To Flash or Not To Flash: A Case Study on Vine Foliar Disease Detection

Robert-Anton Konievic and Levente Tamás
Department of Automation
Technical University of Cluj-Napoca
Cluj-Napoca, 400114, Romania
{Robert.Konievic, Levente.Tamas}@aut.utcluj.ro

Keresztes Barna
Department CNRS, Bordeaux INP, IMS, UMR 5218
University Bordeaux
Talence F-33400, France
barna.keresztes@ims-bordeaux.fr

***Abstract*—This paper investigates the detection of black rot disease in grapevines using both flash and natural outdoor imagery, emphasising real-world field conditions over controlled laboratory environments. Leveraging a dataset captured from an *Unmanned Ground Vehicle* (UGV), the study evaluates the robustness of deep learning models under few-shot learning scenarios using flash and natural light. A hybrid architecture combining a Feature Pyramid Network is employed, with models initially trained on annotated source data and adapted to unlabeled target domains through transfer learning. The results demonstrate that few-shot learning performance in visual disease detection is significantly enhanced when adaptation is guided by a prior model trained on broader symptomatology, underscoring the importance of generalised representations in few-shot conditions.**

***Index Terms*—Visual Disease Detection, Precision Viticulture**

I. Introduction

Precision agriculture is an increasingly evolving field which now benefits from the attention of *Machine Learning* (ML) applications, which aim to boost the yield, quality and sustainability of plants and plantations. The grapevine industry is one of the largest, producing fruits for general consumption or creating beverages [1]. However, grapevines are susceptible to a range of foliar pathogens, such as downy mildew, powdery mildew, and black rot, which manifest visibly on leaves before affecting the entire plant, allowing humans to react and support the plant in overcoming the disease. Hence, early and accurate detection of visual symptoms is crucial for minimising economic losses and reducing excessive agrochemical usage.

Traditional disease identification in vineyards relies on scouting by trained experts [2]. This approach is time-consuming, prone to human error, and not scalable for large-scale or precision agriculture. In recent years, advancements in *Computer Vision* (CV) and ML applications targeting enhancements in agriculture have emerged [3]–[6].

Deploying CV models in real agricultural environments presents a significant challenge, as these models are highly sensitive to illumination, background, or appearance changes. These factors often lead to substantial performance degradation when models trained under controlled conditions are applied to field data. To address this, it is essential to develop models that are sufficiently robust to detect disease symptoms under diverse environmental conditions accurately. One promising solution is the application of *Domain Adaptation* (DA) techniques. More specifically, DA can be employed to achieve better results even with low amounts of data by minimising the domain shift, making it particularly well-suited for agricultural scenarios where labelled data are limited or costly to obtain.

Comparing natural and flash imagery in outdoor data is an emerging subdomain for *Visual Disease Detection* (VDD) in grapevines. Flash [7] based imaging offers a more controlled visual input by minimising some of the most challenging variables in CV, notably illumination changes. By reducing environmental noise, flash images can enhance the stability and reliability of feature extraction, thereby improving the overall robustness of CV models.

The use of *Unmanned Ground Vehicle* (UGVs) and *Unmanned Aerial Vehicles* (UAVs) has become increasingly prominent in precision viticulture for monitoring grapevine health and detecting foliar pathogens [8]. UAVs enable rapid, high-resolution image acquisition over large vineyard areas, providing a non-invasive and scalable alternative to manual inspection. When integrated with computer vision and machine learning models, UAV-based imagery supports real-time, in-situ disease detection, thereby enhancing both the efficiency and accuracy of vineyard management practices.

This study presents a DA approach that improves the precision of symptom detection in grapevine leaves to identify black rot disease, utilising few-shot learning principles. In the broader symptom detection phase, the model was trained to recognise general foliar pathogens, treating all visibly affected leaves as instances of disease without distinguishing between pathogen types. We aim to target only black rot disease to assess which set of data is more reliable between natural and flash camera imagery, using limited labelled data. The main contributions are summarised as follows:

1) We train a base model on a UAV dataset [8] which detects different foliar diseases as one disease, creating a symptom detection model.

This work was supported by the program of the Romanian Authority for Scientific Research, grant number PN-IV-P7-7.1-PTE-2024-0105.

