Computers and Electronics in Agriculture 227 (2024) 109574

Contents lists available at ScienceDirect

Computers and Electronics in Agriculture

journal homepage: www.elsevier.com/locate/compag

Original papers

Advancing plant disease classification: A robust and generalized approach with transformer-fused convolution and Wasserstein domain adaptation

Muhammad Hanif Tunio [a,d], Jian ping Li [a,*], Xiaoyang Zeng [a], Awais Ahmed [a], Syed Attique Shah [b], Hisam-Uddin Shaikh [c], Ghulam Ali Mallah [d], Imam Abdullahi Yahya [e]

[a] School of Computer Science and Engineering, University of Electronic Science and Technology of China - UESTC, Sichuan, 611731, China
[b] Faculty of Computing, Engineering and the Built Environment, Birmingham City University, STEAMhouse, B4 7RQ, Birmingham, United Kingdom
[c] Department of Mathematics, Shah Abdul Latif University, Khairpur 66111, Sindh, Pakistan
[d] Institute of Computer Science, Shah Abdul Latif University, Khairpur 66111, Sindh, Pakistan
[e] Faculty of Computing, Bayero University Kano, Nigeria

ARTICLE INFO

Keywords:
Plant disease classification
Wasserstein distance
Transformer fused convolution
Unsupervised domain adaptation

ABSTRACT

Plant diseases pose significant threats to agricultural productivity and food security. Owing to a scarcity of field environment datasets, the prevailing plant disease classification approaches, trained on laboratory-controlled datasets, often grapple with achieving optimal performance in real-world environments. We proposed a novel and robust framework for Unsupervised Domain Adaptation (UDA), employing an adversarial learning approach with a Wasserstein distance-informed algorithm to learn domain invariant feature representations capable of generalizing more diverse features. This approach incorporates insights from a labeled source domain and adopts an unlabeled target domain by minimizing the distribution discrepancies between domains. Recently, mobile vision transformer (MViT)-based methods have been applied to UDA due to their ability to capture long-distance feature dependencies. However, these methods overlook the fact that MViT lacks effectiveness in extracting local feature details. The proposed framework combines the advantages of convolutional neural networks (CNNs) and MViTs, integrating local features extracted by CNNs with global features captured by MViTs. This fusion of local and global representations enhances transferability and feature discriminability within the domains. Furthermore, we incorporate a feature-fusing method to align channel dimensions and enhance the local details of the global representation. Extensive experiments using three plant disease datasets demonstrate the effectiveness and efficiency of our approach, yielding significant improvements in classification performance with 13.67%, compared to state-of-the-art (SOTA) and baseline methods. Our framework offers a promising solution for robust and efficient plant disease classification, providing valuable insights for sustainable agriculture and crop management.

1. Introduction

Agriculture is the fundamental cornerstone of national income and is pivotal in shaping a country's economic vitality and prosperity. Plant diseases substantially influence food production, emerging as a prominent factor with considerable ramifications for agricultural yields. Research[1] reports indicate that approximately 20%–40% of the total crop yield is annually compromised due to pests and plant diseases, to ensure food security in the forthcoming years, there is a pressing need for a 50%-60% (Vishnoi et al., 2022), the UN's Food and Agriculture Organization proposes a 70% increase in food supply by 2050 to feed the growing global population (Tilman et al., 2011; Moupojou et al., 2023). The timely and accurate diagnosis of plant diseases holds immense importance in augmenting production capacity in sustainable agriculture.

Plant disease monitoring and analysis were primarily conducted manually by domain specialists. Thereafter, various monitoring devices were developed. However, their adoption adds complexity and incurs additional hardware expenses for farmers, posing challenges for those with limited income to afford such devices (Vishnoi et al., 2021). Nevertheless, the multitude of diseases influenced by various

* Corresponding author.
E-mail addresses: hanif.tunio@salu.edu.pk (M.H. Tunio), jpli2222@uestc.edu.cn (J.p. Li), 202011081605@std.uestc.edu.cn (X. Zeng), 202014080105@std.uestc.edu.cn (A. Ahmed), syedattique.shah@bcu.ac.uk (S.A. Shah), hisam.shaikh@salu.edu.pk (H.-U. Shaikh), ghulam.ali@slau.edu.pk (G.A. Mallah), ayimam.it@buk.edu.ng (I.A. Yahya).

[1] https://www.cabi.org/projects/global-burden-of-crop-loss/

https://doi.org/10.1016/j.compag.2024.109574
Received 30 January 2024; Received in revised form 30 September 2024; Accepted 18 October 2024
Available online 21 November 2024

factors, including plant species, environmental conditions, and farming practices in the premature stage, pose challenges even for expert pathologists in diagnosing such diseases at a large scale. Visual observation and conventional instrument-based methods, such as microscope-based observation, for timely plant disease detection. It has been acknowledged that these methods lack speed, rendering them inadequate for timely disease detection (Sarkar et al., 2023). Furthermore, with the continuous advancements in Artificial Intelligence (AI) and the emergence of Machine Learning (ML) DL methods, integrating these techniques has become a promising alternative to conventional plant disease recognition techniques, leading to superior outcomes and remarkable advancements in the field (Wang et al., 2022).

Most contemporary research efforts in this domain have focused on constructing advanced automatic plant disease recognition systems (Ferentinos, 2018; Too et al., 2019). Support Vector Machines (SVMs) have gained popularity as a widely employed classification model (Yigit et al., 2019). In a study by Ratnasari et al. (2014), color and texture features were extracted from sugarcane leaf using RGB images acquired in laboratory conditions. Subsequently, an SVM was applied, resulting in a notable 80% classification accuracy for identifying sugarcane spot disease. Furthermore, other conventional ML methods, such as Random Forest (RF) (Basavaiah and Arlene Anthony, 2020), Genetic Algorithm Optimized SVM (Zhang et al., 2015), K-nearest neighbor (KNN) (Hossain et al., 2019), and backpropagation neural networks (Chanda and Biswas, 2019) have been widely employed for plant disease classification across various crop species using laboratory-controlled datasets.

Although plant disease recognition under field conditions has received limited attention from researchers in recent years, (Chen et al., 2020; Boulent et al., 2019; Fuentes et al., 2020). The study (Gui et al., 2021) proposed an improved CNN approach compared the baseline model's performance on field-PV and PlantVillage datasets, revealing a significant decline (from 99.72% to 41.81%) due to the challenge of complex backgrounds in field plant datasets. The studies (Mohanty et al., 2016; Raja et al., 2018) identified two limitations in their approach to situ plant disease recognition. Firstly, the structured and simplified nature of the PlantVillage dataset needs to adequately reflect real-world challenges. Secondly, their current method is limited to classifying single leaves under controlled conditions. Despite demonstrating an impressive accuracy of 99.35% on the PlantVillage test set, the recognition accuracy experienced a significant decline to 31.4% when confronted with field environment datasets (Ferentinos, 2018). CNN model was trained using a dataset of 87,848 images, including 37.3% collected under field conditions and the rest from PlantVillage. Evaluation of the model on images obtained under field conditions yielded an accuracy of 33.27%, demonstrating a decrease compared to laboratory conditions. Furthermore, exploring domain adaptation techniques has gained attention in the context of the field and real environment plant disease detection, aiming to address domain shift challenges and improve model adaptability to varying laboratory and situ conditions.

To address the aforementioned limitations, this study presents an approach that uses the strengths of both CNNs and MViTs models in domain adaptation. Our proposed method fuses the local patterns and intricate details extracted by CNNs with global features captured by the mobile vision transformer to achieve adversarial domain adaptation. However, we combine CNN-based local features with transformer-based global representations to minimize the domain data variation using Wasserstein algorithm. This integration of CNN and MViT classifies the disease class from all the images. We employ a feature-fusing method to combine CNN-based and transformer-based features that align the channel dimensions through 1×1 convolutions, upsampling and downsampling strategies. This approach improves the global representation by integrating the local features (Yuan et al., 2021; Dosovitskiy et al., 2020). In our approach, we also integrate conditional adversarial domain adaptation (CDAN) (Long et al., 2018) with the transformer fused convolution method. Instead of using the adversarial loss, we employ the loss from CDAN to evaluate the performance in the context of Domain Adaptation. Additionally, we introduce an uncertainty penalty loss (UPL). UPL aims to restrict high prediction scores for incorrect categories by prioritizing uncertain samples with lower cross-entropy losses. Following the theory proposed in Ben-David et al. (2010), a smaller A-distance indicates a more effective alignment of distributions. We also used the Wasserstein distance Loss to enhance the robustness. In contrast, the Wasserstein distance loss aims to provide a more effective and stable measure of dissimilarity between probability distributions, particularly in domain adaptation. Our novel approach's overall domain adoption framework is depicted in Fig. 5. Here, we provide a summary of our key contributions.

- The study introduces a novel robust framework for the field environment plant disease classification. The framework combines the strengths of MViTs and CNNs by integrating local features extracted by CNNs with global features captured by MViTs, which results in enhanced performance.
- We employed an adversarial learning approach with a Wasserstein distance-informed algorithm to learn domain invariant feature representations. Our contribution minimizes distribution discrepancies between labeled source and unlabeled target domains, enabling the generalization of diverse features. This process enhances the robustness of the model by reducing uncertainty in incorrect classes and improving certainty in correct classes.
- The performance of the proposed model is validated and compared against various baseline and SOTA methods. The results demonstrate that the proposed framework significantly reduces the disparity between the two domains, showcasing remarkable improvements on field environment datasets.

The research paper is organized cohesively, beginning with Section 1, which introduces the significance and current challenges in the domain. Section 2 explains the related work. Section 3 provides detailed and comprehensive information on the materials and the methods used in the study. In Section 4, comprehensive results are presented and accompanied by an in-depth discussion. Finally, Section 5 concludes the paper and highlights potential future directions for further research.

2. Related work

An artificial intelligence and machine learning in agricultural automation has witnessed significant advancements and diverse approaches to addressing the challenges of plant disease identification, paving the way for a deeper understanding and improved solutions in this domain.

The Deep Learning-based approaches have shown remarkable progress, leading to significant advancements in plant disease classification. Batchuluun et al. (2022) introduces a method that combines CNN with explainable artificial intelligence (XAI) for accurate and interpretable disease classification. Sladojevic et al. (2016) achieved an accuracy of 96.3% using CNN for classifying 13 leaf disease classes based on 2,589 images. Their work demonstrated the potential of CNNs in accurate leaf disease classification. Furthermore, in the study, GoogleNet (Mukherjee et al., 2017) and its variations were utilized by Brahimi et al. (2017) for diagnosing leaf diseases in apple plants, which achieved an accuracy of 85.04% in identifying four disease classes. However, the task-specific style of modules for feature extraction may result in lower accuracy and limited adaptability of ML methods when utilized to adopt a new task. Transfer Learning can mitigate lower accuracy and adaptability limitations caused by task-specific design styles in feature extraction modules when applied to different tasks (Tunio et al., 2024).

