



Benchmarking Convolutional Neural Network Architectures for Potato Leaf Disease Identification



Yigitcan Cakmak^{1*}, Ramazan Yazgan²

¹ Department of Computer Engineering, Faculty of Engineering, Igdir University, 76000 Igdir, Turkey

² Department of Mathematics, Faculty of Science, Yuzuncu Yil University, 65080 Van, Turkey

* Correspondence: Yigitcan Cakmak (ygtcncakmak@gmail.com)

Received: 07-02-2025

Revised: 09-05-2025

Accepted: 09-20-2025

Citation: Y. Cakmak and R. Yazgan, "Benchmarking convolutional neural network architectures for potato leaf disease identification," *Nonlinear Sci. Intell. Appl.*, vol. 1, no. 1, pp. 18–26, 2025. <https://doi.org/10.56578/nsia010102>.



© 2025 by the author(s). Licensee Acadlore Publishing Services Limited, Hong Kong. This article can be downloaded for free, and reused and quoted with a citation of the original published version, under the CC BY 4.0 license.

Abstract: Potato production is critically influenced by foliar diseases such as Early Blight and Late Blight, which continue to threaten global food security. Although visual inspection remains widely used, such assessments are subjective, time-consuming, and difficult to scale, creating a pressing need for automated and reliable diagnostic frameworks. In this study, the classification performance and computational efficiency of four state-of-the-art Convolutional Neural Network (CNN) architectures—the Residual Network with 50 layers (ResNet-50), Densely Connected Network with 169 layers (DenseNet-169), EfficientNetV2-B3, and InceptionV3—were systematically benchmarked for the identification of healthy potato leaves and those affected by Early Blight or Late Blight using the publicly available PlantVillage dataset. Accuracy, precision, recall, and F1 score were employed to characterize predictive performance, while parameter count and giga floating-point operations per second (GFLOPS) were used to assess computational efficiency. High-level classification capability was consistently achieved across all models, with overall accuracies ranging from 98% to 99%. DenseNet-169 achieved the highest classification accuracy at 99% with fewer than 13 million parameters, and EfficientNetV2-B3 attained 98% accuracy while exhibiting the lowest GFLOPS requirement. The results indicate that architectures designed for parameter efficiency and feature reuse, such as DenseNet-169 and EfficientNetV2-B3, provide accuracy that is comparable to or surpasses that of less efficient baseline models while offering significant advantages in resource efficiency. These findings reinforce the strong potential of lightweight and high-performance CNN architectures to support scalable, real-time agricultural disease diagnostic systems, particularly in regions where computational resources and technical expertise may be limited.

Keywords: Potato diseases; DL, Plant disease classification; Computational efficiency

1 Introduction

As the fourth most important food crop globally, potato (*Solanum tuberosum*) is a key component of global food security [1, 2]. Potatoes are a major source of nutrition for over one billion people, demonstrating the importance of producing and supplying potatoes reliably [3]. Nevertheless, potato production is constantly threatened by several phytopathological conditions, particularly fungal diseases such as Early Blight (*Alternaria solani*) and Late Blight (*Phytophthora infestans*) [4]. These are very damaging diseases that are capable of causing significant losses in yield and economic losses in production, thereby disrupting a major part of the global food supply network [5]. The detection and control of potato diseases has traditionally been dependent on visual inspection by agronomists and growers [6, 7]. The manual approach is inherently problematic because it is subjective, labor-intensive, and also highly prone to human error [8, 9]. The quality of diagnosis is highly dependent on the expertise of the individual, lacks repeatability, and is especially unreliable during the early stages of infection, when symptoms may be ambiguous or easily mistaken for those associated with nutritional deficiencies [10, 11]. Misdiagnoses often lead to delays or improper episodes of fungicides that can increase costs and create a real threat to the environment [12, 13].

The significant constraints of traditional diagnostics in agriculture have highlighted an urgent need for automated, accurate, and scalable systems [14–16]. Paralleling this challenge, deep learning (DL), particularly CNNs, has achieved transformative success in other domains demanding high-stakes pattern recognition. This validation is

exceptionally strong in medical diagnostics, where CNNs have demonstrated a robust ability to discriminate subtle pathologies across diverse and complex modalities [17], from neurological tumors in brain scans to malignancies in breast [18–20] and lung imagery [21]. This proven capacity to autonomously learn and extract hierarchical, discriminative features from such varied visual data is now converging with computer vision to create a new paradigm in precision agriculture. These architectures are consequently becoming the foundation for automated plant disease classification, providing the robust tools needed for accurate diagnosis from complex agricultural imagery, such as in potato leaves [22, 23].

