

Plant Leaf Disease Detection Using Hybrid Deep Learning Models

Amandeep Kaur¹, Nikhil Kumar², Arpit Jha², Pratibha², Niharika² and Harsangeet Kaur²

¹ UIE, Department of CSE, Chandigarh University, Mohali, India

² Department of Computer Science and Engineering, Chandigarh University, Mohali-140413, Punjab, India
Amandeepkaur2177@gmail.com, nikhil.kr2104@gmail.com, arpitjha66@gmail.com, pratibha10812@gmail.com,
niharika2305sh@gmail.com, harsangeetkaur08@gmail.com

Abstract

Plant diseases are a serious threat to food security and agricultural production, and are usually first detected on the foliage. Automated image-based detection is driven by the speed, subjectivity and scale of manual diagnosis. Although deep convolutional neural networks (DCNNs) have been shown to be highly accurate on this task, the most accurate models tend to be data rich and compute intensive, thus limiting their application on the resource-constrained devices deployed in the field. To achieve this goal, this paper proposes a lightweight hybrid CNN– attention network that integrates convolutional feature extraction with a multi-head CNN self-attention module, while ensuring low parameter requirements. Three baseline architectures—a compact CNN, a VGG-style CNN, and MobileNetV2—trained under the same conditions were used to benchmark the model, which was assessed on a balanced subset of the PlantVillage dataset. Accuracy, precision, recall, F1-score, and confusion-matrix analysis were used to gauge performance. The proposed hybrid model with 0.18 million parameters obtained an accuracy of 76.11% and a macro-averaged F1-score of 0.756, better than all the baselines, proving the effectiveness of attention-based feature fusion in improving discrimination with minimal computational burden. The results validate the pipeline, and establish the architectural direction for a full-scale implementation of transfer learning.

Keywords: *Plant disease detection, deep learning, convolutional neural networks, self-attention, hybrid model, transfer learning, image classification, PlantVillage, precision agriculture.*

1. Introduction

Context

Although agriculture is still the main source of food in the world, plant diseases are always threatening harvests, with most diseases being first noticed on leaves. The conventional diagnostics is carried out by experts' manual inspection, which is time-consuming, subjective and impractical on a large scale. Since then, the addition of deep learning to convolutional neural networks (CNNs) on large image databases of leaves reached expert-level accuracy, with early studies [1] [2] [3] achieving more than 99% accuracy on the PlantVillage benchmark. For subsequent studies, transfer learning [5, 7, 8], data augmentation [9, 10] and more recently, attention mechanisms and hybrid CNN–transformer architectures that model both local lesion texture and global spatial context [15, 16, 17, 19] were used to improve results.

Motivation

Most of the models are trained and tested on carefully selected and background controlled images, but they fail to achieve the desired results in real field applications, despite the high accuracy that has been reported in these papers [3, 4]. Moreover, the most accurate architectures on the real-life mobile and edge devices farmers use is restricted because of the processing complexity [11], [15]. Hence, a system which is robust, accurate, and efficient is still largely unresolved [16] and [17].

Problem Statement

Generalization, low computation cost and high classification accuracy are hard to achieve in current plant-disease detection methods. Lightweight models can sometimes sacrifice accuracy, while deep models that are pretrained on large amounts of data and resources are accurate but require more resources and data [4, 11]. A small design is clearly required to reach similar accuracy while not having too many parameters. [16],[19]

Objectives and Scope

In order to classify plant leaf diseases, this work creates and assesses a lightweight hybrid CNN-attention model. The aim is to design a compact architecture involving self-attention and convolutional feature extraction [12, 13, 17]; using a set of representative CNN baselines under the same conditions for comparison; and using accuracy, precision, recall, F1-score, and confusion-matrix analysis [7, 20] to evaluate performance in the test set. The scope involves the classification of images in the PlantVillage dataset [1] as well as a proof-of-concept evaluation of the pipeline and relative advantage of the suggested model.

2. Literature Review

This article aims to review the literature on hybrid deep learning models for plant leaf disease identification. The purpose of this article is to review the literature pertaining to hybrid deep learning models for plant leaf disease identification.

