

Article

Plant Leaf Disease Detection Using Deep Learning: A Multi-Dataset Approach

Manjunatha Shettigere Krishna ¹, Pedro Machado ¹, Richard I. Otuka ¹, Salisu W. Yahaya ¹,
Filipe Neves dos Santos ² and Isibor Kennedy Ihianle ^{1,*}

¹ Department of Computer Science, Nottingham Trent University, Nottingham NG1 4FQ, UK; manjunatha.shettigerekrishtna2023@my.ntu.ac.uk (M.S.K.); pedro.machado@ntu.ac.uk (P.M.); richard.otuka@ntu.ac.uk (R.I.O.); salisu.yahaya@ntu.ac.uk (S.W.Y.)

² INESC TEC, Campus da FEUP, Rua Dr Roberto Frias, 4200-465 Porto, Portugal; fbsantos@inesctec.pt

* Correspondence: isibor.ihianle@ntu.ac.uk

Abstract: Agricultural productivity is increasingly threatened by plant diseases, which can spread rapidly and lead to significant crop losses if not identified early. Detecting plant diseases accurately in diverse and uncontrolled environments remains challenging, as most current detection methods rely heavily on lab-captured images that may not generalise well to real-world settings. This paper aims to develop models capable of accurately identifying plant diseases across diverse conditions, overcoming the limitations of existing methods. A combined dataset was utilised, incorporating the PlantDoc dataset with web-sourced images of plants from online platforms. State-of-the-art convolutional neural network (CNN) architectures, including EfficientNet-B0, EfficientNet-B3, ResNet50, and DenseNet201, were employed and fine-tuned for plant leaf disease classification. A key contribution of this work is the application of enhanced data augmentation techniques, such as adding Gaussian noise, to improve model generalisation. The results demonstrated varied performance across the datasets. When trained and tested on the PlantDoc dataset, EfficientNet-B3 achieved an accuracy of 73.31%. In cross-dataset evaluation, where the model was trained on PlantDoc and tested on a web-sourced dataset, EfficientNet-B3 reached 76.77% accuracy. The best performance was achieved with the combination of the PlanDoc and web-sourced datasets resulting in an accuracy of 80.19% indicating very good generalisation in diverse conditions. Class-wise F1-scores consistently exceeded 90% for diseases such as apple rust leaf and grape leaf across all models, demonstrating the effectiveness of this approach for plant disease detection.

Keywords: plant disease; machine learning (ML); deep learning (DL); precision farming; DenseNet; ResNet; EfficientNet; convolutional neural network (CNN)



Academic Editor: Pietro Cipresso

Received: 8 November 2024

Revised: 27 December 2024

Accepted: 7 January 2025

Published: 15 January 2025

Citation: Krishna, M.S.; Machado, P.; Otuka, R.I.; Yahaya, S.W.; Neves dos Santos, F.; Ihianle, I.K. Plant Leaf Disease Detection Using Deep Learning: A Multi-Dataset Approach. *J* **2025**, *8*, 4. <https://doi.org/10.3390/j8010004>

Copyright: © 2025 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Plant diseases pose a significant threat to agricultural productivity and food security around the world. Crops such as wheat, rice, maize, and potatoes are staple foods for a large portion of the world's population, and their yield and quality are heavily affected by various plant diseases [1–3]. Plant diseases are responsible for an estimated 20–40% reduction in yields for major food and cash crops, causing significant economic and social consequences. Globally, the annual loss due to plant diseases is valued at approximately \$220 billion [3]. The Food and Agriculture Organization (FAO) reports that nearly 690 million people faced hunger worldwide in 2019, a figure expected to increase due to the impact of the

COVID-19 pandemic [4]. The accurate and timely detection of these diseases is crucial for implementing effective management strategies and minimising crop losses.

The development of affordable and accessible plant disease detection tools has the potential to empower farmers, reduce crop losses, and enhance livelihoods. Traditionally, plant disease detection has relied on manual inspection by trained experts, a process that is time-consuming, labour-intensive, and prone to human error. With advancements in digital technologies, there is increasing interest in autonomous systems that utilise computer vision and machine learning to detect and monitor plant diseases. They offer rapid, reliable, and cost-effective solutions for large-scale agricultural applications and precision agriculture [5–7].

Deep learning techniques have been effective in object detection and image classification. CNNs perform well in object detection, image classification, and segmentation by automatically learning hierarchical features from raw image data, making them well-suited for capturing complex patterns associated with plant diseases [5,8,9]. Attention-based models further enhance performance by focusing on relevant features while suppressing irrelevant ones. For example, a model integrating a path-based two-layer tree classifier, replacing the traditional softmax, organises plant species hierarchically in a coarse-to-fine manner and has shown significant improvements in large-scale species recognition tasks [10]. Similarly, an Incremental Region Proposal Network addresses challenges such as background noise and overlapping features using a cascaded two-layer architecture. This approach enables accurate apple leaf disease detection, improving the efficiency of disease monitoring in real-world environments [11]. Furthermore, integrating plant disease detection with other precision agriculture technologies—including weather forecasting, soil monitoring, and crop yield prediction—could provide farmers with actionable insights, facilitating data-driven decision-making, optimising resource allocation, and supporting sustainable farming practices [12,13].

The development of deep learning-based plant disease detection techniques and models presents transformative opportunities for agriculture, contributing significantly to global food security and addressing multiple Sustainable Development Goals (SDGs), including SDG #1 (No Poverty) by enhancing agricultural productivity and economic stability for farmers, SDG #2 (Zero Hunger) through improved crop yields and food availability, SDG #3 (Good Health and Well-being) by reducing the risk of food shortages and malnutrition, and SDG #12 (Responsible Consumption and Production) by minimising food waste through early disease detection and management [14]. These techniques could enable informed decision-making, reduce crop losses, and improve farmer livelihoods by providing accurate, timely, and cost-effective disease diagnoses. However, several challenges must be addressed to fully realise their potential, including dealing with data diversity, integration to precision agriculture, overcoming socio-economic barriers to technology access, and addressing cultural perceptions affecting adoption. This paper aims to explore some of these challenges by addressing the following questions: What factors influence the performance of deep learning models for plant disease detection? How can issues related to data diversity be effectively managed? What accuracy thresholds are necessary for practical implementation in real-world farming contexts? To answer these questions, this paper develops and evaluates deep learning models for plant disease detection using diverse datasets, advanced CNN architectures, and strategies to enhance resilience and performance. The main contributions include:

- A new approach to augmenting datasets by using web-sourced data from platforms like Google, enhancing the diversity and realism of plant disease images.

- An evaluation of various strategies for fusing existing datasets with online-sourced data, aimed at improving the generalisation, resilience and adaptability of disease detection models.
- A comprehensive assessment of several advanced CNN architectures, including EfficientNet-B0, EfficientNet-B3, ResNet50, and DenseNet201, using the newly built dataset.
- The introduction of a constructed dataset, combining PlantDoc and web-sourced data, made publicly available to foster further research and development in plant disease detection.

This article is organised into several sections. Section 2 provides an overview of related work, followed by Section 3, which outlines the proposed methodology. Section 4 presents the experiments and discussions of the results, and finally, Section 5 concludes the paper.

2. Related Works

Plant leaf diseases have traditionally been detected and analysed manually in agricultural practices. While effective in certain cases, these methods are often expensive, labour-intensive, time-consuming, and highly subjective, leading to inconsistent results and inefficiencies [15,16]. With technological advances, researchers have increasingly turned to technology-driven and automated solutions to improve the accuracy and speed of plant disease detection. Recently, numerous techniques have emerged utilising computer vision, image processing, pattern recognition, and machine learning to detect and classify plant diseases more effectively [3,7,16]. These techniques analyse key image features such as texture, colour, shape, and size to detect anomalies associated with specific plant diseases. Furthermore, various models have been developed to automatically assess plant health, greatly reducing the limitations posed by manual techniques [3,6,7,16,17]. These techniques have shown great promise in overcoming the challenges of traditional methods, offering more consistent, objective, and scalable solutions.

Machine learning (ML) and deep learning (DL) techniques have gained significant traction in plant disease detection, offering a more efficient and accurate alternative to traditional manual methods [3,6,7,17,18]. Studies have shown the potential of machine learning (ML) techniques in the classification and segmentation of plant lesions from digital images. For example, traditional ML algorithms including the C4.5 classifier, tree bagger, and support vector machines (SVMs) have been applied to a wide range of tasks, including identifying plant lesions, predicting crop yields, and optimising plant growth [19–22]. Akbar et al. [21] proposed how these algorithms can detect disease patterns at early stages, making them particularly effective in preventing outbreaks. Similarly, Yoosefzadeh-Najafabadi et al. [22] and Cedric et al. [23] demonstrated the application of the tree bagger and C4.5 classifiers in disease identification and pest detection, proving their utility in maximising agricultural yield.

Jeny et al. [24] developed an SVM-based model to identify two types of diseases in cucumber leaves using a statistical pattern recognition approach to segment healthy and infected areas. A related SVM-based model for rice plants was presented by Yao et al. [25] in a different study. The authors applied a colour transformation technique to differentiate between infected and healthy regions. Accurate classification was made possible by the model's extraction of important characteristics from the sick patches, including colour, texture, and form. Sharma et al. [26] introduced a model for plant disease detection based on image processing that enhanced image quality by combining a multi-layer perception technique with a pre-processing step after segmentation of the sick leaf portions using K-means clustering. The authors were able to provide a more thorough examination of plant health by the extraction of colour and texture data from each cluster. The effectiveness of

the ML algorithms in plant disease detection is further reflected in their reported accuracy rates. For instance, support vector machines achieved an accuracy of 87.6% when trained on colour, texture, and shape features [27]. Other models, such as K-nearest neighbours (KNNs) and decision trees, have similarly demonstrated improved accuracy of 97.00% [28] and decision trees reaching 95.26% [29], further demonstrating the effectiveness of these techniques. Naive Bayes classifiers, when applied to colour and texture features, have also proven effective, achieving a 97% accuracy rate [30]. While traditional ML methods have shown promising results, they have limitations in terms of feature representation and scalability. Handcrafted features may not capture the complex patterns and variations present in leaf images, leading to suboptimal performance [31]. Moreover, these methods often require extensive feature engineering and domain expertise, making them less adaptable to new diseases or plant species [32].

