The PV-ALE Dataset: Enhancing Apple Leaf Disease Classification Through Transfer Learning with Convolutional Neural Networks

Joseph Damilola Akinyemi[1][0000−0003−3121−4231] and Kolawole John Adebayo[2][0000−0001−7126−7026]

[1] University of York, York, United Kingdom
[2] Dublin City University, Ireland
joseph.akinyemi@york.ac.uk
kolawolejohn.adebayo@dcu.ie

Abstract. As the global food security landscape continues to evolve, the need for accurate and reliable crop disease diagnosis has never been more pressing. To address global food security concerns, we extend the widely used PlantVillage dataset with additional apple leaf disease classes, enhancing diversity and complexity. Experimental evaluations on both original and extended datasets reveal that existing models struggle with the new additions, highlighting the need for more robust and generalizable computer vision models. Test F1-scores of 99.63% and 97.87% were obtained on the original and extended datasets, respectively. Our study provides a more challenging and diverse benchmark, paving the way for the development of accurate and reliable models for identifying apple leaf diseases under varying imaging conditions. The expanded dataset is available on Kaggle, enabling future research to build upon our findings.

Keywords: Apple Disease · CNN · Deep Learning · PlantVillage Dataset.

1 Introduction

Plant diseases pose a significant threat to plant life, crop yield, and food security, with far-reaching consequences for global food systems. The diverse nature of these diseases, which vary across plant species, necessitates specialized domain knowledge for effective detection and management. [1]. Moreover, timely intervention to manage plant diseases and the need for efficient diagnostic methods have become increasingly pressing in the face of climate change, which continues to disrupt agricultural productivity. [9]. Conventionally, farmers utilise visual cues to identify infected crops. Even though a time-tested and efficient approach, it is time-consuming, labour-intensive and prone to inconsistent diagnoses due to human error or subjective interpretations, leading to the possibility of misclassifying unhealthy plants as healthy and vice versa [6].

Apple trees are susceptible to a range of diseases including Powdery Mildew (PM), Apple Scab (AS), Alternaria leaf spot, Apple rust, etc., which can result

in severe yield losses if left untreated [10]. The timely identification of these diseases at their early stages is crucial for implementing effective control measures, which can mitigate the economic and environmental impacts of these diseases [7]. Recent advances in machine learning and deep learning have created new opportunities for automating plant disease detection, enabling the development of more accurate and efficient diagnostic tools [12,8].
We propose a deep learning-based approach for detecting apple leaf diseases from images, leveraging the economic importance of apple crops and the availability of high-quality, well-annotated open-source datasets. Our system aggregates 5 unique apple leaf disease types and 1 healthy class from consolidated existing and new datasets.

The primary contributions of our work include:

1. A comprehensive consolidated dataset of apple leaf images annotated with corresponding disease labels.
2. An efficient CNN architecture tailored specifically for the task of multi-class classification within the context of apple leaf disease detection.
3. Rigorous evaluation metrics assessing model performance under various scenarios including class imbalance.
4. Demonstration of superior accuracy compared to existing methods highlighting its potential application in real-world settings.

The remaining parts of the paper are organized as follows: Sect. 2 presents a detailed review of the existing literature on apple leaf disease classification, Sect. 3 describes the dataset and method used in our work for apple leaf disease classification, Sect. 4 presents our experiments, results and comparative analysis and Sect. 5 concludes the paper.

2 Literature Review

Machine Learning has been used for plant disease detection with a reasonable degree of success in several studies [13,1,11]. As in many other domains, Deep Learning algorithms have exceeded traditional Machine Learning algorithms in Apple disease detection and classification [2].
Jiang et al. [6] used an Inception network with Rainbow concatenation for Apple disease classification task on five classes of apple diseases (Alternaria leaf spot, Brown spot, Mosaic, Grey spot, and Rust). The experiments were conducted on 26,377 images of the Apple Leaf Disease Dataset (ALDD) dataset and reported a mean average accuracy of 78.8% on a hold-out set of the ALDD. The authors of [16] employed a DenseNet-121 deep convolution network, formulating the problem as a multi-label classification task. The imbalanced nature of most apple disease datasets means that some disease classes will have very small probabilities, so the authors introduced a focus loss function instead of entropy loss, thereby increasing the performance of their model from 92.01% to 93.71%. In solving the class imbalance, Tian et al. [14] proposed using CycleGAN for data augmentation, thereby increasing the size and diversity of the

dataset. Moreover, the authors employed DenseNet to optimize certain layers of the YOLO-V3 model. To conduct experiments, they collected 140 apple fruit images (increased to 700 images through augmentation) and obtained an accuracy of 95.57%.

