

PLANXMAMBA: A hybrid CNN – MAMBA model for plant disease recognition

Tuan Nguyen Huu

Huy Phan Le

Hieu Nguyen Chi

Duong Nguyen Tung

Anh Nguyen Thai

Quynh Dao Thi Thuy

quynhdt@ptit.edu.vn

Research Article

Keywords: plant disease recognition, PlantXMamba, state space model, CNN, Mamba, smart agriculture, image classification

Posted Date: August 21st, 2025

DOI: <https://doi.org/10.21203/rs.3.rs-7362851/v1>

License:  This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Additional Declarations: No competing interests reported.

PLANXMAMBA: A hybrid CNN – MAMBA model for plant disease recognition

Tuan Nguyen Huu¹, Huy Phan Le^{2†}, Hieu Nguyen Chi^{2†},
Duong Nguyen Tung^{2†}, Anh Nguyen Thai^{2†},
Quynh Dao Thi Thuy^{3*†}

¹Teaching Assistant at Information Technology Faculty, PTIT, Ha Noi,
Viet Nam.

²Information Technology student, PTIT, Ha Noi, Viet Nam.

^{3*}Lecturer at Information Technology Faculty, PTIT, Ha Noi, Viet Nam.

*Corresponding author(s). E-mail(s): quynhdt@ptit.edu.vn;

Contributing authors: tuannh@ptit.edu.vn;

huypl.b22at149@stu.ptit.edu.vn; hieunc.b22dt114@stu.ptit.edu.vn;

duongnt.b22vt114@stu.ptit.edu.vn; anhnt.b22vt028@stu.ptit.edu.vn;

[†]These authors contributed equally to this work.

Abstract

Image-based plant disease recognition plays a pivotal role in smart agriculture, facilitating early detection and effective management of crop diseases. Existing approaches primarily employ Convolutional Neural Networks (CNNs) or Vision Transformers (ViTs) to extract discriminative visual features and perform disease classification, achieving encouraging outcomes. However, these models often struggle to adequately model long-range spatial dependencies and sequence-level information while maintaining lightweight architectures suitable for deployment on mobile or edge devices. In this study, we introduce PlantXMamba, an efficient hybrid model that synergistically integrates CNNs with a structured State Space Model (SSM), termed Mamba, to simultaneously capture local and global contextual features. The architecture is designed to enhance accuracy, interpretability, and computational efficiency. Comprehensive experiments were conducted on multiple benchmark datasets—including Maize, Rice, Apple, Embrapa, and PlantVillage—to demonstrate the effectiveness and generalizability of the proposed approach.

Keywords: plant disease recognition, PlantXMamba, state space model, CNN, Mamba, smart agriculture, image classification

1 Introduction

The widespread adoption of deep learning techniques over the past decade, particularly Convolutional Neural Networks (CNNs), has significantly advanced the field of image-based plant disease recognition [1–9]. CNN-based models such as VGG, ResNet, and Inception (GoogleNet), as well as more lightweight alternatives like MobileNet, often augmented with attention mechanisms, have been extensively utilized to detect and classify disease symptoms in crops like maize and rice [1–4] [5]. These models excel at learning localized visual patterns, such as necrotic spots, chlorosis, and marginal blight, which are essential for accurate diagnosis. Despite their effectiveness in capturing fine-grained visual cues, conventional CNN architectures exhibit inherent limitations in modeling global contextual dependencies—an important aspect when symptoms are spatially diffuse or irregularly distributed across the foliar surface. Additionally, their substantial computational complexity often hinders real-time deployment in field environments with limited processing capabilities.

To address these challenges, Transformer-based models, especially the Vision Transformer (ViT) [10–12], have gained traction. By replacing convolutional operations with self-attention mechanisms, ViTs effectively model long-range dependencies between image patches and have shown promising results in complex visual recognition tasks [11, 13–15]. However, their reliance on large-scale training datasets and high computational cost renders them less suitable for lightweight or real-time agricultural applications.

