UNEC Journal of Computer Science and Digital Technologies
Volume 2, № 1, 2026, pp. 5-15
https://doi.org/10.30546/UNECCSDT.2026.001.210

Section:
Artificial Intelligence & Machine Learning

Efficient and Accurate Potato Disease Classification Using Lightweight Vision Transformers: A Comparative Benchmark Against a Deep CNN Architecture

Javanshir Zeynalov[1]*, Yiğitcan Çakmak[2]
[1]Faculty of Architecture and Engineering, Nakhchivan State University, Nakhchivan, Azerbaijan*
[2]Faculty of Engineering, Igdir University, Igdir, Türkiye
[1]0009-0002-4985-6371, cavansirzeynalov@ndu.edu.az
[2]0009-0008-7227-9182, ygtcncakmak@gmail.com

Abstract

Potato is among the most important staple crops underpinning global food security; however, its productivity remains highly vulnerable to destructive foliar diseases, particularly Early Blight and Late Blight, which can cause substantial yield losses when not detected and managed in a timely manner. Conventional disease diagnosis largely depends on visual inspection by farmers or specialists, a process that is time-consuming, expertise-dependent, and prone to subjective interpretation, especially under field conditions where symptoms may overlap or appear at early infection stages. To address these limitations, this study proposes and evaluates an automated deep learning-based framework for classifying potato leaf images into three categories: healthy, Early Blight, and Late Blight, using the publicly available PlantVillage dataset. A comparative assessment was conducted between a well-established convolutional architecture, ResNet-101, and two Vision Transformer-based models, namely Swin Base and MobileViT v2. The models were evaluated in terms of classification effectiveness using accuracy, precision, recall, and F1-score, while their computational practicality was examined through parameter count and GFLOPs. The experimental findings indicate that although all architectures achieved strong diagnostic performance, the transformer-based models consistently surpassed the conventional CNN baseline. Among them, MobileViT v2 delivered the best overall performance, reaching a test accuracy of 99.69% while maintaining a highly compact architecture with only 4.39 million parameters. This combination of high predictive accuracy and low computational demand suggests that lightweight Vision Transformer models offer a more practical and efficient alternative to deeper CNN-based approaches for potato disease recognition. These results underline the potential of such architectures for deployment in mobile, embedded, or other resource-constrained agricultural diagnostic systems, supporting more timely disease management and contributing to sustainable precision farming practices.

Keywords: Potato diseases, Deep learning, Vision Transformer (ViT), Plant disease classification, Computational efficiency

Received:	*Revised:*	*Accepted:*	*Published:*
20/05/2026	*03/06/2026*	*06/06/2026*	*17/06/2026*

1. Introduction

Potato (*Solanum tuberosum*) ranks among the most widely cultivated and consumed food crops worldwide, following maize, wheat, and rice in global importance, and it occupies a central position in sustaining the international food supply chain [1]. Owing to its high

productivity, broad adaptability, and substantial nutritional value, potatoes serve as a major staple for more than one billion people and contribute directly to global food security [2]. Despite this importance, potato production remains highly susceptible to a range of destructive diseases, many of which are caused by fungal and bacterial pathogens and can severely reduce both yield and crop quality [3]. Among these, Early blight (*Alternaria solani*) and Late blight (*Phytophthora infestans*) are particularly damaging, as uncontrolled infections may lead to extensive economic losses and, in severe cases, the failure of entire harvests [4].

In conventional agricultural practice, the detection and management of potato diseases have largely relied on visual inspection performed by farmers, agronomists, or plant protection specialists [5]. Although widely used, this approach is inherently labor-intensive, subjective, and vulnerable to diagnostic errors, particularly during the early stages of infection when visual symptoms are subtle, overlapping, or easily mistaken for abiotic stress and nutrient-related disorders [6]. As a result, diagnostic reliability often depends heavily on the experience of the evaluator, leading to variability across observers and field conditions. These limitations may delay timely intervention or cause inappropriate fungicide application, increasing production costs while also intensifying environmental risks associated with excessive or unnecessary chemical use [7]. Therefore, the development of automated, accurate, and scalable diagnostic systems has become an increasingly important requirement for effective potato disease management [8].