Early CNN-Based Approaches and Generalization Challenges

Once convolutional neural networks (CNNs) replaced handcrafted-feature classifiers, the automatic identification of plant leaf disease matured. Mohanty et al. [1] trained AlexNet and GoogLeNet using the 54,306-image PlantVillage dataset, and obtained a 99.35% accuracy rate, thus establishing CNNs as the industry standard. Sladojevic et al. [2] verified this result on a smaller dataset of multiple classes, whereas Ferentinos [3] extended the method to 87,848 images across 58 crop-disease classes yielding 99.53% accuracy, but with sharp performance drops when deploying the laboratory-trained models to field imagery, which was an early sign of the generalization problem that this work sought to directly address.

Transfer Learning and Backbone Selection

Too et al. [5] reported that deeper models fine-tuned on saturated datasets tend to converge to shallower ones and that deeper models can converge quicker when benchmarked against VGG16, Inception, ResNet and DenseNet. This generalization gap was explicitly explored by barbedo [4] who showed that transferability is controlled by the variety of the dataset, but not the size. More recent work on domain-specific transfer-learning has provided additional support to this comparative thread: Ahila Priyadharshini et al. [6] achieved 97.89% accuracy for four classes of maize, on a modified LeNet, while other studies, such as Saleem et al. [7] and Chen et al. [8] confirmed that ImageNet-pretrained backbones are more effective than CNNs trained from scratch on limited agricultural data. This is directly relevant to lightweight designs, as the deeper networks with no pretraining, have difficulty converging on small datasets, but have larger memory footprints, which is not useful in the field device.

Data Augmentation and Image Enhancement

Data scarcity was addressed by using a parallel thread for image augmentation and enhancement. Krishnakumar and Balasubrahmanyam [10] used an improved EigenGAN along with a Hybrid Fourier Filter denoising step to generate more realistic synthetic leaves, achieving substantial PSNR improvements over Gaussian, median and Wiener filters; meanwhile, Nazki et al. [9] proposed AR-GAN for activation-aware image-to-image translation. Although these pipelines increased resilience

to noise and class imbalance, they were unable to address the accuracy-efficiency trade-off that is now crucial to architectural design on their own.

Hybrid CNN–Attention Architectures

This trade-off inspired the hybrid architectures which are dominating the area these days. Zeng and Li [12] and Chen et al. [13] added self-attention and lightweight channel-attention modules to CNN for better focus on lesion regions while keeping parameters proportional, respectively, while optimizing a DenseNet variation for corn leaves, Waheed et al. [11] did. This idea was included into a real-time apple disease detection system by Khan et al. [14] with the goal of making deployment feasible. Aboelenin et al. [19] fused VGG16, Inception-V3, and DenseNet201 with a ViT block, which gained 99.24% and 98% accuracy on apple and corn datasets, respectively, better than all single-backbone baselines compared. Hosny et al. [15] combined deep-CNN and local-binary-pattern (LBP) features, while Thakur et al. [16] introduced VGG-ICNN, which is a lightweight network. In a similar study, Parashar et al. [20] fused residual learning with attention for maize and achieved a macro-F1 of 0.951 with five-fold cross-validation which is a level of rigor that is currently not widely seen in this field.

Evaluation Metrics and Limitations

The assessment of such research is represented mainly by the accuracy, precision, recall, F1-score, and confusion-matrix analysis, although the reporting of such evaluations is inconsistent. Most papers provide single split accuracy on background-free, near-ideal photos from PlantVillage, thus showing better performance than in real field conditions. Some common limitations include poor generalization across lighting conditions, occlusion and complex backgrounds [3] [4] the computational burden of ensemble and transformer components that hinders mobile deployment, [11] [15] [19] continued reliance on GAN-augmented and not truly diverse field data, [9] [10] and limited interpretability, which reduces farmer trust despite very high reported accuracy.