Furthermore, various studies utilized Transfer Learning techniques to enhance the classification accuracy and robustness of the model,

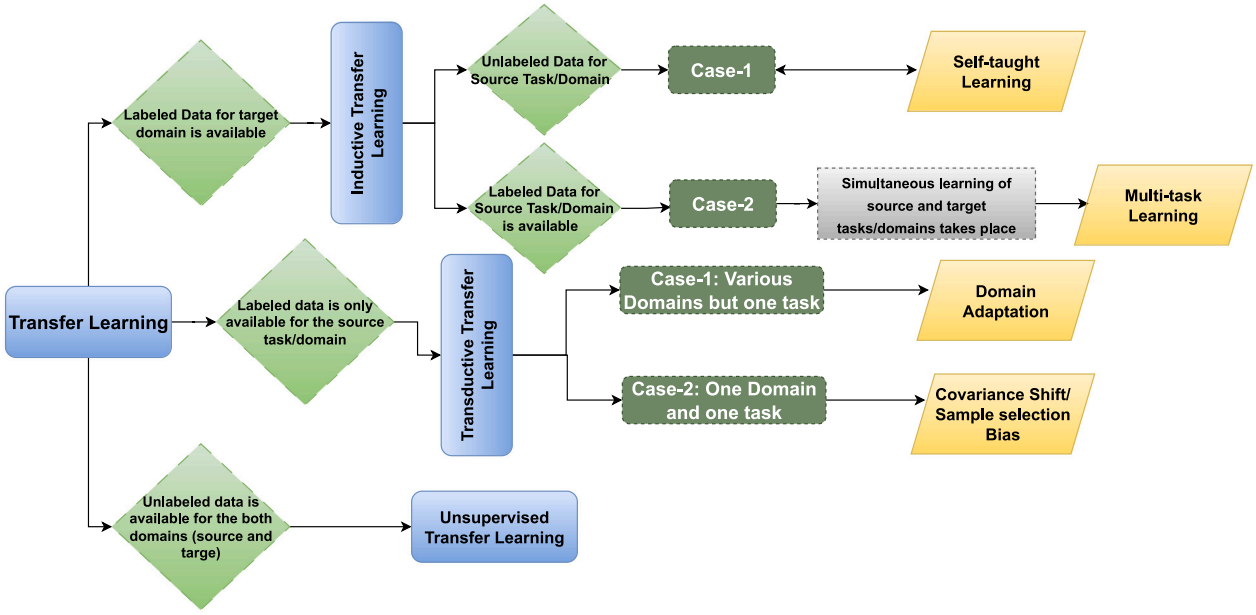


Fig. 1. Comprehensive overview of transfer learning strategies.

particularly for datasets with limited samples (Ramcharan et al., 2017). The other studies increased the depth with additional convolution and ReLU layers (Pereira et al., 2019), as well as introducing complex structures like the inception architecture (Szegedy et al., 2015) for extracting multi-scale features (Li et al., 2020). Deeper and more complex layers in deep learning methods often lead to increased parameters and computational requirements, which can present practical challenges for their application. To address this, Tang et al. (2020) developed a ShuffleNetV2-based lightweight model (Ma et al., 2018), achieving 95.28% accuracy while compressing the size to 5.3 MB. A lightweight object detection model was developed and trained for in situ diagnosis by Ramcharan et al. (2019), Liu et al. (2016), with MobileNet (Howard et al., 2017) as the backbone. Fig. 1 shows the various transfer learning settings (Pan and Yang, 2009). The Multilevel Feature Fusion Network (MFFN) combined with an adaptive attention mechanism (Sunil et al., 2023b) has emerged as a promising approach for tomato plant disease classification, incorporating channel, spatial, and pixel attention to enhance feature extraction. However, it is essential to address the study's limitations, which relied on a subset of the tomato plant dataset (PlantVillage). This narrow focus may raise concerns regarding the generalizability of the framework to other datasets and plant species. To provide a comprehensive understanding of plant disease detection, a thorough analysis of 160 research studies (Sunil et al., 2023a) revealed that approximately 57 studies specifically addressed the detection of plant diseases with multiple plants. Nevertheless, it is crucial to consider the challenges associated with multiple plant diseases, such as imbalanced data within multiple plant disease datasets and the varying nature of data samples caused by different diseases, which can impact model training and generalizability. Addressing limitations and developing robust models enables accurate and reliable detection of plant diseases across diverse species.

2.1. Domain adaptation

Recently, there has been extensive research on adaptive methods in the field of deep networks (Long et al., 2016; Ghifary et al., 2016). Deep networks have shown superior performance compared to shallow adaptive methods. Other studies have also focused on improving model performance by designing discriminators to learn domain-invariant features (Tzeng et al., 2017; Yu et al., 2019). However, only a few studies have applied Unsupervised Domain Adaptation (UDA) methods to the domain of plant and agriculture (Wu et al., 2023). Ganin et al. (2016) proposed a UDA method for transferring density map estimation in counting plant organs. Ganin and Lempitsky (2015) presented a model based on unsupervised adversarial adaptation for estimating leaves from one dataset to another. Both of these methods directly used UDA for plant and agriculture-related tasks.

2.2. Semantic alignment

Numerous nonadversarial and adversarial methods have successfully achieved global alignment in Unsupervised Domain Adaptation (UDA). However, these methods often need to pay more attention to the class-level multimodal structures in the data, leading to potential semantic misalignment. To address this limitation, researchers have emphasized semantic alignment to incorporate fine-grained class-level structures in various tasks appropriately. For instance, Wang and Zhang (2020) proposes a self-adaptive reweighted adversarial approach that improves interclass separation and intraclass compactness for class-level alignment. Deep Subdomain Adaptation Network (DSAN) (Zhu et al., 2020) addresses image classification by adaptively learning fine-grained information in subdomains. Transferable Attention for Domain Adaptation (Wang et al., 2019) focuses on transferability and adaptive modeling of transferable regions or images. Although these semantic alignment methods have shown efficacy, they often need to consider the important differences between domains, resulting in unsatisfactory transfer learning.

In domain adaptation, the vision transformers (ViT) has demonstrated its utility (Yang et al., 2023; Xu et al., 2021; Yang et al., 2021). As an illustration, source-free-domain-adoption (Yang et al., 2021) incorporates the transformer as an attention module within the convolutional network. The study conducted by Xu et al. (2021) on the cross-domain transformer (CDtrans) reveals that the transformer exhibits robustness towards noisy inputs and demonstrates the potential for achieving feature alignment through its utilization. Unfortunately, recent observations in the literature have indicated that the ViT needs to pay attention to the details of local features (Peng et al., 2021), potentially leading to a decrease in discriminability between background and foreground elements when relying solely on the transformer. Alternatively, CNN-based approaches have successfully utilized UDA, progressively capturing local representation through layer-by-layer analysis using local sensory fields. However, capturing the global

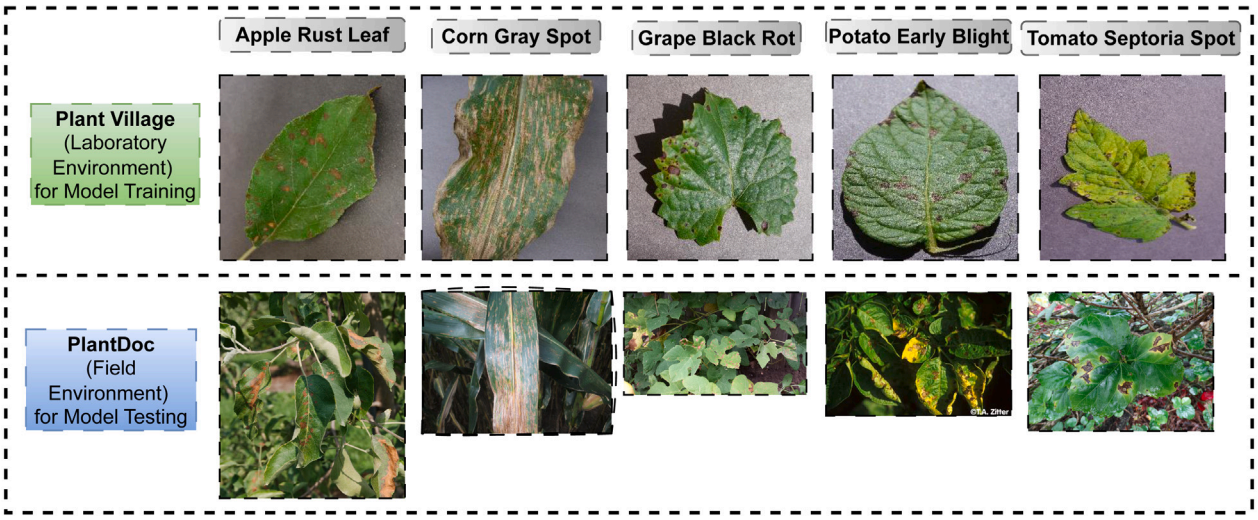


Fig. 2. Subset of plant disease images, captured under controlled laboratory & real-world field conditions.

Table 1
Statistics of the experimental datasets.

Dataset name	No. of samples	No. of classes
PlantVillage	54,304	38
PlantDoc	2598	27
AC-PlantDoc	8456	29

representation remains challenging. Transformer and CNN-based models possess their strengths and weaknesses. The existing studies need more combination of these two approaches in the context of domain adaptation for plant disease applications, aiming to extract the essential elements while eliminating unnecessary components.

3. Materials and methods

3.1. Datasets

In this study, we utilized three datasets: PlantVillage,[2] PlantDoc,[3] and AC-PD (our augmented dataset). These datasets were chosen to encompass various plant-related images and ensure a comprehensive representation. Table 1 provides detailed statistics for each dataset, including sample size and the number of classes.

PlantVillage dataset (PVD) is publicly accessible and is an epidemiological tool designed explicitly for evaluating automated plant disease identification systems. PlantVillage dataset consists of 54,304 images of 256X256 size, representing 38 distinct categories of healthy and diseased plant leaves (Hughes et al., 2015). Each sample within the PlantVillage dataset is provided in three versions: Color, Segmented (black background), and Grayscale. It is important to note that all three versions of images obtained from the PlantVillage dataset are captured in a controlled laboratory environment, which differs from the actual field conditions. Due to its ample sample size and popularity in plant disease detection, we employ the PVD as the source domain dataset in our study. Depending on the particular transfer task and target domain dataset, we carefully selected 7160 samples from the 38 categories within the PVD dataset.

To evaluate the proposed framework's performance on the target model in tackling the difficulties posed by demanding field conditions, we employ PlantDoc dataset, introduced by Singh et al. (2020), is a meticulously annotated collection of 2598 plant disease images obtained from various sources on the internet. This dataset encompasses 27 distinct crop–disease categories. The dataset is curated by selecting images from actual plant growth settings to focus specifically on plant diseases within real field environments. Consequently, the revised dataset is valuable for obtaining plant disease data under field conditions. We augmented and cropped the PlantDoc dataset for the target model. Fig. 2 depicts a subset of images from PlantVillage and PlantDoc datasets.

The images in the PlantDoc dataset contain noise. To mitigate the noise, we used Gaussian and median filtering preprocessing techniques to enhance the quality and improve the accuracy of our framework. Furthermore, the data augmentation and cropping was performed on the original PlantDoc dataset, which initially consisted of 2,598 with multiple regions of interest (RoI), including diseased and healthy leaves. The augmentation and cropping process uses the annotation boxes for the specific RoI in the original PlantDoc images. These annotated regions were extracted, resulting in an augmented and cropped version of the PlantDoc dataset containing 10,958 samples. We thoroughly examined the annotation boxes in the PlantDoc dataset to ensure data accuracy. As some of these annotations were found to be incorrect, such as instances where annotation boxes were mistakenly labeled on apples instead of apple disease leaves, we removed images that exhibited obvious annotation errors. As a result, our final selection comprised 8456 images with 29 classes shown in Fig. 3 deemed accurate and suitable for our study, referred to as AC-PD Dataset. We divided our dataset into class subsets during the training and testing to ensure balanced and class-specific data representation. Each class subset is processed to facilitate efficient batch processing. To further enhance the training process, this study introduced custom data' which allows for a weighted random selection of data from different classes during training. This ensures the training process maintains diversity and mitigates biases towards dominant classes. Additionally, we incorporated normalization transforms to standardize input data. The resulting data were instrumental in achieving a balanced and robust training and testing framework for our machine-learning models. Furthermore, the Statistics of the Training and Testing split are depicted in Table 2.