979-8-3315-5896-3/25/$31.00 ©2025 IEEE

2) We describe and implement a novel DA technique to mitigate the effect of domain shift between UAV and UGV data using few-shot principles.
3) We present an evaluation of flash and natural imagery of outdoor grapevine plants from Bordeaux, France, affected only by black rot disease, thereby transitioning our symptom detection model into a disease detection model.

Section II discusses related works in the fields of DA and precision agriculture, section III presents the proposed method with the dataset details and experimental setup, while section IV highlights the results and discusses limitations of our work, concluding with section V.

II. Related Work

Visual inspection techniques constitute the foundation of numerous CV tasks in precision agriculture, such as image matching, object detection [3]–[5], and instance segmentation [9]–[11]. However, while many models demonstrate high accuracy under laboratory or curated benchmark datasets, their performance often degrades when applied to real-world vineyard settings characterised by variability in lighting, occlusion, leaf orientation, and background conditions [11]–[13].

This performance gap is primarily attributed to the problem of domain shift, where the statistical properties of the training data (source domain) differ from those encountered in a new, unobserved environment (target domain). To mitigate this issue, DA [14] techniques are increasingly employed in agricultural vision tasks. DA facilitates the transfer of learned features from a pre-trained model, typically developed on a large-scale dataset, to a new domain with different data distributions. In the context of grapevine disease detection, such techniques are particularly pertinent, as vineyard conditions can vary significantly across regions, seasons, and imaging systems.

By integrating transfer learning with unsupervised [15]–[17] or semi-supervised [18] DA methods, fine-tuning of *convolutional neural networks* (CNNs) can improve model robustness under diverse field conditions. This approach enhances the model's capacity to generalise, making it more applicable for deployment in precision viticulture where real-time, in situ disease recognition is essential.

In the context of flash and natural image pairs, Sun et al. [7] demonstrated that combining these two modalities can significantly aid in image segmentation tasks. Flash illumination enhances the visibility of objects closer to the camera by increasing their exposure, thereby facilitating the separation of foreground elements from the background. This principle showcases our motivation for incorporating flash imagery in grapevine leaf disease detection, as it helps mitigate challenges posed by inconsistent natural lighting conditions. This notion will also be beneficial in our research as well.

A. Domain Adaptation in Precision Agriculture

ML techniques have recently shown promise for certain tasks. For instance, Shubhra and Stavness [5] combined a convolutional network to count leaves with a deconvolutional network for initial segmentation, and Giuffrida et al. [19] created a network that directly estimates leaf counts from unstructured 2D images. Another useful solution to domain shift is dataset augmentation, where Ubbens et al. [18] add rendered 3D synthetic graphics to their training set. Similar to this, *Generative Adversarial Networks* (GANs) have been used to improve fruit recognition in 3D crop modelling by Fei et al. [20], while Roscher et al. [21] synthesise grape images for pretraining their berry detector and use supervised DA to increase model accuracy.

Karim et al. [22] investigate leaf blight, esca, black rot, and healthy leaves, employing data augmentation to produce more images. They test the effectiveness of the MobileNetV3Large network utilising transfer learning approaches for deploying it on a Jetson Nano. Real-time Grad-CAMs were employed to improve the system's performance. On a dataset of 1806 photos that were lab-collected, leaf by leaf, Kaushik and Sharma [23] are investigating black rot, esca, leaf blight, and healthy leaves. Kunduracioglu and Pacal [24] propose a technique that utilises two datasets: one for detecting grapevine leaf disease and the other for classifying grapevines into various species. They correctly identify the plant's class or disease by comparing different CNN models with different *Vision Transformers* (ViTs).

Wang et al. [25] suggest a Fourier DA-based approach to grapevine leaf disease detection that uses the farthest point sampling method to lessen the effect of randomness. Blight, black rot, and black measles are the diseases being studied in a leaf-by-leaf, lab setting environment. Using a dataset of grapevine leaf diseases, Venkatachalam et al. [26] compare deep learning methods for detecting grapevine leaf disease. To increase data diversity, enhance performance, and reduce overfitting, the study included labels for black rot, esca, leaf blight, and healthy leaves.