Informed by this technological basis, the current study builds and evaluates a DL system for the classification of three key potato leaf classes, namely healthy, Early Blight, and Late Blight, using the public PlantVillage dataset. This study undertakes a systematic comparative study of four leading and diverse Convolutional Neural Network (CNN) architectures: Residual Network with 50 layers (ResNet-50), Densely Connected Network with 169 layers (DenseNet-169), EfficientNetV2-B3, and InceptionV3. The performance of these models is analyzed using standard performance metrics, providing information regarding potential effectiveness for real-world use at scale in agriculture. The key contributions of this study are briefly summarized as follows:

- A robust DL framework is developed for the three-class classification of potato leaf diseases (healthy, Early Blight, and Late Blight) utilizing the public PlantVillage dataset.
- A comprehensive benchmark analysis is conducted, comparing the performance of four distinct and influential CNN architectures: ResNet-50, DenseNet-169, EfficientNetV2-B3, and InceptionV3.
- The models are rigorously evaluated using a suite of metrics, including accuracy, precision, recall, and F1 score, to identify the most effective architecture for this specific diagnostic task.
- This research provides a comparative assessment that aids in selecting computationally efficient and accurate models, thereby supporting the development of accessible diagnostic tools for sustainable agriculture.

2 Related Work

Wang and Su [24] presented a detailed review covering DL applications throughout the potato production chain and grouped these uses into key areas like crop health management, yield prediction, and resource management. Their work examined various models, including CNNs and Recurrent Neural Networks and detailed their roles in tasks from pest detection to price forecasting. It was concluded that while DL offers major benefits for improving efficiency and productivity, major challenges remain about the availability of diverse datasets and the practical deployment of these technologies in real agricultural settings. Selvi et al. [25] developed CropViT, a computationally very efficient Vision Transformer architecture designed specifically for high-throughput plant disease diagnosis. Their experimental work involved a direct comparison of CropViT against a conventional CNN model on the PlantVillage dataset, with a focus on nine different plant species. The results indicated that CropViT achieved an average accuracy of 98.64% and significantly outperformed the traditional CNN. This highlights the strong potential of transformer-based approaches in agricultural diagnostics.

Dutta et al. [26] developed a specialized CNN architecture for automatically detecting and classifying potato blight diseases in their early stages. The proposed model was compared against common architectures like ResNet-50, VGG16, and GoogLeNet using a dataset consisting of healthy, Early Blight, and Late Blight samples. Their findings showed that the custom CNN model reached an accuracy of 98%. This demonstrates the value of tailored DL solutions for specific tasks in agricultural phytopathology. Bajpai et al. [27] proposed an architectural augmentation to the Swin Transformer model to improve the detection accuracy of potato leaf diseases, specifically Early Blight and Late Blight. Their modification involved adding a custom sequential head module consisting of linear, ReLU, and dropout layers to the standard Swin Transformer to improve feature representation and reduce overfitting. When evaluated on a custom dataset, the improved model achieved 99.38% accuracy. This confirmed the effectiveness of architectural refinements in boosting generalization performance for agricultural computer vision applications.

Zhang et al. [28] introduced an optimized VGG16 architecture, called VGG16S, to address both computational efficiency and diagnostic accuracy in potato disease detection. The optimization strategy involved replacing dense layers with global average pooling, integrating the Convolutional Block Attention Module, and using the leaky ReLU activation function. This multi-part approach reduced the model's parameter complexity to just one-tenth of the original VGG16 while still reaching an accuracy of 97.87%. This shows that lightweight, optimized architectures can offer major benefits in both accuracy and efficiency. Sharma and Sharma [29] explored using Recurrent Neural Networks for classifying healthy and diseased potato leaves from the PlantVillage dataset, which breaks from common CNN-based methods. Their proposed architecture used Long Short-Term Memory units for feature extraction and was compared against CNN and Feedforward Neural Network models. The experimental results demonstrated that the Recurrent Neural Network model reached an accuracy of 92.7%. This suggests that Recurrent Neural Network architectures, with their capacity for temporal sequence processing, can offer a competitive advantage in image-based classification tasks.

Zoralioğlu and Polat [30] conducted a comparative analysis to examine the important role of data augmentation

and class balancing in potato disease detection. They evaluated three different architectures (i.e., a custom 5-layer CNN, EfficientNetB2, and ConvNeXtSmall) on both the original (imbalanced) and balanced (augmented) versions of the PlantVillage dataset. Their findings highlighted the important relationship between data distribution and model performance: The custom CNN performed best on imbalanced data, while EfficientNetB2 achieved 99.89% accuracy on the balanced data. This clearly indicates the necessity of data balancing strategies for reaching the full potential of advanced DL models.

3 Materials and Methods

3.1 Dataset and Data Preprocessing

The PlantVillage dataset, a large public dataset created for classifying plant diseases via plant leaf images [31], was used for this study. For this study, only the potato leaf subset was selected, which consists of three classes: two disease classes (Early Blight and Late Blight) and one healthy leaf class. Figure 1 shows examples of the visual characteristics of leaves in each class which this study intended for the models to identify.