Recent studies have proposed various deep learning-based approaches for plant disease detection. For instance, Tian et al. [13] introduced a Multi-scale Dense classification network that achieved state-of-the-art classification accuracies of 94.31% and 94.74% on a dataset of 11 classes, including healthy and diseased apple fruits and leaves. The study employed Cycle-GAN for data augmentation to address the challenge of insufficient images. Similarly, the authors in [1] proposed a three-stage approach for plant disease detection, achieving an accuracy of 99.98% on the PlantVillage dataset [12] using DenseNet [5].

Other deep learning-based studies have mostly involved various network architectures on different datasets using different evaluation strategies, thus making it difficult to make direct comparisons among studies. For instance, the authors in [7] proposed a 2-stage approach using the Xception model to first extract low-level features and then Faster-RCNN to localize the diseased image region. However, their method achieved 88% accuracy which seems subpar compared to simpler CNN-based studies such as [2,8]

The reviewed studies underscore the potential of deep learning algorithms and related techniques in apple disease detection. However, there remains a need for further research to enhance the accuracy and robustness of these models. First, the existing apple leaf datasets, while valuable, present certain limitations that necessitate further research. On primary gap is the variability in disease classes across different datasets. Most datasets focus on a different set of disease classes, and there is a scarcity of comprehensive datasets encompassing a wide range of these different disease classes. Moreover, most of the existing datasets are not very large and the few large ones are not publicly available. These limitations restrict the generalizability of the models trained on these datasets, as they may not perform well when confronted with diseases outside those in their training set.

This paper aims to fill these gaps by extending a well-used existing dataset of apple leaf diseases using validated manual data collection and augmentation techniques to create a larger and more comprehensive dataset. By doing so, we hope to enhance the generalizability and robustness of apple disease detection models. This research is thus a significant step towards harnessing the power of deep learning to address the challenges in apple disease detection, promoting sustainable agriculture.

3 Methodology

3.1 Dataset

Despite the significant research efforts in tackling apple disease detection, the available datasets often present certain limitations that hinder comprehensive

and robust model training. These datasets are typically small-sized, imbalanced, and contain a limited number of disease classes [8]. Moreover, they are not often publicly available, primarily due to the substantial expert effort required for their collection and annotation.

One such dataset is PlantVillage [12], the most commonly used dataset in this domain. It contains only 3,171 images of apple leaves, more than half of which are healthy leaves. The remaining half is distributed among just three disease classes: *rust*, *black rot*, and *scab*. The high-resolution, single-leaf images in the PlantVillage dataset, while useful, present less of a challenge and are less applicable in real-world situations, which often involve variable image quality, multiple leaves, and complex backgrounds.

To address these challenges, we collected apple leaf images for two additional disease classes: *Alternaria leaf spot* and *Powdery Mildew*. These images were collected via Internet image search using the disease names as keywords. Initially, we gathered 79 and 183 images for *Alternaria leaf spot* and *Powdery Mildew*, respectively. To enhance the data pre-processing and preparation steps, and ensure the quality and reliability of our dataset, we followed the guideline below:

1. Image Download: Download only images containing apples and apple leaves. This ensures that our dataset is specific to our research focus and reduces the likelihood of including irrelevant images.
2. Caption Verification: Verify the image caption to ensure it bears the disease name with respect to the query. This step helps to confirm that the image is indeed related to the disease class it is supposed to represent.
3. URL Verification: Verify the image URL to ensure the caption is trustworthy. The URL should be from a publication or website content of an agricultural-focused research performing organization, laboratory or nursery. By doing so, we can increase the likelihood of obtaining accurate and reliable images.
4. Image Matching: Implement additional verification to ensure that the image matches or looks like other images from the same class/pool. This step helps to maintain consistency within each disease class and reduce the risk of mislabeling.
5. Human Validation: To further ensure the quality of our dataset, we also use human annotators to validate each image and its corresponding disease class to ensure their accuracy. This step adds an extra layer of verification and helps to identify and correct any potential errors or inconsistencies in our dataset. For instance, where some images bearing tags related to the disease class were not visually related to the disease class, those were removed.
6. Image Cropping: We cropped images to obtain multiple single-leaf images where possible. This step increased the number of images, as some images with leaf clusters generated more single-leaf images. However, many images of leaf clusters were still retained to increase the complexity and diversity of the dataset.
7. Background Removal: We cropped the images to remove the background, focusing on the apple leaves and their diseases.

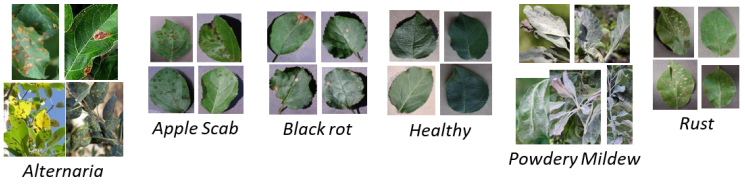

Fig. 1. Sample images of apple leaves in the dataset.

We posit that adhering to this guideline significantly improves the quality, reliability, and diversity of our dataset, ultimately contributing to the development of more accurate and robust apple disease detection models. Fig. 1 presents representative samples of images from the two newly introduced disease classes, namely (*Alternaria* and *Powdery Mildew*), alongside samples from the pre-existing classes in the PlantVillage dataset. Upon completion of the image collection and pre-processing stages, we obtained 85 images for *Alternaria* and 127 images for *Powdery Mildew*. As observed in Fig. 1, the newly incorporated classes introduce several challenges, such as variable-sized and small images, low resolution, the presence of both single-leaf and leaf-cluster images, complex backgrounds, and a limited number of images in each class. However, these characteristics contribute to a more challenging and realistic representation of real-world conditions, thereby enhancing the applicability and generalizability of our apple leaf disease detection models.

After the initial data curation process, we integrated the newly collected disease classes with the existing four classes of the PlantVillage dataset, resulting in a consolidated dataset of 3,383 images. To address the significant class imbalance, we divided the dataset into a 70% training set and a 30% testing set. We then applied various data augmentation techniques, such as angled image rotation, light intensity variation, and zooming, to balance the training set data. This augmentation process yielded 20,808 images fairly evenly distributed across the six classes, forming the PlantVillage Apple Leaves Extended (PV-ALE) dataset. PV-ALE can be accessed on Kaggle.

The dataset statistics presented in Table 1 reveal a significant bias towards healthy leaves. Despite being significantly smaller, the two new classes constitute an essential addition of new disease classes to the entire dataset. Table 1 also demonstrates how the augmented images helped us achieve a fairly balanced distribution of instances per class, resulting in a training set with approximately 3,500 images in each class.

The PV-ALE dataset holds significance not only for the challenges it presents but also for expanding the existing classes in the most widely-used apple disease dataset, PlantVillage. Crucially, implementing class balancing and data augmentation methods generates a more diverse set of samples, satisfying the training and evaluation requirements for future research. Moreover, by open-sourcing the PV-ALE dataset, we aim to establish a more comprehensive benchmark with a broader range of classes and samples. This will facilitate more accurate com-

parisons among future studies, as many researchers have previously published experiments on various closed datasets, making it extremely difficult to objectively compare existing systems in the literature. In the subsequent sections, we offer a detailed account of our primary deep-learning classification models and the established baselines.