Unsupervised Domain Adaptation (UDA) in agriculture on a large scale is a task that Fuentes et al. [15] made progress by proposing a DA approach for classifying tomato diseases that successfully adapts to environmental changes between new greenhouse scenarios. Their findings are displayed for each of the three farm combinations. Wu et al. [16] present a novel method that uses UDA and successfully resolves the issue of significant difference between domains. Yan et al. [17] also propose a UDA between two distant and low-correlated domains to classify plant disease severity. Additionally, cross-species plant disease diagnosis was conducted on strawberries, cassava, and grapes. A leaf spot attention technique is used in a deep learning-based architecture presented by Yu and Son in [27].

Closest to our work is the work of Zhao et al. [4], which presents a method for leaf disease diagnosis that considers labelled data and a constrained distribution shift, and is the closest to our study. They use a self-supervised approach and require a certain quantity of labelled data in the target domain image collection.

In this study, we propose a DA framework for the de-

Fig. 1. Examples of challenging annotated grapevine foliar pathogens of interest using bounding boxes on an outdoor UAV image, left image is manually labelled while the right one is predicted by the base model (best viewed in colour)

tection of grapevine leaf diseases under real-world vineyard conditions. Leveraging a dataset of UAV imagery [8], our approach addresses adaptation to *Unmanned Ground Vehicle* (UGV) flash, natural imagery and few-shot learning principles. Specifically, we train a deep learning model using labelled images from a well-annotated source domain and adapt it to a labelled or unlabeled target domain using DA techniques. This methodology enables the model to generalise effectively to new environments without requiring additional manual annotation, thereby ensuring scalable and robust disease detection in operational vineyard settings.

III. Proposed Method

Representative examples from the employed dataset [8] are illustrated in Figure 1, where we compare a manually annotated leaf image (left) with a corresponding prediction generated by a model trained on August data, see section IV, and applied to an image (right), highlighting the model's cross-temporal inference capacity.

For the detection of diseased grapevine leaves, we employ a *Feature Pyramid Network* (FPN) [28], which is well-suited for multi-scale feature extraction in complex visual scenes. Additionally, we use the YOLO11 architecture, chosen for its superior detection performance and enhanced feature representation. This version features an anchor-free split head design, optimising the trade-off between detection accuracy and computational efficiency. Its lightweight yet powerful architecture makes it particularly suitable for future deployment in embedded systems and UAV-based real-time applications.

Model performance is quantitatively assessed using the *mean Average Precision* (mAP) metric, reported as a percentage. As defined in Eq. 1, mAP is computed over all classes C, with each class-wise average precision AP_k calculated as the area under the corresponding precision-recall curve. We report two standard evaluation thresholds: mAP_{50}, calculated at an *Intersection over Union* (IoU) threshold of 0.5, and $mAP_{50:95}$, which averages mAP scores over IoU thresholds ranging from 0.5 to 0.95 in 0.05 increments. These metrics provide a robust and interpretable measure of the detector's accuracy and consistency across varying prediction strictness levels.

$$mAP = \frac{1}{C} \times \sum_{k=1}^{C} AP_k \tag{1}$$

A. Methodology

The training process followed a supervised learning strategy in which the complete image dataset was split into two subsets: approximately 80% of the data was used for training, while the remaining 20% was reserved for validation or testing. This partitioning was consistent across all experiments.

The dataset [8] used in this study offers three resolution variants:

1) original high-resolution (3849×2160) images.
2) directly downsampled versions of the (3849×2160) images at 640×480 pixels.
3) a sliding window format where each (3849×2160) image is divided into four quadrants, each individually downsampled to 640×480 pixels.

The images chosen for use are the original ones, which were not resized, and only a sliding window approach was adopted to make them square-sized, better fitting our models' requirements.

Training was conducted over 300 epochs using manually labelled images from the source domain. For domain adaptation experiments, training resumed from the pre-trained source model's best weights and continued for an additional 200 epochs. Full domain recalibration has been implemented, given the task's specifics, which means all layers were left trainable during the adaptation process. Across all unsupervised cases, manual annotations were applied exclusively to the source domain, while the target domain remained unlabeled. Its predictions were evaluated based on the adapted models' inference performance.

B. Dataset Overview

The base dataset used in this study comprises over 9,000 images, primarily from Romanian vineyards, with a smaller proportion collected from other European regions, making it the largest known collection of vineyard imagery captured in real agricultural settings, outside of laboratory environments.