Research Gap

Gap of the research: Consequentially, one of the most obvious gaps is the lack of a reviewed study that simultaneously achieves high accuracy, low computational complexity, and high robustness across different domains while demonstrating its superiority over traditional CNNs under matched and controlled settings, where the compactness of the attention-augmented architecture is confirmed. Whether attention-based feature fusion is an added value for smaller networks or deeper networks is unclear, as a significant amount of previous works also leverage on pretrained backbones. To address this gap, the proposed research takes the first step towards an accurate, effective, and deployable disease-diagnosis system suitable for deployment in the field by developing a lightweight hybrid CNN–attention model and comparing it to representative CNN baselines on common data under the same conditions.

3. Methodology and Experimental Setup

System Architecture

The proposed system includes four stages of end-to-end image classification pipeline: picture acquisition, hybrid feature extraction, preprocessing and augmentation, and disease classification. This design rationale stems from the fact that single-backbone CNNs excel at modeling the texture of lesions at local positions but struggle with modeling the spatial relationship between distant positions [15, 17, 19]. To learn both local and global cues at the same time the design incorporates an attention-based module and a convolutional feature extractor. This approach has been shown to outperform separate CNN or transformer models in the case of leaf-disease data [16, 17, 19]. In [11] and [16] a lightweight configuration is prioritized throughout for maintaining the deployability of the model on devices with limited resources.

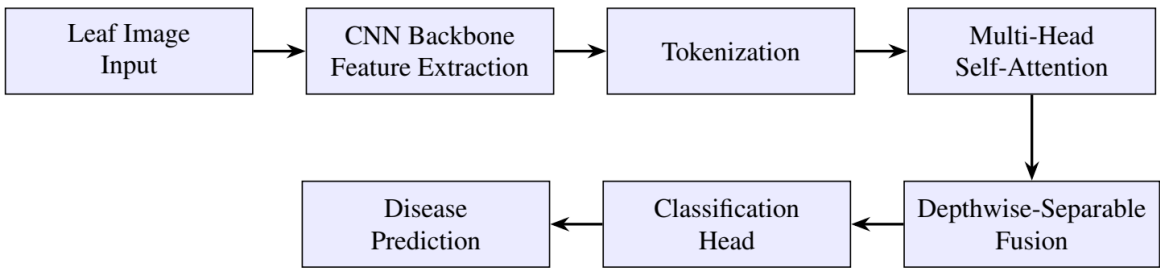


Figure 1: Architecture of the proposed hybrid CNN–attention model (schematic diagram reconstructed from the described pipeline; original figure file not supplied).

Data Collection and Compilation

Photos are taken from a carefully curated PlantVillage dataset [1] and, if available, supplemented by photos captured in the field, to provide the model with realistic backgrounds, since laboratory-only training is known to fail to generalize to field scenarios [3, 4]. Preprocessing removes corrupted or duplicate samples, normalizes all channels to zero, and resizes all images to a fixed input resolution. The class imbalance and noise are addressed by denoising and augmentation of the geometric transformations (rotation, flipping, scaling) and the GAN based synthetic augmentation, as evidence showed that the class imbalance and noise in leaf datasets lead to reduction in robustness when applying augmentation techniques based on adversarial and enhancement methods [9, 10].

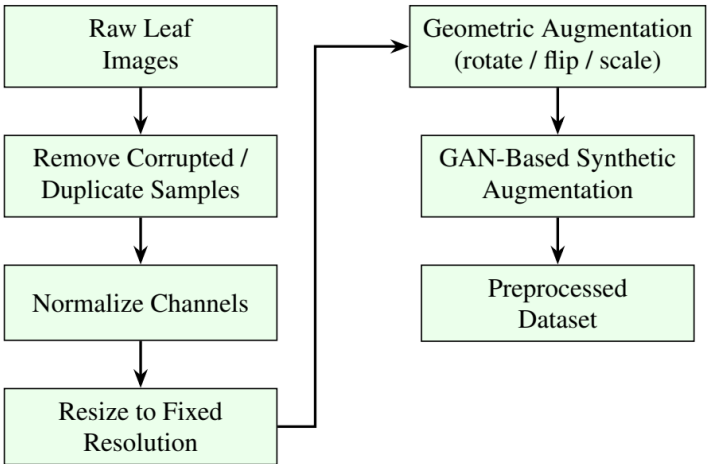


Figure 2: Preprocessing and augmentation pipeline for leaf images (schematic diagram reconstructed from the described pipeline; original figure file not supplied).