Table 1. Dataset statistics

Classes	original set	training set	test set	augmented training set
Alternaria	85	60	25	3420
Apple Scab	630	441	189	3528
Black rot	621	435	186	3480
Cedar Apple rust	275	192	83	3456
Healthy	1645	1151	494	3453
Powdery mildew	127	89	38	3471
Totals	3383	2368	1015	20808

3.2 Apple disease classification

The efficacy and utility of the proposed PV-ALE dataset were comprehensively evaluated through a series of experiments conducted on both the extended (PV-ALE) and the original (PV-AL) PlantVillage datasets. To assess the impact of the two newly incorporated classes on performance and to determine the potential contribution of the PV-ALE dataset in advancing research on apple leaf disease detection, we employed two distinct deep learning models for training and validation on each dataset, resulting in a total of four experiments.

The task of identifying or classifying apple leaf diseases based on their leaf images was formulated as a multiclass classification problem. Deep learning techniques were leveraged for both feature extraction and classification throughout the experiments. In one set of experiments, we adopted a transfer learning approach by fine-tuning a ResNet50 architecture [4] pre-trained on the ImageNet dataset [3]. The ResNet model was utilized to extract salient features from the input leaf images, which were subsequently classified based on the extracted features. In the other set of experiments, we constructed a simple 7-layer Convolutional Neural Network (CNN) from scratch for the same purpose.

The selection of these models was motivated by the proven capability of CNNs and ResNet50 in extracting discriminative features from images, which can be highly descriptive and instrumental in distinguishing images based on subtle differences. In the context of this study, these subtle differences pertained to the specific disease types affecting the apple leaves.

In our transfer learning approach, we employed the *ResNet50* model pre-trained on the ImageNet dataset as the backbone for feature extraction. The original classification layer was replaced with a custom head comprising a Global Average Pooling layer, a Flatten layer, and five Fully Connected layers interspersed with

three Dropout layers, resulting in a total of over 24 million trainable parameters. The number of units in the Fully Connected (Dense) layers was empirically determined during hyperparameter tuning, with the final layer having six units corresponding to the six classes for the extended dataset or four units for the original four-class PlantVillage dataset. The Dropout layers randomly omitted either 30% or 20% of the neurons from the preceding layers. The model was trained using the Adam optimizer with a base learning rate of *5e-5*, a categorical cross-entropy loss function, and a *softmax* activation function for the final Dense layer, while rectified linear unit activations were employed for the remaining Dense layers.

Our custom 7-layer CNN architecture consisted of two *2D* convolutional layers, a Flatten layer, and four Fully Connected (Dense) layers, complemented by intermittent pooling and Dropout layers to extract salient features and mitigate overfitting during training. With over 252 million trainable parameters, this model had significantly more parameters than our transfer learning approach, primarily due to the training of the entire CNN, including the convolutional layers with numerous 3×3 filters, from scratch. The CNN was trained using the Adam optimizer with a *5e-5* base learning rate, a categorical cross-entropy loss function, and rectified linear unit activation functions for all Dense layers except the final layer, which employed a Softmax activation function. Across all experiments, the input images were resized to 225×225 pixels to conform to the ResNet input layer requirements.

To prevent overfitting and data leakage, we employed three strategies: a clear separation of training and test data, Dropout layers, and the Early Stopping technique. The test sets were strictly isolated from the training sets and models during training. For validation purposes, 10% of the training set was held out as a validation set, allowing for fine-tuning and performance evaluation without exposing the models to the test set, thus preventing overfitting. This approach ensured a reliable assessment of the models' generalization capabilities on the test set. Additionally, Dropout layers were incorporated into each network architecture, randomly omitting *20%* to *30%* of the neurons from the preceding layer before propagating to the next layer. Furthermore, Early Stopping was implemented to monitor the validation loss and terminate training after 10 (for ResNet) or 5 (for CNN) consecutive epochs without any further reduction in validation loss. These measures collectively aimed to produce well-generalized models capable of robust performance on the test set. The difference in Early Stopping tolerance values between the ResNet and CNN models was to provide the ResNet with a higher chance of convergence and to reduce the CNN's training time, given its larger parameter count.