From this set of images 4,473 are from August, and they have been chosen to be used as a base for our testing. An additional set of private data from Bordeaux, France, was included to facilitate the natural-flash study presented in this work. In this image, a few-shot study has been conducted, using image sets comprising 98 images of plants affected only by black rot disease, 49 for the natural image set, and another 49 for the flash image set.

Figure 2 showcases sample images from the flash and natural image dataset, featuring grapevines affected exclusively by black rot disease. Due to the limited number of annotated samples, this dataset provides an ideal setting for evaluating few-shot learning in conjunction with DA techniques, allowing us to assess model performance under data-scarce conditions.

C. Experimental setup

In all experiments, models were trained for 300 epochs using manually labelled images drawn from the specified source dataset partitions, as indicated in the *Source* column of the accompanying tables. When a *Target* domain is listed, an adaptation phase is introduced. In this phase, training resumed for an additional 200 epochs using the best weights from the source model as the initialisation point.

D. DA theoretical background

DA is a specialised branch of transfer learning that addresses the challenge of deploying machine learning models across differing but related data distributions. In particular, it seeks to transfer knowledge from a source domain, denoted as θ_S, where annotated data are abundant, to a target domain, θ_T, which may contain limited or no labelled data. This setting is especially relevant in real-world applications, such as precision agriculture, where collecting labelled data across all environmental or temporal conditions is impractical.

The distributions θ_S and θ_T represent the statistical characteristics of the source and target domains, respectively. While both govern the mapping from input features to output labels, they diverge in terms of data availability and domain-specific statistical properties, often manifesting as distributional shifts.

The source domain distribution θ_S is typically estimated from a labelled dataset $\mathcal{D}_S = \{(x_i^S, y_i^S)\}_{i=1}^{n_S}$, where x_i^S are input samples drawn from the marginal distribution $P_S(X)$, and y_i^S are the corresponding labels sampled from the conditional distribution $P_S(Y|X)$. Consequently, θ_S approximates the joint distribution $P_S(X, Y)$, either explicitly, in the case of generative models, or implicitly, via learned model parameters in discriminative settings.

In contrast, the target domain θ_T is inferred from an unlabelled or partially labelled dataset $\mathcal{D}_T = \{x_j^T\}_{j=1}^{n_T}$, where only the marginal distribution $P_T(X)$ is accessible, while $P_T(Y|X)$ is either unknown or sparsely defined. This imbalance imposes significant challenges for direct supervised learning in the target domain.

The divergence between θ_S and θ_T arises from various types of domain shifts:

- Covariate shift: $P_S(X) \neq P_T(X)$ while $P_S(Y|X) \approx P_T(Y|X)$,
- Label shift: $P_S(Y) \neq P_T(Y)$,
- Concept shift: $P_S(Y|X) \neq P_T(Y|X)$.

Both θ_S and θ_T may be viewed as parameter sets governing domain-specific distributions $P(X, Y; \theta)$. The central objective of DA is to minimise the distributional discrepancy between θ_S and θ_T so that a model trained on the source domain can generalise effectively to the target domain. Formally, this is expressed in Eq. 2:

$$\min_{\theta_{S,T}} d(\theta_S, \theta_T) \tag{2}$$

Here, $d(\cdot)$ represents a divergence measure, such as *Maximum Mean Discrepancy* (MMD) or domain adversarial loss, quantifying the gap between the source and target feature distributions. In practical terms, DA methods fall into several categories depending on the availability of target labels:

1) Unsupervised DA: Assumes no labelled data in the target domain.
2) Few-shot or Semi-supervised DA: Leverages a small number of labelled samples from the target domain, often in conjunction with unsupervised objectives.

In this study, the proposed DA strategy is model-based, meaning it centres on adapting the parameters of a deep learning model, specifically a CNN, so that it learns domain-invariant features. This approach is particularly beneficial in agricultural vision tasks, where models trained on one region or season often fail to generalise to others due to domain shifts induced by environmental variability, lighting changes, and leaf morphology.

By combining few-shot learning principles with domain adaptation, the model can be effectively tuned using a limited amount of labelled target data, significantly improving its performance in scenarios characterised by data scarcity and cross-domain generalisation challenges.