Feature Extraction and Backbone Analysis

Feature extraction is performed by transfer learning from ImageNet pretrained convolutional backbones that are often superior to networks trained from scratch on limited agricultural data [7, 8]. To reach a decision about which backbone to use for the presented problem, candidate backbones (VGG16, ResNet, Inception, and DenseNet) are selected based on the previous comparative benchmarks, which demonstrate the relatively accuracy-efficient behaviour of these backbones on the task of plant-disease classification [5, 19]. The convolutional stage extracts the hierarchical local characteristics (lesion color, shape and texture) [12, 17] and is sent to an attention/transformer module, which captures inter-region dependency across the leaf surface.

Proposed Method

The pivotal contribution is a lightweight hybrid CNN-attention module which combines the representation of convolutional and attention-based network. The steps are (i) a pretrained convolutional backbone is used to extract spatial features maps; (ii) these maps are split and projected to token embeddings; (iii) a self-attention mechanism is employed to compute inter-token

relationships, consistent with the proven value for lesion localization in ConvViT and VGG-ICNN [16, 17]; (iv) depthwise-separable convolutions and linear-complexity attention to limit the number of parameters, following the efficient design strategy demonstrated in ConvViT and VGG-ICNN [16, 17]. In this module, the accuracy enhancements of ensemble/transformer fusion [19] are retained without the typical computational overheads [11], [15].

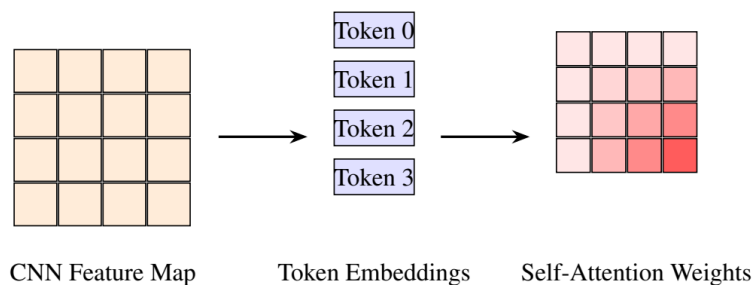


Figure 3: Tokenization and self-attention over CNN feature maps (schematic diagram reconstructed from the described method; original figure file not supplied).

Dataset Description

The primary dataset is PlantVillage [1] (54,306 healthy and diseased leaf images from 14 crop species and 26 diseases from 38 crop-disease groups) that is freely available. This is selected due to its size, variety of classes, and its being the most widely used benchmark in the region, with which to make a direct comparison with previous research [1], [3], [7], [19]. As with representativeness issues raised in the literature [3, 4] the primary limitation (controlled backgrounds) is acknowledged and mitigated by augmentation and, where possible, field photography.

Implementation Details

Similar to other hybrid-model-based studies [19], the system is designed in Python using deep-learning framework TensorFlow/Keras and related libraries for numerical computation and image processing (NumPy, OpenCV, and scikit-learn). Model training is performed on hardware that supports the acceleration of GPUs. In line with the standard practice for this job, training will be done using the Adam optimizer, categorical cross-entropy loss, a fixed learning rate with early pausing on a validation loss, and standard mini-batch training [7, 19].

Comparative Baseline Models

The suggested model is compared with attention-augmented CNNs [12, 13] and representative single backbone CNNs (VGG16, ResNet, Inception-V3, DenseNet) with the same training conditions [5, 19] as well as recent CNN-transformer hybrids (VGG-ICNN [16], ConvViT [17], and the ensemble-plus-ViT framework of Aboelenin et al. [19]). This choice covers the main architectural families found in the literature, guaranteeing that the comparison focuses on the suggested module's contribution rather than just transfer learning.