Recent advances at the intersection of computer vision and deep learning have introduced powerful data-driven approaches for addressing long-standing challenges in precision agriculture [9,10]. These developments have been particularly influential in automated plant disease recognition, where image-based deep learning models have shown strong potential for accurate and scalable crop health assessment [11–14]. In this context, Convolutional Neural Networks (CNNs) have been widely adopted as a dominant methodological framework due to their ability to learn hierarchical and discriminative visual representations directly from image data [15,16]. Their strong performance across diverse agricultural imaging tasks has established CNNs as a reliable baseline for plant disease classification [17–19]. More recently, Vision Transformers (ViTs) have emerged as a compelling alternative architecture by replacing purely convolutional feature extraction with self-attention mechanisms capable of modeling long-range dependencies across image regions. Following their success in general-purpose computer vision, ViT-based models are increasingly being explored for domain-specific applications, including agricultural image analysis and crop disease diagnosis [20].

Within this context, the present study develops and evaluates a deep learning-based framework for the automated classification of three potato leaf conditions—healthy, Early blight, and Late blight—using the publicly available PlantVillage dataset. A systematic comparative analysis is conducted between a representative deep CNN architecture, ResNet-101, and two transformer-based models, Swin Base and MobileViT v2, to assess both predictive performance and practical computational suitability. Model effectiveness is evaluated using standard classification metrics, including accuracy, precision, recall, and F1-score, while computational efficiency is considered through model complexity indicators. By comparing established convolutional learning with more recent transformer-based paradigms, this study aims to provide evidence on the suitability of lightweight and high-performing architectures for practical potato disease diagnosis. The resulting findings are expected to support the development of accessible, efficient, and scalable decision-support tools for precision crop protection, thereby contributing to more sustainable potato production and improved agricultural resilience.

2. Related Work

Wang and Su (2024) [21] provided a broad review of deep learning applications across the potato production chain, organizing existing studies around major operational domains such as crop health monitoring, yield estimation, and resource management. Their review covered a

range of architectures, including Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs), and discussed their use in tasks extending from pest and disease detection to market-oriented applications such as price forecasting. Although the study emphasized the considerable potential of deep learning to improve productivity, decision-making, and operational efficiency in potato production, it also highlighted persistent barriers to real-world adoption, particularly the limited availability of diverse datasets and the difficulty of deploying these systems under practical field conditions.

Selvi et al. (2024) [22] introduced CropViT, a lightweight Vision Transformer architecture designed for efficient crop disease diagnosis. The model was fine-tuned and evaluated on nine crop categories from the PlantVillage dataset and was directly compared with a conventional CNN baseline. Their results showed that CropViT achieved a mean accuracy of 98.64%, outperforming the CNN model and demonstrating the capacity of transformer-based architectures to deliver high diagnostic performance while maintaining computational efficiency. This work provided important evidence that compact transformer models can serve as practical alternatives to traditional convolutional approaches in agricultural disease recognition.

Dutta et al. (2024) [23] developed a customized CNN model for the early detection and classification of potato blight diseases using leaf images. Their study focused on distinguishing healthy leaves from Early blight and Late blight cases and compared the proposed architecture with established deep learning models, including ResNet-50, VGG-16, and GoogLeNet. The customized CNN achieved an accuracy of 98%, surpassing the benchmark models and indicating that task-specific architectural design can improve performance in potato disease classification. Their findings further support the relevance of domain-adapted deep learning models for agricultural phytopathology applications.

Bajpai et al. (2024) [24] improved the standard Swin Transformer architecture for potato leaf disease detection by incorporating a customized sequential classification head consisting of Linear, ReLU, and Dropout layers. This modification was intended to strengthen feature representation and reduce overfitting during classification. Evaluated on a custom potato leaf dataset including Early blight and Late blight categories, the enhanced Swin Transformer reached an accuracy of 99.38%. The study demonstrated that targeted modifications to transformer-based models can improve generalization and classification accuracy in specialized agricultural computer vision tasks.

Zhang et al. (2025) [25] addressed both model efficiency and diagnostic accuracy in potato leaf disease recognition by proposing VGG16S, an optimized variant of the original VGG16 architecture. Their approach combined several architectural refinements, including the replacement of fully connected layers with global average pooling, the integration of the CBAM attention mechanism, and the adoption of the Leaky ReLU activation function. These changes substantially reduced the number of parameters to approximately one-tenth of the original VGG16 model while increasing classification accuracy to 97.87% on their proprietary dataset. Their findings illustrate the value of architectural compression and attention-based enhancement for developing more efficient plant disease diagnosis systems.