Recently, the Mamba architecture, a member of the State Space Model (SSM) family [16–21], has emerged as a viable alternative. Mamba leverages selective state-space transitions to capture temporal and spatial dependencies with linear time complexity, offering an appealing trade-off between modeling capacity and computational efficiency [22–24]. Motivated by the complementary strengths of CNNs and SSMs, this paper proposes PlantXMamba, a hybrid architecture that unifies hierarchical local feature extraction and global contextual modeling for enhanced plant disease recognition.

2 Related work

In recent years, the advancement and consolidation of deep learning techniques—particularly Convolutional Neural Networks (CNNs)—have significantly contributed to the evolution of image-based systems for plant disease diagnosis. CNN architectures have demonstrated exceptional capability in learning fine-grained local visual patterns and in accurately identifying pathological symptoms manifested on foliar surfaces.

Numerous foundational studies have highlighted the effectiveness of CNN-based models. For instance, Chen et al. [1] [2] developed lightweight CNNs integrated with attention mechanisms, achieving classification accuracies of approximately 98.5% on maize and rice datasets. Zhao et al. [3] introduced RIC-Net, a sophisticated architecture that combines Inception and Residual modules alongside attention layers, reaching a state-of-the-art accuracy of 99.55%. Similarly, MobileNet-V2, enhanced with attention mechanisms [4], reported performance in the range of 97.3% to 99.1% on the PlantVillage dataset [25]. Moreover, transfer learning methods that utilize

deep architectures such as ResNet and DenseNet have achieved high accuracy in crop-specific disease classification tasks, particularly in rice [26] and maize [27].

Despite these advances, conventional CNNs inherently struggle to capture global contextual semantics across spatial regions—an essential capability for identifying symptoms that are spatially diffuse or heterogeneously distributed. To address this limitation, Vision Transformers (ViTs) [10] have emerged as a transformative architecture in computer vision. By discarding convolutional inductive biases in favor of self-attention mechanisms, ViTs effectively model long-range dependencies across image patches, thus enhancing performance in tasks requiring global perceptual understanding.

Within the domain of plant disease detection, both pure ViT-based models and hybrid CNN–ViT architectures have been explored. For example, PlantXViT [13], a hybrid explainable model, reported classification accuracies ranging from 92.6% to 98.3% across Apple, Maize, Rice, PlantVillage and Embrapa datasets [25, 26, 28, 29]. Other works have leveraged ViTs alongside CNN backbones such as VGG16 and DenseNet, achieving classification accuracies of 99.24% and 98.0% for apple and maize leaf diseases, respectively [30]. Additionally, the Data-efficient Image Transformer (DeiT) has shown superior performance in classifying mango leaf diseases, surpassing CNN counterparts with comparable parameter budgets by attaining an accuracy of 99.75% [31].

Concurrently, a promising alternative to Transformer-based models has emerged in the form of State Space Models (SSMs), particularly the Mamba architecture [16] [20] [32] [21]. Mamba introduces a novel sequential modeling paradigm that utilizes selective state transitions to capture long-range dependencies with linear computational complexity. Its efficient design makes it well-suited for spatiotemporal feature modeling and high-throughput inference on GPU/TPU accelerators. Noteworthy applications include YOLO-BSMamba [17], an enhanced YOLOv8s model equipped with Mamba modules for in-field tomato leaf disease detection. By incorporating the HCMamba variant, the model achieved significant gains in accuracy, recall, and mean average precision (mAP), with improvements of 3–5% over baseline models [33].

Architectural variants of Mamba have also been explored for plant disease classification. ConMamba [34] proposed a self-supervised framework utilizing the Vision Mamba Encoder (VME), which effectively captures global inter-patch dependencies, outperforming traditional CNNs across benchmark datasets. VMamba [35] further introduced a Selective Spatial Scanning module (SS2D) tailored for plant leaf disease recognition, and demonstrated superior predictive accuracy and training efficiency compared to models such as ResNet-50, ViT, and Swin Transformer.

Motivated by these advancements, the present study introduces PlantXMamba, a novel hybrid architecture that fuses the hierarchical local feature extraction of CNNs with the global modeling capabilities of Mamba. The architecture is systematically evaluated in three configurations:

- VGG16 backbone.
- InceptionNet backbone.
- A composite VGG16 + InceptionNet backbone.