3.2. MViT segment fused with CNNs segment

The fusion of Mobile Vision Transformers (MViTs) and Convolutional Neural Networks (CNNs) in the proposed method is achieved by aligning the channel dimensions of the features extracted from the

[2] https://github.com/gabrieldgf4/PlantVillage-Dataset
[3] https://github.com/pratikkayal/PlantDoc-Dataset

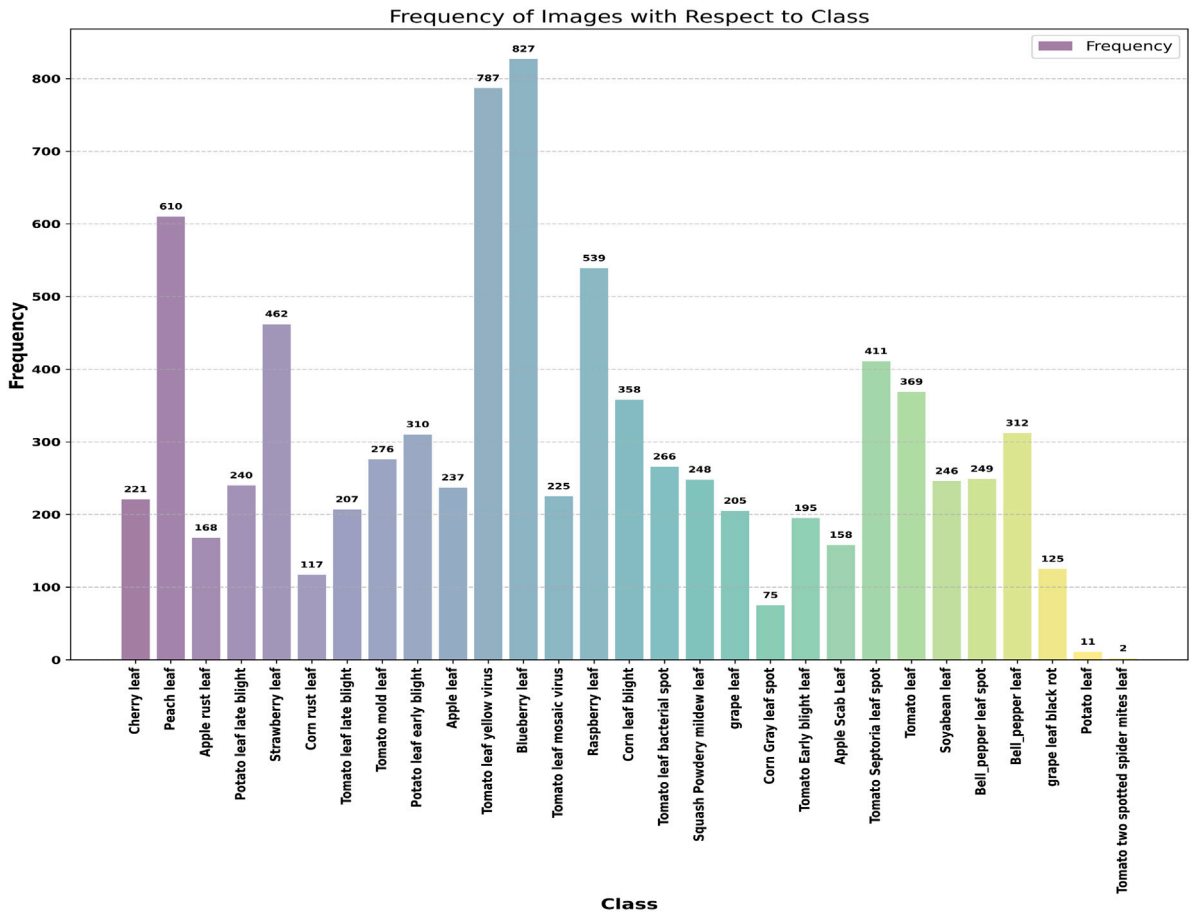


Fig. 3. Statistics of augmented cropped PlantDoc (AC-PD) Dataset.

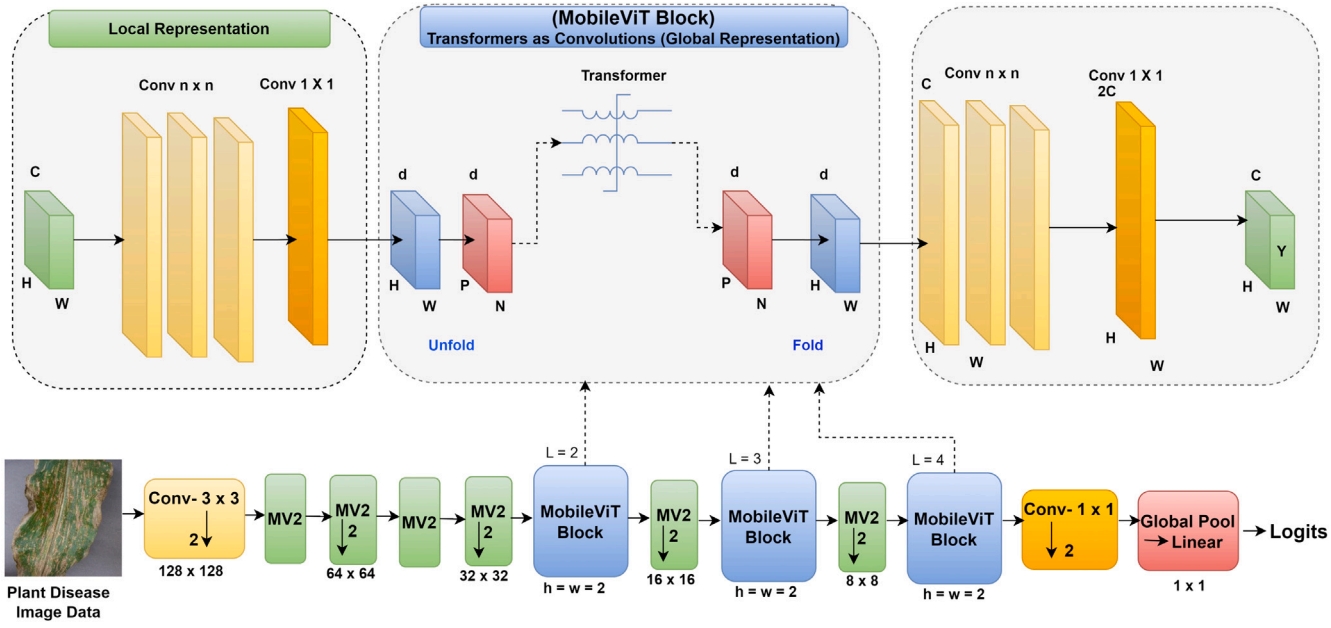


Fig. 4. Integration of local CNN and global MViT features representation for enhanced feature discriminability.

CNN and MViT architectures. Fusion presents a range of compelling benefits. Firstly, it enables the framework to capture comprehensive local and global information. MViTs capture high-level semantics and global context, allowing them to effectively grasp plant disease images' characteristics. In contrast, CNNs can capture the images' local patterns and intricate details. By combining their strengths, the framework achieves a more holistic understanding of visual content, enhancing the representation capacity. Secondly, this fusion improves the robustness

and adaptability of the model. CNNs provide resilience against local variations and domain-specific details, while MViTs capture domain-invariant representations. This combination ensures better generalization to unseen target domain samples. Additionally, experimental results demonstrate significant performance gains that validate the performance of the proposed framework in unsupervised domain adaptation tasks. The proposed framework outperforms existing methods by employing fusion to enhance discrimination and adaptability. Hence, the fusion of MViTs and CNNs yields a powerful approach that effectively addresses the limitations of each technique, leading to superior performance and improved domain adaptation capabilities.

3.2.1. MobileViT block

The MobileViT block replaces localized convolutions with global processing using MViT, minimizing the number of parameters, as shown in Fig. 4. In a formal sense, the MobileViT algorithm processes the input tensor as $X \in \mathbb{R}^{H\times W\times C}$, where X belongs to the set of real numbers. It applies a standard size $(n\times n)$ convolutional layer followed by a point-wise convolutional layer of size (1×1), resulting in an output tensor, $XL \in \mathbb{R}^{H\times W\times d}$, where d is a higher dimension than C. The convolutional layer with dimensions $(n \times n)$ captures spatial details of a local context. On the other hand, the point-wise convolutional layer enhances the tensor by acquiring linear combinations of the input channels, thereby expanding it into a higher-dimensional space. MobileViT unfolds the XL tensor into N non-overlapping flattened patches, enabling global representation learning with spatial inductive bias denoted as $(XU \in \mathbb{R}^{P\times N\times d})$, where $(P = wh)$ and $(N = \frac{HW}{P})$. The relationships between patches are encoded using transformers to obtain $(XG \in \mathbb{R}^{P\times N\times d})$as expressed in Eq. (1):

$$XG(p) = \text{Transformer}(XU(p)), \quad 1 \le p \le P \tag{1}$$

$$Attention = Softmax(Q \cdot K) \tag{2}$$

$$\begin{aligned} \text{Q} &= \text{Seq}(\text{Conv}(W_q, X)) \\ \text{K} &= \text{Seq}(\text{Conv}(W_k, X)) \\ \text{V} &= \text{Seq}(\text{Conv}(W_v, X)) \end{aligned} \tag{3}$$

$$X_{patch} = \text{Img}(Attention \cdot V) \tag{4}$$

The MobileViT architecture preserves both the patch order and spatial order of pixels within each patch, allowing the tensor $X_G \in \mathbb{R}^{P\times N\times d}$ to be folded into $X_F \in \mathbb{R}^{H\times W\times d}$. X_F is then projected into a lower-dimensional space of dimension C using a point-wise convolution and combined with the original tensor X through concatenation. The fused features are further processed by an additional $n\times n$ convolutional layer. Moreover, integrating convolutions and transformers in the MobileViT block allows transformers to function as convolutions. This design facilitates the efficient implementation of convolutions and transformers, making the MobileViT architecture suitable for deployment on various devices without additional effort. The Fig. 6 illustrates an example of a grid consisting of nine rows and nine columns, where each cell represents a patch containing a 3×3 grid of pixels. Within the MobileViT block, every pixel, such as the red pixel in this example, attends to other pixels, specifically the green pixels located at corresponding central positions in different patches. The transformers can achieve this attention mechanism. Since the green pixels have already incorporated information about neighboring pixels through convolutions, as shown in the top-right patch, the red pixel can effectively encode information from all pixels in the plant disease image. The blue and yellow grids in the figure represent patches and individual pixels.

The advantages of using MobileViT include the preservation of patch and spatial order, an effective receptive field size of $H\times W$, seamless integration with other components, and straightforward implementation of convolutions and transformers. These advantages enable the MobileViT block to combine and use the benefits of both convolutions and transformers for plant disease classification.

3.2.2. CNNs segment:

We adopt the feature pyramid structure in CNNs for domain adaptation. This structure entails decreasing the resolution of feature maps while increasing the channel number with network depth. A bottleneck module is employed, which includes a 1×1 down-sampling convolution. The bottleneck consists of a 3×3 spatial convolution, a 1×1 up-sampling convolution, and a residual connection that connects the input and output. We devise a concurrent network structure to leverage local and global representations simultaneously. In MViT fused convolution, we exploit the complementary nature of the two style features by continuously feeding back the global background from the transformer branch to the feature map, enhancing the CNN branch's global perception ability. Similarly, the CNN branch gradually returns the local features to the patch embedding, enhancing the local details of the transformer branch. This iterative process facilitates effective interaction and fusion, significantly enhancing both the local details of global features and the global perception of local features.