Figure 1. Sample images of potato leaves for healthy, Early Blight, and late Blight Classes

To ensure a complete and unbiased assessment of the models, the data was carefully separated into three distinct subsets of data: 70% for training, 15% for validation, and 15% for the final evaluation of the models. The distribution of images in the different subsets of data (with the count of samples for each class listed for training, validation, and testing) can be seen in Table 1. This means the models are trained on the majority of the data, optimized on a separate validation dataset, and then finally tested on completely separate data.

Table 1. Distribution of classes across data splits

Class	Train (70%)	Validation (15%)	Test (15%)	Total
Early Blight	700	150	150	1000
Healthy	106	22	24	152
Late Blight	700	150	150	1000
Total	1506	322	324	2152

As the first necessary step, a consistent standardized data preprocessing pipeline was set up. All images were resized to a consistent 224×224 pixel size, based on the input size dimensions of the pre-trained networks used in this work. Pixel values for images were standardized to floating-point numbers in the range of [0, 1], which is a common practice to stabilize and speed up convergence during training. To reduce the chances of overfitting and maximize model generalizability, a data augmentation approach was used only for the purpose of training the dataset. Data augmentation in this study entailed the random application of transformations, including horizontal flipping, rotations, and zooming, for training images dynamically during training to generate an artificial pictorial variety for the training dataset—without changing validation or testing datasets [32, 33].

3.2 Model Architecture

This study comparatively evaluates four distinct CNN architectures, each representing significant advancements in DL for computer vision tasks. These models were selected for their diverse structural philosophies and proven performance across various image recognition benchmarks. The ResNet architecture, particularly ResNet-50, presented the concept of residual learning to solve the issue of degradation associated with training extremely

deep networks. The core idea concerns the use of identity shortcut connections (i.e., skip connections) that allow gradients to skip one or multiple layers. This permits the network to learn a residual function with respect to the inputs from the earlier layers, which supports easier optimization and enables deeper networks to be built without losing performance to vanishing gradients. The architecture of ResNet-50 includes an initial convolution layer, a max-pooling layer, aggregated stacked residual blocks that contain 1×1 , 3×3 , and 1×1 convolutions, a global average pooling layer and lastly a fully connected classification layer [34].

The Densely Connected Convolutional Network (DenseNet) displays maximum information flow between layers with a unique connectivity model that connects every layer directly to every other layer in a deep feed-forward manner (represented here with 169 layers). DenseNet connects feature maps from all previous layers together at each layer, while ResNets take a weighted sum of features (the skipped connections), which allows feature reuse and utilizes deeper neural networks. Thus, dense connectivity results in models that are often smaller and contain fewer parameters than similarly deep ResNets, while also alleviating the vanishing gradient problem. Each layer outputs features which are concatenated to all previous layers at specific depths in the network, referred to as “dense blocks,” followed by layers that perform batch normalization and average pooling, with the intention of reducing the feature map size, referred to as “transition layers” [35].

EfficientNetV2 is the next version of EfficientNets designed to address not only accuracy and parameter efficiency but also improved training speed. It is based on the compound scaling mechanism in the original EfficientNet that uniformly scales the depth, width, and resolution of the network, and it has a few additions, such as Fused-MBConv blocks (fusing the depthwise and 1×1 convolutions into one regular convolution in the initial layers) and a progressive training scheme that adjusts image size and regularization during training. EfficientNetV2-B3 is a configuration of EfficientNetV2 that attempts to achieve a strong trade-off between cost and performance in a variety of applications where speed and accuracy matter [36].

The InceptionV3 structure is a member of the GoogLeNet family and is noted for its distinctive Inception module. The purpose of this module is to allow learning from multiple scales at once in a deliberate attempt to achieve parallelism within a single layer. Inception achieves this parallelism by using convolutional filters of different sizes (1×1 , 3×3 , and 5×5) operating in parallel, along with max pooling. The InceptionV3 model made improvements such as factorizing larger convolutions into smaller convolutions (e.g., replacing a 5×5 convolution with two 3×3 convolutions) and using asymmetric convolution configurations (e.g., using a $1 \times n$ convolution followed by an $n \times 1$ convolution) with the intent of reducing computational cost while preserving the representational capacity of the model. InceptionV3 also introduced the use of batch normalization and label smoothing regularization to enhance stability during training and maximize generalization [37].