IV. Results

In this section, we present a comprehensive evaluation of our grapevine disease detection model. The goal is to quantify the robustness of the model to real-world domain shifts, using both standard supervised training and UDA strategies. Performance is evaluated using the mAP metric on manually labelled validation sets.

Bold is used in all result tables to highlight the best-performing model for each distinct validation set, chosen from among the two configurations, source-only training and DA with all layers trainable.

The base model performance which is trained on August UAV data from the dataset [8] has a performance of $64.7\%[mAP_{50}]$.

Fig. 2. Examples of natural and flash UGV images used for few-shot learning

A. Few-Shot Learning

Building upon the previous symptomatology-based results from the August data, where the model was trained to detect general foliar disease symptoms regardless of specific pathogen types, we now shift our focus to a targeted detection task. Specifically, we explore the model's capacity to recognise black rot disease in grapevines from Bordeaux, France, employing a supervised domain adaptation strategy under a few-shot learning scenario.

In this experiment, the dataset comprises image pairs of grapevines captured under both natural and flash lighting conditions from the perspective of an *Unmanned Ground Vehicle* (UGV). The flash/natural image pairs were selected, used either independently or in combination, with careful handling to avoid cross-contamination. Representative examples of these high-resolution images, used in the training process,

TABLE I
RESULTS[1] ON FEW-SHOT LEARNING ON 5MPX FLASH/NATURAL IMAGES FOR GRAPEVINE LEAF DISEASE DETECTION

Source	Target	Validated	mAP50[%]
Flash&Natural	∅	Flash&Natural	48.3
Natural	∅	Natural	26.6
Natural	∅	Flash	20.5
Flash	∅	Flash	**54.8**
Flash	∅	Natural	26.7
August	Flash&Natural	Flash&Natural	66.7
August	Natural	Natural	64.6
August	Natural	Flash	57.7
August	Flash	Flash	**68.7**
August	Flash	Natural	48.3

TABLE II
RESULTS[1] ON UNSUPERVISED ADAPTATION ON 5MPX FLASH/NATURAL IMAGES FOR GRAPEVINE LEAF DISEASE DETECTION

Source	Target	Validated	mAP50[%]
August	Natural	Natural	38.4
August	Natural	Flash	42.7
August	Flash	Flash	**53.2**
August	Flash	Natural	40.0

are shown in Figure 2.

Initial training results based on the annotated dataset are reported in the first half of Table I. The second half of the table presents results from a supervised adaptation setup, where the best-performing symptomatology model suggests that pretraining on broader disease classes can significantly enhance adaptation to specific pathogens under conditions of limited data. A generalised symptomatology model offers measurable benefits in few-shot scenarios, outperforming training from scratch on the flash/natural image pairs. This indicates that pretraining on broader disease classes can significantly enhance adaptation to specific pathogens under limited data conditions.

To further evaluate the effectiveness of UDA in a few-shot context, an additional experimental setup was implemented. Specifically, the best-performing August-trained model, adapted either to flash or natural lighting conditions, as reported in Table I, was used as the source model to generate pseudo-labels on its complementary image set (i.e., flash to natural, or vice versa). These pseudo-labels then served as the basis for unsupervised adaptation to the target domain.

The results of this UDA procedure are summarised in Table II. This setup enables us to examine the model's performance when transferring between visually distinct yet semantically similar imaging conditions, testing its robustness to lighting variability and demonstrating its adaptability even in low-data regimes.

B. Results Interpretation using DA knowledge

The results presented in Tables I and II can be more deeply understood by grounding them in the theoretical framework

[1] where source is the image set used for training, target is the image set used for adaptation, validation is the image set the model is composed of, source and target are tested on, **bold** is the best result for each source-target image pair and ∅ denotes no adaptation

Fig. 3. Examples of annotated black rot disease using bounding boxes on an outdoor UGV image, left image is manually labelled while the right one is predicted by the best model (best viewed in colour)

of DA. As described in subsection III-D, the central goal of DA is to reduce the discrepancy between the source domain distribution θ_S and the target domain distribution θ_T, thereby enabling the model to generalize effectively to unseen or under-represented data regimes.