MViT fused convolutions comprise a stem module, dual branches, and two classifiers (FC layers) for the dual branches. The stem module performs a 7×7 convolution with a stride of 2 followed by 3×3 max pooling with a stride of 2 to extract initial local features (such as boundary and texture information). The dual branch receives these features as input, with the CNN and transformer branches consisting of convolution and transformer blocks, respectively. This concurrent architecture guarantees each branch's optimal preservation of local features and global representations. As a bridging module, MViT fused convolutions integrate the CNN branch's local features with the transformer's global representations, ensuring seamless fusion.

3.2.3. Addressing class imbalance with NFRegNet

The PlantVillage and PlantDoc are highly imbalanced datasets (Sunil et al., 2023b; Sharma and Sharma, 2024; Chung, 2024; Yao et al., 2024). This research investigated normalizer-free architectures and took inspiration (Brock et al., 2021) to overcome the class imbalance issue. For example, When all samples in a batch belong to the same class, it creates class imbalance and can pose challenges during the early stages of training. Batch normalization is ineffective in such cases and requires a larger batch size. Homogeneous batches result in highly similar features, causing instability in batch normalization and potentially leading to training failure. To overcome these issues, we use an NFRegNet (Normalizer-Free Regularized Network) based image classification model that diverges from conventional normalization techniques like batch normalization. Instead, NFRegNet employs L2 regularization methods to address class imbalance and batch similarity concerns, improving accuracy and robustness.

By adding this regularization term to the loss function during training, RegNet encourages smaller weights and helps prevent overfitting, promoting better generalization to unseen data. RegNet models balance accuracy and efficiency in computer vision applications through regularization techniques as explained below:

1. L1 Regularization (Loss regularization): L1 regularization adds the weights' absolute values to the loss function, which promotes sparsity in the model. It may influence feature selection and lessen the significance of less essential features.

$$L1_R = Loss + \lambda * \|W\|^1 \tag{5}$$

Here, $\|w\|^1$ is the L1 norm of the weight vector w, and λ is the regularization parameter controlling regularization strength.

2. L2 Regularization (Ridge Regularization): The squared magnitudes of the weights are added to the loss function using L2 regularization. Smoother models are promoted by encouraging smaller weights and preventing significant weight values.

$$L2_R = Loss + \lambda * \|W\|^2 \tag{6}$$

Here, λ is the regularization parameter and $\|w\|^2$ denotes the weight vector $wL2$ norm (Euclidean norm).

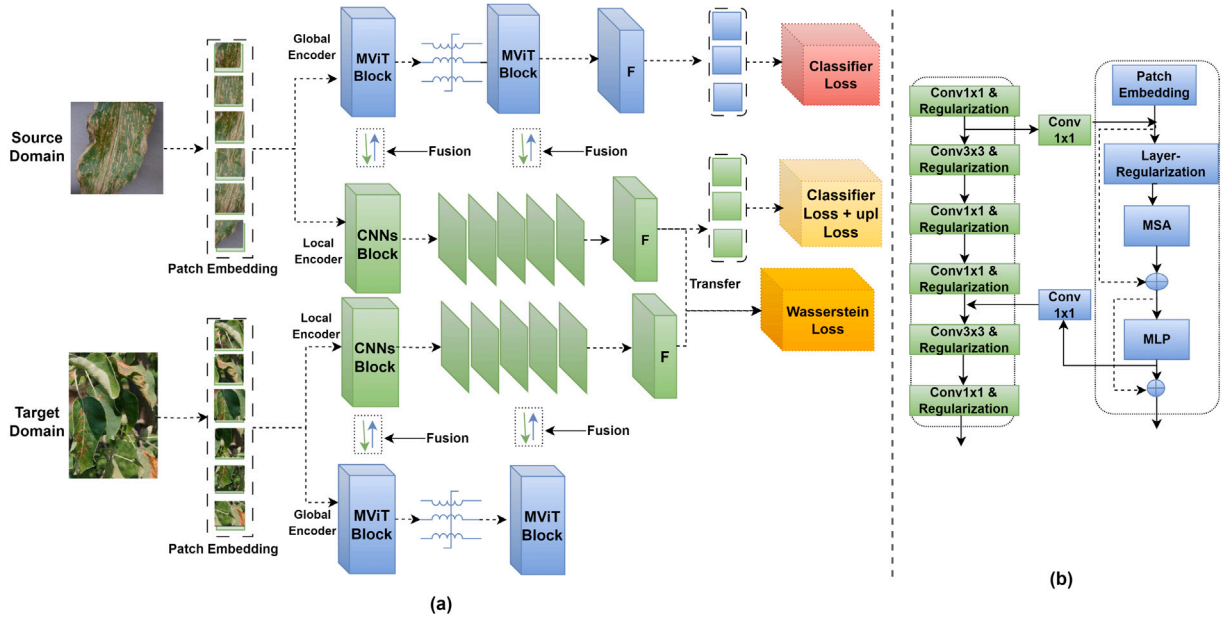


Fig. 5. MViT and CNNs based domain adoption framework. upl shows uncertainty-penalty-loss that can mitigate the uncertainty of the classifier. Where F represents the extracted features.

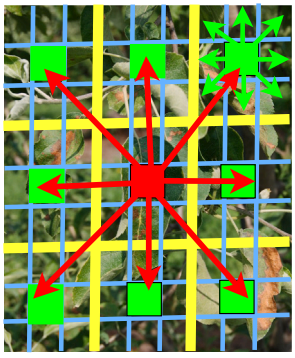


Fig. 6. In MobileVit: Each pixel looks other pixels in plant disease image.

3. Elastic Net Regularization: L1 and L2 regularization are combined in Elastic Net to maximize the advantages of both approaches. It promotes the clustering of associated traits as well as sparsity.

$$EN_R = Loss + \lambda_1 * \|W\|^1 + \lambda_2 * \|W\|^2 \quad (7)$$

Here, the L1 norm is represented by $\|W\|^1$, the L2 norm by $\|W\|^1$, and the regularization parameters λ_1 and λ_2 govern the intensities of the corresponding regularizations.

3.3. Wasserstein distance guided (domain adaptation)

In the context of UDA, we examine two domains: the source domain $S = (x_i^s, y_i^s)_{i=1}^{N_s}$ where $y_i^s \in (1, 2..N)$ denotes the source label of the corresponding samples x_i^s, and the target domain $T = (x_j^t)_{j=1}^{N_t}$, which comprises unlabeled N^t samples. Both domains acknowledge a shared identical label space but have distinct conditional distributions, $\mathbb{P}_{x^s}$ and $\mathbb{P}_{x^t}$, respectively. The objective is to train a transferable classifier $\eta(x)$ that minimizes the target risk by applying the Transformer Fused Convolution (TFC) as shown in Fig. 5 $\in_t$, representing the probability of encountering a mismatch between $\eta(x)$ and the true label y in the target domain D_t. This objective is accomplished by using all the provided data.

In most Domain Adaptations, CNN features are aligned by considering their distributions, encompassing the entire image feature. The CNN-based methods capture only local features. Sometimes, it may consist of irrelevant information (Ji et al., 2021), such as cluttered background details, and employing only CNN-based feature distribution alignment leads to a relatively high misclassification rate, causing adverse effects. Recently, transformer-based methods have gained increasing attention. These methods utilize transformers to capture global features, yet they overlook the finer details of local features. A more effective approach involves combining the strengths of both CNN and transformer techniques (Peng et al., 2021). We propose a novel approach to address the challenge of unsupervised domain adaptation arising from differing data distributions between two domains. Our method focuses on learning feature representations that remain invariant to domain changes. We accomplish this by employing adversarial training and minimizing the empirical Wasserstein distance between the representations of the source and target domains. In this research, we observe that the Kullback–Leibler (KL) (Kashyap et al., 2020) divergence between the target domain and the generated data from the source domain tends to be smaller for modes that are more prevalent in the target domain but less common in the source domain. However, considering the complex and diverse nature of the plant leaf data in our field environment, we find that the Wasserstein distance offers better performance. By utilizing the Wasserstein distance, we can obtain more diverse target features that closely resemble the reality of the target domain.

Our adversarial representation learning includes a neural network-based feature extractor. This feature extractor aims to learn domain-invariant feature representations from both domains. For any instance $x \in \mathbb{R}^m$ originating from either domain, our feature extractor is trained to learn a function $f_g : \mathbb{R}^m \Rightarrow \mathbb{R}^d$, which maps the instance to a d-dimensional representation using the network parameter θg. We employ the domain critic technique proposed by Arjovsky et al. (2017) to minimize the disparity between the two domains. The main objective of the domain critic is to estimate the Wasserstein distance between the distributions of the source and target representations. We compute the feature representation $h = f_g(x)$ using the feature extractor to accomplish this. The domain critic then learns $f_w : \mathbb{R}^d \Rightarrow \mathbb{R}$, parameterized by θ_w, which maps the feature representation to a real number. Finally, the Wasserstein distance between the representation

distributions $\mathbb{P}_{h^s}$ and $\mathbb{P}_{h^t}$, here $h^s = f_g(x^s)$ and $h^s = f_g(x^t)$ computed as:

$$W_1(\mathbb{P}_{h^s}, \mathbb{P}_{h^t}) = sup_{\|F_w\|_L \le 1} \mathbb{E}_{P_{h^s}}[f_w(h)] - \mathbb{E}_{P_{h^t}}[f_w(h)] = sup_{\|F_w\|_L \le 1} \mathbb{E}_{P_{x^t}}[f_w(f_g(x))] - \mathbb{E}_{P_{x^t}}[f_w(f_g(x))] \quad (8)$$

When the parameterized family of domain critic functions f_w satisfies the 1-Lipschitz condition, we effectively estimated the empirical Wasserstein distance by optimizing the domain critic loss function L_{wd} concerning the parameter θ_w.

$$L_{wd}(x^s, x^t) = \frac{1}{N^s} \sum_{x^s \in X^s} f_w(f_g(x^s)) - \frac{1}{N^s} \sum_{x^t \in X^t} f_w(f_g(x^t)) \quad (9)$$

The issue of enforcing the Lipschitz constraint arises here. Arjovsky et al. (2017) proposed weight clipping, but Gulrajani et al. (2017) suggested a more reasonable approach: using a gradient penalty L_{grad} concerning the domain critic parameter θ_w, to address the problem of capacity under-use and gradient vanishing problems.

$$L_{grad}(\hat{h}) = (\|\nabla_{\hat{h}} f_w(\hat{h})\|_2 - 1)^2 \quad (10)$$

To ensure effective gradient penalization, we broaden the scope of feature representations $\hat{h}$ to encompass random points along the direct path connecting source and target representation pairs. This expanded definition permits us to estimate the empirical Wasserstein distance by resolving the problem.

$$Max_{\theta_w}[L_{wd} - \gamma L_{grad}] \quad (11)$$

Here, γ represents the balancing coefficient. Given that the Wasserstein distance is differentiable and continuous in most cases, our approach involves initially training the domain critic until it reaches optimal performance. Subsequently, the feature extractor network can learn feature representations that mitigate domain discrepancies by fixing the domain critic's optimal parameter values and minimizing the Wasserstein distance estimator. At this point, solving the minimax problem accomplished the representation learning process as:

$$Min_{\theta_g} Max_{\theta_w}[L_{wd} - \gamma L_{grad}] \quad (12)$$

During the optimization of the minimum operation, it is important to set γ to 0 to prevent the gradient penalty from influencing the process of learning feature representations. By iteratively improving the feature representations to reduce the Wasserstein distance, the adversarial objective eventually facilitates the learning of feature representations invariant to domain variations.