Sharma and Sharma (2024) [26] investigated the use of Recurrent Neural Networks for classifying healthy and diseased potato leaves from RGB images in the PlantVillage dataset. Their proposed model employed Long Short-Term Memory units for feature extraction and was compared with both a CNN and a Feedforward Neural Network. The RNN-based architecture achieved an accuracy of 92.7%, outperforming the alternative models evaluated in the study. Although convolutional architectures remain more common in image-based plant disease classification, their results suggest that sequence-oriented models may still offer useful representational capabilities when adapted to visual classification tasks.

Zoralioğlu and Polat (2024) [27] examined the influence of data augmentation and class balancing on potato disease classification performance. They evaluated a custom 5-layer CNN, EfficientNetB2, and ConvNeXtSmall on the PlantVillage dataset under both imbalanced and

augmented balanced conditions. Their results revealed that model performance was strongly affected by the underlying data distribution. While the custom CNN performed best on the original imbalanced dataset, EfficientNetB2 achieved the highest accuracy of 99.89% after augmentation and class balancing. This finding highlights the importance of preprocessing and dataset balancing strategies for enabling advanced deep learning architectures to reach their full potential in plant disease detection.

Although previous studies have demonstrated the effectiveness of deep learning models for potato disease classification, several limitations remain evident in the existing literature. Many studies focus primarily on maximizing classification accuracy, while computational efficiency and deployment feasibility are often treated as secondary considerations. In addition, prior works frequently evaluate either CNN-based models or transformer-based architectures in isolation, limiting the ability to draw a balanced comparison between established convolutional methods and emerging attention-based approaches. Considering the practical requirements of precision agriculture, especially the need for accurate models that can operate on mobile or resource-constrained devices, there is still a need for comparative evaluations that jointly examine predictive performance and computational complexity. To address this gap, the present study systematically benchmarks ResNet-101, Swin Base, and MobileViT v2 on the PlantVillage potato leaf dataset, providing a performance–efficiency perspective for identifying architectures suitable for practical potato disease diagnosis.

3. Materials and Methods

3.1 Dataset and Data Preprocessing

This study employed the publicly available PlantVillage dataset, a widely used image repository for plant disease recognition tasks. From this dataset, the potato leaf subset was selected, comprising three diagnostic categories: healthy leaves, Early blight, and Late blight [28]. Representative samples from each class are presented in Figure 1, illustrating the visual characteristics associated with the corresponding leaf conditions and providing an overview of the classification targets addressed in this study. To enable a reliable and unbiased assessment of model performance, the dataset was divided into training, validation, and test subsets using a 70:15:15 ratio. The detailed distribution of images across these subsets is reported in Table 1, which summarizes the number of samples assigned to each class within the training, validation, and testing partitions.

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

A consistent preprocessing pipeline was applied prior to model training to ensure input standardization across all architectures. Each image was resized to 224 × 224 pixels, matching the required input resolution of the pre-trained models used in this study. Pixel intensity values were subsequently scaled to the [0, 1] range, a normalization step intended to improve numerical stability and support more efficient convergence during training. To reduce the risk of overfitting and enhance the models' ability to generalize beyond the training data, data augmentation was applied only to the training subset.

Figure 1. Sample Images of Potato Leaves for Healthy, Early Blight, and Late Blight Classes

This augmentation process generated additional variability through random transformations, including horizontal flipping, rotation, and zooming, thereby exposing the models to a broader range of plausible visual variations while preserving the underlying class labels [29,30].

3.2 Foundational Principles of Deep Learning Models

3.2.1 Convolutional Neural Networks (CNNs)

Convolutional Neural Networks (CNNs) are a class of deep learning architectures specifically designed to process structured grid-like data, with images being one of their most prominent application domains. Their effectiveness stems from a hierarchical feature learning mechanism, in which lower layers capture simple visual patterns such as edges and textures, while deeper layers progressively encode more complex and semantically meaningful representations. The core component of a CNN is the convolutional layer, where learnable kernels are applied across the input image to produce feature maps that highlight relevant spatial patterns. These convolutional operations are commonly followed by pooling layers, which reduce the spatial resolution of the feature maps and thereby lower computational cost while improving robustness to minor spatial variations. After successive stages of convolution and pooling, the extracted high-level features are passed to fully connected layers or classification heads, which transform the learned representations into final class predictions.