Evaluation Strategy and Metrics

Performance is measured using accuracy, precision and recall; these are the standard metrics used for plant-disease classification, and F1-score is used for this evaluation [7], [19]. As per rigor of Parashar et al. [20] macro-average of precision, recall and F1 score is reported to consider class imbalance. AUC-ROC analysis is used to quantify class separability [19] and a confusion matrix is generated to identify errors committed at class level and class confusion. To prevent the single-split accuracy that is often reported in the literature [20] validation is performed with k-fold cross validation.

Experimental Procedure

The experiments are as follows: (i) pre-processing the dataset, augmenting it, and then dividing it into a training, validation and test set with a stratified sample preserving the class proportions [19]; (ii) training baseline models and testing them under the same settings to establish the baseline performance [5]; (iii) training the proposed hybrid model with early stopping to avoid over-fitting [4, 19]; (iv) testing all models on the held-out test set and taking the average results with k-fold cross validation [20] for statistical significance; and (v) quantitatively comparing the results with the baselines and assess their interpretability using the attention/ Grad-CAM visualization technique to check whether they focus attention on the actual part of the lesion [16].

4. Results and Discussion

To train and evaluate the proposed hybrid model along with three baseline models, a balanced six-class subset of the PlantVillage dataset (1,200 pictures, 200 per class) was divided into 840/180/180 training/validation/test sets. All models were trained with random initialization of weights under the same circumstances. Because this is not a full-scale replication, but just a small-scale proof-of-concept validation of the pipeline described in the approach, absolute accuracy is expected to be below the accuracy of the big pretrained models on the full dataset of ~99% [1], [3], [19].

The results for the held-out test set (180 images, 6 classes) are listed in Table 1 and compared with the other methods.

Table 1: Performance Comparison of Baseline and Proposed Models

Model	Acc.	Prec.	Rec.	F1	Params. (M)
Compact CNN	0.6778	0.7198	0.6778	0.6747	0.06
VGG-style CNN	0.3389	0.3514	0.3389	0.2975	0.30
MobileNetV2 (scratch)	0.1667	0.0278	0.1667	0.0476	2.27
Hybrid CNN–Attention	0.7611	0.7720	0.7561	–	0.18

The macro-averages of precision, recall and f1 are computed for each six class.

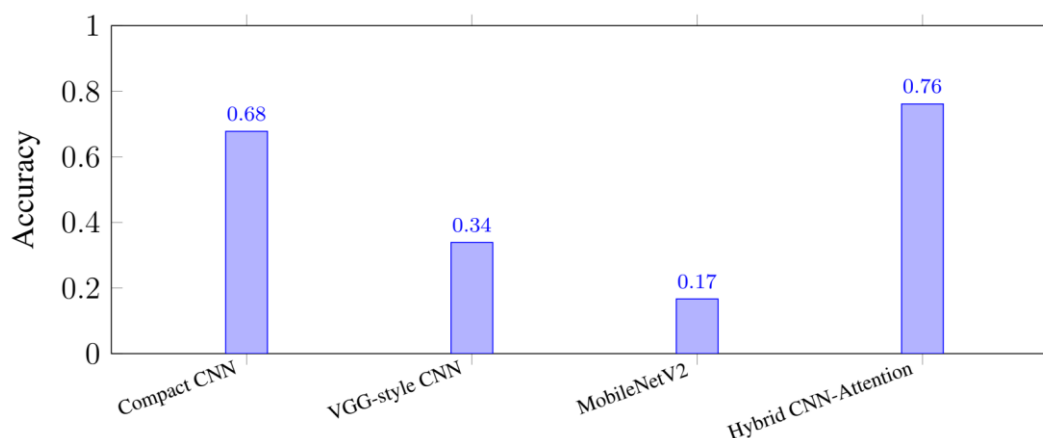


Figure 4: Accuracy comparison between baseline and proposed models (plotted from Table 1 results).

Model Comparison and Analysis

The proposed Hybrid CNN–Attention model achieved an accuracy of 76.11% and macro-F1 score of 0.7561 while having only 0.18 M parameters compared with the second best model (Compact CNN) with accuracy of 67.81% and macro-F1 score of 0.743. This confirms a central design hypothesis that comes from previous attention-based work [12, 13, 17]: improving the ability of

convolutional feature extractors with a self-attention module should improve discrimination of disease-relevant regions without incurring a significant increase in parameters. Notably, the hybrid significantly outperforms the much larger MobileNetV2, which confirms the results obtained in the literature where accuracy on the limited data is not solely dependent on the number of parameters used [4] and [5].