Combining With Discriminator As mentioned earlier, our ultimate objective is to develop a high-performance classifier for the target domain. However, the unsupervised nature of Wasserstein distance-informed representation learning (WDIRL) might lead to less discriminative domain-invariant representations. To address this, we incorporate supervision signals from the source domain data, similar to DANN (Ganin et al., 2016), into the representation learning process. Algorithm 1 provides a detailed description of this combination. The discriminator is added as an additional layer after the feature extractor network. WDIRL ensures the transferability of learned representations, so the shared discriminator can be directly utilized for target domain prediction once training is complete. The objective of the discriminator, denoted as $f_c : \mathbb{R}^d \Rightarrow \mathbb{R}^l$, is to compute softmax predictions with parameter θ_c, where l represents the number of classes. Given the labeled source data, the discriminator loss function is defined as the cross-entropy between the predicted probability distribution and the one-hot encoding of the class labels.

$$L_c(x^s, y^s) = -\frac{1}{N^s} \sum_{i=1}^{N^s} \sum_{k=1}^{l} 1(y_i^s = k) log f_c(f_g(x_i^s))_k \quad (13)$$

The indicator function, denoted as $1(y_i^s = k)$, determines whether y_i^s is equal to k. The term $f_c(f_g(x_i^s))_k$ represents the kth dimension value of the distribution $f_c(f_g(x_i^s))$.

Algorithm 1: Wasserstein Distance Informed Algorithm Combined with Discriminator

Input: Laboratory dataset (source domain) X^s, Real environment dataset (target domain) X^t
Output: Trained model for plant disease classification in the target domain
Initialize the feature extractor network F and the classifier network C;
Initialize the discriminator network D;
Initialize hyperparameters: learning rate, batch size, number of training epochs, etc.;
for *each training epoch* **do**
  for *each batch in the laboratory dataset* **do**
    Sample a minibatch of source domain samples: $\mathcal{X}^s \sim$ Laboratory Dataset;
    Compute the Wasserstein distance between the source and target domain feature distributions: $WD(\mathcal{X}^s, \mathcal{X}^t)$;
    Update the feature extractor network F and the classifier network C using the source domain samples and the Wasserstein distance: $F, C \leftarrow$ Update$(F, C, \mathcal{X}_s, WD(\mathcal{X}^s, \mathcal{X}^t))$;
    Update the discriminator network D using the source and the target domains samples: $D \leftarrow$ Update$(D, \mathcal{X}^s, \mathcal{X}^t)$;
  end
  for *each batch in the real environment dataset* **do**
    Sample a batch of target domain samples: $\mathcal{X}^t \sim$ Real Environment Dataset;
    Update the feature extractor network F and the classifier network C using the target domain samples and the Wasserstein distance: $F, C \leftarrow$ Update$(F, C, \mathcal{X}^t, WD(\mathcal{X}^s, \mathcal{X}^t))$;
    Update the discriminator network D using the source and the target domains samples: $D \leftarrow$ Update$(D, \mathcal{X}^s, \mathcal{X}^t)$;
  end
end
return Trained model F and C for plant disease classification in the target domain;

Uncertainty Penalty Loss (UPL) Previous studies on transfer learning (Tzeng et al., 2017; Xia et al., 2021; Liu et al., 2021) have primarily aimed to enhance feature transferability by developing various strategies for aligning features. However, a significant drawback of these approaches is their tendency to neglect the distinguishability of features, relying solely on conventional cross-entropy loss for feature learning in the annotated source domain. Consequently, even if the issue of feature shift is successfully addressed, the classifier's performance on the target domain may still be suboptimal. For instance, consider two scenarios where the source domain outputs are [0.5, 0.4, 0.3] and [0.5, 0.35, 0.35], respectively. Although they yield the same cross-entropy loss, the first [0.5, 0.4, 0.3] exhibits higher uncertainty than the second [0.5, 0.35, 0.35]. Based on this, we speculate that by neutralizing the influences of incorrect predictions, neural networks can achieve improved generalization. This is because the presence of a highly predicted incorrect class challenging the correct class is less likely. To address this, we propose utilizing an uncertainty penalty loss, which focuses on enhancing the model's likelihood of predicting the correct class accurately (Chen et al., 2019). The mathematical formulation is presented as follows:

$$L_{upl} = -\frac{1}{N_s} \sum_{i=1}^{N_s} \sum_{j=1, j \neq g}^{C_s} \left(\frac{\hat{Y}_{ij}}{1 - \hat{Y}_{ig}}\right) \log\left(\frac{\hat{Y}_{ij}}{1 - \hat{Y}_{ig}}\right) \quad (14)$$

Here, g represents the correct class within the source domain, while N_s denotes the total number of samples. The rationale behind this design is to introduce an averaging technique that diminishes the impact

Table 2
Experimental Settings.

Parameter	Value
Training Dataset Size	7160+7175(PVD+AC-PD)=14335 images
Validation Dataset Size	1281 (AC-PD) images
Test Dataset Size	452 (Original PlantDoc) images
Preprocessing	Resize to 224 × 224 pixels
Data Augmentation	Dynamic horizontal flip, rotation
Backbone Architecture	MobileNetV2
Optimization Algorithm	Adam
Learning Rate	0.0003
Batch Size	32
Number of Epochs	50

of incorrect predictions. Consequently, this approach minimizes the prediction probability associated with incorrect classes while maximizing the correct classes' prediction probability. By incorporating the discriminator and upl loss and combining these elements, we achieve our final objective function.

$$Min_{\theta_g,\theta_c}[L_c - \lambda Max_{\theta_w}[L_{wd} - \gamma L_{grad} + \alpha Lupl]] \tag{15}$$

The coefficient λ balances discriminative and transferable feature learning. Setting γ to 0 during optimization maintains the integrity of the representation learning process. However, α represents the weighting coefficient for the uncertainty penalty loss, allowing the control of its influence on the overall loss.

3.4. Experimental settings

The Table 2 presents the details of the experimental settings used in our study. These settings define the dataset, training data size, validation data size, test data size, preprocessing steps, data augmentation techniques, backbone architecture, optimization algorithm, learning rate, batch size, and number of epochs. Deciding the number of epochs, We carefully considered model's time complexity per epoch, dataset size, and convergence patterns. We found that 50 epochs struck a reasonable balance between training adequacy and avoiding overfitting, as increasing epochs offered diminishing returns without substantial benefits and imposed excessive computational burdens.

These settings were carefully selected to ensure a comprehensive evaluation of our proposed approach.

3.5. Model evaluation metrics

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}} \tag{16a}$$

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \tag{16b}$$

$$\text{Recall (TP Rate)} = \frac{\text{TP}}{\text{TP} + \text{FN}} \tag{16c}$$

$$\text{Specificity (TN Rate)} = \frac{\text{TN}}{\text{TN} + \text{FP}} \tag{16d}$$

$$\text{F1} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \tag{16e}$$

$$\text{Balanced Accuracy} = \frac{\text{TPR} + \text{TNR}}{2} \tag{16f}$$

$$\text{Cohen's Kappa} = \frac{\text{Po} - \text{Pe}}{1 - \text{Pe}} \tag{16g}$$

$$\text{A\_AD (\%)} = \frac{1}{n}\sum_{i=1}^{n}(\text{A\_Proposed} - \text{A\_SOTA}) \tag{16h}$$

The framework's performance was evaluated using the PlantDoc dataset on the target domains. We employ a comprehensive set of evaluation metrics, including Accuracy, Balanced Accuracy, Macro and Micro mechanisms of Precision, Recall, F1 score, Cohen's Kappa, and time. TP (True Positive) depicts the frequency of truly identified positive instances, and FP (False Positive) depicts the frequency of falsely identified positive instances. TN (True Negative) depicts the frequency of truly identified negative instances and FN (False Negative) depicts the frequency of falsely identified negative instances. Furthermore, TPR (True Positive Rate), also known as Sensitivity or Recall, is calculated as the ratio of TP to the sum of TP and FN, and TNR (True Negative Rate), also known as Specificity, is calculated as the ratio of TN to the sum of TN and FP. These metrics provide insights into the model's performance, enabling us to assess its effectiveness in accurately classifying plant diseases and mitigating class imbalance.

In equation (16 g) P_o is the observed agreement between the model's predictions and the ground truth, while P_e is the expected agreement due to chance. Furthermore, in Eq. (16h), n is the number of SOTA models, A_AD is the average accuracy difference, A_Proposed shows the Accuracy of the proposed approach, and A_SOTA shows the accuracy of the corresponding state-of-the-art study.

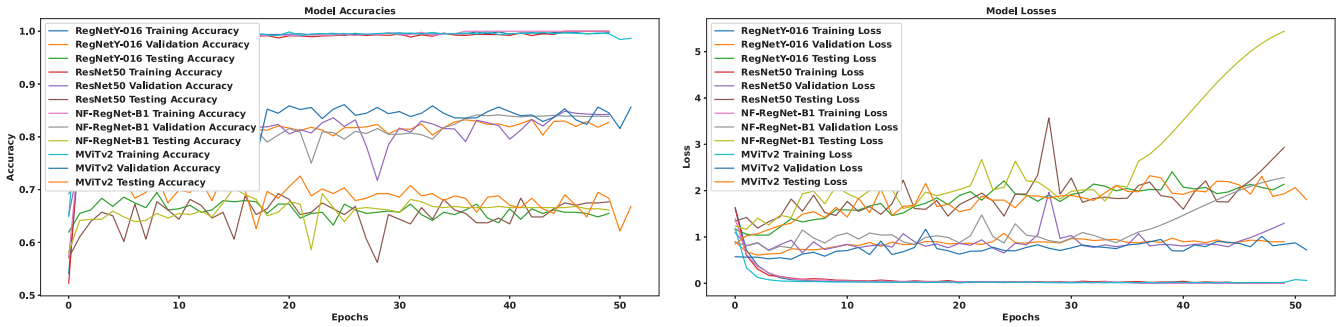
4. Results and discussion

The performance evaluation of different baseline models on the target domain, excluding the Wasserstein distance representation learning, is presented in this section. Table 3 provides a comprehensive overview of the performance of various baseline models and our proposed model. A detailed analysis of the results shows that our proposed framework outperforms the baseline models regarding accuracy and loss metrics. Additionally, Table 4 compares the performance of our proposed framework with the SOTA in the field. The results highlight the significant improvements achieved by our approach. These findings substantiate the efficacy of our proposed model in addressing the challenges of the target domain and advancing the current SOTA.

4.1. Performance analysis of proposed framework with baseline models

We rigorously assess the effectiveness of the proposed framework by conducting and comparing a meticulous performance evaluation against a diverse range of state-of-the-art baseline models. The baseline models included ResNet18 (He et al., 2016), ResNet34 (He et al., 2016), ECAResNet-light (Wang et al., 2020), ReXNet-100 (Han et al., 2020), RegNetY-016 (Radosavovic et al., 2020), NF-ResNet50 (Brock et al., 2021), NF-RegNet-B1 (Brock et al., 2021), MViT-v2 (multi-scale ViT) (Li et al., 2022), and ConViT-small (d'Ascoli et al., 2021). By evaluating their performance on the PlantDoc dataset, which represents the target domain, without adversarial learning techniques, we comprehensively analyze and compare the notable advancements achieved by our proposed framework. The performance evaluation of different baseline models on the target domain, Without the Wasserstein distance representation learning, is presented in this section.