3.3 Transfer Learning

In order to improve model convergence and prediction ability, the transfer learning approach was used in this study. The selected CNN architectures (ResNet-50, DenseNet-169, EfficientNetV2-B3, and InceptionV3) were initialized with weights trained on the large-scale ImageNet dataset. This takes advantage of the rich, hierarchical features learned from millions of varied images, producing strong inductive bias for the target task. For each pre-trained model, the original final classification layer, typically for the 1000 classes in ImageNet, was removed and replaced with a new fully connected dense layer specific to the three-class taxonomy of the potato leaf disease classification problem (healthy, Early Blight, and Late Blight).

A fine-tuning protocol consisting of two stages was implemented in the training procedure. In the first stage, the parameters of the pre-trained feature extraction backbone were frozen, and only the newly added classification layer was trained. This allows the classifier to become accustomed to the features generated by the frozen backbone. In stage three, the whole network was trained end-to-end, frequently with a lower learning rate, which refined and specialized the pre-trained features for the specifics of the potato leaf dataset. This approach, combined with data augmentation applied to the training set, helped generalize the model and lessen the risk of overfitting.

3.4 Experimental Design and Training Protocol

To facilitate a fair and straightforward comparison of the DL architectures selected for this study, a structured experimental protocol was developed. All models were implemented in the Python programming environment using the TensorFlow framework. Both training and inference were completed on a high-performance computer workstation with an NVIDIA GeForce RTX 5090 GPU (32 GB VRAM). All experiments in this study were conducted using the same training procedure to create consistency. The Adam optimization algorithm was used to optimize the model’s parameters. Then, for the fine-tuning (i.e., the second part of the training where the entire network is retrained), the learning rate remained at 1×10^{-4} . A mini-batch size of 16 was utilized in all training. It was also predetermined to run for a maximum of 100 epochs.

To counteract overfitting and encourage generalizability, the criterion of early stopping was incorporated. This criterion observed the validation loss at the end of each epoch to stop training if it had not improved after 10 epochs

(patience = 10). When training was completed (whether at the maximum allowable epochs or because of early stopping), the model’s weights from the epoch with the lowest validation loss were saved for final evaluation on the held-out test dataset.

3.5 Performance Evaluation Metrics

The performance of each of the DL models was evaluated on the independent test dataset. A full complement of standard performance metrics was used for evaluation. Accuracy is the key performance measure and reflects the overall ratio of correctly classified examples compared to the total number of examples in the test set. It provides an overall measure of predictive accuracy.

To better understand the class-level performance of the model and potential issues arising from class imbalance, precision, recall, and F1 score were also included in the evaluation. Precision indicates the proportion of positive predictions that are truly positive, also referred to as the positive predictive value. Recall is sometimes called sensitivity, or the true positive rate, and indicates the overall ability of the model to capture all actual positive examples that belong to the class. The F1 score is a measure that indicates the harmonic mean of precision and recall and is valuable for class imbalance scenarios. The mathematical definitions of these metrics are provided below.

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \quad (1)$$

$$\text{Precision} = \frac{TP}{TP + FP} \quad (2)$$

$$\text{Recall} = \frac{TP}{TP + FN} \quad (3)$$

$$\text{F1 score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \quad (4)$$

where, TP , TN , FP and FN denote true positive, true negative, false positive, and false negative, respectively. For the multi-class classification problem solved in this project, all metrics were calculated separately for each class (healthy, Early Blight, and Late Blight) and subsequently macro-averaged to generate an overall performance score for the model where all classes were weighted equally irrespective of their sample size.

4 Results and Discussion

The experimental stage aimed to facilitate a comprehensive assessment of the four learning models, in particular evaluating their performance on the unseen potato leaf disease test dataset. A comprehensive assessment was conducted involving the extraction of quantitative results that represented standard classification performance measures as well as complexity measures. The four primary evaluation measures of accuracy, precision, recall, and F1 score were put in place to ensure an unbiased and reasonable assessment of each model’s generalization ability. In conjunction, each of the model’s computational complexities was documented by measuring its parameters (Params) and giga floating-point operations per second (GFLOPS). The summarized findings, which inform this report and analysis, are organized in Table 2.

Table 2. Performance evaluation results of deep learning models

Models	Accuracy	Precision	Recall	F1 score	Params	GFLOPS	Inference Time (ms)
ResNet-50	0.98	0.96	0.96	0.96	23.51 M	8.26	6,5754
DenseNet-169	0.99	0.99	0.99	0.99	12.49 M	6.72	10,121
EfficientNetV2-B3	0.98	0.95	0.98	0.96	12.83 M	3.04	4,0404
InceptionV3	0.98	0.9750	0.9750	0.9750	21.79 M	5.67	6,0674

Note: GFLOPS, giga floating-point operations per second; ResNet-50, Residual Network with 50 layers; DenseNet-169, Densely Connected Network with 169 layers.