In this context, the August-trained model serves as the source model θ_S, trained on diverse foliar disease symptoms. When adapted to the flash or natural imagery domains, each representing θ_T under different lighting conditions, the model attempts to minimise the divergence $d(\theta_S, \theta_T)$ via supervised or unsupervised strategies. The superiority of the supervised August-to-Flash configuration $68.7\%[mAP_{50}]$, highlights the effectiveness of DA when a strong base model is combined with limited target supervision.

From a DA perspective, the strong performance of the adapted August model suggests that its feature representations are partially domain-invariant, allowing successful transfer to the target domain despite significant differences in lighting and viewpoint. This aligns with theoretical expectations by initialising adaptation with a model trained on general symptomatology, and adapting it via few-shot learning, the system achieves a robust balance between source generalisation and target specificity.

Moreover, the results in Table II further demonstrate the potential of unsupervised DA in few-shot settings. Using pseudo-labels generated from the source model, the system performs unsupervised adaptation across flash and natural domains, yielding moderate improvements even without additional labels. This reflects the practical value of minimising domain divergence $d(\theta_S, \theta_T)$ through alignment of feature distributions, especially when $P_T(Y|X)$ is sparsely defined.

Figure 3 visually illustrates the success of this approach, showing that the adapted model is capable of closely replicating human annotations in real outdoor vineyard imagery, despite limited supervision and domain variability. These results validate the theoretical underpinnings of DA and underscore its applicability in precision agriculture, where models must generalise across heterogeneous real-world conditions using minimal labelled data.

C. Discussion

Several key insights emerge from the experimental findings, particularly in relation to image pre-processing strategies and few-shot learning in visual plant disease detection.

First, the impact of image resizing in deep learning workflows warrants careful consideration. In visual detection tasks, such as identifying grapevine leaf disease, aggressive downscaling can significantly degrade model performance. Important high-frequency features, such as small lesions, subtle discolourations, or edge contours, may be lost, rendering the model unable to correctly interpret the input or associate it with the appropriate label. To mitigate this, the sliding window approach proves advantageous. Rather than rescaling the entire image, this method crops high-resolution images into smaller segments, allowing the model to process them efficiently while preserving the original spatial and visual detail. This approach is particularly suitable when adapting input sizes to GPU or model constraints without sacrificing crucial disease-related information.

Second, in the context of few-shot learning, the results indicate that initialising with a general-purpose base model trained on symptomatology rather than a specific disease can yield notable performance improvements. As shown in Table I, even when the pre-trained model was not explicitly optimised for detecting black rot, its prior exposure to a broad spectrum of foliar disease features provided a valuable representational foundation for downstream adaptation. This supports the use of symptom-centric pretraining as a practical strategy in low-data scenarios, especially when precise annotations for specific diseases are scarce or difficult to obtain.

D. Limitations

One of the primary limitations of this study is related to the dataset, which comprises real-world vineyard imagery captured under uncontrolled field conditions. Unlike datasets curated in laboratory environments [29], where individual leaves are photographed against uniform backgrounds, the images used here are subject to various sources of environmental noise. These include inconsistent lighting, complex and cluttered backgrounds, variable leaf orientations, and occlusions from surrounding vegetation. Such variability introduces additional challenges for accurate disease detection and increases the difficulty of learning robust visual patterns.

Figure 4 illustrates a representative failure case that highlights the effects of domain shift on model performance. In this example, the model, trained on images collected from Romanian vineyards, was evaluated on data featuring a different type of foliar disease not included in the original training set. As shown in the right-hand image, the model incorrectly identifies disease symptoms, producing false positive predictions. The corresponding manual annotation on the left-hand image reveals the absence of such disease categories in the training data. These misclassifications can be attributed to visual similarity between the new disease symptoms and previously learned patterns, leading to semantic confusion. Such cases underscore the need for more comprehensive datasets that capture a wider range of disease types, as well as improved domain generalisation strategies.

V. Conclusions

This work presents a domain adaptation approach for the automated detection of grapevine leaf diseases in uncontrolled, real-world vineyard environments. By leveraging a model trained on labelled source domains and adapted to labelled or unlabeled target domains, we explored the effectiveness of transfer learning strategies, with a focus on flash and natural imagery.

The results underscore the complexity of disease identification in natural agricultural settings. Unlike controlled laboratory imagery, field-collected data introduces a range of visual noise and variability, which can adversely impact model generalisation.