This hybrid design aims to balance representational richness with architectural compactness, making it suitable for real-time deployment in resource-constrained agricultural settings without compromising diagnostic performance.

3 Methodology

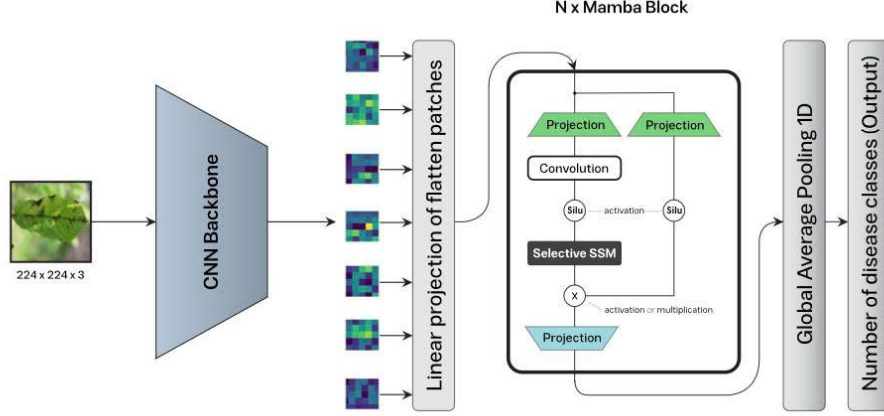


Fig. 1: Architecture of PlantXMamba

The process of plant disease recognition from images typically involves three key stages: image feature extraction, spatial relationship modeling, and disease classification. Previous approaches have primarily relied on conventional Convolutional Neural Networks (CNNs) or Transformer-based architectures for feature learning. However, these methods often face limitations in effectively modeling long-range dependencies while maintaining computational efficiency. In this study, we propose PlantXMamba, a novel architecture that integrates CNNs with a Structured State Space Model (Mamba) to simultaneously capture both local features and global spatial relationships in leaf images. As illustrated in Figure 1, the proposed model consists of three main phases: extracting visual features from input images using CNNs, modeling spatial dependencies through Mamba blocks, and classifying leaf diseases based on the aggregated features.

The PlantXMamba model is designed to handle input images with flexible resolutions, depending on the chosen CNN backbone architecture. The backbone is configurable with various pretrained CNN architectures such as VGG16 or Inception, allowing adaptation to the characteristics and requirements of different datasets. The CNN backbone is responsible for reducing the spatial dimensions of the input image (e.g., from $224 \times 224 \times 3$) to a compressed spatial representation (typically $56 \times 56 \times 512$), which serves as the input to the subsequent patch embedding and Mamba blocks.

The model retains the ability to process images of varying resolutions, which is advantageous for deployment in real-world applications. In our experiments, images are normalized to fixed sizes such as 224×224 or 299×299 . The resulting feature tensor is then partitioned into 121 patches of size 5×5 and linearly projected into a sequence with shape (121, 16). This sequence is fed into a stack of N Mamba blocks to extract local features. Specifically, PlantXMamba employs N = 4 stacked Mamba blocks to perform this task. Finally, a 1D pooling layer and a fully connected layer are applied to aggregate the features and produce the final classification output.

Here, the loss for all the n samples in a batch was calculated using y_i as the actual label and $\hat{y}_i$ as the predicted value of the i -th sample.

$$Loss = -\frac{1}{n} \sum_{i=1}^n y_i \log(\hat{y}_i) \quad (1)$$

To obtain a comprehensive understanding of the model’s performance, we employed a set of standard classification evaluation metrics. These metrics assess various aspects of the model’s predictive capability.

Accuracy: Used as an overall indicator, accuracy measures the proportion of correct predictions over the entire dataset.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (2)$$

Precision: The ratio of true positive samples to the total number of samples predicted as positive.