Table 3 depicts the effectiveness in the context of the PD dataset on the target domain. Here, ResNet18 achieves an accuracy of 0.63 and a balanced accuracy of 0.81, showing its capability to classify the data accurately in the field environment. Similarly, ECAResNet-light and RegNetY-016 exhibit competitive results with accuracies of 0.66 and 0.69, respectively, and balanced accuracies above 0.80. These models also demonstrate good precision, recall, and F1 scores in macro and micro pairs, indicating their ability to classify the data using the PD dataset on the target domain. NF-ResNet50 and NF-RegNet-B1 achieve moderate accuracies of 0.61 and 0.66, respectively, but their F1 scores suggest room for improvement. The MViT-v2 without adversarial shows an accuracy of 0.68 and a balanced accuracy of 0.84. Despite all this, there is room for improvement in all paired metrics. Furthermore, ConViT-small demonstrates more promising results than all other baseline models, with an accuracy of 0.69 and a balanced accuracy of 0.84. These baseline architectures are strong references for performance comparison in the proposed research paper.



(a) Accuracy (b) Loss

Fig. 7. Training, validation, and test accuracy and loss of various baseline models.

Table 3
Performance Evaluation of various baseline Architectures without an adversarial and proposed model on target domain using PD Datasets. The Precision, Recall, and F1 results are presented in (Macro and Micro) pairs.

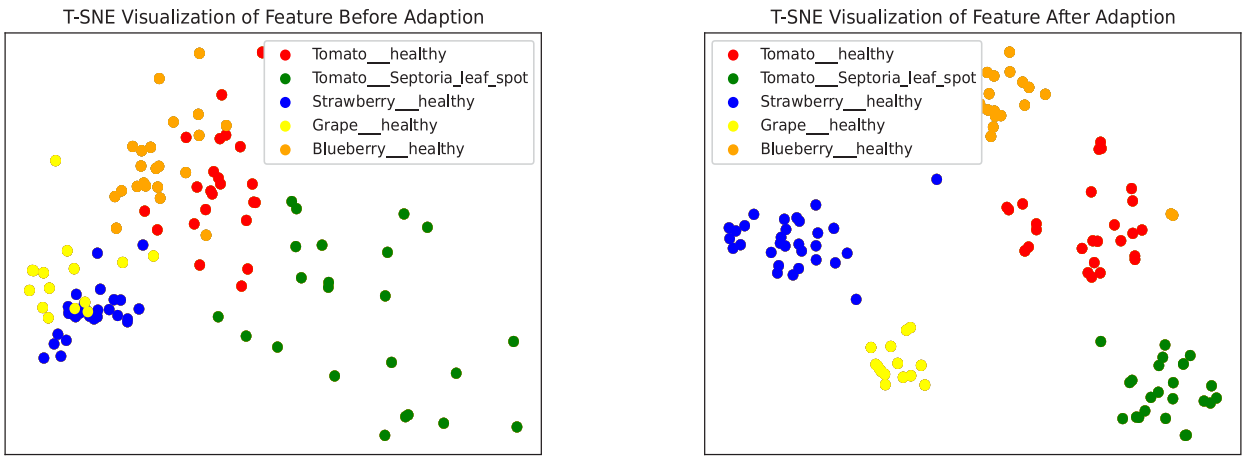
Model-Ref.	Acc	Bal Acc	Precision	Recall	F1	Cohen's Kappa	Time
ResNet18 (He et al., 2016)	0.63	0.81	(0.64,0.63)	(0.63, 0.63)	(0.61, 0,63)	0.61	82.1
ResNet34 (He et al., 2016)	0.60	0.79	(0.63, 0.60)	(0.60, 0.60)	(0.58, 0.60)	0.58	79.7
ECAResNet-light (Wang et al., 2020)	0.66	0.81	(0.65, 0.66)	(0.64, 0.66)	(0.63, 0.66)	0.65	77.8
ReXNet-100 (Han et al., 2020)	0.64	0.81	(0.64, 0.64)	(0.64, 0.64)	(0.62, 0.64)	0.62	81.3
RegNetY-016 (Radosavovic et al., 2020)	0.69	0.83	(0.67, 0.69)	(0.67, 0.69)	(0.66, 0.69)	0.68	78.9
NF-ResNet50 (Brock et al., 2021)	0.61	0.80	(0.60, 0.61)	(0.62, 0.61)	(0.57, 0.61)	0.59	77.3
NF-RegNet-B1 (Brock et al., 2021)	0.66	0.81	(0.65, 0.66)	(0.64, 0.66)	(0.63, 0.66)	0.64	76.2
MViT-v2 (multi-scale ViT) (Li et al., 2022)	0.68	0.83	(0.68, 0.68)	(0.68, 0.68)	(0.66, 0.68)	66	80.28
ConViT-small (d'Ascoli et al., 2021)	0.69	0.84	(0.69, 0.68)	(0.69, 0.70)	(0.68, 0.70)	68	74.4
Proposed Model	**0.70**	**0.84**	(0.68, 0.70)	(0.69, 0.70)	(0.67, 0.70)	**69**	**29.27**

Table 4
Performance Comparison of Proposed Model with current State-of-the-art (SOTA) Models. [The last column displays the accuracy difference between the Proposed and SOTA models. The last value in the column represents the average increase in accuracy achieved by the proposed model].

Ref.study & YoP	Datset	Backbone	Accuracy (%)	Accuracy difference (%)
Singh et al. (2020)	PlantDoc	InceptionResNet V2	29.73	40.27
Menon et al. (2021)	PlantDoc	MobileNetV2	66.74	3.26
Sama et al. (2023)	PlantDoc	ViT small_8	66.74	3.26
Wu et al. (2023)	PlantDoc	MSUN	56.06	13.94
Ziba et al. (2023)	PlantDoc	MobileNetV2	57.55	12.45
Moupojou et al. (2023)	PlantDoc	MobileNet	60.14	9.86
Proposed Model	PlantDoc	MobileNetV2	**70.0**	**13.67**

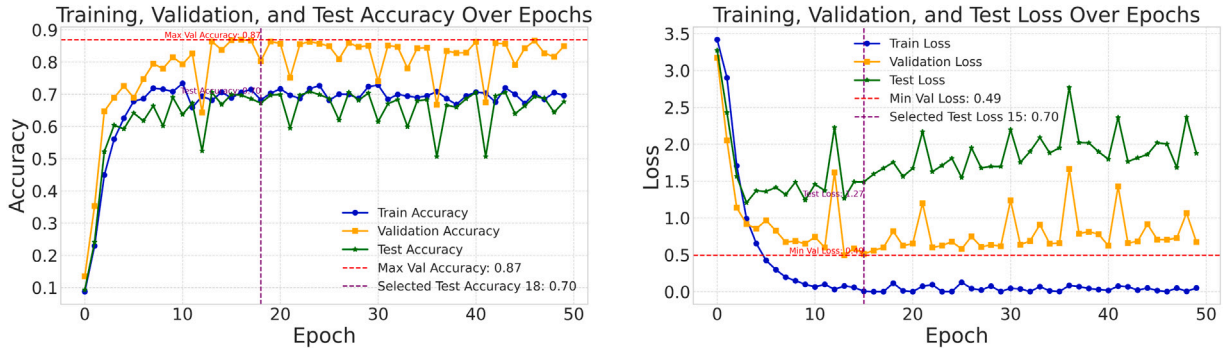
The proposed model in Table 3 exhibits notable performance compared to the baseline architectures. The proposed model outperforms several baselines with an accuracy of 0.70 and a balanced accuracy of 0.84. Its precision, recall, and F1 scores in both macro and micro pairs (0.68, 0.70), (0.69, 0.70), and (0.67, 0.70), respectively, illustrate its ability to achieve high performance and capture class-specific characteristics effectively for field environment plant disease dataset. Furthermore, Cohen's Kappa coefficient of 0.69 indicates substantial agreement between the predicted and true labels. An important advantage of the proposed model is its significantly reduced time of 29.27. However, the traditional approaches for handling class imbalance during model training involve batch normalization, which requires computing the mean and standard deviation of each feature over the mini-batch, which is time-consuming, especially when dealing with imbalanced classes in the dataset, where the minority class has significantly fewer samples, each mini-batch likely contain a small number of minority class instances. This results in the mean and standard deviation being dominated by the majority class, leading to ineffective and long computational normalization for the minority class.

To avoid the high computational cost and effectively handle the class imbalanced issue, we used NFRegNet, which does not rely on batch normalization but offers an alternative regularization strategy. Using NFRegNet, the instability caused by homogeneous batches resulting from class imbalance was mitigated in our model by assigning implicit weights to samples based on their class distribution. This effectively provides a form of adaptive normalization that addresses class imbalance without requiring explicit batch normalization, which requires high computations. Furthermore, a series of figures are included in this section to enhance the understanding of the baseline models' performance and our proposed framework. Fig. 7(a) and (b) presents a comprehensive visualization of various baseline models' accuracy and loss metrics. These figures allow for a detailed illustration of the baseline models' training, validation, and test accuracy throughout training epochs. We observed that all models achieved an increasing trend in accuracy as the training progressed. Notably, ResNet18 and ResNet34 showed comparable performance, with a maximum validation accuracy of 0.78, and exhibited loss patterns with a minimum validation loss of 0.83 and 0.76, respectively. However, ECAResNet-light models performed reasonably well, achieving maximum validation accuracies of 0.83 and 0.63 minimum validation loss. Among the models tested, MViT-v2 achieved the highest accuracy, with a maximum validation accuracy of 0.86 and a minimum validation loss of 0.52. NF-RegNet-B1 and ReXNet-100 also demonstrated strong performance, reaching maximum validation accuracies of 0.84 and 0.81, respectively, and minimum validation loss of 0.71 and 0.78, respectively. The remaining models, RegNetY-016, NF-ResNet50, and ConViT-small, achieved maximum validation accuracies ranging from 0.81 to 0.85 and minimum validation losses from 0.65 to 0.61, respectively.



(a) Features Before Fusion and Adaptation (b) Features After Fusion and Adaptation

Fig. 8. Feature visualization before and after fusion and adaptation.



(a) Proposed Model Accuracy (b) Proposed Framework Loss

Fig. 9. Training, validation, and test accuracy and losses of proposed framework.

Moreover, Fig. 9 provides a focused analysis of our proposed model's maximum validation accuracy of 0.87 and minimum validation loss of 0.49, showing superior performance of our approach. Fig. 10 provides a visual representation of the impact of domain adaptation on our model's performance. It demonstrates the significant improvement achieved through the incorporation of the fusion, handling the class imbalance, and crucial aspects of the Wasserstein algorithm-based technique. The figure showcases that the model exhibits a significant reduction in loss values over the training epochs, indicating enhanced convergence and improved predictive capabilities. This reduction in loss values highlights the effectiveness of domain adaptation in aligning feature distributions and reducing the domain shift. Consequently, the model becomes more adept at capturing the underlying patterns and characteristics of the target domain, resulting in improved accuracy. The resulting image visually reinforces the positive impact of domain adaptation, highlighting its crucial role in enhancing the performance of our model.

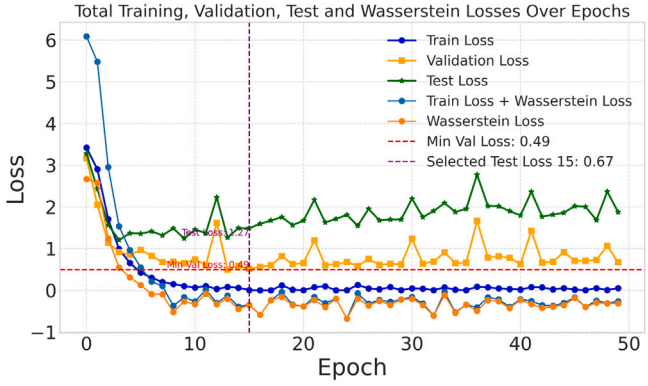


Fig. 10. Impact of domain adaptation.