3.2.2 Vision Transformers (ViTs)

Vision Transformers (ViTs) depart from the conventional convolution-based design by adapting the Transformer architecture, originally developed for sequence modeling in natural language processing, to visual recognition tasks. Rather than relying on local receptive fields to process pixel neighborhoods, a ViT first divides an input image into a set of fixed-size, non-overlapping patches and treats these patches as a sequence of visual tokens. Each patch is flattened and projected into a latent embedding space, while learnable positional embeddings are added to preserve spatial information that would otherwise be weakened during tokenization. The resulting token sequence is then passed through multiple Transformer encoder layers, where multi-head self-attention allows the model to estimate relationships among all image patches simultaneously. This mechanism enables ViTs to capture long-range dependencies and global contextual patterns that may be difficult to model using purely local

convolutional operations. The final representation is subsequently processed by a classification head, which maps the learned image-level features to the corresponding disease category.

3.3 Transfer Learning and Data Augmentation Strategy

Transfer learning was adopted to improve convergence efficiency and strengthen classification performance. The selected architectures were initialized with ImageNet pre-trained weights, enabling the models to benefit from generic visual representations learned from large-scale and diverse image collections. For each architecture, the original classification layer was removed and replaced with a new task-specific output layer configured for the three potato leaf categories considered in this study. The training procedure was implemented in two stages. In the first stage, the pre-trained feature extraction backbone was kept frozen, and only the newly added classification layer was trained to adapt the model to the target task. In the second stage, the full network was fine-tuned end-to-end using a low learning rate, allowing the pre-trained representations to be gradually adjusted to the visual characteristics of potato leaf diseases. This transfer learning strategy was complemented by data augmentation, which increased the diversity of the training samples and helped reduce overfitting, thereby improving the robustness and generalization capacity of the models.

3.4 Experimental Design and Training Protocol

A standardized experimental protocol was designed to ensure a fair and reproducible comparison among the selected deep learning architectures. All models were implemented in Python using the TensorFlow framework, and both training and inference procedures were carried out on a high-performance workstation equipped with an NVIDIA GeForce RTX 5090 GPU. To maintain consistency across experiments, the same training configuration was applied to all models. Model parameters were optimized using the Adam optimizer, with a learning rate of 1×10^{-4} during the fine-tuning stage. Training was performed with a mini-batch size of 16 for a maximum of 100 epochs. To limit overfitting and prevent unnecessary training, early stopping was applied with a patience value of 10 epochs, terminating the process when no improvement in validation loss was observed over consecutive epochs. The checkpoint corresponding to the lowest validation loss was retained and used for the final evaluation on the independent test set.

3.5 Performance Evaluation Metrics

The classification performance of each model was evaluated on the independent hold-out test set using a set of widely accepted evaluation metrics. Accuracy was used as the primary indicator of overall predictive performance, reflecting the proportion of correctly classified samples among all test instances. However, to obtain a more detailed understanding of model behavior, Precision, Recall, and F1-score were also computed. Precision measures the reliability of positive predictions by indicating the proportion of correctly identified positive samples among all samples predicted as positive. Recall, also referred to as sensitivity, quantifies the model's ability to detect all relevant positive instances within a given class. The F1-score combines Precision and Recall through their harmonic mean, providing a balanced performance measure that is particularly useful when class distributions are uneven. The mathematical formulations of these metrics are provided as follows:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{1}$$

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \tag{2}$$

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \tag{3}$$

$$\text{F1} - \text{Score} = 2 * \frac{\text{Precision} * \text{Recall}}{\text{Precision} + \text{Recall}} \quad (4)$$

In these equations, TP, TN, FP, and FN represent the numbers of true positive, true negative, false positive, and false negative predictions, respectively. Since the present task involves a multi-class classification setting, the metrics were first calculated separately for each class and then aggregated using macro-averaging. This approach assigns equal importance to each class and produces a single overall score that reflects the model's classification performance across all potato leaf categories.

4. Results and Discussion

The experimental evaluation was conducted to systematically assess and compare the performance of the selected deep learning architectures in the task of potato leaf disease classification. Quantitative findings, including both classification of performance metrics and computational complexity indicators, are presented to provide a comprehensive basis for model comparison. Accuracy, precision, recall, and F1-score were calculated on the independent test set, thereby enabling an objective evaluation of each model's generalization capability on previously unseen data. The complete results are summarized in Table 2 and form the basis for the comparative analysis discussed in the following section.