4 Results

Previous works on apple leaf disease identification or classification have often reported high accuracy rates due to the low variability present in the datasets. Most datasets contain high-resolution single-leaf images captured against plain

backgrounds. While this simplifies image processing, it fails to represent real-world scenarios where individual leaves are seldom examined against a plain background to determine plant diseases, especially in large-scale farming operations. The PV-ALE dataset collected for this work addresses this limitation by incorporating two new classes containing low-resolution images, leaf clusters, and complex backgrounds (often including other leaves or garden plants).
In this section, we report the results of our experiments using both the original PlantVillage dataset and the PV-ALE dataset. To ensure a fair comparison, the same network architectures and parameter settings were employed for both datasets. Additionally, we spot-checked the reported accuracies of some previous works on specific classes of apple leaf diseases included in our dataset and compared them with our results. All experiments were performed on Kaggle's cloud-based GPU scripting environment, utilizing a GPU P100 with 16GB GPU Memory and 32GB RAM.

4.1 Results on the PV-ALE dataset

Transfer learning results on the PV-ALE dataset. The Transfer Learning (TL) method on the PlantVillage-ALE (PV-ALE) dataset demonstrated very promising results. The training and validation accuracy and loss curves for the ResNet50 model are shown in Fig. 2. Fig. 2a depicts the loss, while Fig. 2b illustrates the accuracy. Both figures indicate a smooth training progression with no evidence of overfitting, as the validation performance is at par with the training performance throughout the training process. It can be observed that from the very first epoch, both training and validation accuracy surpassed 90%. This is made possible by the depth of the ResNet architecture, which enables it to learn rich image features, and the added top layers effectively adapted those features to discriminate between the different diseases shown in the leaf images. Table 2 shows the results obtained by our TL method (ResNet50) on the PV-ALE and PV-AL datasets (rows 1 & 3), which are impressive.

Table 2. Precision (Prec.), Recall (Rec.), F1-scores (F1) and Accuracy on PV-ALE and PV-AL datasets

Dataset	Model	Precision (%)	Recall (%)	F1-score (%)	Accuracy (%)
PV-ALE	ResNet50	99.06	96. 82	97.87	99.11
PV-ALE	CNN	99.01	88.94	89.77	94.98
PV-AL	ResNet50	99.55	99.72	99.63	99.58
PV-AL	CNN	94.37	95.82	95.05	95.59

CNN results on the PV-ALE dataset. Similar to the trend observed with the TL technique, the simple CNN architecture we constructed also performed reasonably well on the training, validation and test datasets. As shown in figures Fig. 2c and Fig. 2d, the CNN training progressed well, though not as smoothly as the ResNet model, yet the validation progression generally matched the training

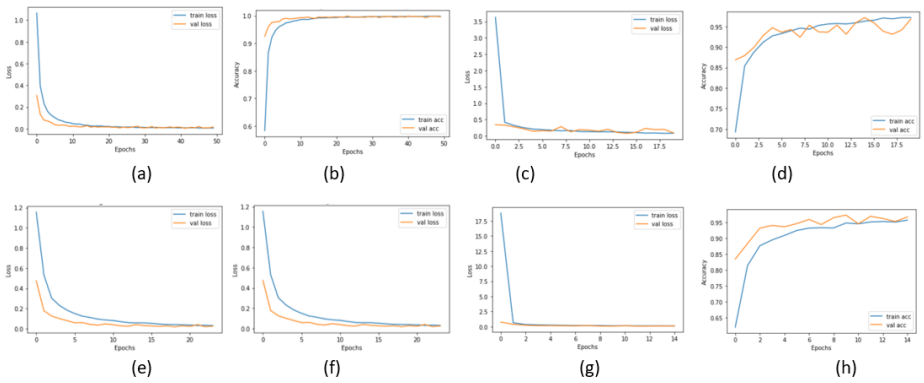

Fig. 2. Loss and accuracy on PV-ALE (top row) and PV-AL (bottom row) datasets

progression. As presented in Table 2 (row 2), the CNN model achieves ≈95% accuracy which is ≈4% lower than that of the ResNet50 model. This small difference indicates that the CNN architecture is generally well-suited to the problem of identifying apple diseases from apple leaf images.