4.2. t-SNE feature visualization before and after adaptation

To further validate the effectiveness of the proposed framework, we employed t-distributed stochastic neighbor embedding (t-SNE) as a visualization method using the subset of the total training set, as depicted in Figs. 8(a) and 8(b). In Fig. 8(a), the t-SNE plot demonstrates the impact of domain adaptation and the utilization impact of future fusion techniques on feature representation. The scattered and mixed distribution of points from different domains and subsets of classes reveals a lack of alignment and transferability in the extracted features. Additionally, the accuracy and loss plots from Fig. 7 indicate that the baseline method, which directly transfers from the source to the target domain without domain adaptation, fails to achieve effective performance due to domain shift. The feature distributions of the two domains and the intra-domain class features are not distinguished. Although existing methods perform better, they still struggle with the

ambiguous boundary between different feature classes and the scattered distribution within the same feature class, leading to inaccurate classification. This limitation arises from their failure to address the extensive domain divergence between laboratory and field images, which is unique to our plant disease task.

However, in Fig. 8(b), the t-SNE plot exhibits a more coherent and clustered distribution of points after domain adaptation and the incorporation of future fusion techniques. The improved alignment of extracted features across domains and subsets of classes is evident through the distinct colors representing different domains and subsets. This demonstrates the effectiveness of our domain adaptation approach and feature fusion in bridging the gap between domains, enhancing the transferability of features, and enabling better generalization across diverse datasets and subsets of classes within the utilized dataset. In contrast to previous methods, our proposed framework not only aligns the features of each class between the source and target domains but also distinguishes the features of different classes from the target domain by aligning the subdomains, i.e., the embedded features belonging to each class. Consequently, our method effectively addresses the significant domain shift observed in our plant disease task.

4.3. Performance comparison of proposed model with SOTA

The performance comparison presented in Table 4 showcases the superiority of the proposed model when compared to the current state-of-the-art (SOTA) models in the domain. The **Ref.Study & YoP** column shows each study's reference and year of publication. The **Dataset** column indicates the utilization of the PlantDoc dataset by each study for the classification task, employing different backbone architectures. The **Accuracy (%)** column showcases the performance achieved by each previous study. The **Accuracy Difference (%)** column quantifies the improvement in accuracy achieved by our proposed approach over the referenced SOTA study. It indicates the performance gain obtained by our methodology compared to the previous sOTA, the last value of the column representing the **Average Accuracy Difference** achieved by the proposed model. We believe this column reinforces the significance and impact of our research by quantifying the improvement achieved over the SOTA. Singh et al. (2020) employed InceptionResNet V2 as the backbone model and achieved an accuracy of 29.73%. Menon et al. (2021) utilized MobileNetV2, a popular lightweight architecture, and achieved a notable improvement with an accuracy of 66.74%. Similarly, Sama et al. (2023) adopted the Vision Transformer (ViT) architecture with a small_8 configuration, also achieving an accuracy of 66.74%. Wu et al. (2023) utilized the MSUN as the backbone model and obtained an accuracy of 56.06%. Ziba et al. (2023) applied MobileNetV2 as backbone achieving an accuracy of 57.55%. Moupojou et al. (2023) utilized MobileNet and obtained an accuracy of 60.14%. These studies left room to achieve the desired accuracy on the PlantDoc dataset, emphasizing the importance of advancing robust classification capabilities for the field environment datasets.

The proposed model outperforms these SOTA models by a significant margin, attaining an impressive accuracy of 70.0% and 13.67% average accuracy difference on the PlantDoc dataset. The proposed model, utilizing MobileNetV2 as the backbone architecture, demonstrates its capability to use the network's depth and width to effectively capture and learn intricate features from plant disease images. Our proposed framework for plant disease classification demonstrates exceptional performance, surpassing previous benchmarks and establishing a new standard for accuracy in the field.

This significant advancement can be attributed to several key strengths of our approach. Firstly, the fusion of CNNs and MViTs enhances our model's accuracy and utilizes their respective strengths. This combination enables our framework to capture local and global information, improving model performance. Secondly, we address the class imbalance issue in the training dataset by employing NFRegNet, which mitigates the impact of imbalanced classes during model training. By using NFRegNet, we overcome the instability caused by homogeneous batches resulting from class imbalance, enhancing the stability and effectiveness of training. Additionally, we incorporate the Wasserstein algorithm-based domain adaptation technique, which focuses on learning domain-invariant feature representations by minimizing the empirical Wasserstein distance between the source and target domains' representations. This adaptation process leads to improved accuracy by obtaining target features that closely resemble the characteristics of the target domain. These strengths collectively contribute to the superior performance of our proposed framework in plant disease classification.

5. Conclusion

In conclusion, this research study introduces a novel and robust framework for plant disease classification that addresses the limitations of existing approaches trained on laboratory datasets, which struggle to achieve high accuracy on real-world environment datasets. By addressing domain shift through unsupervised domain adaptation (UDA), our framework learns domain-invariant feature representations by utilizing Wasserstein distance adversarial learning and combining the strengths of convolutional neural networks (CNNs) and mobile vision transformers (MViTs). This fusion of local and global features enhances the transferability and discriminability within different domains, resulting in more accurate and reliable plant disease recognition. Extensive experiments conducted on plant disease datasets demonstrate the superior performance of the proposed framework with an accuracy of 70.0%, balanced accuracy of 84.0%, precision (micro 70.0%), Recall (micro 70.0%), F1 (micro 70.0%), Cohen's kappa of 69.0%, and outperforming the baseline and state-of-the-art methods with an increase in accuracy of 13.67%. This research contributes to accurate and timely plant disease diagnosis in real environments for sustainable agriculture and food security by offering a promising solution to overcome the challenges faced by traditional methods. However, it is important to acknowledge the limitations of the diverging of training and testing loss and reliance on labeled data for source domain, as well as the need for evaluation on larger real-time field datasets. Considering the dynamic nature of plant diseases and changing environmental conditions, it is necessary to develop adaptive and self-learning systems that continuously update their knowledge and adapt to new challenges. Future research can focus on IoT, reinforcement learning domain adaptation techniques, further tuning regularization techniques for emerging deep learning architectures, and investigating incremental learning for real-time field environment plant disease detection.

CRediT authorship contribution statement

Muhammad Hanif Tunio: Writing – review & editing, Writing – original draft, Visualization, Validation, Software, Resources, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Jian ping Li:** Supervision, Project administration, Conceptualization. **Xiaoyang Zeng:** Writing – review & editing, Conceptualization. **Awais Ahmed:** Writing – review & editing. **Syed Attique Shah:** Supervision, Writing – review & editing. **Hisam-Uddin Shaikh:** Writing – review & editing. **Ghulam Ali Mallah:** Writing – review & editing. **Imam Abdullahi Yahya:** Writing – review & editing.

Declaration of competing interest

None Declared.

Acknowledgments

The authors acknowledge that this research paper received financial support from Birmingham City University, United Kingdom.

Data availability

Datasets Links are given in the manuscripts Footnotes.

References

Arjovsky, M., Chintala, S., Bottou, L., 2017. Wasserstein generative adversarial networks. In: International Conference on Machine Learning. PMLR, pp. 214–223.

Basavaiah, J., Arlene Anthony, A., 2020. Tomato leaf disease classification using multiple feature extraction techniques. Wirel. Pers. Commun. 115 (1), 633–651.

Batchuluun, G., Nam, S.H., Park, K.R., 2022. Deep learning-based plant classification and crop disease classification by thermal camera. J. King Saud Univ.-Computer Inf. Sci. 34 (10), 10474–10486.

Ben-David, S., Blitzer, J., Crammer, K., Kulesza, A., Pereira, F., Vaughan, J.W., 2010. A theory of learning from different domains. Mach. Learn. 79, 151–175.

Boulent, J., Foucher, S., Théau, J., St-Charles, P.-L., 2019. Convolutional neural networks for the automatic identification of plant diseases. Front. Plant Sci. 10, 941.

Brahimi, M., Boukhalfa, K., Moussaoui, A., 2017. Deep learning for tomato diseases: classification and symptoms visualization. Appl. Artif. Intell. 31 (4), 299–315.

Brock, A., De, S., Smith, S.L., Simonyan, K., 2021. High-performance large-scale image recognition without normalization. In: International Conference on Machine Learning. PMLR, pp. 1059–1071.

Chanda, M., Biswas, M., 2019. Plant disease identification and classification using back-propagation neural network with particle swarm optimization. In: 2019 3rd International Conference on Trends in Electronics and Informatics. ICOEI, IEEE, pp. 1029–1036.

Chen, J., Chen, J., Zhang, D., Sun, Y., Nanehkaran, Y.A., 2020. Using deep transfer learning for image-based plant disease identification. Comput. Electron. Agric. 173, 105393.

Chen, H.-Y., Liang, J.-H., Chang, S.-C., Pan, J.-Y., Chen, Y.-T., Wei, W., Juan, D.-C., 2019. Improving adversarial robustness via guided complement entropy. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 4881–4889.

Chung, B., 2024. Addressing data imbalance in plant disease recognition through contrastive learning. In: 2024 IEEE 3rd International Conference on AI in Cybersecurity. ICAIC, IEEE, pp. 1–6.

d'Ascoli, S., Touvron, H., Leavitt, M.L., Morcos, A.S., Biroli, G., Sagun, L., 2021. Convit: Improving vision transformers with soft convolutional inductive biases. In: International Conference on Machine Learning. PMLR, pp. 2286–2296.

Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., Dehghani, M., Minderer, M., Heigold, G., Gelly, S., et al., 2020. An image is worth 16x16 words: Transformers for image recognition at scale. arXiv preprint arXiv:2010.11929.

Ferentinos, K.P., 2018. Deep learning models for plant disease detection and diagnosis. Comput. Electron. Agric. 145, 311–318.

Fuentes, A., Yoon, S., Park, D.S., 2020. Deep learning-based techniques for plant diseases recognition in real-field scenarios. In: Advanced Concepts for Intelligent Vision Systems: 20th International Conference, ACIVS 2020, Auckland, New Zealand, February 10–14, 2020, Proceedings 20. Springer, pp. 3–14.

Ganin, Y., Lempitsky, V., 2015. Unsupervised domain adaptation by backpropagation. In: International Conference on Machine Learning. PMLR, pp. 1180–1189.

Ganin, Y., Ustinova, E., Ajakan, H., Germain, P., Larochelle, H., Laviolette, F., March, M., Lempitsky, V., 2016. Domain-adversarial training of neural networks. J. Mach. Learn. Res. 17 (59), 1–35.

Ghifary, M., Kleijn, W.B., Zhang, M., Balduzzi, D., Li, W., 2016. Deep reconstruction-classification networks for unsupervised domain adaptation. In: Computer Vision–ECCV 2016: 14th European Conference, Amsterdam, the Netherlands, October 11–14, 2016, Proceedings, Part IV 14. Springer, pp. 597–613.

Gui, P., Dang, W., Zhu, F., Zhao, Q., 2021. Towards automatic field plant disease recognition. Comput. Electron. Agric. 191, 106523.

Gulrajani, I., Ahmed, F., Arjovsky, M., Dumoulin, V., Courville, A.C., 2017. Improved training of wasserstein gans. Adv. Neural Inf. Process. Syst. 30.