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

Recall: The ratio of true positive samples to the total number of actual positive samples.

$$Recall = \frac{TP}{TP + FP} \quad (4)$$

F1-Score: It relates Precision and Recall by calculating the harmonic mean between them.

$$F1\text{-Score} = 2 \times \frac{Precision \times Recall}{Precision + Recall} \quad (5)$$

Area Under the Curve (AUC): AUC is the area covered by the receiver characteristic operator ROC curve. The ROC is calculated using the plot of True Positive Rate (TPR) against the False Positive Rate (FPR).

$$TPR = \frac{TP}{TP + FN} \quad (6)$$

$$\text{FPR} = \frac{FP}{FP + TN} \quad (7)$$

Cohen’s Kappa Score: Cohen’s kappa coefficient or score is a probability based measure where the outcome is the level of agreement parties for classification problems.

$$\kappa = \frac{p_o - p_e}{1 - p_e} \quad (8)$$

Where p_o is the relative agreement probability and p_e is the hypothetical agreement probability between the parties.

4 Experimental Results

4.1 Datasets

Table 1: Datasets overview

Dataset	# of classes	# of images
Apple [29]	4	1821
Embrapa [28]	93	46376
Maize [26]	4	1999
Plant Village [25]	38	54305
Rice [26]	5	560

Experiments were carried out on five publicly available datasets, each representing different geographical regions and varying in scale, balance, and crop diversity. The objective was to train and evaluate the proposed model across a broad spectrum of plant diseases and imaging conditions, thereby assessing its generalizability and robustness.

The datasets encompass both small, balanced collections (approximately 400–500 images) and large, imbalanced datasets (exceeding 40,000 images). Specifically, the PlantVillage dataset comprises 54,305 images across 38 disease classes, while the Embrapa dataset includes 46,376 images categorized into 93 classes—both containing data from multiple plant species with foliar symptoms. In contrast, the Apple, Maize, and Rice datasets are single-species collections utilized for targeted evaluation on specific crop types.

A summary of dataset characteristics is provided in Table 1, and representative image samples from each dataset are illustrated in Figure 2, depicting a range of disease manifestations such as apple scab, Embrapa brusone, maize gray leaf spot, potato early blight, and rice bacterial streak.

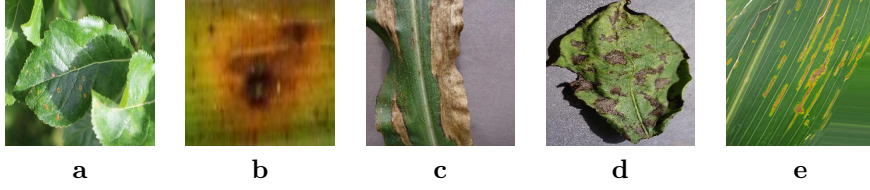


Fig. 2: Sample images from all datasets: (a) scab (Apple), (b) brusone (Embrapa), (c) gray leaf spot (Maize), (d) potato early blight (PlantVillage), (e) bacterial leaf streak (Rice).

4.2 Experimental Setup

All experiments were carried out in a high performance computational environment equipped with an NVIDIA A100 GPU (40GB memory), an Intel Xeon processor (32 cores), and 128GB of RAM. The software environment was configured using Python 3.8, PyTorch 1.12, and the CUDA 11.4 acceleration toolkit.

During training, the input images were resized to uniform dimensions (e.g., 224×224 or 299×299), depending on the configuration of the CNN backbone. The model was trained using the Adam optimizer with a learning rate of 0.0001 and a batch size of 16. Training performance was monitored using multi-class cross-entropy loss, with model selection guided by validation accuracy

4.3 Results and analysis

4.3.1 Performance of Backbone Configurations

A comparative evaluation of PlantXMamba under various CNN backbones is presented in Table 2. On the large-scale datasets (PlantVillage and Embrapa), the VGG16 backbone consistently yielded superior performance across most evaluation metrics, including Accuracy, Precision, Recall, F1-Score, and Cohen’s Kappa.

For the smaller datasets, Rice and Maize achieved the highest accuracy with the InceptionNet backbone, while the Apple dataset exhibited the best overall performance when using the combined VGG16 + InceptionNet configuration. These findings suggest that different backbone configurations offer complementary strengths depending on the dataset size and complexity. The consistent improvements observed across multiple evaluation metrics further support the robustness and adaptability of the PlantXMamba architecture under varying backbone strategies.