DL approaches are effective in the detection of plant diseases, offering significant advantages over traditional ML methods. CNNs have in particular consistently outperformed traditional ML methods in terms of accuracy and generalisation ability largely due to their ability to automatically extract discriminative features directly from raw pixel data. The performance of CNNs can be attributed to their ability to learn discriminative features directly from raw pixel data, capturing both local and global patterns in leaf images [5,15,17,18,33]. DL techniques have proven particularly effective in handling high-dimensional and complex datasets, such as large-scale image datasets, making them suitable for plant disease detection tasks [5,15]. Unlike traditional manual classification techniques, which struggle to extract deep and intricate features from images, DL models automatically identify and learn abstract patterns and subtle details from images [34]. As such, they are able to handle complex plant images more efficiently and deliver more better performance. For instance, Deep Belief Networks (DBNs), a type of unsupervised DL model composed of stacked Restricted Boltzmann Machines (RBMs), have been successfully applied to detecting plant lesions and pest infestations [35–38]. DBNs extract essential features from plant leaf images and have achieved accuracy rates between 96% and 97.5% for classifying diseases and pests [38]. Furthermore, Recurrent Neural Networks (RNNs) have also demonstrated good performance in achieving high accuracy in identifying plant lesions across various species [18]. DL models are a reliable option for large-scale plant lesion identification since they do not require manual feature engineering and can efficiently identify and categorise various disease signs from photos [39,40]. Mohanty, Hughes, and Salathé [41] conducted one of the earliest AlexNet CNN-based work to analyse a dataset containing 54,306 images of healthy and diseased plant leaves. The model achieved 99.27% accuracy in the detection of 14 different crop species and 26 distinct diseases, demonstrating the potential of DL methods in the detection of plant diseases. Too et al. [42] conducted an in-depth comparison of various deep-learning architectures for plant disease classification. The authors evaluated AlexNet, VGGNet, Inception, ResNet, and DenseNet on the widely used PlantVillage dataset, consisting of 54,309 images of healthy and diseased leaves from 14 crop species [43]. DenseNet121 best compared to other models with 99.75% in accuracy. Recent advances in CNN architectures have also led to improved plant disease detection approaches. Karthik et al. [44] proposed a custom CNN architecture utilising residual connections and attention mechanisms to enhance feature learning and disease localisation. The CNN model was trained on the PlantVillage dataset and achieved an accuracy of 98% demonstrating the potential of hybrid CNN architectures for plant health monitoring. Similarly, Arsenovic et al. [45] designed a custom CNN architecture based on AlexNet to classify 18 different plant diseases using a custom dataset. Despite achieving a lower accuracy of 93.67%, the authors emphasised the critical role of data quality and diversity in improving DL model performance for plant disease classification. Atila et al. [46] proposed

EfficientNet-B3, a lightweight CNN architecture optimised for computational efficiency. EfficientNet-B3 performed better with lower processing needs than previous CNN models when trained on the PlantVillage dataset. Transfer learning, a technique where a model developed for one task is reused as the starting point for a model on a different but related task, has also been successfully applied to plant disease detection. Elhassouny and Smarandache [47], utilised MobileNet-a lightweight CNN architecture, pre-trained on a large dataset, and fine-tuned it on a custom dataset of 10 distinct plant diseases. The model worked by extracting the features from the pre-trained model to an enhanced performance demonstrating the effectiveness of transfer learning in adapting models to new tasks, even when data availability is limited.

While significant progress has been made in using DL techniques for plant disease detection, several challenges remain. The challenges include the need for high-quality, diverse datasets, and improving model generalisation and scalability. Issues related to symptom variability, image background noise, and practical deployment in agricultural settings also requires further exploration.

3. Proposed Methodology

This paper proposes accurate detection of plant diseases in diverse and uncontrolled image environments using advanced CNN architectures. The proposed architecture depicted in Figure 1 has a modular and scalable architecture, utilising cutting-edge deep learning techniques. It is made of four modules, each contributing in the detection pipeline. The first is a data preprocessing module which involves preparing and cleaning the input data. This is followed by the deep learning module for model training and fine-tuning. As part of this, the models are trained and evaluated to assess their performance using various metrics, ensuring both effectiveness and generalisability.

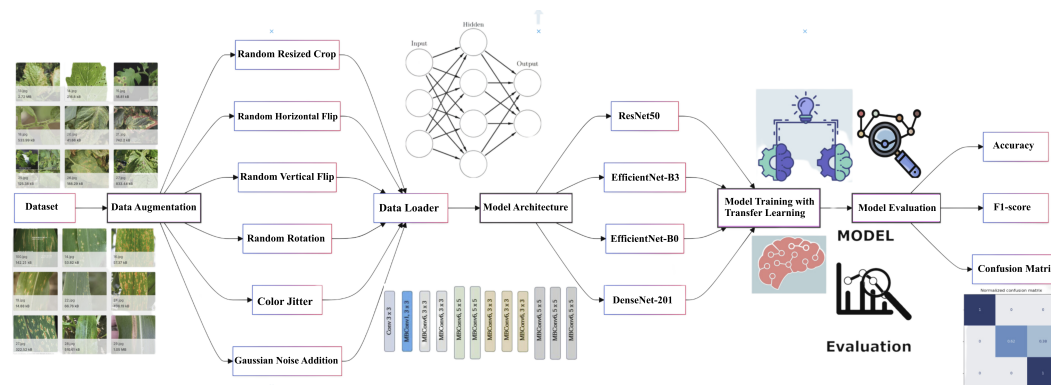


Figure 1. The proposed architecture of the plant disease detection system, illustrating the modular design with four key components: data preprocessing, model architecture, training, and evaluation modules.

3.1. Dataset and Data Collection Process

To enhance model generalisation and improve performance, a novel approach to dataset creation by combining multiple sources. The primary datasets include the PlantDoc dataset (PlantDoc dataset: <https://github.com/pratikkayal/PlantDoc-Dataset>, last accessed 1 November 2024), and a new dataset (Web-sourced dataset: <https://doi.org/10.5281/zenodo.14051480>). The proposed approach provides broader coverage of plant disease scenarios, balancing controlled and real-world conditions essential for generalisation of ML models for plant disease detection.

3.1.1. PlantDoc Dataset

The PlantDoc dataset consists of high-quality images from 27 plant disease classes—see Figure 2 for the distribution and make-up of the dataset. It has been widely used in plant disease detection research as it is well-curated and consistently labelled images—also see Figure 3a,b for sample images of the dataset. A potential limitation of the dataset is that most images were collected in controlled environments, which could hinder a model's ability to generalise its performance to the varied and intricate ways plant diseases manifest in the fields.

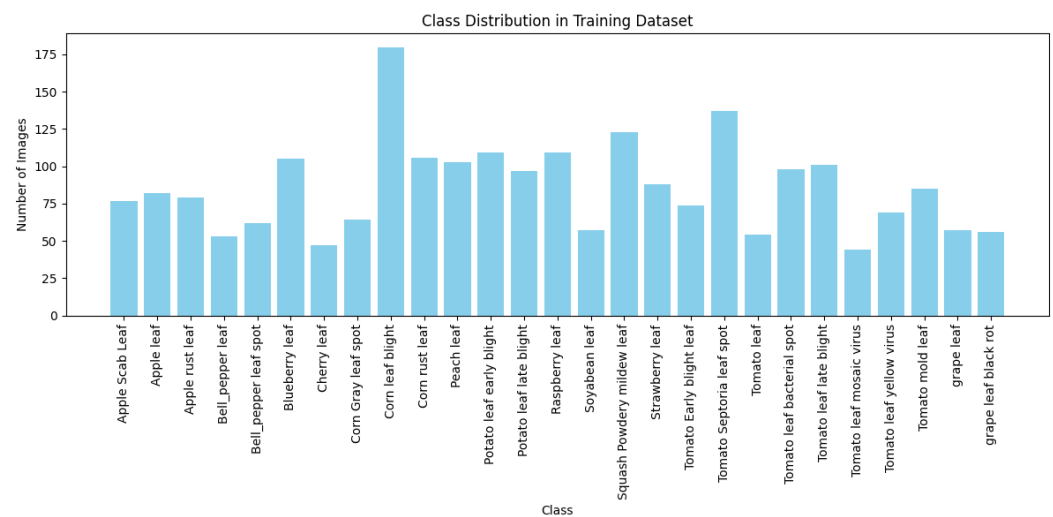
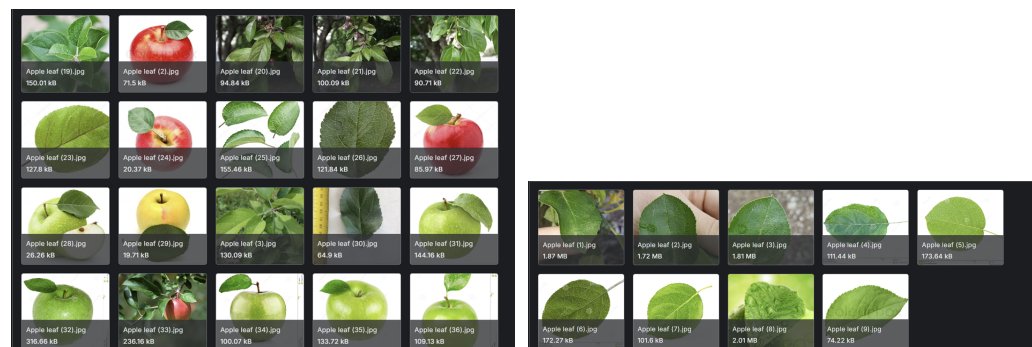


Figure 2. Class distribution of PlantDoc dataset.



(a) Train subset of apple leaf class

(b) Test subset of apple leaf class

Figure 3. PlantDoc apple leaf class.

3.1.2. Web-Sourced Dataset

To overcome the limitations with existing datasets in PlantDoc, a new dataset was compiled by scraping images from various online platforms, including Google Images, using a custom API. A custom file scraper was developed to utilise Google's Custom Search API for scraping images of plant leaves exhibiting various disease symptoms. With the file scraper, a large diverse set of leaf images from multiple websites were collected. The initial output from the scraping process contained a mixture of relevant and irrelevant content. As a result, it was essential to filter out images that were faulty, irrelevant, or of low quality. This process ensured that only high-quality, relevant images remained in the dataset. The final web-sourced dataset comprised of approximately 3000 images representing 16 plant disease classes, as illustrated in Figures 4 and 5. The figures illustrate the instances from the training and test subsets, capturing a broad spectrum of real-world

conditions. It includes variations in lighting, background complexity, plant growth stages, and environmental factors.

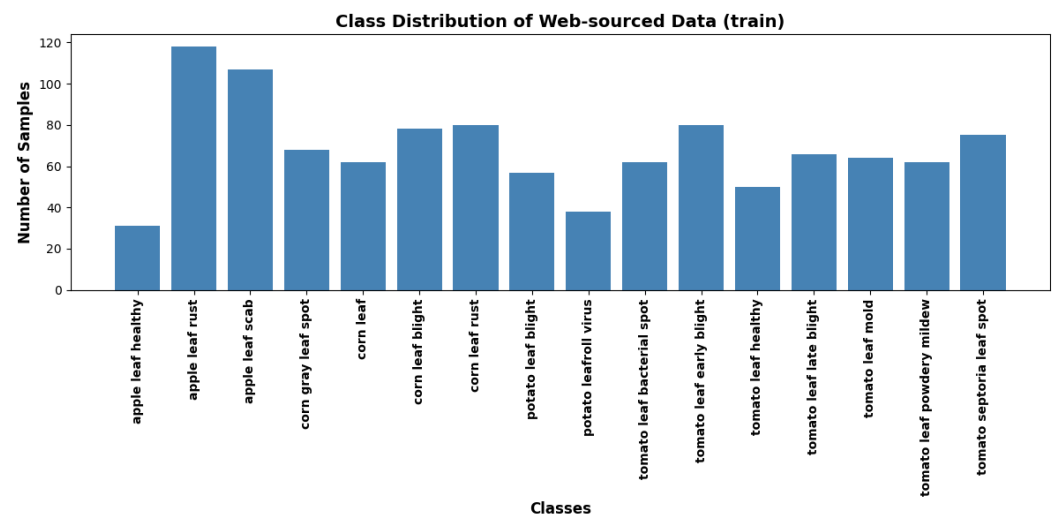


Figure 4. Class distribution of web-sourced data (train).