Of the architectures examined, ResNet-50 provided a strong baseline, reaching a baseline accuracy of 98%. The precision, recall, and F1 score were all measured consistently to be 96%, demonstrating that deep Residual Networks possess strong extractive properties. However, this model is the most demanding in terms of computational

complexity, with 8.26 GFLOPS of computation and 23.51 million parameters, resulting in an inference time of 6.58 ms. InceptionV3 achieved a comparable accuracy of 98%, along with precision, recall, and F1 score of 97.50%, respectively. This model was situated midway in terms of complexity with 21.79 million parameters, 5.67 GFLOPS of computation, and an inference time of 6.07 ms.

DenseNet-169 was the highest-performing model on the basis of pure classification capability, with an impressive accuracy of 99% and consistent performance across all metrics: precision = 99%, recall = 99%, and F1 score = 99%. The breakdown of performance relative to the three categories in the classification task is expressed in the confusion matrix in Figure 2, for example. Notably, this best-in-class performance came with very low compute requirements, utilizing only 12.49 million parameters and 6.72 GFLOPS of computation. However, despite its parameter efficiency, it recorded the highest latency with an inference time of 10.12 ms, likely due to the complex memory access patterns of dense connections compared to the ResNet-50 baseline.

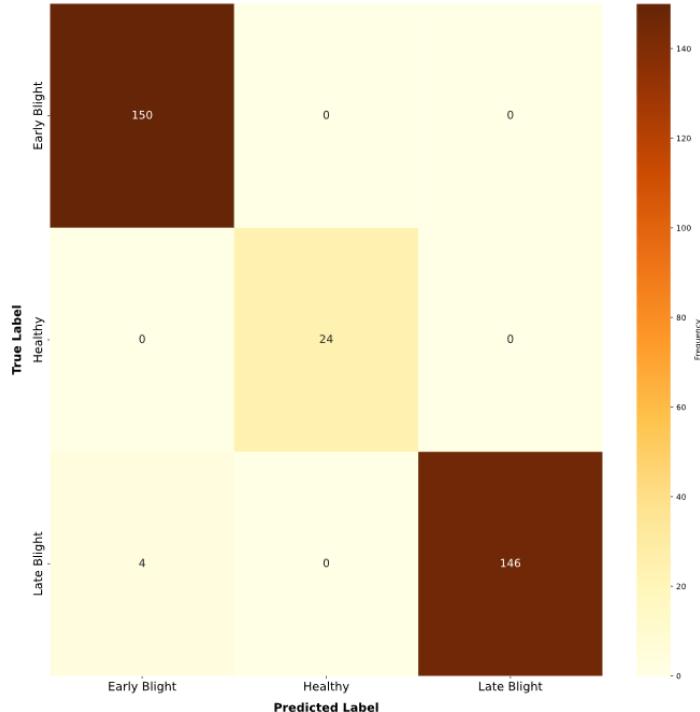


Figure 2. Confusion matrix showing the classification results of the Densely Connected Network with 169 layers model on the test dataset

EfficientNetV2-B3 exhibited high diagnostic performance with a 98% accuracy. It achieved a precision of 95%, a recall of 98%, and an F1 score of 96%. However, EfficientNetV2-B3's most prominent feature is its overall computational efficiency; it required only 3.04 GFLOPS and achieved the fastest inference speed of 4.04 ms, making it the most computationally feasible model tested. In addition, EfficientNetV2-B3 achieved a low parameter count of 12.83 million, which is comparable to DenseNet-169. Overall, the accuracy combined with the efficiency of EfficientNetV2-B3 is certainly an attractive option for resource-scarce environments.

The four models were all compared and showed relatively high classification performance for this task with final accuracies between 98% and 99%. The most major difference was efficiency. DenseNet-169 demonstrated the highest overall accuracy while providing significantly fewer parameters and GFLOPS than the ResNet-50 and InceptionV3 models. While EfficientNetV2-B3 achieved similar performance in terms of a 98% accuracy, it displayed superior computational performance with a relatively low GFLOPS requirement and the lowest inference latency among all evaluated architectures. Results revealed that the DenseNet and EfficientNet family of architectures could classify potato disease images with state-of-the-art accuracy while minimizing computational resources in a way that exceeds the performance of the established baseline models. The relatively high classification accuracy and computational efficiency of DenseNet-169 and EfficientNetV2-B3 demonstrate that they could be well-suited for practical use in an agricultural context.

While the experimental results underscore the potential of these CNN architectures, particularly the efficiency-accuracy balance of EfficientNetV2-B3, several limitations warrant attention. A primary constraint is the reliance on the PlantVillage dataset, which comprises images captured in controlled settings with homogenous backgrounds.