Table 2: PlantXMamba’s performance with different backbones (in %)

Backbone	Accuracy	Precision	Recall	F1 Score	Kappa
PlantVillage Dataset					
VGG16 + InceptionNet	98.71%	98.74%	98.71%	98.72%	98.65%
VGG16	98.76%	98.82%	98.76%	98.77%	98.70%
InceptionNet	96.26%	96.49%	96.26%	96.24%	96.08%
Embrapa Dataset					
VGG16 + InceptionNet	94.34%	94.34%	94.32%	94.32%	93.26%
VGG16	98.72%	98.72%	98.72%	98.72%	98.68%
InceptionNet	98.65%	98.65%	98.65%	98.65%	98.55%
Apple Dataset					
VGG16 + InceptionNet	97.08%	97.43%	97.08%	97.20%	95.81%
VGG16	95.99%	96.58%	95.99%	96.21%	94.25%
InceptionNet	96.35%	96.08%	96.35%	96.15%	94.71%
Rice Dataset					
VGG16 + InceptionNet	99.00%	99.05%	99.00%	99.00%	98.74%
VGG16	98.00%	98.17%	98.00%	98.02%	97.48%
InceptionNet	99.00%	99.05%	99.00%	99.00%	98.74%
Maize Dataset					
VGG16 + InceptionNet	97.75%	97.69%	97.75%	97.75%	97.00%
VGG16	96.50%	96.60%	96.50%	96.50%	95.33%
InceptionNet	98.50%	98.51%	98.50%	98.50%	98.00%

4.3.2 Compare with other models:

PlantXMamba was compared with five plant disease classification models using CNN [1–4, 36] and a combination model between CNN and ViT [13]. The models selected are all modern architectures that integrate attention mechanisms and have a compact structure. Each model represents a different approach to leaf diseases detection, so the comparative results are highly reference and directional. Table 3 presents results comparing the accuracy of the models, with PlantXMamba achieving accuracy of 97.08%, 89.34%, 98.50%, 98.76% and 99.00% on five datasets Apple, Embrapa, Maize, PlantVillage, and Rice, respectively.

5 Conclusion

In this study, we proposed and developed PlantXMamba — a hybrid architecture combining Convolutional Neural Networks (CNNs) and Mamba blocks—to address the problem of plant leaf disease classification. The model leverages the effective local feature extraction capabilities of CNNs with the capacity of Mamba to model long-range dependencies within the image space, thereby enhancing accuracy in computer vision tasks.