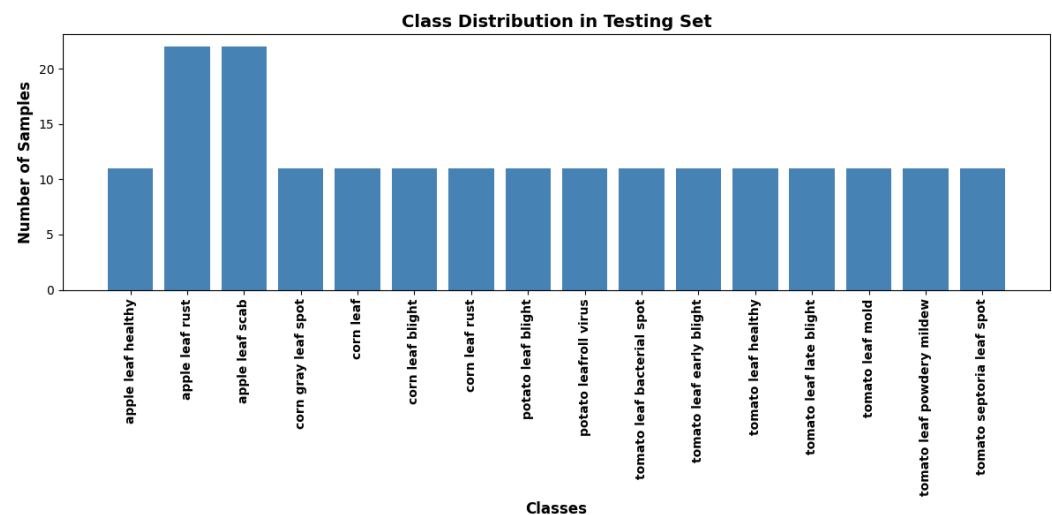


Figure 5. Class distribution of web-sourced dataset (test).

3.1.3. Combined Dataset to Enhance Diversity

To achieve a more comprehensive representation of real-world agricultural environments, the PlantDoc and web-sourced datasets were merged into a unified dataset. This combined dataset enhances model generalisation by exposing it to a wider array of image characteristics, disease manifestations, and background variations. It is comprised of approximately 5600 images across 27 distinct plant disease classes, as illustrated in Figures 6 and 7, offering a more realistic and diverse training set for model development.

While datasets like PlantDoc and PlantVillage are popular in plant disease detection research, they present several limitations that necessitate a more refined approach. A notable issue in the PlantDoc dataset is the inconsistency in image content across the training and testing sets. For example, the “Apple leaf” class features numerous images of whole apples in the training set, while the test set contains only close-up images of leaves. This inconsistency hinders a model’s ability to accurately detect diseases and necessitates a consistent focus on relevant plant parts within the dataset. PlantVillage typically consist of images captured in controlled environments, characterised by uniform backgrounds and optimal lighting conditions. The settings limit the model’s ability to

generalise to field conditions, where plants may display disease symptoms influenced by various environmental factors. The combined dataset features a blend of high-quality, controlled images from PlantDoc and real-world images sourced from the web, exhibiting variations in background complexity, lighting conditions, and image quality. Images from the combined dataset can be seen in Figure 8.

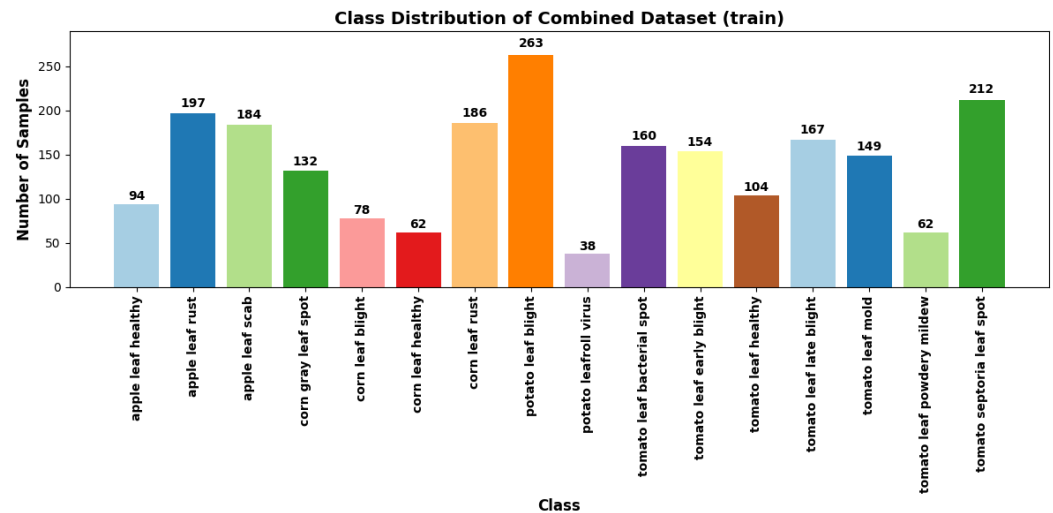


Figure 6. Class distribution of combined dataset (train).

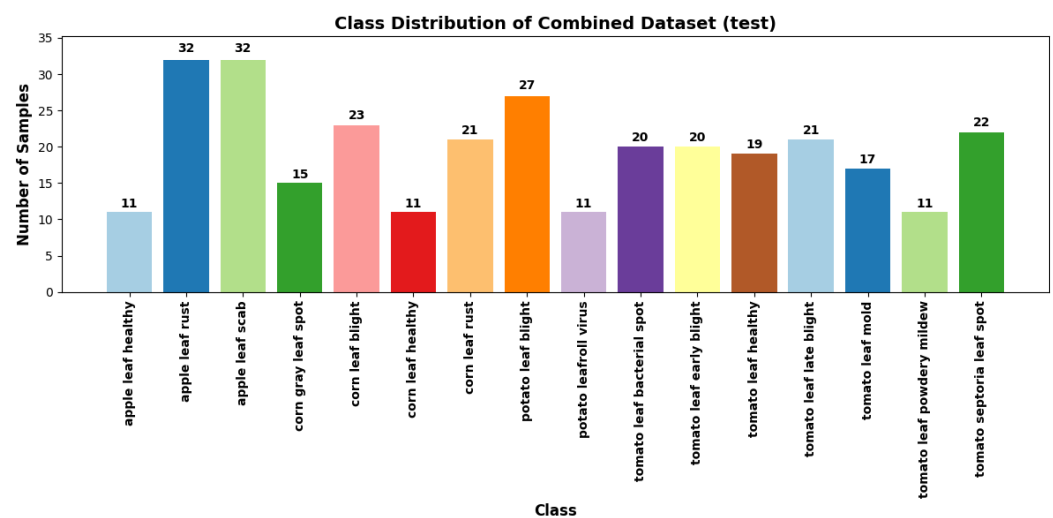


Figure 7. Class distribution of combined dataset (test).

3.2. Data Processing

To process the combined PlantDoc and web-sourced datasets, images were resized to 224×224 pixels. This standard size, compatible with pre-trained models like ResNet and EfficientNet, ensures consistent input dimensions while preserving sufficient detail for disease detection and streamlines the learning process. The images were subsequently normalised using ImageNet statistics, pre-training of the models done to enhance performance and accelerate training convergence.

Data augmentation techniques were crucial for enhancing model resilience and adaptability. Random resized cropping was applied to simulate variability in leaf positioning and camera distance, thereby increasing the model's adaptability. Random horizontal and vertical flips were included as plant diseases are generally invariant to orientation, further diversifying the dataset without compromising realism. Random rotations of up to 20 degrees mimicked natural variations while preserving discernible disease features.

Colour jitter, with modest adjustments to account for lighting fluctuations, was implemented. Gaussian noise with a mean of 0 and a standard deviation of 0.05 was added to simulate real-world image degradation, striking a balance between image degradation and the preservation of critical features.

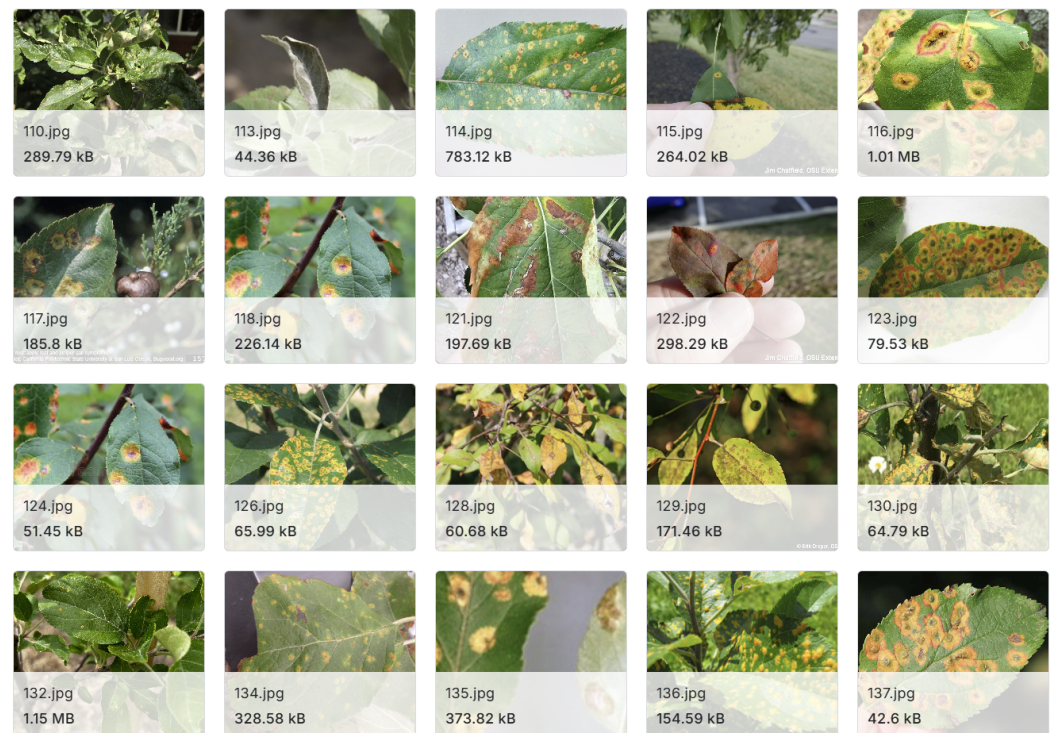


Figure 8. Apple leaf rust (combined dataset).

3.3. Model Architecture

The pipeline, as illustrated in Figure 9, includes image processing techniques, data augmentation, model training, testing, and evaluation. Four distinct CNN architectures from PyTorch's torchvision models are utilised, each contributing unique capabilities to enhance the performance of plant disease classification. The pipeline outlines the image processing steps, the augmentation strategies implemented to enhance model generalisation, and the training and evaluation processes to assess performance. The CNN architectures include the following:

ResNet50 [48], consisting of 50 layers and residual connections to mitigate the vanishing gradient problem common in deep networks. The ResNet50 architecture integrates bottleneck blocks with each block made of a 1×1 convolution to reduce the channel size, a 3×3 convolution for spatial features, and another 1×1 convolution to restore the channel size. Key features include batch normalisation after each convolution, which enhances training stability and speed, and global average pooling that replaces fully connected layers, reducing the number of parameters and thus computational complexity. The architecture also consists of four residual stages with varying numbers of bottleneck blocks: Layer1 with 3 blocks, Layer2 with 4 blocks, Layer3 with 6 blocks, and Layer4 with 3 blocks, followed by a final global average pooling and a fully connected layer. The output layer is adjusted to align with the specific classification task by replacing the default fully connected layer with an nn.Linear layer tailored to the number of plant disease classes.