Consequently, the models' robustness against the visual complexity of real-world agricultural environments characterized by variable lighting, shadowing, and cluttered backgrounds remains to be fully validated. Furthermore, although the inference times reported in this study highlight the relative efficiency of the models, these metrics were derived from a high-performance workstation equipped with an NVIDIA GeForce RTX 5090. In practical agricultural scenarios, deployment often targets low-power hardware. Therefore, future research will focus on bridging this gap by evaluating model latency and energy consumption on resource-constrained edge devices, such as the Raspberry Pi or Jetson Nano, to ensure viability for in-field deployment. Additionally, expanding the diagnostic scope beyond the current three classes to encompass a broader spectrum of potato pathologies will be a critical step toward developing a comprehensive decision support system for farmers. Furthermore, exploring Vision Transformers and hybrid CNN-Transformer architectures represents a significant avenue for future research. Investigating these advanced models could further enhance classification performance, particularly in handling the high variability and complex patterns inherent in field-acquired agricultural imagery.

5 Conclusion

This study evaluated four CNN architectures, i.e., ResNet-50, DenseNet-169, EfficientNetV2-B3, and InceptionV3, to investigate the feasibility of automated classification of potato leaf diseases using images from the PlantVillage dataset. Through the experimental results, this study confirmed the diagnostic ability of all models, which yielded test accuracies between 98% and 99%. DenseNet-169 did produce the best-performing model at a 99% test accuracy, though all models performed well. Results show that recent architectures such as DenseNet-169 and especially EfficientNetV2-B3 performed diagnostically well using far fewer parameters and a lower computational load (GFLOPS) in comparison to ResNet-50 and InceptionV3. These results suggest that efficient yet highly accurate models could serve as beneficial and effective alternatives for automated plant disease diagnosis. Efficient models present opportunities suitable for field use in resource-constrained agriculture to facilitate diagnosis and support sustainable agriculture. Future work may involve field execution, and if successful, the potential for the development of a broadly available, accessible decision support tool for farmers.

Author Contributions

Conceptualization, Y.C. and R.Y.; methodology, Y.C.; software, Y.C.; validation, Y.C. and R.Y.; formal analysis, Y.C.; investigation, Y.C.; data curation, Y.C.; writing—original draft preparation, Y.C.; writing—review and editing, Y.C. and R.Y.; visualization, Y.C.; supervision, R.Y. All authors have read and agreed to the published version of the manuscript.

Data Availability

The dataset analyzed for this study is the public PlantVillage dataset, which is available on Kaggle: <https://www.kaggle.com/datasets/emmarex/plantdisease>.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Declaration of Generative AI and AI-Assisted Technologies in the Writing Process

During the preparation of this work, the author(s) used artificial intelligence tools in order to improve the readability and language quality of the manuscript. After using this tool, the author(s) reviewed and edited the content as needed and take(s) full responsibility for the content of the publication.

References

- [1] P. Kalpana, R. Anandan, A. G. Hussien, H. Migdady, and L. Abualigah, "Plant disease recognition using residual convolutional enlightened Swin transformer networks," *Sci. Rep.*, vol. 14, no. 1, p. 8660, 2024. <https://doi.org/10.1038/s41598-024-56393-8>
- [2] A. Kaur, G. S. Randhawa, F. Abbas, M. Ali, T. J. Esau, A. A. Farooque, and R. Singh, "Artificial intelligence driven smart farming for accurate detection of potato diseases: A systematic review," *IEEE Access*, vol. 12, pp. 193 902–193 922, 2024. <https://doi.org/10.1109/ACCESS.2024.3510456>
- [3] X. Chen, Y. Li, and Z. Zhang, "Effvit-potatonet: An efficientvit-based model for potato leaf disease classification," in *2025 International Conference on Digital Analysis and Processing, Intelligent Computation (DAPIC)*, Incheon, Korea, Republic of, 2025, pp. 84–88. <https://doi.org/10.1109/DAPIC66097.2025.00022>
- [4] S. Kamal, P. Sharma, P. K. Gupta, M. K. Siddiqui, A. Singh, and A. Dutt, "DVTXAI: A novel deep vision transformer with an explainable AI-based framework and its application in agriculture," *J. Supercomput.*, vol. 81, no. 1, p. 280, 2025. <https://doi.org/10.1007/s11227-024-06494-y>