Table 2. *Performance Evaluation Results of Deep Learning Models*

Models	Accuracy	Precision	Recall	F1-score	Params	GFLOPs
MobileViT v2 [31]	0.9969	0.9978	0.9861	0.9918	4.39M	2.8234
ResNet-101 [32]	0.9846	0.9772	0.9772	0.9772	42.51M	15.7288
Swin Base [33]	0.9938	0.9956	0.9839	0.9896	86.75M	30.3375

ResNet-101 was used as the representative CNN-based baseline in the comparative evaluation. The model achieved an accuracy of 0.9846, with precision, recall, and F1-score values of 0.9772, indicating a strong capacity to extract discriminative visual features from potato leaf images. This result confirms the effectiveness of deep residual learning for plant disease classification. Nevertheless, the model's computational cost remains relatively high, with 42.51 million parameters and 15.7288 GFLOPs, making it less suitable for resource-limited deployment scenarios when compared with more compact architectures. Although ResNet-101 provided a robust baseline, its performance was consistently exceeded by the transformer-based models, suggesting that attention-driven architectures may offer greater representational effectiveness for this classification task.

The Vision Transformer-based models achieved superior overall performance compared with the conventional CNN architecture. Swin Base reached an accuracy of 0.9938 and a precision of 0.9956, demonstrating the effectiveness of its hierarchical structure and shifted-window self-attention mechanism in capturing disease-relevant visual patterns. However, this high predictive performance was accompanied by the greatest computational demand among the evaluated models, requiring 86.75 million parameters and 30.3375 GFLOPs. By contrast, MobileViT v2 provided the most favorable balance between accuracy and efficiency. It achieved the highest accuracy of 0.9969, together with a precision of 0.9978 and an F1-score of 0.9918, while maintaining a substantially smaller computational footprint. The confusion matrix presented in Figure 2 further supports this result by showing that MobileViT v2 correctly classified all Early Blight and Late Blight samples, with 150 correct predictions in each category. Only one misclassification was observed in the Healthy class, where a single image was predicted as Late Blight, resulting in 23 correctly classified samples out of 24. This near-perfect classification pattern indicates that MobileViT v2 can effectively distinguish subtle visual differences among healthy and diseased potato leaves while preserving computational efficiency.