Han, D., Yun, S., Heo, B., Yoo, Y., 2020. Rexnet: Diminishing representational bottleneck on convolutional neural network. p. 1, arXiv preprint arXiv:2007.00992, 6.

He, K., Zhang, X., Ren, S., Sun, J., 2016. Deep residual learning for image recognition. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 770–778.

Hossain, E., Hossain, M.F., Rahaman, M.A., 2019. A color and texture based approach for the detection and classification of plant leaf disease using KNN classifier. In: 2019 International Conference on Electrical, Computer and Communication Engineering. ECCE, IEEE, pp. 1–6.

Howard, A.G., Zhu, M., Chen, B., Kalenichenko, D., Wang, W., Weyand, T., Andreetto, M., Adam, H., 2017. Mobilenets: Efficient convolutional neural networks for mobile vision applications. arXiv preprint arXiv:1704.04861.

Hughes, D., Salathé, M., et al., 2015. An open access repository of images on plant health to enable the development of mobile disease diagnostics. arXiv preprint arXiv:1511.08060.

Ji, Y., Zhang, H., Zhang, Z., Liu, M., 2021. CNN-based encoder-decoder networks for salient object detection: A comprehensive review and recent advances. Inform. Sci. 546, 835–857.

Kashyap, A.R., Hazarika, D., Kan, M.-Y., Zimmermann, R., 2020. Domain divergences: A survey and empirical analysis. arXiv preprint arXiv:2010.12198.

Li, Y., Wu, C.-Y., Fan, H., Mangalam, K., Xiong, B., Malik, J., Feichtenhofer, C., 2022. Mvitv2: Improved multiscale vision transformers for classification and detection. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 4804–4814.

Li, Z., Yang, Y., Li, Y., Guo, R., Yang, J., Yue, J., 2020. A solanaceae disease recognition model based on SE-inception. Comput. Electron. Agric. 178, 105792.

Liu, W., Anguelov, D., Erhan, D., Szegedy, C., Reed, S., Fu, C.-Y., Berg, A.C., 2016. Ssd: Single shot multibox detector. In: Computer Vision–ECCV 2016: 14th European Conference, Amsterdam, the Netherlands, October 11–14, 2016, Proceedings, Part I 14. Springer, pp. 21–37.

Liu, X., Li, S., Ge, Y., Ye, P., You, J., Lu, J., 2021. Recursively conditional gaussian for ordinal unsupervised domain adaptation. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 764–773.

Long, M., Cao, Z., Wang, J., Jordan, M.I., 2018. Conditional adversarial domain adaptation. Adv. Neural Inf. Process. Syst. 31.

Long, M., Zhu, H., Wang, J., Jordan, M.I., 2016. Unsupervised domain adaptation with residual transfer networks. Adv. Neural Inf. Process. Syst. 29.

Ma, N., Zhang, X., Zheng, H.-T., Sun, J., 2018. Shufflenet v2: Practical guidelines for efficient cnn architecture design. In: Proceedings of the European Conference on Computer Vision. ECCV, pp. 116–131.

Menon, V., Ashwin, V., Deepa, R.K., 2021. Plant disease detection using cnn and transfer learning. In: 2021 International Conference on Communication, Control and Information Sciences, Vol. 1. ICCISc, IEEE, pp. 1–6.

Mohanty, S.P., Hughes, D.P., Salathé, M., 2016. Using deep learning for image-based plant disease detection. Front. Plant Sci. 7, 1419.

Moupojou, E., Tagne, A., Retraint, F., Tadonkemwa, A., Wilfried, D., Tapamo, H., Nkenlifack, M., 2023. FieldPlant: A dataset of field plant images for plant disease detection and classification with deep learning. IEEE Access 11, 35398–35410.

Mukherjee, S., Kumar, P., Saini, R., Roy, P.P., Dogra, D.P., Kim, B.-G., et al., 2017. Plant disease identification using deep neural networks. J. Multimedia Inf. Syst. 4 (4), 233–238.

Pan, S.J., Yang, Q., 2009. A survey on transfer learning. IEEE Trans. Knowl. Data Eng. 22 (10), 1345–1359.

Peng, Z., Huang, W., Gu, S., Xie, L., Wang, Y., Jiao, J., Ye, Q., 2021. Conformer: Local features coupling global representations for visual recognition. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 367–376.

Pereira, C.S., Morais, R., Reis, M.J., 2019. Deep learning techniques for grape plant species identification in natural images. Sensors 19 (22), 4850.

Radosavovic, I., Kosaraju, R.P., Girshick, R., He, K., Dollár, P., 2020. Designing network design spaces. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 10428–10436.

Raja, M.U., Mukhtar, T., Shaheen, F.A., Bodlah, I., Jamal, A., Fatima, B., Ismail, M., Shah, I., 2018. Climate change and its impact on plant health: a Pakistan's prospective. Plant Protect. 2 (2), 51–56.

Ramcharan, A., Baranowski, K., McCloskey, P., Ahmed, B., Legg, J., Hughes, D.P., 2017. Deep learning for image-based cassava disease detection. Front. Plant Sci. 8, 1852.

Ramcharan, A., McCloskey, P., Baranowski, K., Mbilinyi, N., Mrisho, L., Ndalahwa, M., Legg, J., Hughes, D.P., 2019. A mobile-based deep learning model for cassava disease diagnosis. Front. Plant Sci. 10, 425916.

Ratnasari, E.K., Mentari, M., Dewi, R.K., Ginardi, R.H., 2014. Sugarcane leaf disease detection and severity estimation based on segmented spots image. In: Proceedings of International Conference on Information, Communication Technology and System (ICTS) 2014. IEEE, pp. 93–98.

Sama, N., David, E., Rossetti, S., Antona, A., Franchetti, B., Pirri, F., 2023. A new large dataset and a transfer learning methodology for plant phenotyping in vertical farms. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 540–551.

Sarkar, C., Gupta, D., Gupta, U., Hazarika, B.B., 2023. Leaf disease detection using machine learning and deep learning: Review and challenges. Appl. Soft Comput. 110534.

Sharma, P., Sharma, A., 2024. A novel plant disease diagnosis framework by integrating semi-supervised and ensemble learning. J. Plant Dis. Protect. 131 (1), 177–198.

Singh, D., Jain, N., Jain, P., Kayal, P., Kumawat, S., Batra, N., 2020. PlantDoc: A dataset for visual plant disease detection. In: Proceedings of the 7th ACM IKDD CoDS and 25th COMAD. pp. 249–253.

Sladojevic, S., Arsenovic, M., Anderla, A., Culibrk, D., Stefanovic, D., et al., 2016. Deep neural networks based recognition of plant diseases by leaf image classification. Comput. Intell. Neurosci. 2016.

Sunil, C., Jaidhar, C., Patil, N., 2023a. Systematic study on deep learning-based plant disease detection or classification. Artif. Intell. Rev. 56 (12), 14955–15052.

Sunil, C., Jaidhar, C., Patil, N., 2023b. Tomato plant disease classification using multi-level feature fusion with adaptive channel spatial and pixel attention mechanism. Expert Syst. Appl. 228, 120381.

Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., Erhan, D., Vanhoucke, V., Rabinovich, A., 2015. Going deeper with convolutions. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 1–9.

Tang, Z., Yang, J., Li, Z., Qi, F., 2020. Grape disease image classification based on lightweight convolution neural networks and channelwise attention. Comput. Electron. Agric. 178, 105735.

Tilman, D., Balzer, C., Hill, J., Befort, B.L., 2011. Global food demand and the sustainable intensification of agriculture. Proc. Natl. Acad. Sci. 108 (50), 20260–20264.

Too, E.C., Yujian, L., Njuki, S., Yingchun, L., 2019. A comparative study of fine-tuning deep learning models for plant disease identification. Comput. Electron. Agric. 161, 272–279.

Tunio, M.H., Li, J.P., Zeng, X., Akhtar, F., Shah, S.A., Ahmed, A., Yang, Y., Heyat, M.B.B., 2024. Meta-knowledge guided Bayesian optimization framework for robust crop yield estimation. J. King Saud Univ.-Comput. Inf. Sci. 101895.

Tzeng, E., Hoffman, J., Saenko, K., Darrell, T., 2017. Adversarial discriminative domain adaptation. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 7167–7176.

Vishnoi, V.K., Kumar, K., Kumar, B., 2021. Plant disease detection using computational intelligence and image processing. J. Plant Diseases Protect. 128, 19–53.

Vishnoi, V.K., Kumar, K., Kumar, B., Mohan, S., Khan, A.A., 2022. Detection of apple plant diseases using leaf images through convolutional neural network. IEEE Access 11, 6594–6609.

Wang, X., Li, L., Ye, W., Long, M., Wang, J., 2019. Transferable attention for domain adaptation. In: Proceedings of the AAAI Conference on Artificial Intelligence, Vol. 33, No. 01. pp. 5345–5352.

Wang, D., Wang, J., Ren, Z., Li, W., 2022. DHBP: A dual-stream hierarchical bilinear pooling model for plant disease multi-task classification. Comput. Electron. Agric. 195, 106788.

Wang, Q., Wu, B., Zhu, P., Li, P., Zuo, W., Hu, Q., 2020. ECA-net: Efficient channel attention for deep convolutional neural networks. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 11534–11542.

Wang, S., Zhang, L., 2020. Self-adaptive re-weighted adversarial domain adaptation. arXiv preprint arXiv:2006.00223.

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

Xia, H., Zhao, H., Ding, Z., 2021. Adaptive adversarial network for source-free domain adaptation. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 9010–9019.

Xu, T., Chen, W., Wang, P., Wang, F., Li, H., Jin, R., 2021. Cdtrans: Cross-domain transformer for unsupervised domain adaptation. arXiv preprint arXiv:2109.06165.

Yang, J., Liu, J., Xu, N., Huang, J., 2023. Tvt: Transferable vision transformer for unsupervised domain adaptation. In: Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision. pp. 520–530.

Yang, G., Tang, H., Zhong, Z., Ding, M., Shao, L., Sebe, N., Ricci, E., 2021. Transformer-based source-free domain adaptation. arXiv preprint arXiv:2105.14138.

Yao, J., Tran, S.N., Garg, S., Sawyer, S., 2024. Deep learning for plant identification and disease classification from leaf images: Multi-prediction approaches. ACM Comput. Surv. 56 (6), 1–37.

Yigit, E., Sabanci, K., Toktas, A., Kayabasi, A., 2019. A study on visual features of leaves in plant identification using artificial intelligence techniques. Comput. Electron. Agric. 156, 369–377.

Yu, C., Wang, J., Chen, Y., Huang, M., 2019. Transfer learning with dynamic adversarial adaptation network. In: 2019 IEEE International Conference on Data Mining. ICDM, IEEE, pp. 778–786.

Yuan, L., Chen, Y., Wang, T., Yu, W., Shi, Y., Jiang, Z.-H., Tay, F.E., Feng, J., Yan, S., 2021. Tokens-to-token vit: Training vision transformers from scratch on imagenet. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 558–567.

Zhang, Z., He, X., Sun, X., Guo, L., Wang, J., Wang, F., 2015. Image recognition of maize leaf disease based on GA-SVM. Chem. Eng. Trans. 46, 199–204.

Zhu, Y., Zhuang, F., Wang, J., Ke, G., Chen, J., Bian, J., Xiong, H., He, Q., 2020. Deep subdomain adaptation network for image classification. IEEE Trans. Neural Netw. Learn. Syst. 32 (4), 1713–1722.

Ziba, M., Przewłoka, K., Grela, M., Szkoła, K., Kuta, M., 2023. Image recognition of plants and plant diseases with transfer learning and feature compression. In: International Conference on Computational Science. Springer, pp. 204–211.