4.2 Results on the Original PV-AL Dataset

The experiments conducted on the original PlantVillage Apple Leaf (PV-AL) dataset, comprising four classes (three disease classes and one healthy class), reveal that it is less challenging than the PV-ALE dataset. Both the ResNet and CNN models demonstrate this observation, as shown in Fig. 2 (bottom row). The validation accuracy and loss consistently outperform the training accuracy and loss for most of the training epochs. Despite an early stopping tolerance of 10 epochs, the ResNet model converges quickly after 24 epochs (Fig. 2e and 2f), which is only halfway through the number of epochs required for the same ResNet model to stop training on the PV-ALE dataset. Table 2 (rows 3 & 4) shows that ResNet achieves an accuracy of 99.58% and an F1 score of 99.63% while CNN achieves an accuracy of 95.59% and an F1 score of 95.05% on the PV-AL dataset. The training progression of the CNN model indicates a good fit (Fig. 2g and 2h) as the validation performance surpasses the training performance throughout the training process. Without the Early Stopping setting, the CNN could potentially reach even higher accuracy in a few more epochs.

4.3 Comparing results on both datasets

Fig. 3 illustrates the precision, recall and F1-score values obtained using the ResNet50 and CNN models on each class of both datasets. Generally, the *Alternaria leaf spot* class (light blue bar) is the most challenging to identify while the *Cedar Apple Rust* class (yellow bar) is almost perfectly identified in all cases. One possible explanation for this observation is the number of samples for each

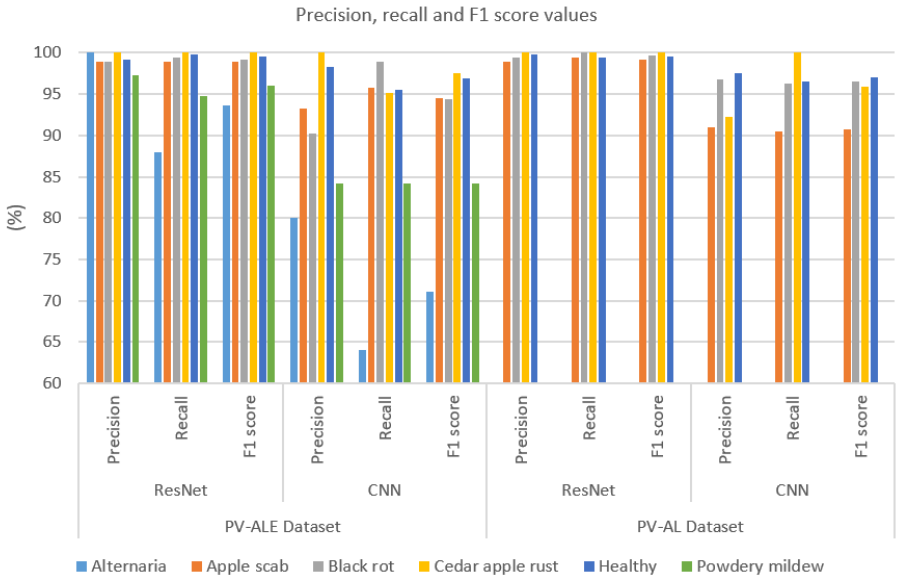

Fig. 3. Precision, recall and F1-score values on each class of both datasets

class in the test set; the *Alternaria* class has the fewest samples, while the *Cedar Apple Rust* class has the highest number of samples. However, this discrepancy in class distribution is an accurate reflection of the actual data and could not be circumvented. Notably, this result is not a function of the training data distribution, as the augmented training data is fairly balanced across all classes. Even though the *Alternaria* class has the least number of samples (3420) in the augmented training set, the *Cedar Apple Rust* class, which has the third lowest number of samples (3456), exhibits the best performance as shown in Fig. 3. This suggests that the presented results on the test set are more influenced by the visual cues responsible for each class rather than the class size. The network seems to recognize the *Cedar Apple Rust* disease more easily than any other class. Interestingly, while one might expect the *Healthy* class to be most easily recognised, this is not the case. This pattern is consistent throughout the experiments, with the *Rust* class being the most accurately recognized, followed closely by the *Healthy* and *Black Rot* classes. This implies that the visual features on the *cedar apple rust* leaves are easier for the networks to identify compared to other classes.