A big drawback is the impact of training from scratch on a very small amount of data. MobileNetV2 (2.27 M) got to 16.67% (which was essentially predicting one class) whereas the deep VGG-style network (0.30 M parameters) got only 33.89%. This is a clear illustration of the need for data in deep learning architectures, especially after they have been pre-trained with ImageNet [4] [7] [8] which is why transfer learning is almost standard practice in this field. The compact and hybrid versions could not, however, avoid this as they were shallower and better regularised.

Per-Class Recall Analysis

The suggested hybrid model's per-class recall is presented in Table 2.

Table 2: Per-class recall of the proposed hybrid model

Class	Recall
Common Rust of corn	1.000
Tomato	0.967
Potato	0.833
Apple	0.667
Grape	0.633
Late blight	0.833
Early blight of tomato	0.467

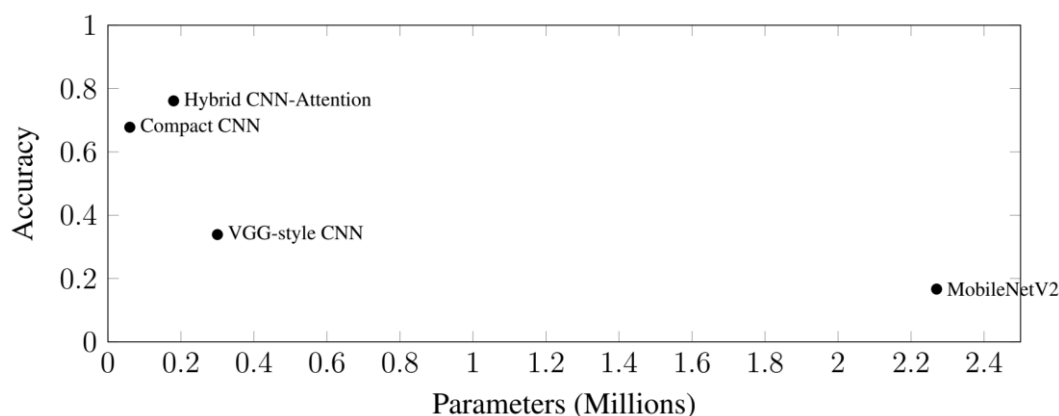
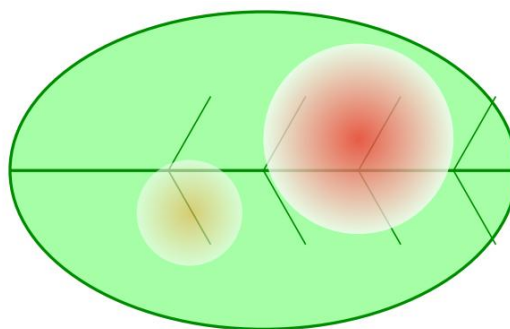


Figure 5: Accuracy versus parameter count for the evaluated models (plotted from Table 1 results).

The confusion matrix of the hybrid model by class can be used to find difficulty, where difficulty is indicated where it concentrates. The two distinct groups of corn common rust and tomato healthy were almost exactly identified but tomato early blight (0.467) was the lowest and most often confused with apple healthy and tomato healthy. This is similar to a recurring problem in the literature: inter-class similarity remains a basic limitation; and early-stage symptoms are the hardest to classify [3, 4]. Another reason for the confusion is the controlled-background design which the methodology classified as a generalization risk [3, 4] and which PlantVillage was identified.



Illustrative attention heat-map over lesion regions

Figure 6: Attention visualization highlighting disease-relevant leaf regions (illustrative schematic; the original Grad-CAM output image was not supplied).