The **EfficientNet-B3** [49] architecture uses Mobile Inverted Bottleneck Convolution (MBConv) blocks with Squeeze-and-Excitation (SE) blocks. These blocks consist of a pointwise 1×1 convolution, a depth-wise 3×3 convolution, and another 1×1 pointwise convolution, followed by an SE module that adaptively recalibrates channel-wise feature

responses. The EfficientNet-B3 as designed has a deeper architecture compared to the smaller EfficientNet-B0, with more MBConv blocks and a larger input resolution. This scaling approach allows for better model performance at the cost of increased computational requirements, making it suitable for tasks where accuracy is necessary as a metric. As with ResNet50, the final layer is customised to output the correct number of plant disease classes.

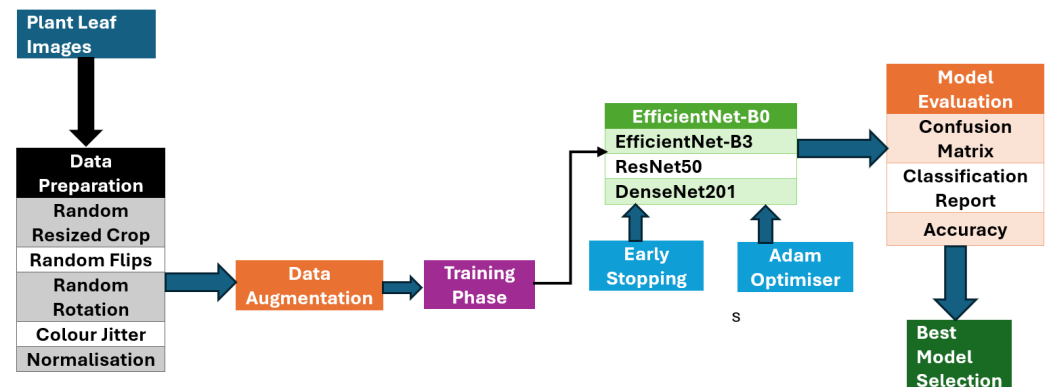


Figure 9. Pipeline illustrating the image processing techniques, data augmentation, model training, testing, and evaluation steps.

The **DenseNet201** [50] architecture has dense connections with each layer receiving inputs from all preceding layers, facilitating better gradient flow and reducing the risk of overfitting. The architecture consists of four dense blocks, each with multiple layers where the output of each layer is concatenated with its input. Each layer within the dense block follows the structure of batch normalisation, ReLU activation, a 1×1 convolution, another batch normalisation, ReLU, and a 3×3 convolution. Transition layers between dense blocks provide downsampling through 1×1 convolutions and 2×2 average pooling. The output of the final dense block is passed through global average pooling followed by a fully connected layer, which is customised for the plant disease classification task.

The **EfficientNet-B0** [49] architecture serves as the baseline for the proposed approach and comes with a more computationally efficient design suitable for environments with constrained resources. The architecture retains the essential features of EfficientNet, such as MBConv blocks and SE modules, but operates with a smaller network size. The model structure follows a similar pattern to EfficientNet-B3, but with fewer layers and a reduced network size, making it ideal for tasks that require faster computation and lower resource consumption without trading off accuracy.

Experiments were conducted in parallel to compare model performance using both original and images with added Gaussian noise. This allows for a direct comparison of model performance under different conditions, especially in noisy environments. The selection of Gaussian noise with a mean of 0 and a standard deviation of 0.05 is based on the need to simulate a realistic amount of noise without overwhelming the models. Gaussian noise with a mean of 0 ensures that the noise distribution does not introduce any bias, while a standard deviation of 0.05 strikes a balance between adding sufficient perturbation to test model resilience and not excessively distorting the images, which would limit the models' ability to learn meaningful patterns.

A batch size of 64 is selected to balance the memory constraints with the stability of gradient estimates during training. This batch size ensures that the model can process enough data at once to optimise the learning process, while also preventing excessive memory use that could slow down training. To address the class imbalance present in the web-sourced dataset, inverse frequency weighting was applied to the loss function. This approach enhances the model sensitivity to under-represented classes, ensuring that

the performance across all classes remains equitable, particularly for plant disease classes with fewer samples. A custom function, `create_model`, was developed to instantiate the models with an appropriate output layer corresponding to the number of plant disease classes. The final fully connected layer was replaced to ensure alignment with the specific classification task.

4. Experiments and Results

The experiments and results are structured around three key evaluations, each aimed at testing different dimensions of the model's performance and its ability to generalise across diverse datasets. These experiments were designed to evaluate the resilience and adaptability of the model across diverse datasets, including both controlled and real-world image variations. The overall objective was to measure the effectiveness of the noise augmentation technique, the impact of data augmentation strategies, and the model's ability to generalise across different datasets.

The first experiment involves within-dataset evaluation, where the model is trained and tested separately on the PlantDoc dataset. The experiment provides a baseline measure of performance on controlled (PlantDoc) and real-world (web-sourced) image collections, helping to assess how well the model handles variability in image quality and background conditions. In the second experiment, a cross-dataset evaluation is conducted. The model is trained on the PlantDoc and tested on the other web-sourced dataset, and vice versa. This experiment is to demonstrate the model's generalisation when exposed to data from a different source, which simulates real-world deployment scenarios where the model may encounter unseen variations in image characteristics. In the third experiment, the combined dataset is evaluated by training the model on the merged dataset (PlantDoc + web-sourced) and testing it on the web-sourced dataset alone. The approach was designed to explore whether the increased diversity of training data enhances the model's performance on real-world images by providing exposure to both controlled and uncontrolled conditions during training.

Performance metrics such as accuracy, precision, recall, and F1-score are used to evaluate the model performance. However, for the sake of consistency, the experimental results presented in this paper primarily focuses on accuracy and F1-score as performance metrics, supplemented by confusion matrices. Accuracy has been chosen as the primary metric because it provides a clear and straightforward measure of the model's overall effectiveness in correctly classifying images, making it suitable for comparing performance across different datasets and scenarios. The F1-score complements accuracy by balancing precision and recall. This is especially important in disease detection tasks, where false negatives can lead to missed detections, and false positives can result in unnecessary interventions. The confusion matrix further provides insights into specific misclassification patterns, highlighting strengths and weaknesses across different classes.

4.1. Experiment 1: Training and Testing with PlantDoc Dataset

The performance of four pre-trained models, ResNet50, EfficientNet-B3, DenseNet201, and EfficientNet-B0 were evaluated on the PlantDoc dataset in this experiment. The evaluation was conducted under two conditions: without noise augmentation and with the addition of Gaussian noise to the input images. This approach enabled an assessment of both the baseline performance of the models and their generalisability to image perturbations.

As shown in Table 1 and Figure 10, EfficientNet-B0 performed best without noise, which can be attributed to their compound scaling method. This method uniformly scales network width, depth, and resolution, allowing EfficientNet to maintain an optimal balance between model size and accuracy, outperforming the traditional scaling approaches with

the ResNet and DenseNet architectures. ResNet50, despite its deep architecture, showed the lowest performance among the models. DenseNet201, with its dense connectivity pattern, performed well, although not as effectively as the EfficientNet models. While the dense connections facilitate feature reuse and alleviate the vanishing gradient problem, they may have introduced redundancy in the feature maps, potentially limiting its performance compared to the more efficient architectures of the EfficientNet models.

Table 1. Model performance on PlantDoc dataset.

Model	Accuracy (Without Noise)	Accuracy (With Noise)
ResNet50	0.5508	0.6229
EfficientNet-B3	0.7163	0.7119
DenseNet201	0.7034	0.6695
EfficientNet-B0	0.7203	0.6737

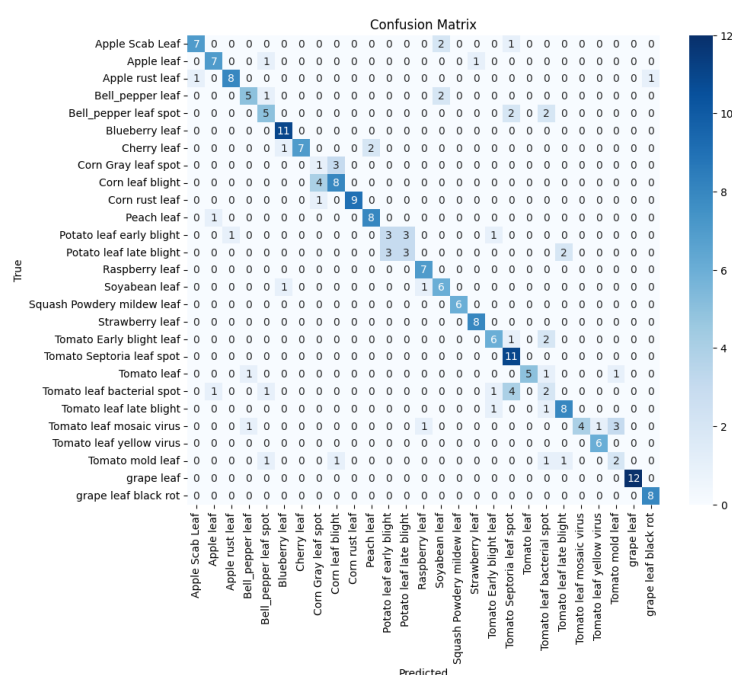


Figure 10. Confusion matrix for EfficientNet-B0 (PlantDoc).

To further understand model behaviour, class-wise analysis was performed using F1-scores as presented in Table 2. The results show that several classes, including raspberry leaf, squash powdery mildew leaf, and grape leaf, demonstrated consistently high F1-scores across all models which can be attributed to the distinctive features even under noise perturbations. EfficientNet models, particularly B3 and B0, exhibited greater resilience to noise, as evidenced by their stable performance in categories like corn rust leaf and strawberry leaf, maintaining F1-scores above 0.90 in both conditions. In contrast, ResNet50 showed more variability, with notable drops for tomato leaf bacterial spot and corn grey leaf spot, reflecting its sensitivity to noisy inputs. Interestingly, certain classes, such as apple rust leaf, improved with noise, suggesting that Gaussian noise may have acted as a form of augmentation, enhancing feature generalisation. These observations reinforce the capability of EfficientNet models, supporting their suitability for plant disease classification tasks in challenging conditions.

Table 2. Class-wise F1-scores for each model, without noise/with noise.