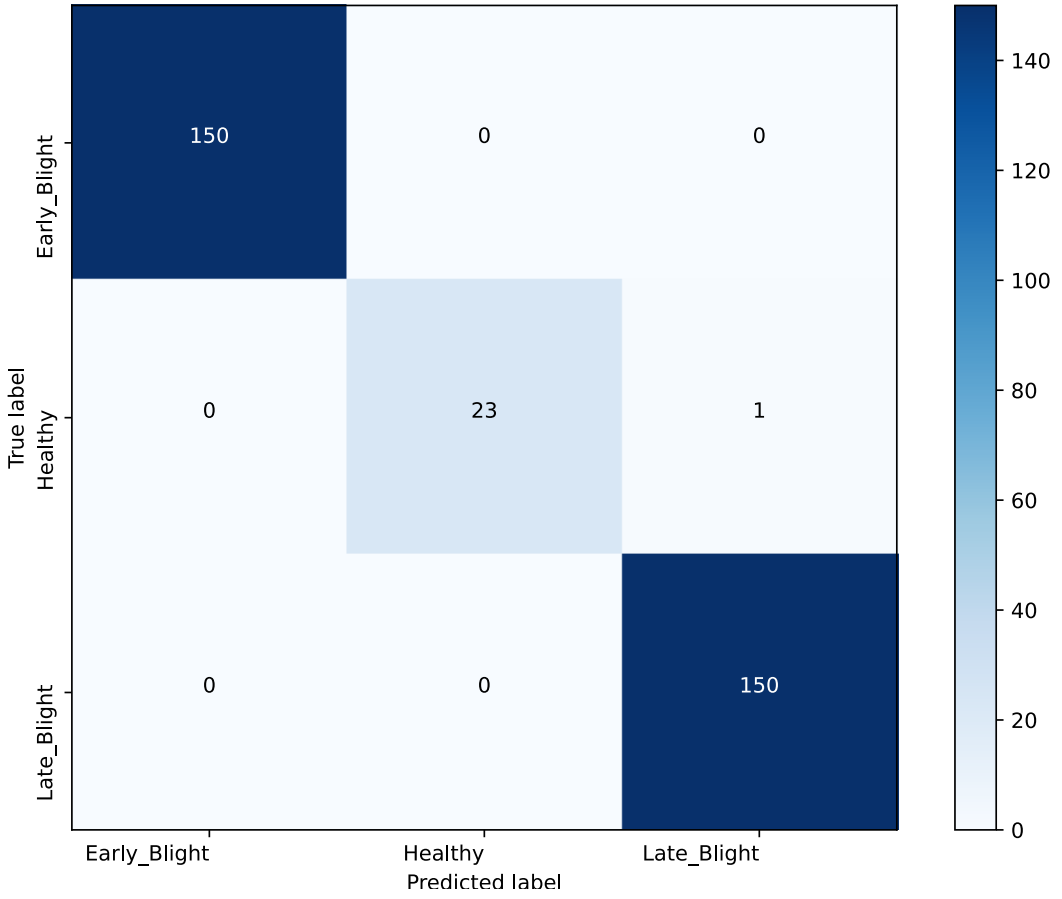

Figure 2. Confusion Matrix Showing the Classification Results of the Mobilevit v2 Model on the Test Dataset

Overall, the comparative findings demonstrate that Vision Transformer-based architectures provided superior performance for potato leaf disease classification when compared with the conventional CNN baseline. Although ResNet-101 achieved strong results and confirmed the effectiveness of deep residual learning, both Swin Base and MobileViT v2 consistently outperformed it across the main evaluation metrics. The most notable result was observed for MobileViT v2, which achieved the best overall accuracy of 0.9969 while also requiring the lowest computational resources among all evaluated models. With only 4.39 million parameters and 2.8234 GFLOPs, MobileViT v2 was substantially more compact than ResNet-101 and Swin Base, using approximately one-tenth and one-twentieth of their parameter counts, respectively. This favorable balance between predictive accuracy and computational efficiency makes MobileViT v2 particularly suitable for practical deployment in resource-constrained agricultural environments, including mobile-based diagnostic platforms and on-farm decision-support systems. These results indicate that lightweight transformer architectures can deliver highly accurate disease recognition without imposing excessive computational demands, offering a promising direction for scalable and field-ready applications in precision agriculture.

5. Conclusion

This study presented a comparative evaluation of CNN- and Vision Transformer-based architectures for the automated classification of potato leaf diseases, with the objective of identifying a model that can provide both high diagnostic accuracy and computational efficiency for practical agricultural deployment. Using the PlantVillage potato leaf dataset under a standardized experimental protocol, ResNet-101 was assessed as a representative deep CNN baseline, while Swin Base and MobileViT v2 were evaluated as transformer-based

alternatives. The results showed that although ResNet-101 achieved a strong baseline accuracy of 98.46%, both Vision Transformer models delivered superior classification performance. Among all evaluated architectures, MobileViT v2 achieved the best overall result, reaching 99.69% accuracy on the test set while requiring only 4.39 million parameters. This finding demonstrates that lightweight transformer-based models can offer a favorable balance between predictive performance and computational cost, making them particularly suitable for mobile applications, embedded platforms, and low-power on-farm diagnostic systems. The ability of such models to support rapid and reliable disease identification may contribute to earlier intervention, more efficient resource use, reduced unnecessary chemical application, and more sustainable potato production. Future work should focus on validating the proposed approach using more diverse field-acquired datasets that include variations in illumination, background complexity, disease severity, and geographic conditions, as well as developing user-oriented deployment frameworks to translate the model into a practical decision-support tool for farmers and agricultural specialists.