Table 3: Comparison of the work with other state-of-the-art

Approach	Accuracy	Precision	Recall	F1 Score	Kappa
Apple Dataset					
Karthik et al. [36]	62.37	62.84	61.83	62.35	0.45
Chen et al. [1]	55.91	56.35	54.84	55.58	0.36
Chen et al. [4]	83.33	83.7	82.8	83.25	0.76
Chen et al. [2]	83.33	85.47	82.26	83.83	0.76
Zhao et al. [3]	88.71	89.62	88.17	88.89	0.84
Thankur et al. [13]	93.55	93.55	93.55	93.55	0.91
PlantXMamba	97.08	97.43	97.08	97.20	95.81
Embrapa Dataset					
Karthik et al. [36]	80.29	83.07	78.38	80.6	0.82
Chen et al. [1]	80.95	85.08	76.02	80.3	0.78
Chen et al. [4]	73.63	80.12	67.86	73.48	0.73
Chen et al. [2]	74.88	82.93	66.06	73.18	0.75
Zhao et al. [3]	88.48	89.13	87.14	88.28	0.88
Thankur et al. [13]	89.24	91.17	88.27	89.7	0.89
PlantXMamba	89.34	89.27	89.34	89.05	88.93
Maize Dataset					
Karthik et al. [36]	54.32	59.72	53.09	56.21	0.39
Chen et al. [1]	49.38	50.85	37.04	42.86	0.33
Chen et al. [4]	87.65	87.65	87.65	87.65	0.84
Chen et al. [2]	88.89	91.03	87.65	89.31	0.85
Zhao et al. [3]	77.78	78.75	77.78	78.26	0.7
Thankur et al. [13]	92.59	93.67	91.36	92.5	0.9
PlantXMamba	98.50	98.51	98.50	98.50	98.00
PlantVillage Dataset					
Karthik et al. [36]	95.83	96.2	95.6	95.89	0.96
Chen et al. [1]	87.94	89.59	86.71	88.07	0.85
Chen et al. [4]	74.63	80.92	69.64	75.03	0.74
Chen et al. [2]	96.68	97.49	95.83	96.64	0.97
Zhao et al. [3]	97.28	97.49	97.06	97.27	97.15
Thankur et al. [13]	98.86	98.9	98.81	98.85	0.99
PlantXMamba	98.76	98.82	98.76	98.77	98.70
Rice Dataset					
Karthik et al. [36]	71.67	79.55	58.33	69.31	0.64
Chen et al. [1]	76.67	83.02	73.33	77.87	0.71
Chen et al. [4]	96.67	96.67	96.67	96.67	0.95
Chen et al. [2]	95	96.61	96.61	96.61	0.94
Zhao et al. [3]	91.67	91.67	91.67	91.67	89.27
Thankur et al. [13]	98.33	98.33	98.33	98.33	0.98
PlantXMamba	99.00	99.05	99.00	99.00	98.74

Experimental results on five public datasets demonstrate that PlantXMamba achieves outstanding performance compared to ViT-based models on small to medium-sized datasets, including Rice (99.00%), PlantVillage (98.76%), Maize (98.50%), and Apple (97.08%). Notably, on the more complex and diverse Embrapa dataset, the

model still attained an accuracy of 89.34%, showcasing its stability and practical application potential.

However, PlantXMamba requires further improvement to enhance its ability to model complex global-level relationships, particularly in the context of highly imbalanced data. The evaluation metrics indicate that PlantXMamba is an effective, lightweight, and easily deployable approach. Future research directions include: optimizing the architecture to improve performance on larger datasets; applying data augmentation techniques and weighted loss functions to address class imbalance; and exploring self-supervised learning to better leverage unlabeled datasets.

Acknowledgement

This research is funded by the Posts and Telecommunications Institute of Technology (PTIT). The authors would like to thank PTIT for the financial support.

Declarations

This section presents declarations related to this study.

- Conflict of interest/Competing interests: No conflict of interest exists in the submission of this manuscript, and manuscript is approved by all authors for publication.
- Ethics approval: Consent was obtained from all participants prior to their involvements in the study, and they were informed of their right to withdraw at any time without consequence.
- Consent to participate: All authors agreed to participate in the construction and development of this research topic.
- Consent to publication: All authors agreed to make this study public.
- Availability of data and materials: The paper uses public datasets, namely QMUL-Chair-V2 and QMUL-Shoe-V2, and follows the guidelines provided by the respective dataset authors for usage
- Authors' contributions: All authors contributed to study's conception and design. All authors read and approved the final manuscript.

Funding

This work was supported by Posts and Telecommunications Institute of Technology

References

- [1] Chen, J., Wang, W., Zhang, D., Zeb, A., Nanekaran, Y.A.: Attention embedded lightweight network for maize disease recognition. Plant Pathology, 1483–1491 (2020) <https://doi.org/10.1111/ppa.13322>