Class	ResNet50	EfficientNet-B3	DenseNet201	EfficientNet-B0
Apple scab leaf	0.7000/0.7000	0.7500/0.8421	0.8889/0.8889	0.7778/0.7778
Apple leaf	0.6364/0.5455	0.5000/0.5000	0.8000/0.8000	0.7778/0.7778
Apple rust leaf	0.6667/0.8421	0.8889/0.9474	0.9524/0.9524	0.9474/0.8421
Bell pepper leaf	0.5882/0.3750	0.7778/0.7778	0.6667/0.6667	0.6250/0.7059
Bell pepper leaf spot	0.4286/0.3750	0.6667/0.7778	0.5714/0.5714	0.6316/0.6316
Blueberry leaf	0.5000/0.5455	0.7368/0.7619	0.6957/0.6957	0.8696/0.8696
Cherry leaf	0.0000/0.1538	0.7000/0.7000	0.5333/0.5333	0.7059/0.5882
Corn grey leaf spot	0.2000/0.2222	0.2222/0.3636	0.1667/0.1667	0.5714/0.3333
Corn leaf blight	0.6087/0.6087	0.7500/0.6316	0.4762/0.4762	0.9231/0.5263
Corn rust leaf	0.9000/0.8000	0.9474/0.9474	0.9000/0.9000	0.9474/0.9524
Peach leaf	0.6957/0.6400	0.8000/0.8000	0.7059/0.7059	0.7000/0.7000
Potato leaf early blight	0.5882/0.3750	0.5882/0.5714	0.5333/0.5333	0.6667/0.5556
Potato leaf late blight	0.3077/0.3158	0.6250/0.4615	0.4286/0.4286	0.5333/0.4286
Raspberry leaf	1.0000/0.8750	0.9333/0.9333	1.0000/1.0000	1.0000/0.8750
Soya bean leaf	0.6364/0.5714	0.7059/0.7059	0.7059/0.7059	0.8235/0.6154
Squash powdery mildew leaf	0.9091/0.6250	1.0000/1.0000	1.0000/1.0000	0.9091/0.9091
Strawberry leaf	0.8889/0.9412	1.0000/0.9412	0.9412/0.9412	1.0000/0.9412
Tomato early blight leaf	0.2667/0.4706	0.7059/0.5556	0.5882/0.5882	0.6000/0.6000
Tomato Septoria leaf spot	0.5556/0.5263	0.6667/0.6897	0.6875/0.6875	0.6667/0.6780
Tomato leaf	0.5714/0.5882	0.6250/0.5714	0.6667/0.6667	0.6667/0.5263
Tomato leaf bacterial spot	0.2500/0.1818	0.0000/0.2857	0.1538/0.1538	0.1176/0.2500
Tomato leaf late blight	0.4000/0.5789	0.6250/0.7273	0.7000/0.7000	0.7000/0.7619
Tomato leaf mosaic virus	0.1818/0.3333	0.8235/0.5714	0.1818/0.1818	0.1667/0.1667
Tomato leaf yellow virus	0.7143/0.7692	0.9091/0.7692	0.9091/0.9091	0.6667/0.8333
Tomato mould leaf	0.4286/0.6667	0.5000/0.7407	0.4706/0.4706	0.3750/0.5333
Grape leaf	0.9091/0.9565	1.0000/0.9231	0.9600/0.9600	1.0000/0.9231
Grape leaf black rot	0.7500/0.9333	0.9333/0.9333	1.0000/1.0000	1.0000/0.9333

4.2. Experiment 2: Training with PlantDoc Dataset and Testing with Web-Sourced Dataset

In Experiment 2, model generalization was evaluated by training on the PlantDoc dataset and subsequently testing on the web-sourced data. This evaluation, as presented in Table 3 and Figure 11a,b, aimed to assess the models' ability to transfer knowledge between datasets and their resilience to image perturbations.

Table 3. Model performance on web-sourced data (trained on PlantDoc).

Model	Accuracy (Without Noise)	Accuracy (With Noise)
ResNet50	0.6515	0.6364
EfficientNet-B3	0.7677	0.7424
DenseNet201	0.7323	0.7020
EfficientNet-B0	0.7374	0.7525

EfficientNet-B0 performed nearly as well as EfficientNet-B3 in the absence of noise. Remarkably, it showed the best performance when noise was added, underscoring the capability of the EfficientNet architecture, even at smaller scales, and its ability to generalise well, particularly with image perturbations. DenseNet201 achieved an accuracy of 73.23% without noise, demonstrating strong performance. However, it experienced the largest performance drop when noise was introduced, suggesting that its dense architecture might be more susceptible to image perturbations. ResNet50 with an accuracy of 65.15% without noise, performed worse compared to other models. The results showed the smallest decrease in performance when noise was added, indicating a certain level of generalisability, though at a lower overall accuracy compared to the other models.



(a) EfficientNet-B3 (b) EfficientNet-B0 (with noise)
Figure 11. Confusion matrix for PlantDoc → web-sourced dataset.

The class-wise analysis with F1-scores, as presented in Table 4 shows the performances of the different plant diseases. Apple leaf rust and apple leaf scab consistently showed high F1-scores across all models and noise conditions, indicating distinctive and transferable features. EfficientNet-B3 demonstrated strong performance in classifying potato leafroll virus and tomato leaf healthy, while EfficientNet-B0 performed very well in identifying potato leaf blight and tomato leaf late blight, maintaining stable F1-scores even with noise. In contrast, challenging classes included corn leaf blight and corn leaf healthy, which displayed significantly lower F1-scores, reflecting substantial differences between the PlantDoc and web-sourced datasets. DenseNet201 showed sensitivity to noise in some classes like Tomato leaf early blight. ResNet50 achieved relatively lower accuracy overall and performed well for specific classes such as apple leaf rust and corn leaf rust, suggesting that it may still be effective for certain disease categories despite its general limitations.

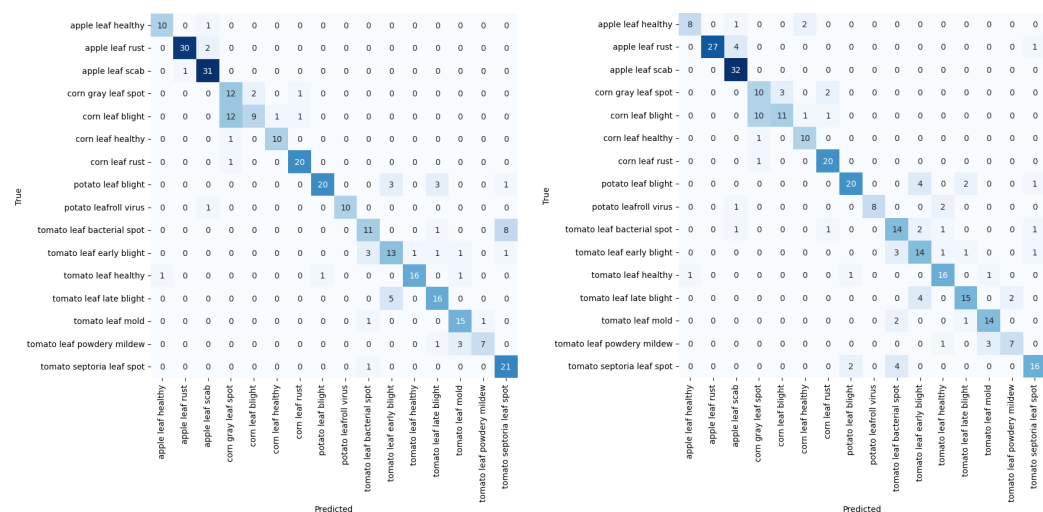
Table 4. Class-wise F1-scores for each model, without/with noise.

Class	ResNet50	EfficientNet-B3	DenseNet201	EfficientNet-B0
Apple leaf healthy	0.7059/0.6667	0.8421/0.8000	0.8421/0.7778	0.8000/0.8182
Apple leaf rust	0.9048/0.8511	0.9524/0.9302	0.9302/0.9130	0.9091/0.9091
Apple leaf scab	0.8750/0.7826	0.8696/0.8750	0.8936/0.8936	0.8696/0.8696
Corn grey leaf spot	0.7407/0.6364	0.7407/0.5238	0.6923/0.5854	0.6400/0.6400
Corn leaf blight	0.0000/0.0000	0.1176/0.1176	0.0000/0.0000	0.1111/0.1111
Corn leaf healthy	0.0800/0.0952	0.0952/0.0952	0.0000/0.0800	0.0870/0.0870
Corn leaf rust	0.9091/0.8696	0.8696/0.8696	0.8800/0.8696	0.8333/0.8333
Potato leaf blight	0.6667/0.7059	0.7619/0.7619	0.7619/0.7826	0.8571/0.8571
Potato leafroll virus	0.4000/0.5833	0.9000/0.8571	0.6667/0.8571	0.9000/0.9000
Tomato leaf bacterial spot	0.5000/0.3810	0.6667/0.7000	0.6667/0.4865	0.6667/0.6667
Tomato leaf early blight	0.4000/0.5263	0.5556/0.5556	0.5714/0.5714	0.5882/0.5882
Tomato leaf healthy	0.6667/0.5714	0.8696/0.9000	0.7500/0.8333	0.8421/0.8421
Tomato leaf late blight	0.6667/0.5641	0.7273/0.7273	0.7778/0.6667	0.8000/0.8000
Tomato leaf mould	0.8571/0.6667	0.7407/0.7407	0.8696/0.8696	0.8800/0.8800
Tomato leaf powdery mildew	0.7778/0.7059	0.8421/0.7778	0.9000/0.9000	0.9000/0.9000
Tomato Septoria leaf spot	0.6429/0.5373	0.6897/0.7200	0.7200/0.6909	0.7200/0.7200

4.3. Experiment 3: Training with Combined Dataset and PlantDoc, Test with Web-Sourced Dataset

The performance of the models was evaluated after training on a combined dataset from PlantDoc and the web-sourced dataset, and tested on a sample of the web-sourced dataset. See Table 5 and Figure 12a,b for the experimental results for model performance

when training with the combined PlantDoc and web-sourced datasets and testing with the web-sourced data.



(a) EfficientNet-B3

(b) EfficientNet-B3 (with noise)

Figure 12. Confusion matrix (trained on combined, tested on web-sourced).

EfficientNet-B3 performed exceptionally well in both noise-free and noisy conditions, demonstrating effective feature learning and adaptability to image degradation caused by noise. EfficientNet-B0 followed closely, particularly in the noise-free scenario, demonstrating the model's efficiency. DenseNet201 achieved a strong accuracy of 77.00% without noise, indicating good performance and good resistance to noise. ResNet50, while improving to 67.41% accuracy without noise due to more training data, performed worse than the other models in overall performance.

Table 5. Model performance when trained on combined dataset (PlantDoc and web-sourced) and tested on web-sourced data.

Model	Accuracy (Without Noise)	Accuracy (With Noise)
ResNet50	0.6741	0.6486
EfficientNet-B3	0.8019	0.7732
DenseNet201	0.7700	0.7636
EfficientNet-B0	0.7987	0.7677

The impact of noise on model performance showed that all models experienced a decrease in accuracy when noise was introduced, but the drop was less pronounced compared to previous experiments. The improvement suggests that the combined dataset helped the models learn more features, contributing to adaptability and resilience to noise. EfficientNet-B3 performed best under noisy conditions, demonstrating its ability to maintain strong feature learning even in the presence of perturbations. However, ResNet50 displayed the largest relative performance drop, indicating that it remains more sensitive to noise compared to the other models.

A class-wise performance analysis as shown in Table 6 shows that certain disease classes were consistently well-classified by all the models, including apple diseases (healthy, rust, scab), corn leaf healthy, and corn leaf rust, which yielded high F1-scores across both noise conditions. The results suggest that these diseases have distinctive features that are well captured by the combined dataset. Furthermore, challenging classes, such as corn grey leaf spot and corn leaf blight, saw performance improvements but continued to pose classification difficulties, even with the enhanced dataset. EfficientNet-B3 demonstrated

consistently high performance for most classes, both with and without noise. EfficientNet-B0 performed well in particular classes, such as tomato leaf powdery mildew, where it achieved perfect F1 scores. DenseNet201 showed better performance, particularly on several tomato diseases in the absence of noise. ResNet50, while generally underperforming in comparison to the other models, showed marked improvements from earlier experiments, especially for classes like apple leaf healthy and corn leaf healthy.

Table 6. Class-wise F1-scores for each model, without noise.