References

1. P. Kalpana, R. Anandan, A.G. Hussien, H. Migdady, L. Abualigah, Plant disease recognition using residual convolutional enlightened Swin transformer networks, Sci. Rep. 14 (2024). https://doi.org/10.1038/s41598-024-56393-8.
2. S. Kamal, P. Sharma, P.K. Gupta, M.K. Siddiqui, A. Singh, A. Dutt, DVTXAI: a novel deep vision transformer with an explainable AI-based framework and its application in agriculture, Journal of Supercomputing 81 (2025). https://doi.org/10.1007/s11227-024-06494-y.
3. X. Chen, Y. Li, Z. Zhang, Effvit-Potatonet: An Efficientvit-Based Model for Potato Leaf Disease Classification, in: Proceedings - 2025 International Conference on Digital Analysis and Processing, Intelligent Computation, DAPIC 2025, Institute of Electrical and Electronics Engineers Inc., 2025: pp. 84–88. https://doi.org/10.1109/DAPIC66097.2025.00022.
4. F.O. Isinkaye, M.O. Olusanya, A.A. Akinyelu, A multi-class hybrid variational autoencoder and vision transformer model for enhanced plant disease identification, Intelligent Systems with Applications 26 (2025). https://doi.org/10.1016/j.iswa.2025.200490.
5. J.H. Sinamenye, A. Chatterjee, R. Shrestha, Potato plant disease detection: leveraging hybrid deep learning models, BMC Plant Biol. 25 (2025). https://doi.org/10.1186/s12870-025-06679-4.
6. A. Bajpai, S. Sahu, N. Tiwari, V. Srivastava, S. Yadav, An Efficient Approach for Potato Leaf Disease Classification Using Cascaded CNN-Transformers, in: International IEEE Conference Proceedings, IS, Institute of Electrical and Electronics Engineers Inc., 2024. https://doi.org/10.1109/IS61756.2024.10705224.
7. S. Austin, A. Barua, S.N. Haider, F.L. Niha, M. Faisal, S.M. Shawon, Precision Classification of Potato Diseases Using Transformer-Enhanced CNNs, in: 2025 International Conference on Quantum Photonics, Artificial Intelligence, and Networking (QPAIN), IEEE, 2025: pp. 1–6. https://doi.org/10.1109/QPAIN66474.2025.11172153.
8. S. Adhikari, Advancements in Agricultural Technology: Vision Transformer-Based Potato Leaf Disease Classification, Journal of Soft Computing Paradigm 6 (2024) 169–185. https://doi.org/10.36548/jscp.2024.2.005.
9. Y. Cakmak, N. Pacal, Deep Learning for Automated Breast Cancer Detection in Ultrasound: A Comparative Study of Four CNN Architectures, Artificial Intelligence in Applied Sciences 1 (2025) 13–19. https://doi.org/10.69882/ADBA.AI.2025073.
10. Y. Cakmak, Machine Learning Approaches for Enhanced Diagnosis of Hematological Disorders, Computational Systems and Artificial Intelligence 1 (2025) 8–14. https://doi.org/10.69882/ADBA.CSAI.2025072.

11. M.E. Ahmed, H.H. Tuhin, M.A. Kayum, Md.J. Islam, Md.T. Ahad, A Vision Transformer Approach to Potato Leaf Disease Detection, in: 2025 International Conference on Quantum Photonics, Artificial Intelligence, and Networking (QPAIN), IEEE, 2025: pp. 1–5. https://doi.org/10.1109/QPAIN66474.2025.11172146.
12. J. Zeynalov, Y. Çakmak, İ. Paçal, Automated Apple Leaf Disease Classification Using Deep Convolutional Neural Networks: A Comparative Study on the Plant Village Dataset, Journal of Computer Science and Digital Technologies 1 (2025) 5–17. https://doi.org/10.61640/jcsdt.2025.0601.
13. I. Pacal, G. Işık, Utilizing convolutional neural networks and vision transformers for precise corn leaf disease identification, Neural Comput. Appl. 37 (2024) 2479–2496. https://doi.org/10.1007/S00521-024-10769-Z/TABLES/5.
14. I. Kunduracioglu, I. Pacal, Advancements in deep learning for accurate classification of grape leaves and diagnosis of grape diseases, Journal of Plant Diseases and Protection 131 (2024) 1061–1080. https://doi.org/10.1007/S41348-024-00896-Z/TABLES/7.
15. Y. Cakmak, A. Maman, Deep Learning for Early Diagnosis of Lung Cancer, Computational Systems and Artificial Intelligence 1 (2025) 20–25. https://doi.org/10.69882/ADBA.CSAI.2025074.
16. Y. Cakmak, J. Zeynalov, A Comparative Analysis of Convolutional Neural Network Architectures for Breast Cancer Classification from Mammograms, Artificial Intelligence in Applied Sciences 1 (2025) 28–34. https://doi.org/10.69882/ADBA.AI.2025075.
17. N.K. Tiwari, S.S. Rajput, Enhancing Potato Leaf Disease Detection Using Super-Resolution and Multi-path Multi-attention Transformers, Potato Res. (2025). https://doi.org/10.1007/s11540-025-09890-w.
18. 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. 238 (2024) 122099. https://doi.org/10.1016/J.ESWA.2023.122099.
19. I. Pacal, I. Kunduracioglu, M.H. Alma, M. Deveci, S. Kadry, J. Nedoma, V. Slany, R. Martinek, A systematic review of deep learning techniques for plant diseases, Artificial Intelligence Review 2024 57:11 57 (2024) 1–39. https://doi.org/10.1007/S10462-024-10944-7.
20. A. Nandana, N. M, V.E. R, A Transformer-based Ensemble Model for Plant Disease Detection, in: 2025 5th International Conference on Soft Computing for Security Applications (ICSCSA), IEEE, 2025: pp. 833–839. https://doi.org/10.1109/ICSCSA66339.2025.11170705.
21. R.F. Wang, W.H. Su, The Application of Deep Learning in the Whole Potato Production Chain: A Comprehensive Review, Agriculture (Switzerland) 14 (2024). https://doi.org/10.3390/agriculture14081225.
22. G. Chemmalar Selvi, H.J. Charan, D. Kumar, CropViT: A light-weight Transformer Model for Crop Disease Detection, in: 2024 3rd International Conference on Artificial Intelligence for Internet of Things, AIIoT 2024, Institute of Electrical and Electronics Engineers Inc., 2024. https://doi.org/10.1109/AIIoT58432.2024.10574729.
23. S. Dutta, S.G. Neogi, 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 2024, Institute of Electrical and Electronics Engineers Inc., 2024. https://doi.org/10.1109/INSPECT63485.2024.10896181.
24. A. Bajpai, N. Tiwari, P. Rajput, S. Sahu, D. Singh, Enhanced Potato Leaf Disease Detection via Modified Swin Transformer Architecture, in: 2024 15th International Conference on Computing Communication and Networking Technologies, ICCCNT 2024, Institute of Electrical and Electronics Engineers Inc., 2024. https://doi.org/10.1109/ICCCNT61001.2024.10724512.