- [2] Chen, J., Zeb, A., Zhang, D., Nanekharan, A.A.: Identification of rice plant diseases using lightweight attention networks. *Expert Systems with Applications* **169** (2021) <https://doi.org/10.1016/j.eswa.2020.114514>
- [3] Zhao, Y., Sun, C., Xu, X., Chen, J.: Ric-net: A plant disease classification model based on the fusion of inception and residual structure and embedded attention mechanism. *Computers and Electronics in Agriculture* **193** (2022) <https://doi.org/10.1016/j.compag.2021.106644>
- [4] Chen, J., Zhang, D., Suzauddola, M., Zeb, A.: Identifying crop diseases using attention embedded mobilenet-v2 model. *Applied Soft Computing* **113** (2021) <https://doi.org/10.1016/j.asoc.2021.107901>
- [5] Abade, A., Ferreira, P.A., Vidal, F.D.B.: Plant diseases recognition on images using convolutional neural networks: A systematic review. *Computers and Electronics in Agriculture* **185** (2021) <https://doi.org/10.1016/j.compag.2021.106125>
- [6] Agarwal, M., Gupta, S.K., Biswas, K.K.: Development of efficient cnn model for tomato crop disease identification. *Sustainable Computing Informatics and Systems* **28** (2020) <https://doi.org/10.1016/j.suscom.2020.100407>
- [7] Sun, X., Li, G., Qu, P., Xie, X., Pan, X., Zhang, W.: Research on plant disease identification based on cnn. *Cognitive Robotics* **2** (2022) <https://doi.org/10.1016/j.cogr.2022.07.001>
- [8] Shrestha, G., Deepshikha, Das, M., Dey, N.: Plant disease detection using cnn. *Proceedings of the 2020 IEEE Applied Signal Processing Conference (ASPCON)* (2020)
- [9] Zhao, Y., Sun, C., Xu, X., Chen, J.: Plant disease detection using cnn. *Proceedings of the ITM Web of Conferences* **03049** (2022) <https://doi.org/10.1016/j.compag.2021.106644>
- [10] Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., Dehghani, M., Minderer, M., Heigold, G., Gelly, S., Uszkoreit, J., Houlsby, N.: An image is worth 16×16 words: Transformers for image recognition at scale. In: *Proceedings of the International Conference on Learning Representations (ICLR)* (2021). <https://doi.org/10.48550/arXiv.2010.11929>
- [11] Prashanthi, B., Krishna, A.V.P., Rao, C.M.: Levit: Leaf disease identification and classification using an enhanced vision transformers (vit) model. *Multi-media Tools and Applications* **84**, 23313–23344 (2025) <https://doi.org/10.1007/s11042-024-19866-6>
- [12] Yu, S., Xie, L., Huangn, Q.: Inception convolutional vision transformers for plant disease identification. *Internet of Things* **21** (2023) <https://doi.org/10.1016/j.iot.2022.100650>

- [13] Thakur, P.S., Khanna, P., Sheorey, T., Ojha, A.: Explainable vision transformer enabled convolutional neural network for plant disease identification: Plan-txvit. arXiv preprint arXiv:2207.07919 (2022) <https://doi.org/10.48550/arXiv.2207.07919>
- [14] Xu, M., Yoon, S., Jeong, Y., Park, D.S.: Transfer learning for versatile plant disease recognition with limited data. *Frontiers in Plant Science* **13** (2022) <https://doi.org/10.3389/fpls.2022.1010981>
- [15] Srivastava, M., Sisaudia, V., Meena, J.: Agritl-vit: A vision transformer model with attention techniques for classification of plant leaf disease. *Expert Systems with Applications* **294** (2025) <https://doi.org/10.1016/j.eswa.2025.128793>
- [16] Gu, A., Dao, T.: Mamba: Linear-time sequence modeling with selective state spaces. arXiv preprint arXiv:2312.00752 (2023) <https://doi.org/10.48550/arXiv.2312.00752>
- [17] Yuan, C., Li, S., Wang, K., Liu, Q., Li, W., Zhao, W., Guo, G., al.: Mamba-yolo-ml: A state-space model-based approach for mulberry leaf disease detection. *Plants* **14** (2025)
- [18] Wang, X., Tang, S.H., Ariffin, M.K.A.B.M., Ismail, M.I.S.B., al.: Leafmamba: A novel iot-integrated network for accurate and efficient plant leaf disease detection. *Alexandria Engineering Journal* **64** (2025)
- [19] Tang, Z., Zhu, Y., Huang, M., Zhu, J.: A study on a wheat image recognition system based on transformer and mamba networks with light enhancement. In: *Proceedings of the 2024 5th International Conference on Artificial Intelligence and Computer Engineering (ICAICE)*. IEEE, ??? (2024)
- [20] Quang-Vinh Dinh, M.-D.B.: Mamba: When transformer is not all you need. *AI VIET NAM* (2024)
- [21] Biswas, M., Sayeedur, R., Farhat, T.S., Dipro, N., Aqeed, S.M.: Comparative analysis of attention-based, convolutional, and ssm-based models for multi-domain image classification. BRAC University (2025) <https://doi.org/http://hdl.handle.net/10361/26059>
- [22] Zhang, X., Mu, W.: Gmamba: State space model with convolution for grape leaf disease segmentation. *Computers and Electronics in Agriculture* **225** (2024) <https://doi.org/10.1016/j.compag.2024.109290>
- [23] Shi, D., Li, C., Shi, H., Liang, L.: A hierarchical feature-aware model for accurate tomato blight disease spot detection: U-net with vision mamba and convnext perspective. *Agronomy* **14** (2024) <https://doi.org/10.3390/agronomy14102227>
- [24] Zhu, T., Yan, F., Lv, X., Zhao, H., Wang, Z., Dong, K., Fu, Z., Jia, R., Lv,