Limitations of the proposed approach are related to the dataset scope and disease coverage. False positives resulting from unseen or visually similar diseases highlight the need for broader and more diverse training data.

In future work, these limitations can be addressed in several key directions. First, expanding the dataset through synthetic image generation and domain randomisation would allow exploration of *simulation-to-real* (Sim2Real) adaptation techniques. This could help compensate for under-represented disease categories and improve model robustness. Second, extending the model to other agricultural applications, where larger annotated datasets exist, could validate its adaptability across different crops and visual disease signatures.

References

[1] J. M. Alston and O. Sambucci, "Grapes in the world economy," in *The grape genome*, pp. 1–24, Springer, 2019.

[2] F. Portela, J. J. Sousa, C. Araújo-Paredes, E. Peres, R. Morais, and L. Pádua, "A systematic review on the advancements in remote sensing and proximity tools for grapevine disease detection," *Sensors*, vol. 24, no. 24, p. 8172, 2024.

[3] S. Molnár and L. Tamás, "Close proximity aerial image for precision viticulture. A review," *Journal of Plant Diseases and Protection*, vol. 132, p. 57, Jan. 2025.

[4] R. Zhao, Y. Zhu, and Y. Li, "CLA: A self-supervised contrastive learning method for leaf disease identification with domain adaptation," *Computers and Electronics in Agriculture*, vol. 211, Aug. 2023.

[5] S. Aich and I. Stavness, "Leaf Counting with Deep Con-volutional and Deconvolutional Networks," in *Proc. of the IEEE International*, (Venice, Italy), Conference on Computer Vision Workshops (ICCVW-2017), Oct. 2017.

[6] S. F. Di Gennaro, E. Battiston, S. Di Marco, O. Facini, A. Matese, M. Nocentini, A. Palliotti, and L. Mugnai, "Unmanned Aerial Vehicle (UAV)-based remote sensing to monitor grapevine leaf stripe disease within a vineyard affected by esca complex," *Phytopathologia Mediterranea*, vol. 55, July 2016.

[7] J. Sun, S. B. Kang, Z.-B. Xu, X. Tang, and H.-Y. Shum, "Flash cut: Foreground extraction with flash and no-flash image pairs," in *2007 IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1–8, IEEE, 2007.

[8] D. E. Székely, D. Dobra, A. E. Dobre, V. Domşa, B. G. Drăghici, T.-A. Ileni, R. Konievic, S. Molnár, P. Sucala, E. Zah, A. S. Darabant, A. Sándor, and L. Tamás, "Bacterial-fungicidal vine disease detection with proximal aerial images," *Heliyon*, vol. 10, p. e34017, July 2024.

[9] M. Ren and R. S. Zemel, "End-To-End Instance Segmentation With Recurrent Attention," in *Proc. of the IEEE*, (Honolulu, Hawaii, USA), pp. 6656–6664, Conf. on Computer Vision and Pattern Recognition (CVPR-2017), July 2017.

[10] M. Litrico, D. Talon, S. Battiato, A. Del Bue, M. V. Giuffrida, and P. Morerio, "Uncertainty-guided Open-Set Source-Free Unsupervised Domain Adaptation with Target-private Class Segregation," in *Proceedings of the IEEE*, (Seattle, WA, USA), Causal and Object-Centric Representations for Robotics Workshop at CVPR (CORR-2024), June 2024.

[11] L. Tamas, S. Gubo, and T. Lukić, "Vine diseases detection trials in the carpathian region with proximity aerial images," in *IEEE World Symposium on Applied Machine Intelligence and Informatics (SAMI)*, (Stará Lesná, Slovakia), pp. 29–34, Jan. 2024.

[12] M. Xu, H. Kim, J. Yang, A. Fuentes, Y. Meng, S. Yoon, T. Kim, and D. S. Park, "Embracing limited and imperfect training datasets: opportunities and challenges in plant disease recognition using deep learning," *Frontiers in Plant Science*, vol. 14, p. 1225409, 2023.

[13] M. Xu, J.-E. Park, J. Lee, J. Yang, and S. Yoon, "Plant disease recognition datasets in the age of deep learning: challenges and opportunities," *Frontiers in Plant Science*, vol. 15, p. 1452551, 2024.

