *Frontiers in plant science*, 7: 1419.

Muhammad, A.; Salman, Z.; Lee, K.; and Han, D. 2023. Harnessing the power of diffusion models for plant disease image augmentation. *Frontiers in Plant Science*, 14: 1280496.

Nazki, H.; Yoon, S.; Fuentes, A.; and Park, D. S. 2020. Unsupervised image translation using adversarial networks for improved plant disease recognition. *Computers and Electronics in Agriculture*, 168: 105117.

Rombach, R.; Blattmann, A.; Lorenz, D.; Esser, P.; and Ommer, B. 2022. High-resolution image synthesis with latent diffusion models. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 10684–10695.

Russakovsky, O.; Deng, J.; Su, H.; Krause, J.; Satheesh, S.; Ma, S.; Huang, Z.; Karpathy, A.; Khosla, A.; Bernstein, M.; et al. 2015. Imagenet large scale visual recognition challenge. *International journal of computer vision*, 115: 211–252.

Saikawa, T.; Cap, Q. H.; Kagiwada, S.; Uga, H.; and Iyatomi, H. 2019. Aop: An anti-overfitting pretreatment for practical image-based plant diagnosis. In *2019 IEEE International Conference on Big Data (Big Data)*, 5177–5182. IEEE.

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*.

Suwa, K.; Cap, Q. H.; Kotani, R.; Uga, H.; Kagiwada, S.; and Iyatomi, H. 2019. A comparable study: Intrinsic difficulties of practical plant diagnosis from wide-angle images. In *2019 IEEE International Conference on Big Data (Big Data)*, 5195–5201. IEEE.

Tan, M.; and Le, Q. 2021. Efficientnetv2: Smaller models and faster training. In *International conference on machine learning*, 10096–10106. PMLR.

Tzeng, E.; Hoffman, J.; Saenko, K.; and Darrell, T. 2017. Adversarial discriminative domain adaptation. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, 7167–7176.

Van der Maaten, L.; and Hinton, G. 2008. Visualizing data using t-SNE. *Journal of machine learning research*, 9(11).

Wang, Z.; Dai, Z.; Póczos, B.; and Carbonell, J. 2019. Characterizing and avoiding negative transfer. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 11293–11302.

Wayama, R.; Sasaki, Y.; Kagiwada, S.; Iwasaki, N.; and Iyatomi, H. 2024. Investigation to answer three key questions concerning plant pest identification and development of a practical identification framework. *Computers and Electronics in Agriculture*, 222: 109021.

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.

Yan, K.; Guo, X.; Ji, Z.; and Zhou, X. 2021. Deep transfer learning for cross-species plant disease diagnosis adapting mixed subdomains. *IEEE/ACM transactions on computational biology and bioinformatics*, 20(4): 2555–2564.

Zhao, H.; Des Combes, R. T.; Zhang, K.; and Gordon, G. 2019. On learning invariant representations for domain adaptation. In *International conference on machine learning*, 7523–7532. PMLR.

Zhu, J.-Y.; Park, T.; Isola, P.; and Efros, A. A. 2017. Unpaired image-to-image translation using cycle-consistent adversarial networks. In *Proceedings of the IEEE international conference on computer vision*, 2223–2232.