25. C. Zhang, S. Wang, C. Wang, H. Wang, Y. Du, Z. Zong, Research on a Potato Leaf Disease Diagnosis System Based on Deep Learning, Agriculture (Switzerland) 15 (2025). https://doi.org/10.3390/agriculture15040424.
26. A. Sharma, A. Sharma, Recurrent Neural Network-Based Classification of Potato Leaves using RGB Images, in: Proceedings - 2nd International Conference on Advancement in Computation and Computer Technologies, InCACCT 2024, Institute of Electrical and Electronics Engineers Inc., 2024: pp. 486–491. https://doi.org/10.1109/InCACCT61598.2024.10551226.
27. Y. Zoralioglu, Ö. Polat, Detection of Potato Plant Disease from Leaf Images using Deep Learning Models, in: 2024 Innovations in Intelligent Systems and Applications Conference, ASYU 2024, Institute of Electrical and Electronics Engineers Inc., 2024. https://doi.org/10.1109/ASYU62119.2024.10756961.
28. PlantVillage Dataset, (n.d.). https://www.kaggle.com/datasets/emmarex/plantdisease (accessed October 10, 2025).
29. Z. Wang, P. Wang, K. Liu, P. Wang, Y. Fu, C.-T. Lu, C.C. Aggarwal, J. Pei, Y. Zhou, A Comprehensive Survey on Data Augmentation, (2024). https://arxiv.org/abs/2405.09591v3 (accessed May 28, 2025).
30. A. Mumuni, F. Mumuni, N.K. Gerrar, A Survey of Synthetic Data Augmentation Methods in Machine Vision, Machine Intelligence Research 2024 21:5 21 (2024) 831–869. https://doi.org/10.1007/S11633-022-1411-7.
31. S. Mehta, A. Mohammad, R. Apple, Separable Self-attention for Mobile Vision Transformers, (n.d.). https://github.com/apple/ml-cvnets (accessed July 23, 2025).
32. K. He, X. Zhang, S. Ren, J. Sun, Deep Residual Learning for Image Recognition, (n.d.). http://image-net.org/challenges/LSVRC/2015/ (accessed September 25, 2025).
33. Z. Liu, Y. Lin, Y. Cao, H. Hu, Y. Wei, Z. Zhang, S. Lin, B. Guo, Swin Transformer: Hierarchical Vision Transformer using Shifted Windows, (n.d.). https://github. (accessed September 25, 2025).