[14] G. Csurka, "Domain adaptation for visual applications: A comprehensive survey," 2017.

[15] A. Fuentes, S. Yoon, T. Kim, and D. S. Park, "Open Set Self and Across Domain Adaptation for Tomato Disease Recognition With Deep Learning Techniques," *Frontiers in Plant Science*, vol. 12, Dec. 2021.

[16] X. Wu, X. Fan, P. Luo, S. D. Choudhury, T. Tjahjadi, and C. Hu, "From Laboratory to Field: Unsupervised Domain Adaptation for Plant Disease Recognition in the Wild," *Plant Phenomics*, vol. 5, Mar. 2023.

[17] K. Yan, X. Guo, Z. Ji, and X. Zhou, "Deep transfer learning for cross-species plant disease diagnosis adapting mixed subdomains," *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, vol. 20, Dec. 2021.

[18] J. Ubbens, M. Cieslak, P. Prusinkiewicz, and I. Stavness, "The use of plant models in deep learning: an application to leaf counting in rosette plants," *Plant Methods*, vol. 14, pp. 1–10, Jan. 2018.

[19] M. V. Giuffrida, P. Doerner, and S. A. Tsaftaris, "Pheno-Deep Counter: A unified and versatile deep learning architecture for leaf counting," *The Plant Journal*, vol. 96, pp. 880–890, Aug. 2018.

[20] Z. Fei, A. G. Olenskyj, B. N. Bailey, and M. Earles, "Enlisting 3D crop models and GANs for more data efficient and generalizable fruit detection," in *Proc. of the IEEE/CVF International*, pp. 1269–1277, Int. Conf. on Computer Vision, Mar. 2021.

Fig. 4. Examples of limitations on Bordeaux UAV image (best viewed in colour), manually labelled image on left, and the model prediction on the Cluj-August dataset on right

[21] R. Roscher, K. Herzog, A. Kunkel, A. Kicherer, R. Töpfer, and W. Förstner, "Automated image analysis framework for high-throughput determination of grapevine berry sizes using conditional random fields," *Computers and Electronics in Agriculture*, vol. 100, pp. 148–158, Jan. 2014.

[22] M. J. Karim, M. O. F. Goni, M. Nahiduzzaman, M. Ahsan, J. Haider, and M. Kowalski, "Enhancing agriculture through real-time grape leaf disease classification via an edge device with a lightweight CNN architecture and Grad-CAM," *Scientific Reports*, vol. 14, no. 1, 2024.

[23] P. Kaushik and P. Sharma, "Leveraging EfficientNet for Enhanced Grape Leaf Disease Detection: A Novel Approach to Precision Viticulture," in *2024 International Conference on Cybernation and Computation (CYBERCOM)*, pp. 345–349, IEEE, 2024.

[24] I. Kunduracioglu and I. Pacal, "Advancements in deep learning for accurate classification of grape leaves and diagnosis of grape diseases," *Journal of Plant Diseases and Protection*, vol. 131, no. 3, pp. 1061–1080, 2024.

[25] J. Wang, Q. Wu, T. Liu, Y. Wang, P. Li, T. Yuan, and Z. Ji, "Fourier domain adaptation for the identification of grape leaf diseases," *Applied Sciences*, vol. 14, no. 9, p. 3727, 2024.

[26] C. Venkatachalam, P. Shah, K. R. KM, Y. Kumaran, A. Roy, *et al.*, "Advanced Grape Leaf Disease Diagnosis Using EfficientNetV2L with Data Augmentation and Grad-CAM Visualization in Precision Agriculture," *Procedia Computer Science*, vol. 260, pp. 332–340, 2025.

[27] H.-J. Yu and C.-H. Son, "Leaf Spot Attention Network for Apple Leaf Disease Identification," in *Proc. of the IEEE*, (Seattle, Washington, USA), pp. 229–237, Conf. on Computer Vision and Pattern Recognition Workshops (CVPRW-2020), June 2020.

[28] T.-Y. Lin, P. Dollar, R. Girshick, K. He, B. Hariharan, and S. Belongie, "Feature Pyramid Networks for Object Detection," in *Proceedings of the IEEE*, (Honolulu, Hawaii, USA), pp. 936–944, Conf. on Computer Vision and Pattern Recognition (CVPR-2017), July 2017.

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