Class	ResNet50	EfficientNet-B0	EfficientNet-B3	DenseNet201
Apple leaf healthy	0.8182	0.9091	0.9091	0.8421
Apple leaf rust	0.9032	0.9355	0.9524	0.9032
Apple leaf scab	0.8438	0.9254	0.9254	0.8657
Corn grey leaf spot	0.5854	0.5778	0.5854	0.5238
Corn leaf blight	0.4848	0.4242	0.5294	0.4706
Corn leaf healthy	0.9000	0.8571	0.9091	0.8696
Corn leaf rust	0.8889	0.9268	0.9302	0.9048
Potato leaf blight	0.7742	0.8077	0.8333	0.8148
Potato leafroll virus	0.4286	0.8571	0.9524	0.9524
Tomato leaf bacterial spot	0.1935	0.6316	0.6111	0.3125
Tomato leaf early blight	0.4571	0.7556	0.6341	0.7179
Tomato leaf healthy	0.6842	0.8125	0.8889	0.8649
Tomato leaf late blight	0.5641	0.7179	0.7442	0.8095
Tomato leaf mould	0.6667	0.8571	0.8108	0.8485
Tomato leaf powdery mildew	0.7059	1.0000	0.7368	0.9524
Tomato Septoria leaf spot	0.5373	0.7843	0.7925	0.6897

The performance of the models across different plant types, as presented in Table 7, highlights notable trends in classification performance. Apple diseases consistently exhibited strong results across all models in both noise-free and noisy conditions, suggesting that the combined dataset provided rich and generalisable features for detecting apple-related diseases. Corn diseases, which had been more challenging in previous experiments, showed improvement, particularly with the EfficientNet models, where EfficientNet-B3 delivered the most reliable performance under both conditions.

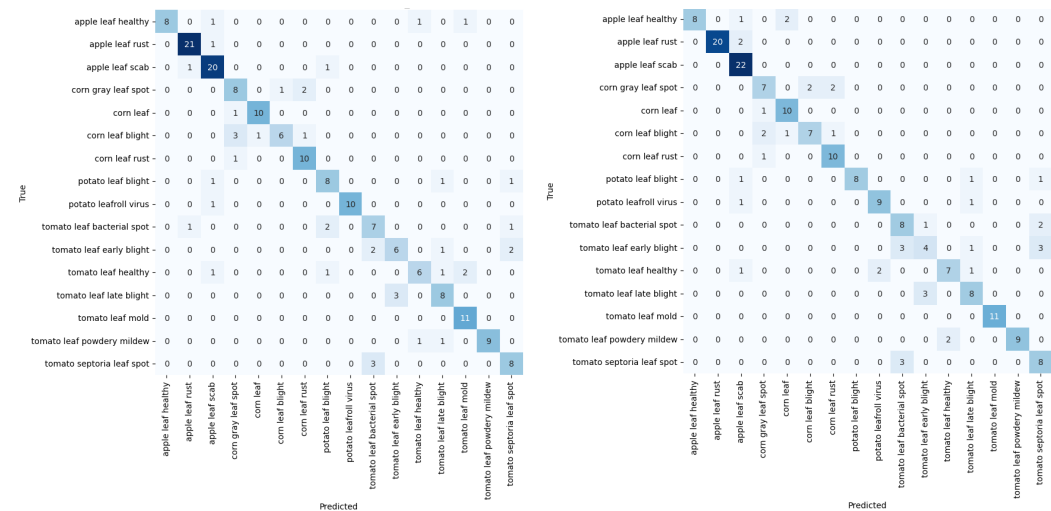
Table 7. Average F1-score by plant type, without noise/with noise.

Plant Type	ResNet50	EfficientNet-B3	DenseNet201	EfficientNet-B0
Apple	0.8551/0.8043	0.9290/0.9290	0.9230/0.8703	0.9233/0.8881
Corn	0.7148/0.5799	0.7385/0.7436	0.6922/0.6913	0.6965/0.7293
Potato	0.6014/0.6863	0.8929/0.8633	0.8836/0.8836	0.8324/0.8324
Tomato	0.5584/0.5829	0.7455/0.7455	0.7422/0.7422	0.7942/0.7942

For potato diseases, there was a marked improvement, with both EfficientNet and DenseNet models achieving high F1 scores and demonstrating effective feature learning and resilience to the presence of noise. These results indicate the models' effectiveness in handling potato disease detection under varying conditions. Tomato diseases were classified with moderate to high accuracy across the models, with EfficientNet-B0 standing out as the best performer. Performance on tomato diseases remained consistent across noise conditions for all models, except for ResNet50, which showed some sensitivity to noise. The results suggest that the models are capable of maintaining their classification capabilities for tomato diseases, regardless of the presence of noise.

4.4. Experiment 4: Training and Testing with Web-Sourced Dataset

The evaluation of model performance, when both trained and tested on the web-sourced dataset, revealed distinct patterns, as shown in Table 8. EfficientNet-B0 performed best without noise, though it exhibited some sensitivity to image perturbations. The confusion matrices in Figure 13a,b provide a detailed breakdown, with EfficientNet-B0's performance compared to EfficientNet-B3 under noisy conditions. EfficientNet-B3 displayed remarkable resilience to noise, indicating its ability to learn noise-invariant features more effectively.



(a) EfficientNet-B0

(b) EfficientNet-B3 (with noise)

Figure 13. Confusion matrix, web-sourced data.

DenseNet201 exhibited improved performance in the presence of noise, suggesting that its architecture may benefit from the regularising effect of noise. In contrast, ResNet50, while showing weaker performance overall, demonstrated better resilience to noise, potentially benefiting from the increased dataset diversity and perturbations introduced during training. This trend suggests that ResNet50's ability to generalise improves when faced with varied input data despite its overall lower accuracy.

Table 8. Model performance on web-sourced dataset (trained and tested on web-sourced dataset).

Model	Accuracy (Without Noise)	Accuracy (With Noise)
ResNet50	0.6616	0.6515
EfficientNet-B3	0.7777	0.7878
DenseNet201	0.7424	0.7828
EfficientNet-B0	0.7878	0.7727

The class-wise analysis of the web-sourced dataset, as shown in Table 9, summarises performance variations across different plant disease categories. For example, apple diseases such as leaf healthy and rust exhibited strong classification results across all models, with EfficientNet-B3 and EfficientNet-B0 showing the highest consistency, even under noisy conditions. In contrast, corn diseases, particularly grey leaf spot and leaf blight, presented challenges for all models, with EfficientNet-B3 generally outperforming the others, especially in noise-free conditions. However, noise appeared to benefit ResNet50, especially with the rust class, where it showed improvements in accuracy, suggesting that noise may support in the model's generalisation ability. Potato diseases, particularly leafroll virus, exhibited inconsistent performance, with DenseNet201 showing better handling of noise compared to other models. Tomato diseases displayed a more nuanced performance, with

EfficientNet-B3 performing better in categories such as leaf mould and early blight, while DenseNet201 performed well across a variety of tomato classes, particularly in noisy environments. On the average performance by plant type shown in Table 10, EfficientNet-B3 consistently outperformed the other models, achieving the highest F1-scores for both apple and potato diseases, even under noisy conditions. In contrast, ResNet50 showed the lowest performance, particularly for potato and tomato diseases, while EfficientNet-B0 exhibited a drop in performance for apple diseases with noise.

Table 9. Class-wise performance on web-sourced dataset, without noise/with noise.

Class	ResNet50	EfficientNet-B3	DenseNet201	EfficientNet-B0
Apple leaf healthy	0.7000/0.8182	0.9091/0.9091	0.8889/0.8889	0.9091/0.8182
Apple leaf rust	0.9062/0.9062	0.9524/0.9524	0.9302/0.9130	0.9091/0.9091
Apple leaf scab	0.8387/0.8387	0.8750/0.8750	0.8444/0.8936	0.8696/0.8696
Corn grey leaf spot	0.5581/0.5581	0.7407/0.6667	0.4444/0.1667	0.6400/0.5714
Corn leaf blight	0.5294/0.5294	0.5946/0.5946	0.5263/0.5263	0.5556/0.6667
Corn leaf healthy	0.8333/0.8333	0.8696/0.8696	0.8571/0.8571	0.8571/0.8571
Corn leaf rust	0.8780/0.8780	0.8696/0.8696	0.8696/0.8696	0.8696/0.8333
Potato leaf blight	0.6667/0.7619	0.7619/0.7826	0.7619/0.8400	0.7273/0.7273
Potato leafroll virus	0.1538/0.5833	0.9000/0.8696	0.7778/0.8182	0.8696/0.9000
Tomato leaf bacterial spot	0.1935/0.3810	0.7000/0.7000	0.4865/0.5217	0.6667/0.6667
Tomato leaf early blight	0.4000/0.5263	0.5556/0.6000	0.5882/0.5882	0.5882/0.5882
Tomato leaf healthy	0.6087/0.5714	0.8000/0.9000	0.5882/0.6667	0.7143/0.7143
Tomato leaf late blight	0.5789/0.6000	0.7273/0.7273	0.7000/0.6977	0.6364/0.7619
Tomato leaf mould	0.7273/0.6667	0.7407/0.7407	0.9565/0.9565	0.9565/0.9565
Tomato leaf powdery mildew	0.9524/0.7778	0.9091/0.7778	0.9091/0.9091	0.9000/0.9000
Tomato Septoria leaf spot	0.6129/0.5926	0.7619/0.6897	0.5926/0.5926	0.7619/0.7200

Table 10. Average F1-score by plant type, without noise/with noise.

Plant Type	ResNet50	EfficientNet-B3	DenseNet201	EfficientNet-B0
Apple	0.8150/0.8544	0.9122/0.9122	0.8878/0.8985	0.8959/0.8656
Corn	0.6997/0.6997	0.7686/0.7501	0.6744/0.6049	0.7306/0.7321
Potato	0.4103/0.6726	0.8310/0.8261	0.7699/0.8291	0.7985/0.8137
Tomato	0.5820/0.5880	0.7368/0.7419	0.6916/0.7046	0.7463/0.7582

4.5. Experiment 5: Training with Web-Sourced Data and Testing with PlantDoc Dataset

This experiment evaluated the models' generalisation capabilities by training them exclusively on the web-sourced dataset and subsequently testing their performance on the PlantDoc dataset. The results, summarised in Table 11, demonstrate that DenseNet201 exhibited consistent performance across both noise-free and noisy conditions. This suggests that DenseNet201 effectively learned features from the web-sourced dataset that generalise well to the distinct characteristics of the PlantDoc dataset.

Table 11. Model performance on PlantDoc dataset (trained on the web-sourced dataset).

Model	Accuracy (Without Noise)	Accuracy (With Noise)
ResNet50	0.5217	0.4957
EfficientNet-B3	0.5826	0.5739
DenseNet201	0.5913	0.5913
EfficientNet-B0	0.5478	0.5652

EfficientNet-B3, while maintaining strong performance, exhibited a weaker performance compared to previous experiments and demonstrated a slight sensitivity to noise in this cross-dataset scenario. EfficientNet-B0, though initially performing lower without noise, showed marginal improvement when noise was introduced, reflecting its adaptability. ResNet50 faced the most significant challenges in this experiment, with the lowest accuracy

overall and a further decline in performance when noise was added, suggesting difficulties in learning transferable features from the web-sourced dataset to the PlantDoc dataset.

The class-wise performance analysis, as shown in Table 12, shows some interesting trends across the models. Corn rust leaf and tomato leaf powdery mildew were consistently well-classified across all models and noise conditions, with EfficientNet-B0 achieving perfect classification for the latter under noisy conditions. However, apple diseases, including apple rust leaf and apple scab leaf, proved challenging for all models, with relatively low classification scores across the board. DenseNet201 performed very well in classifying tomato diseases such as tomato leaf mould and tomato leaf powdery mildew, while EfficientNet-B3 demonstrated superior performance on potato leafroll virus and corn rust leaf. ResNet50, despite generally underperforming, showed notable advantages in classifying healthy tomato leaves and tomato leaf powdery mildew in the absence of noise, reflecting its potential in specific cases.