Given that both datasets were subjected to the same network under identical parameter settings, the performance difference of approximately 2% (F1 scores of 97.78% on PV-ALE and 99.63% on PV-AL, as shown in Table 2 (rows 1 & 3) and the fact that better performance was achieved on the PV-AL dataset in fewer training epochs indicate that the two new classes introduced additional complexities to the original dataset. These findings suggest that the PV-ALE

dataset will prove resourceful for future research in this field. While one could argue that the PV-AL dataset is smaller than the PV-ALE dataset, the difference between the two test sets is only 63 samples, which is not up to the size of any of the two added classes and is about 6% - 7% of the total size of the test set for each dataset. These results reveal the need for more disease classes in the apple leaf disease dataset.

4.4 Analysis and Discussion

Finally, we performed a comparative analysis of the classification accuracy for each class in the PV-ALE dataset with those reported in the literature. It is worth noting that most previous works often employed different disease classes, datasets, and dataset split ratios, making direct comparisons challenging. Therefore, our comparison is solely based on the reported accuracies for corresponding classes found in previous literature against those in our dataset. In cases where the number of class samples in the test set is included, we have stated this information as well. For each previous work, Table 3 reports the best F1 score for each class, as well as the overall F1 score across all classes, bearing in mind that the entire set of classes in the concerned datasets differs.

Table 3. Class-wise comparison of F1 scores (%) of previous works on different apple disease datasets (number of instances per class in brackets).

Work	[16]	[8]	[15]	[7]	**Ours**
Model	Dense-Net	RegNet	CA-ENet	Xcep.	ResNet
Num. classes	6	5	8	9	6
Alternaria	-	-	-	86 (116)	88.9 (25)
Scab	73.7 (82)	98.9 (46)	99.8 (500)	83 (212)	96.9 (189)
Rot	-	-	98.9 (500)	-	99.7 (186)
Rust	87.1 (42)	99.1 (54)	99.9 (500)	-	100 (83)
Powdery Mildew	-	-	-	86 (85)	92.1 (38)
Healthy	98.5 (127)	98 (49)	97.6 (500)	97 (90)	98.9 (494)
All classes	93.7 (493)	99.2 (260)	98.8 (4000)	78.1 (686)	97.9 (1015)

From Table 3, one can observe the high variation in the class distributions as well as sample sizes across different works/datasets; making direct comparisons challenging. However, the table provides the following insights:

1. The relative dataset sizes and number of samples in each class remain relatively small across most studies.
2. Datasets containing larger sample sizes per class tend to achieve better overall performance, regardless of the number of classes.
3. Our dataset remains significantly larger than most other datasets while maintaining approximately 50% coverage of the classes present in other datasets.
4. The two most well-predicted classes are Cedar Rust and Healthy leaves while Alternaria and Powdery Mildew are scarcely represented in most datasets.

It is evident from the comparative analysis that the PV-ALE dataset poses a more challenging and realistic benchmark for apple leaf disease classification. While previous works have reported high accuracies on specific disease classes, their evaluations were often conducted on datasets with limited complexity. The inclusion of low-resolution images, leaf clusters, and complex backgrounds in the PV-ALE dataset introduces diversities that are more representative of the real world. Furthermore, the discrepancies in the reported accuracies across different classes highlight the importance of a comprehensive and diverse dataset. Certain disease classes, such as *Cedar Apple Rust*, appear less challenging due to their distinct visual cues unlike *Alternaria Leaf Spot* which has subtle visual manifestations or limited representation in the dataset.

These observations underscore the need for standardized and comprehensive datasets that encompass a wide range of disease types and imaging conditions. The PV-ALE dataset addresses this need by providing a diverse and challenging benchmark that can facilitate the development of robust and generalizable Computer Vision models for accurate apple leaf disease classification. It is important to note that while direct comparisons with previous works are difficult due to the aforementioned differences in datasets and experimental setups, the comparative analysis serves to highlight the advancements and challenges introduced by the PV-ALE dataset.