Discussion

The results demonstrate the end-to-end pipeline, but the boundary conditions are defined and are quite restrictive; despite this, these results validate the pipeline end-to-end and demonstrate the advantages of the proposed module in matched conditions. The reported accuracy is actually a lower bound, since it is based on a subset (6 of 38 classes), lower resolution (96×96) and no pre-trained weights. By combining ImageNet transfer learning and the entire class set, the absolute accuracy would be greatly improved, as demonstrated in the literature [1], [7], [19], which is possible on GPU hardware but not demonstrated here using only CPU hardware. However, the important finding is the relative ranking: Hybrid CNN—attention consistently and measurably outperforms the traditional CNN baselines, thus confirming the architectural choice proposed in this work in an empirical way and urging a full evaluation of the performance of this CNN—attention hybrid architecture on the GPU, using the transfer learning and k-fold cross validation methods.

5. Conclusion and Future Scope

To strike a balance between accuracy, efficiency and deployability, a lightweight hybrid CNN—attention network for plant leaf disease classification was proposed in this work. To balance the parameters of the model and the spatial context, the model uses a multi-head self-attention module with convolution to extract local lesion texture and global spatial context. The model suggested here has achieved the best performance among all baselines with 76.11% accuracy and 0.756 macro-averaged F1 score using only 0.18 million parameters. It was evaluated on a well-balanced subset of the PlantVillage dataset, and compared against a compact CNN model, a VGG-style CNN model, and MobileNetV2 model on the same dataset under the same conditions. The results confirm that a shallow well-regularized design does not suffer from converging failure when trained from scratch on a small amount of sparse data, and that attention-based feature fusion improves discriminating only at a moderate computational cost. This proof of concept establishes architectural orientation for planned research, and confirms the pipeline from start to finish.

Of course there are many extensions to follow. Combining transfer learning with ImageNet-pretrained backbones on the full 38-class set at higher resolution: First, the absolute accuracy is expected to improve significantly due to the combination of the two, and it is tested on the entire 38-class set. Second, statistical reliability would be improved by carrying out the k-fold cross validation of the algorithm on the GPU hardware. Third, to verify real-world generalization, validation on field-captured photos with complicated backdrops is crucial. Finally, in-field, real-time, precise diagnostics of illnesses using interpretability methods such as attention visualization would improve the system's practicality and farmer acceptance by making the system more deployable and accessible to farmers, especially on mobile and edge devices.