Table 12. Class-wise F1-scores for each model from Experiment 5, evaluated without noise/with noise.

Class	ResNet50	EfficientNet-B3	DenseNet201	EfficientNet-B0
Apple rust leaf	0.0952/0.1053	0.0000/0.0000	0.1111/0.1111	0.0909/0.0909
Apple scab leaf	0.0000/0.2105	0.0952/0.0952	0.3000/0.2857	0.0000/0.1905
Corn grey leaf spot	0.3636/0.3750	0.3333/0.3333	0.2857/0.1667	0.3636/0.3750
Corn leaf blight	0.2667/0.4706	0.1176/0.1176	0.5263/0.5263	0.6000/0.6667
Corn rust leaf	0.8333/0.8000	0.9474/0.9091	0.9474/0.9048	0.9524/0.9268
Potato leaf blight	0.6667/0.6667	0.8276/0.7742	0.7742/0.8148	0.7742/0.8077
Potato leafroll virus	0.4286/0.4286	0.9000/0.8571	0.9524/0.8182	0.8571/0.9000
Tomato leaf bacterial spot	0.2353/0.2353	0.4706/0.6111	0.2857/0.5217	0.2857/0.6667
Tomato leaf early blight	0.4211/0.4211	0.5217/0.6341	0.5882/0.5882	0.5217/0.5882
Tomato leaf healthy	0.9333/0.9333	0.8333/0.9000	0.9333/0.6667	0.7143/0.7143
Tomato leaf late blight	0.4762/0.4706	0.7619/0.7442	0.7000/0.6977	0.7000/0.7619
Tomato mould leaf	0.6667/0.6667	0.7407/0.8108	0.9565/0.9565	0.3750/0.9565
Tomato leaf powdery mildew	0.9091/0.7778	0.8421/0.7368	0.9524/0.9091	0.9000/1.0000

5. Conclusions

This paper investigated the development of effective and generalisable DL models for plant disease detection, aimed at advancing sustainable agricultural practices. Through a series of carefully designed experiments, the performance of state-of-the-art CNN architectures was evaluated across different datasets and conditions, revealing both challenges and opportunities in this field. The findings significantly contribute to the ongoing research in agricultural technology and provide valuable insights for future advancements and real-world applications. A comparison of results from recent studies utilising the PlantDoc dataset, presented in Table 13, demonstrates the effectiveness of the EfficientNet-B3 model.

Table 13. Comparison of results from other studies.

Study	Method	Dataset	Accuracy (%)
This study	EfficientNet-B3	PlantDoc	73.31
This study	EfficientNet-B3	PlantDoc (train) + web-sourced (test)	76.77
This study	EfficientNet-B3	Combined (train) + web-sourced (test)	80.19
Chandra et al. [51]	SVM with twin network of Siamese	PlantDoc	62.76
Moupojou et al. [52]	MobileNet	PlantDoc	60.14
Menon et al. [53]	MobileNet-V2	PlantDoc	69.80

The EfficientNet-B3 model achieved an accuracy of 73.31% when trained and tested on the PlantDoc dataset, surpassing the performance reported in other recent studies utilising PlantDoc [51–53]. Moreover, the experiments involving cross-dataset and combined dataset evaluations demonstrated even higher accuracies, suggesting the potential of these approaches to significantly enhance real-world plant disease detection and foster more resilient agricultural practices.

Key Findings and Their Implications

Model Performance Across Datasets: The experimental results revealed significant variations in model performance across different dataset combinations. When trained and tested on the PlantDoc dataset, EfficientNet-B3 demonstrated the highest accuracy at 72.46%, closely followed by EfficientNet-B0 at 69.49%. These results highlight the effectiveness of these architectures for plant disease detection within controlled, curated datasets. However, cross-dataset evaluations underscored the challenges of generalizing to real-world conditions. Models trained on PlantDoc and tested on the web-sourced data exhibited a slight performance decrease, with EfficientNet-B3 achieving an accuracy of 76.77%. More notably, when models were trained on the web-sourced dataset and tested on PlantDoc, performance dropped significantly, with DenseNet201 achieving the highest accuracy at only 59.13%.

The results emphasise the critical role of diverse and representative datasets in training models capable of generalisation. The observed gap in performance between within-dataset and cross-dataset evaluations suggests that current models may be overfitting specific dataset features rather than learning generalised characteristics of plant diseases. This has important implications for deploying these models in real-world agricultural environments, where images may differ significantly from training data conditions.

Impact of Dataset Combination: The results from Experiment Section 4.3, where models were trained on a combined dataset (PlantDoc and web-sourced images) and tested on the web-sourced data, offer important insights into the benefits of data augmentation. In this scenario, EfficientNet-B3 achieved the highest accuracy at 80.19%, demonstrating that increased dataset diversity enhances the resilience, adaptability and generalisation of the models. These findings underscore the importance of developing large, diverse datasets that capture a wide range of real-world conditions for plant disease detection. Such datasets would allow models to learn more generalisable features, helping to close the gap between controlled laboratory performance and real-world agricultural applications.

Architecture-Specific Insights: The experiments revealed distinct performance patterns across the different architectures. EfficientNet models, particularly EfficientNet-B3 and EfficientNet-B0, consistently demonstrated strong performance across various conditions. This can be attributed to their compound scaling method, which effectively balances network depth, width, and resolution, resulting in highly efficient and accurate models for plant disease detection. DenseNet201 exhibited notable stability, particularly in cross-dataset scenarios. Its dense connectivity facilitates the reuse of features, potentially enabling it to learn more robust and generalisable representations that are less sensitive to dataset-specific variations. This characteristic makes DenseNet201 a strong candidate for real-world applications where image conditions can vary significantly. While ResNet50 generally exhibited weaker performance compared to other architectures, it demonstrated specific strengths in certain disease classifications. However, its relatively weaker performance in cross-dataset scenarios suggests that its feature learning may be more susceptible to dataset-specific characteristics.

Impact of Noise Augmentation: The introduction of Gaussian noise to the images provided valuable insights into the models' resilience and adaptability. Notably, the

effects of noise varied significantly across different models and experiments. For instance, EfficientNet-B0 in Experiment Section 4.4 demonstrated improved performance with the addition of noise, increasing precision from 78.78% to 76.77%. Conversely, ResNet50 in Experiment Section 4.5 experienced a performance decline, with precision dropping from 52.17% to 49.57%. These contrasting outcomes underscore the intricate interplay between data augmentation techniques, model architectures, and the specific features learned by each model. The enhanced performance observed in some cases suggests that noise augmentation can act as a form of regularisation, enabling models to develop more adaptable and generalisable features.

Limitations of the work presented in this paper include the potential sensitivity of the models to noise, especially in noisy environments where various environmental factors could introduce unpredictable variations. Class imbalance within the datasets, as seen in the varying performance across different plant disease classes, may hinder the model's ability to reliably classify all plant types, particularly less-represented diseases. With regard to these, several promising avenues for future research emerge to address these limitations. Domain adaptation techniques can bridge the gap between different datasets, enhancing model performance under diverse real-world conditions. Integrating image data with other sensor data, such as spectral and environmental information, has the potential to significantly improve the accuracy and performance of plant disease detection systems. Moreover, developing continuous learning systems that can adapt to new data and changing conditions is crucial for addressing the challenges of model generalisation in dynamic agricultural environments. To mitigate the impact of class imbalance, future research should explore techniques such as data augmentation, improved sampling strategies, and cost-sensitive learning to ensure more reliable performance across all plant disease classes.

Author Contributions: Conceptualization, M.S.K. and I.K.I.; methodology, M.S.K. and I.K.I.; software, M.S.K.; validation, M.S.K., I.K.I. and P.M.; formal analysis, M.S.K., I.K.I., R.I.O., S.W.Y. and P.M.; investigation, M.S.K. and I.K.I.; resources, M.S.K. and I.K.I.; data curation, M.S.K. and I.K.I.; writing—original draft preparation, M.S.K., P.M., R.I.O., S.W.Y., F.N.d.S. and I.K.I.; writing—review and editing, M.S.K., P.M., R.I.O., S.W.Y., F.N.d.S. and I.K.I.; supervision, I.K.I.; project administration, I.K.I.; funding acquisition, I.K.I. and P.M. All authors have read and agreed to the published version of the manuscript.

Funding: This research was supported by the SmartBerry Project (Project Number: 10071867) and the Department of Computer Science, Nottingham Trent University.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Data used in paper are available at: <https://zenodo.org/records/14051480> (accessed on 14 December 2024).

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Jones, R.A. Global plant virus disease pandemics and epidemics. *Plants* **2021**, *10*, 233. [CrossRef] [PubMed]
2. Liliane, T.N.; Charles, M.S. Factors affecting yield of crops. *Agron.-Clim. Chang. Food Secur.* **2020**, *9*.
3. Savary, S.; Willocquet, L. Modeling the impact of crop diseases on global food security. *Annu. Rev. Phytopathol.* **2020**, *58*, 313–341. [CrossRef]
4. United Nations Children's Fund (UNICEF). The State of Food Security and Nutrition in the World 2020: Transforming Food Systems for Affordable Healthy Diets. Available online: <https://www.unicef.org/media/72676/file/sofi-2020-full-report.pdf> (accessed on 14 December 2024).