- [5] F. O. Isinkaye, M. O. Olusanya, and A. A. Akinyelu, “A multi-class hybrid variational autoencoder and vision transformer model for enhanced plant disease identification,” *Intell. Syst. Appl.*, vol. 26, p. 200490, 2025. <https://doi.org/10.1016/j.iswa.2025.200490>
- [6] S. Adhikari, “Advancements in agricultural technology: Vision transformer-based potato leaf disease classification,” *J. Soft Comput. Paradigm*, vol. 6, no. 2, pp. 169–185, 2024. <https://doi.org/10.36548/jscp.2024.2.005>
- [7] S. Benkaihoul, S. Khadar, Y. Özüpak, E. Aslan, M. M. Almalki, and M. A. Mossa, “Advanced fault classification in induction motors for electric vehicles using a stacking ensemble learning approach,” *World Electr. Veh. J.*, vol. 16, no. 11, p. 614, 2025. <https://doi.org/10.3390/WEVJ16110614>
- [8] S. Austin, A. Barua, S. N. Haider, F. L. Niha, M. Faisal, and S. M. Shawon, “Precision classification of potato diseases using transformer-enhanced CNNs,” in *2025 International Conference on Quantum Photonics, Artificial Intelligence, and Networking (QPAIN)*, Rangpur, Bangladesh, 2025, pp. 1–6. <https://doi.org/10.1109/QPAIN66474.2025.11172153>
- [9] E. Aslan and Y. Özüpak, “Diagnosis and accurate classification of apple leaf diseases using vision transformers,” *Comput. Decis. Mak.: An Int. J.*, vol. 1, pp. 1–12, 2024. <https://doi.org/10.59543/COMDEM.V1I.10039>
- [10] J. H. Sinamenye, A. Chatterjee, and R. Shrestha, “Potato plant disease detection: Leveraging hybrid deep learning models,” *BMC Plant Biol.*, vol. 25, no. 1, p. 647, 2025. <https://doi.org/10.1186/s12870-025-06679-4>
- [11] Y. Özüpak, F. Alpalaz, E. Aslan, and H. Uzel, “Hybrid deep learning model for maize leaf disease classification with explainable AI,” *N. Z. J. Crop Hortic. Sci.*, pp. 1–23, 2025. <https://doi.org/10.1080/01140671.2025.2519570>
- [12] A. Bajpai, S. Sahu, N. Tiwari, V. Srivastava, and S. Yadav, “An efficient approach for potato leaf disease classification using cascaded CNN-transformers,” in *2024 IEEE 12th International Conference on Intelligent Systems (IS)*, Varna, Bulgaria, 2024, pp. 1–6. <https://doi.org/10.1109/IS61756.2024.10705224>
- [13] F. Alpalaz, Y. Özüpak, E. Aslan, and H. Uzel, “Classification of maize leaf diseases with deep learning: Performance evaluation of the proposed model and use of explicable artificial intelligence,” *Chemom. Intell. Lab. Syst.*, p. 105412, 2025. <https://doi.org/10.1016/J.CHEMOLAB.2025.105412>
- [14] J. Zeynalov, Y. Çakmak, and İ. Paçal, “Automated apple leaf disease classification using deep convolutional neural networks: A comparative study on the Plant Village Dataset,” *J. Comput. Sci. Digit. Technol.*, vol. 1, no. 1, pp. 5–17, 2025. <https://doi.org/10.30546/UNECCSDT.2025.01.005>
- [15] I. Pacal and G. Işık, “Utilizing convolutional neural networks and vision transformers for precise corn leaf disease identification,” *Neural Comput. Appl.*, vol. 37, no. 4, pp. 2479–2496, 2025. <https://doi.org/10.1007/s00521-024-10769-z>
- [16] I. Kunduracioglu and I. Pacal, “Advancements in deep learning for accurate classification of grape leaves and diagnosis of grape diseases,” *J. Plant Dis. Prot.*, vol. 131, no. 3, pp. 1061–1080, 2024. <https://doi.org/10.1007/s41348-024-00896-z>
- [17] Y. Çakmak, “Machine learning approaches for enhanced diagnosis of hematological disorders,” *Comput. Syst. Artif. Intell.*, vol. 1, no. 1, pp. 8–14, 2025. <https://doi.org/10.69882/ADBA.CSAI.2025072>
- [18] Y. Cakmak and I. Pacal, “Enhancing breast cancer diagnosis: A comparative evaluation of machine learning algorithms using the Wisconsin Dataset,” *J. Oper. Intell.*, vol. 3, no. 1, pp. 175–196, 2025. <https://doi.org/10.31181/JOPI31202539>
- [19] Y. Çakmak and J. Zeynalov, “A comparative analysis of convolutional neural network architectures for breast cancer classification from mammograms,” *Artif. Intell. Appl. Sci.*, vol. 1, no. 1, pp. 28–34, 2025. <https://doi.org/10.69882/ADBA.AI.2025075>
- [20] Y. Çakmak and N. Pacal, “Deep learning for automated breast cancer detection in ultrasound: A comparative study of four CNN architectures,” *Artif. Intell. Appl. Sci.*, vol. 1, no. 1, pp. 13–19, 2025. <https://doi.org/10.69882/ADBA.AI.2025073>
- [21] Y. Çakmak and A. Maman, “Deep learning for early diagnosis of lung cancer,” *Comput. Syst. Artif. Intell.*, vol. 1, no. 1, pp. 20–25, 2025. <https://doi.org/10.69882/ADBA.CSAI.2025074>
- [22] I. Pacal, I. Kunduracioglu, M. H. Alma, M. Deveci, S. Kadry, J. Nedoma, V. Slany, and R. Martinek, “A systematic review of deep learning techniques for plant diseases,” *Artif. Intell. Rev.*, vol. 57, no. 11, p. 304, 2024. <https://doi.org/10.1007/S10462-024-10944-7>
- [23] I. Pacal, “Enhancing crop productivity and sustainability through disease identification in maize leaves: Exploiting a large dataset with an advanced vision transformer model,” *Expert Syst. Appl.*, vol. 238, p. 122099, 2024. <https://doi.org/10.1016/J.ESWA.2023.122099>
- [24] R. F. Wang and W. H. Su, “The application of deep learning in the whole potato production Chain: A comprehensive review,” *Agriculture*, vol. 14, no. 8, p. 1225, 2024. <https://doi.org/10.3390/agriculture14081225>
- [25] G. C. Selvi, H. J. Charan, and D. Kumar, “CropViT: A light-weight transformer model for crop disease