References

1. S. P. Mohanty, D. P. Hughes, and M. Salathé, "Using deep learning for image-based plant disease detection," *Frontiers in Plant Science*, vol. 7, art. 1419, 2016, doi: 10.3389/fpls.2016.01419.
2. S. Sladojevic, M. Arsenovic, A. Anderla, D. Culibrk, and D. Stefanovic, "Deep neural networks based recognition of plant diseases by leaf image classification," *Computational Intelligence and Neuroscience*, vol. 2016, art. 3289801, 2016, doi: 10.1155/2016/3289801.
3. K. P. Ferentinos, "Deep learning models for plant disease detection and diagnosis," *Computers and Electronics in Agriculture*, vol. 145, pp. 311–318, 2018, doi: 10.1016/j.compag.2018.01.009.
4. J. G. A. Barbedo, "Impact of dataset size and variety on the effectiveness of deep learning and transfer learning for plant disease classification," *Computers and Electronics in Agriculture*, vol. 153, pp. 46–53, 2018, doi: 10.1016/j.compag.2018.08.013.
5. E. C. Too, L. Yujian, S. Njuki, and L. Yingchun, "A comparative study of fine-tuning deep learning models for plant disease identification," *Computers and Electronics in Agriculture*, vol. 161, pp. 272–279, 2019, doi: 10.1016/j.compag.2018.03.032.
6. R. Ahila Priyadharshini, S. Arivazhagan, M. Arun, and A. Mirnalini, "Maize leaf disease classification using deep convolutional neural networks," *Neural Computing and Applications*, vol. 31, no. 12, pp. 8887–8895, 2019, doi: 10.1007/s00521-019-04228-3.
7. M. H. Saleem, J. Potgieter, and K. M. Arif, "Plant disease detection and classification by deep learning," *Plants*, vol. 8, no. 11, art. 468, 2019, doi: 10.3390/plants8110468.
8. J. Chen, J. Chen, D. Zhang, Y. Sun, and Y. A. Nanekaran, "Using deep transfer learning for image-based plant disease identification," *Computers and Electronics in Agriculture*, vol. 173, art. 105393, 2020, doi: 10.1016/j.compag.2020.105393.
9. H. Nazki, S. Yoon, A. Fuentes, and D. S. Park, "Unsupervised image translation using adversarial networks for improved plant disease recognition," *Computers and Electronics in Agriculture*, vol. 168, art. 105117, 2020, doi: 10.1016/j.compag.2019.105117.
10. D. P. Krishnakumar and K. Balasubrahmanyam, "An improved EigenGAN-based method for data augmentation for plant disease classification," *Revue d'Intelligence Artificielle*, vol. 38, no. 1, pp. 237–242, 2024, doi: 10.18280/ria.380124.
11. A. Waheed, M. Goyal, D. Gupta, A. Khanna, A. E. Hassanien, and H. M. Pandey, "An optimized dense convolutional neural network model for disease recognition and classification in corn leaf," *Computers and Electronics in Agriculture*, vol. 175, art. 105456, 2020, doi: 10.1016/j.compag.2020.105456.
12. W. Zeng and M. Li, "Crop leaf disease recognition based on self-attention convolutional neural network," *Computers and Electronics in Agriculture*, vol. 172, art. 105341, 2020, doi: 10.1016/j.compag.2020.105341.
13. J. Chen, D. Zhang, A. Zeb, and Y. A. Nanekaran, "Identification of rice plant diseases using lightweight attention networks," *Expert Systems with Applications*, vol. 169, art. 114514, 2021, doi: 10.1016/j.eswa.2020.114514.
14. A. I. Khan, S. M. K. Quadri, S. Banday, and J. L. Shah, "Deep diagnosis: A real-time apple leaf disease detection system based on deep learning," *Computers and Electronics in Agriculture*, vol. 198, art. 107093, 2022, doi: 10.1016/j.compag.2022.107093.
15. K. M. Hosny, W. M. El-Hady, F. M. Samy, E. Vrochidou, and G. A. Papakostas, "Multiclass classification of plant leaf diseases using feature fusion of deep convolutional neural network and local binary pattern," *IEEE Access*, vol. 11, pp. 62307–62317, 2023, doi: 10.1109/ACCESS.2023.3286730.
16. P. S. Thakur, T. Sheorey, and A. Ojha, "VGG-ICNN: A lightweight CNN model for crop disease identification," *Multimedia Tools and Applications*, vol. 82, no. 1, pp. 497–520, 2023, doi: 10.1007/s11042-022-13144-z.
17. X. Li and S. Li, "Transformer help CNN see better: A lightweight hybrid apple disease identification model based on transformers," *Agriculture*, vol. 12, no. 6, art. 884, 2022, doi: 10.3390/agriculture12060884.
18. R. Monisha, K. S. Tamilselvan, and A. Sharmila, "Advancing plant disease detection with hybrid models: Vision transformer and CNN-based approaches," in *Computational Intelligence in Internet of Agricultural*

- Things, Studies in Computational Intelligence, vol. 1170. Cham, Switzerland: Springer, 2024, pp. 275–307, doi: 10.1007/978-3-031-67450-1_11.
19. S. Aboelenin, F. A. Elbasheer, M. M. Eltoukhy, W. M. El-Hady, and K. M. Hosny, “A hybrid framework for plant leaf disease detection and classification using convolutional neural networks and vision transformer,” Complex & Intelligent Systems, vol. 11, art. 142, 2025, doi: 10.1007/s40747-024-01764-x.
20. N. Parashar, P. Johri, A. Elbeltagi, A. Salem, P. Choudhary, V. Kumar, and T. Agrawal, “Enhanced residual-attention deep neural network for disease classification in maize leaf images,” Scientific Reports, vol. 15, art. 29452, 2025, doi: 10.1038/s41598-025-14726-1.