5. Saleem, M.H.; Potgieter, J.; Arif, K.M. Plant disease classification: A comparative evaluation of convolutional neural networks and deep learning optimizers. *Plants* **2020**, *9*, 1319. [CrossRef]
6. Li, L.; Zhang, S.; Wang, B. Plant disease detection and classification by deep learning—A review. *IEEE Access* **2021**, *9*, 56683–56698. [CrossRef]
7. Kakani, V.; Nguyen, V.H.; Kumar, B.P.; Kim, H.; Pasupuleti, V.R. A critical review on computer vision and artificial intelligence in food industry. *J. Agric. Food Res.* **2020**, *2*, 100033. [CrossRef]
8. Rakhmatulin, I.; Kamilaris, A.; Andreassen, C. Deep neural networks to detect weeds from crops in agricultural environments in real-time: A review. *Remote Sens.* **2021**, *13*, 4486. [CrossRef]
9. Omer, S.M.; Ghafoor, K.Z.; Askar, S.K. Plant disease diagnosing based on deep learning techniques. *Aro- Sci. J. Koya Univ.* **2023**, *11*, 38–47. [CrossRef]
10. Zhang, H.; Yang, J.; Lv, C.; Wei, X.; Han, H.; Liu, B. Incremental RPN: Hierarchical Region Proposal Network for Apple Leaf Disease Detection in Natural Environments. *IEEE/Acm Trans. Comput. Biol. Bioinform.* **2024**. [CrossRef]
11. Zhang, H.; He, G.; Li, F.; Xia, Z.; Liu, B.; Peng, J. Plant taxonomy-guided path-based tree classifier for large-scale plant species identification. *J. Electron. Imaging* **2021**, *30*, 023019. [CrossRef]
12. Getahun, S.; Kefale, H.; Gelaye, Y. Application of Precision Agriculture Technologies for Sustainable Crop Production and Environmental Sustainability: A Systematic Review. *Sci. World J.* **2024**, *2024*, 2126734. [CrossRef] [PubMed]
13. SS, V.C.; Hareendran, A.; Albaaji, G.F. Precision farming for sustainability: An agricultural intelligence model. *Comput. Electron. Agric.* **2024**, *226*, 109386. [CrossRef]
14. United Nations. SDG Briefing Book 2023. United Nations Partnerships; United Nations, 2023. Available online: <https://unpartnerships.un.org/sites/default/files/publications/2024-01/SDG>
15. Tugrul, B.; Elfatimi, E.; Eryigit, R. Convolutional neural networks in detection of plant leaf diseases: A review. *Agriculture* **2022**, *12*, 1192. [CrossRef]
16. Khakimov, A.; Salakhutdinov, I.; Omolikhov, A.; Utaganov, S. Traditional and current-prospective methods of agricultural plant diseases detection: A review. In *IOP Conference Series: Earth and Environmental Science*; IOP Publishing: Bristol, UK, 2022; Volume 951, p. 012002. [CrossRef]
17. Pomykala, J.; de Lemos, F.; Ihianle, I.K.; Adama, D.A.; Machado, P. Deep Learning approach for Classifying Trusses and Runners of Strawberries. In Proceedings of the UK Workshop on Computational Intelligence, Sheffield, UK, 7–9 September 2022; Springer: Berlin/Heidelberg, Germany, 2022; pp. 427–435.
18. Xu, J.; Chen, J.; You, S.; Xiao, Z.; Yang, Y.; Lu, J. Robustness of deep learning models on graphs: A survey. *AI Open* **2021**, *2*, 69–78. [CrossRef]
19. Shoaib, M.; Shah, B.; Ei-Sappagh, S.; Ali, A.; Ullah, A.; Alenezi, F.; Gechev, T.; Hussain, T.; Ali, F. An advanced deep learning models-based plant disease detection: A review of recent research. *Front. Plant Sci.* **2023**, *14*, 1158933.
20. Nduku, L.; Munghemezulu, C.; Mashaba-Munghemezulu, Z.; Masiza, W.; Ratshiedana, P.E.; Kalumba, A.M.; Chirima, J.G. Field-Scale Winter Wheat Growth Prediction Applying Machine Learning Methods with Unmanned Aerial Vehicle Imagery and Soil Properties. *Land* **2024**, *13*, 299. [CrossRef]
21. Akbar, M.; Ullah, M.; Shah, B.; Khan, R.U.; Hussain, T.; Ali, F.; Alenezi, F.; Syed, I.; Kwak, K.S. An effective deep learning approach for the classification of Bacteriosis in peach leave. *Front. Plant Sci.* **2022**, *13*, 1064854. [CrossRef]
22. Yoosefzadeh-Najafabadi, M.; Earl, H.J.; Tulpan, D.; Sulik, J.; Eskandari, M. Application of machine learning algorithms in plant breeding: Predicting yield from hyperspectral reflectance in soybean. *Front. Plant Sci.* **2021**, *11*, 624273. [CrossRef]
23. Cedric, L.S.; Adoni, W.Y.H.; Aworka, R.; Zoueu, J.T.; Mutombo, F.K.; Krichen, M.; Kimpolo, C.L.M. Crops yield prediction based on machine learning models: Case of West African countries. *Smart Agric. Technol.* **2022**, *2*, 100049. [CrossRef]
24. Jeny, A.A.; Junayed, M.S.; Islam, M.B.; Imani, H.; Shah, A.F.M.S. Machine vision-based expert system for automated cucumber diseases recognition and classification. In Proceedings of the 2021 International Conference on Innovations in Intelligent Systems and Applications (INISTA), Kocaeli, Turkey, 25–27 August 2021; IEEE: Piscataway, NJ, USA, 2021; pp. 1–6.
25. Sethy, P.K.; Barpanda, N.K.; Rath, A.K.; Behera, S.K. Deep feature based rice leaf disease identification using support vector machine. *Comput. Electron. Agric.* **2020**, *175*, 105527. [CrossRef]
26. Sharma, S.; Rajesh, E.; Kasi, B.; Govindaraju, S. The Classification Model for Plant Disease Detection with Segmentation and GLCM. In Proceedings of the 1st International Conference on Artificial Intelligence, Communication, IoT, Data Engineering and Security, IACIDS 2023, Lavasa, Pune, India, 23–25 November 2023.
27. Das, D.; Singh, M.; Mohanty, S.S.; Chakravarty, S. Leaf disease detection using support vector machine. In Proceedings of the 2020 International Conference on Communication and Signal Processing (ICCSP), Chennai, India, 28–30 July 2020; IEEE: Piscataway, NJ, USA, 2020; pp. 1036–1040.

28. Singh, J.; Kaur, H. Plant Disease Detection Based on Region-Based Segmentation and KNN Classifier. In Proceedings of the International Conference on ISMAC in Computational Vision and Bio-Engineering 2018 (ISMAC-CVB), Palladam, India, 16–17 May, pp. 1667–1675. [\[CrossRef\]](#)
29. Rajesh, B.; Vardhan, M.V.S.; Sujihelen, L. Leaf disease detection and classification by decision tree. In Proceedings of the 2020 4th International Conference on Trends in Electronics and Informatics (ICOEI)(48184), Tirunelveli, India, 15–17 June 2020; IEEE: Piscataway, NJ, USA, 2020; pp. 705–708.
30. Mohanapriya, K.; Balasubramani, M. Recognition of Unhealthy Plant Leaves Using Naive Bayes Classifier. In Proceedings of the IOP Conference Series: Materials Science and Engineering, Tamil Nadu, India, 12–13 April 2019; Volume 561, p. 012094. [\[CrossRef\]](#)
31. Ngugi, L.C.; Abelwahab, M.; Abo-Zahhad, M. Recent advances in image processing techniques for automated leaf pest and disease recognition—A review. *Inf. Process. Agric.* **2021**, *8*, 27–51. [\[CrossRef\]](#)
32. Lu, Y.; Lu, X.; Zheng, L.; Sun, M.; Chen, S.; Chen, B.; Wang, T.; Yang, J.; Lv, C. Application of multimodal transformer model in intelligent agricultural disease detection and question-answering systems. *Plants* **2024**, *13*, 972. [\[CrossRef\]](#) [\[PubMed\]](#)
33. Gazzard, M.; Hicks, H.; Ihianle, I.K.; Bird, J.J.; Hasan, M.M.; Machado, P. WeedScout: Real-Time Autonomous blackgrass Classification and Mapping using dedicated hardware. *arXiv* **2024**, arXiv:2405.07349.
34. Huynh-The, T.; Hua, C.H.; Tu, N.A.; Kim, D.S. Learning 3D spatiotemporal gait feature by convolutional network for person identification. *Neurocomputing* **2020**, *397*, 192–202. [\[CrossRef\]](#)
35. Wang, C.; Liu, B.; Liu, L.; Zhu, Y.; Hou, J.; Liu, P.; Li, X. A review of deep learning used in the hyperspectral image analysis for agriculture. *Artif. Intell. Rev.* **2021**, *54*, 5205–5253. [\[CrossRef\]](#)
36. Meghraoui, K.; Sebari, I.; Pilz, J.; Ait El Kadi, K.; Bensiali, S. Applied Deep Learning-Based Crop Yield Prediction: A Systematic Analysis of Current Developments and Potential Challenges. *Technologies* **2024**, *12*, 43. [\[CrossRef\]](#)
37. Du, Q.; Wang, Z.; Huang, P.; Zhai, Y.; Yang, X.; Ma, S. Remote Sensing Monitoring of Grassland Locust Density Based on Machine Learning. *Sensors* **2024**, *24*, 3121. [\[CrossRef\]](#)
38. Hasan, R.I.; Yusuf, S.M.; Alzubaidi, L. Review of the state of the art of deep learning for plant diseases: A broad analysis and discussion. *Plants* **2020**, *9*, 1302. [\[CrossRef\]](#)
39. Arcaini, P.; Bombarda, A.; Bonfanti, S.; Gargantini, A. Dealing with robustness of convolutional neural networks for image classification. In Proceedings of the 2020 IEEE International Conference On Artificial Intelligence Testing (AITest), Oxford, UK, 3–6 August 2020; IEEE: Piscataway, NJ, USA, 2020; pp. 7–14.
40. Drenkow, N.; Sani, N.; Shpitser, I.; Unberath, M. A systematic review of robustness in deep learning for computer vision: Mind the gap? *arXiv* **2021**, arXiv:2112.00639.
41. Mohanty, S.P.; Hughes, D.P.; Salathé, M. Using Deep Learning for Image-Based Plant Disease Detection. *Front. Plant Sci.* **2016**, *7*. [\[CrossRef\]](#)
42. Too, E.C.; Yujian, L.; Njuki, S.; Yingchun, L. A comparative study of fine-tuning deep learning models for plant disease identification. *Comput. Electron. Agric.* **2018**, *161*, 272–279. [\[CrossRef\]](#)
43. Hughes, D.P.; Salathé, M. An open access repository of images on plant health to enable the development of mobile disease diagnostics through machine learning and crowdsourcing. *arXiv* **2015**, arXiv:1511.08060. [\[CrossRef\]](#)
44. Karthik, R.; Hariharan, M.; Anand, S.; Mathikshara, P.; Johnson, A.; Menaka, R. Attention embedded residual CNN for disease detection in tomato leaves. *Appl. Soft Comput.* **2020**, *86*, 105933.
45. Arsenovic, M.; Karanovic, M.; Sladojevic, S.; Anderla, A.; Stefanovic, D. Solving current limitations of deep learning based approaches for plant disease detection. *Symmetry* **2019**, *11*, 939. [\[CrossRef\]](#)
46. Atila, Ü.; Uçar, M.; Akyol, K.; Uçar, E. Plant leaf disease classification using EfficientNet deep learning model. *Ecol. Inform.* **2021**, *61*, 101182. [\[CrossRef\]](#)
47. Elhassouny, A.; Smarandache, F. Smart mobile application to recognize tomato leaf diseases using Convolutional Neural Networks. In Proceedings of the 2019 International Conference of Computer Science and Renewable Energies (ICCSRE), Agadir, Morocco, 22–24 July 2019; IEEE: Piscataway, NJ, USA, 2019; pp. 1–4.
48. 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, Las Vegas, NV, USA, 26 June–1 July 2016; pp. 770–778.
49. Tan, M.; Le, Q. Efficientnetv2: Smaller models and faster training. In Proceedings of the International Conference on Machine Learning, Virtual, 18–24 July 2021; pp. 10096–10106.
50. 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, Honolulu, HI, USA, 21–26 July 2017; pp. 4700–4708.
51. Chandra, M.; Redkar, S.; Roy, S.; Patil, P. Classification of Various Plant Diseases Using Deep Siamese Network. Available online: <https://www.researchgate.net/publication/341322315> (accessed on 14 December 2024).

52. Moupojou, E.; Tagne, A.; Retraint, F.; Tadonkemwa, A.; Dongmo, W.; Tapamo, H.; Nkenlifack, M. FieldPlant: A dataset of field plant images for plant disease detection and classification with deep learning. *IEEE Access* **2023**, *11*, 35398–35410. [[CrossRef](#)]
53. Menon, V.; Ashwin, V.; Deepa, R.K. Plant disease detection using CNN and transfer learning. In Proceedings of the 2021 International Conference on Communication, Control and Information Sciences (ICCISc), Idukki, India, 16–18 June 2021; IEEE: Piscataway, NJ, USA, 2021; Volume 1, pp. 1–6.

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.