- detection,” in *2024 3rd International Conference on Artificial Intelligence For Internet of Things (AIoT)*, Vellore, India, 2024, pp. 1–6. <https://doi.org/10.1109/AIoT58432.2024.10574729>
- [26] S. Dutta, S. G. Neogi, and A. Halder, “Automatic early detection of potato blight disease using deep neural networks,” in *2024 IEEE International Conference on Intelligent Signal Processing and Effective Communication Technologies (INSPECT)*, Gwalior, India, 2024, pp. 1–7. <https://doi.org/10.1109/INSPECT63485.2024.10896181>
- [27] A. Bajpai, N. Tiwari, P. Rajput, S. Sahu, and D. Singh, “Enhanced potato leaf disease detection via modified swin transformer architecture,” in *2024 15th International Conference on Computing Communication and Networking Technologies (ICCCNT)*, Kamand, India, 2024, pp. 1–7. <https://doi.org/10.1109/ICCCNT61001.2024.10724512>
- [28] C. Zhang, S. Wang, C. Wang, H. Wang, Y. Du, and Z. Zong, “Research on a potato leaf disease diagnosis system based on deep learning,” *Agriculture*, vol. 15, no. 4, p. 424, 2025. <https://doi.org/10.3390/agriculture15040424>
- [29] A. Sharma and A. Sharma, “Recurrent neural network-based classification of potato leaves using RGB images,” in *2024 2nd International Conference on Advancement in Computation & Computer Technologies (InCACCT)*, Gharuan, India, 2024, pp. 486–491. <https://doi.org/10.1109/InCACCT61598.2024.10551226>
- [30] Y. Zoralioğlu and Ö. Polat, “Detection of potato plant disease from leaf images using deep learning models,” in *2024 Innovations in Intelligent Systems and Applications Conference (ASYU)*, Ankara, Turkiye, 2024, pp. 1–5. <https://doi.org/10.1109/ASYU62119.2024.10756961>
- [31] “PlantVillage Dataset.” <https://www.kaggle.com/datasets/emmarex/plantdisease>
- [32] Z. T. Wang, P. F. Wang, K. P. Liu, P. Y. Wang, Y. J. Fu, C. T. Lu, C. C. Aggarwal, J. Pei, and Y. C. Zhou, “A comprehensive survey on data augmentation,” *arXiv preprint*, no. arXiv: 2405.09591, 2024. <https://doi.org/10.48550/arXiv.2405.09591>
- [33] A. Mumuni, F. Mumuni, and N. K. Gerrar, “A survey of synthetic data augmentation methods in machine vision,” *Mach. Intell. Res.*, vol. 21, no. 5, pp. 831–869, 2024. <https://doi.org/10.1007/S11633-022-1411-7>
- [34] K. He, X. Zhang, S. Ren, and J. Sun, “Deep residual learning for image recognition,” in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016, pp. 770–778. https://openaccess.thecvf.com/content_cvpr_2016/html/He_Deep_Residual_Learning_CVPR_2016_paper.html
- [35] G. Huang, Z. Liu, L. Van der Maaten, and K. Weinberger, “Densely Connected Convolutional Networks (DenseNets).” <https://github.com/liuzhuang13/DenseNet>
- [36] M. Tan and Q. Le, “EfficientNetV2: Smaller models and faster training,” in *International Conference on Machine Learning*, 2021, pp. 10 096–10 106. <https://proceedings.mlr.press/v139/tan21a/tan21a.pdf>
- [37] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, “Rethinking the inception architecture for computer vision,” in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, Las Vegas, NV, USA, 2016, pp. 2818–2826. <https://doi.org/10.1109/CVPR.2016.308>