- C.: A deep learning model for accurate maize disease detection based on state-space attention and feature fusion. *Plants* **13** (2024) <https://doi.org/10.3390/plants13223151>
- [25] Hughes, D., Salathé, M.: An open access repository of images on plant health to enable the development of mobile disease diagnostics. arXiv preprint arXiv:1511.08060 (2015)
 - [26] Chen, J., Chen, J., Zhang, D., Sun, Y., Nanekaran, Y.A.: Using deep transfer learning for image-based plant disease identification. *Computers and Electronics in Agriculture* **173**, 105393 (2020) <https://doi.org/10.1016/j.compag.2020.105393>
 - [27] Thakur, P.S., Sheorey, T., Ojha, A.: Vgg-icnn: A lightweight cnn model for crop disease identification. *Multimedia Tools and Applications* **82**(1), 497–520 (2022) <https://doi.org/10.1007/s11042-022-13144-z>
 - [28] Dai, D., Tang, X., Xia, S., Liu, Y., Wang, G., Chen, Z.: Multi-granularity association learning framework for on-the-fly fine-grained sketch-based image retrieval. *Knowledge-Based Systems* **253**, 109447 (2022) <https://doi.org/10.1016/j.knosys.2022.109447>
 - [29] Thapa, R., Snaveley, N., Belongie, S., Khan, A.: The plant pathology 2020 challenge dataset to classify foliar disease of apples. arXiv preprint arXiv:2004.11958 (2020) <https://doi.org/10.48550/arXiv.2004.11958>
 - [30] Thakur, P.S., Sheorey, T., Ojha, A.: A hybrid cnn–vit framework for robust plant disease classification. *Complex & Intelligent Systems* (2025) <https://doi.org/10.1007/s40747-024-01764-x>
 - [31] Niranjana, A., al.: Automated mango leaf disease classification using data-efficient transformer. *Heliyon* **10**(5), 12392 (2024)
 - [32] Lu, H., Dong, B., Zhu, B., Ma, S.: A survey on deep learning-based object detection for crop monitoring: Pest, yield, weed, and growth applications. *The Visual Computer* (2025) <https://doi.org/10.1007/s00371-025-04022-4>
 - [33] Zhang, Y., al.: Yolov8s-bsmamba: A lightweight tomato leaf disease detection model. *Agronomy* **15**(4), 870 (2025)
 - [34] Wang, Y., Chen, T., Liu, J., Zhang, Y.: Conmamba: Contrastive vision mamba for plant disease detection. arXiv preprint arXiv:2506.03213 (2024)
 - [35] Zhao, L., Sun, Q., Liu, R.: Vmamba for plant leaf disease identification: Design and experiment. *Journal of Artificial Intelligence in Agriculture* (2024). In Press
 - [36] Karthik, R., Hariharan, M., Anand, S., Mathikshara, P., Johnson, A., Menaka, R.:

Attention embedded residual cnn for disease detection in tomato leaves. Applied Soft Computing **86**, 105933 (2020)