5 Conclusion

This work presents PV-ALE, a more diverse and comprehensive Apple Disease dataset over the PlantVillage dataset [12]. Extensive experiments were conducted using transfer learning with ResNet50 and a custom CNN model. The results demonstrated that while both models performed excellently on the two datasets, PV-ALE seemed more challenging. We conducted a class-wise comparison with previous works that employed various datasets, revealing a need for increased diversity in existing datasets. The PV-ALE dataset addresses this need by incorporating two new classes that are underrepresented in other datasets, thereby introducing much-needed diversity and complexity. A major limitation of this work is the relatively small size of the test set, which is a consequence of the size and distribution of the original dataset. In future works, we aim to further enhance the dataset by increasing the size of each class and covering a broader range of apple disease types.

Acknowledgement. Kolawole Adebayo has been supported by Enterprise Ireland's CareerFit-Plus Co-fund and the European Union's Horizon 2020 research and innovation programme Marie Skłodowska-Curie Grant No. 847402.

References

1. Albattah, W., Nawaz, M., Javed, A., Masood, M., Albahli, S.: A novel deep learning method for detection and classification of plant diseases. Complex & Intelligent Systems pp. 1–18 (2022)

2. Bansal, P., Kumar, R., Kumar, S.: Disease detection in apple leaves using deep convolutional neural network. Agriculture **11**(7), 617 (2021)
3. Deng, J., Dong, W., Socher, R., Li, L.J., Li, K., Fei-Fei, L.: Imagenet: A large-scale hierarchical image database. In: Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on. pp. 248–255. IEEE (2009), https://ieeexplore.ieee.org/abstract/document/5206848/
4. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 770–778 (2016)
5. Huang, G., Liu, Z., Van Der Maaten, L., Weinberger, K.Q.: Densely connected convolutional networks. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 4700–4708 (2017)
6. Jiang, P., Chen, Y., Liu, B., He, D., Liang, C.: Real-time detection of apple leaf diseases using deep learning approach based on improved convolutional neural networks. IEEE Access **7**, 59069–59080 (2019)
7. Khan, A.I., Quadri, S., Banday, S., Shah, J.L.: Deep diagnosis: A real-time apple leaf disease detection system based on deep learning. computers and Electronics in Agriculture **198**, 107093 (2022)
8. Li, L., Zhang, S., Wang, B.: Apple leaf disease identification with a small and imbalanced dataset based on lightweight convolutional networks. Sensors **22**(1), 173 (2021)
9. Lineham, V., Thorpe, S., Andrews, N., Kim, Y., Beaini, F.: Food Demand to 2050: Opportunities for Australian Agriculture, Algebraic Description of Agrifood Model. Abares (2012)
10. Ristaino, J.B., Anderson, P.K., Bebber, D.P., Brauman, K.A., Cunniffe, N.J., Fedoroff, N.V., Finegold, C., Garrett, K.A., Gilligan, C.A., Jones, C.M., et al.: The persistent threat of emerging plant disease pandemics to global food security. Proceedings of the National Academy of Sciences **118**(23), e2022239118 (2021)
11. Si, H., Li, M., Li, W., Zhang, G., Wang, M., Li, F., Li, Y.: A dual-branch model integrating cnn and swin transformer for efficient apple leaf disease classification. Agriculture **14**(1), 142 (2024)
12. Thapa, R., Zhang, K., Snavely, N., Belongie, S., Khan, A.: The plant pathology challenge 2020 data set to classify foliar disease of apples. Applications in plant sciences **8**(9), e11390 (2020)
13. Tian, Y., Li, E., Liang, Z.: Diagnosis of typical apple diseases: a deep learning method based on multi-scale dense classification network. Frontiers in Plant Science **12**, 698474 (2021)
14. Tian, Y., Yang, G., Wang, Z., Li, E., Liang, Z., et al.: Detection of apple lesions in orchards based on deep learning methods of cyclegan and yolov3-dense. Journal of Sensors **2019** (2019)
15. Wang, P., Niu, T., Mao, Y., Zhang, Z., Liu, B., He, D.: Identification of apple leaf diseases by improved deep convolutional neural networks with an attention mechanism. Frontiers in Plant Science **12**, 723294 (2021)
16. Zhong, Y., Zhao, M.: Research on deep learning in apple leaf disease recognition. Computers and electronics in agriculture **168**, 105146 (2020)