JOURNAL OF COMPUTER SCIENCE AND INFORMATICS ENGINEERING
VOL. 05, NO. 3, JULY 2026: 436-447
Doi: https://doi.org/10.55537/cosie.v5i3.1834

e-ISSN 2827-8356
p-ISSN 2827-9999

Comparative Analysis of MambaOut Tiny and Vision Transformer for Corn Leaf Disease Classification

Adi Slamet Priyadi*[1], Albert Yakobus Chandra[2]
[1,2]Information Systems Study Program, Faculty of Information Technology, Mercu Buana University Yogyakarta, Yogyakarta, Indonesia

221210012@student.mercubuana-yogya.ac.id[1],albert.ch@mercubuana-yogya.ac.id[2]

Received: June 18, 2026 | Revised: July 20, 2026 | Accepted: July 31, 2026

Abstract

Maize production remains vulnerable to foliar diseases such as Gray Leaf Spot, Common Rust, and Northern Leaf Blight, which can impair growth and reduce crop yields. Manual disease identification is time-consuming, labor-intensive, and reliant on the observer's expertise, necessitating automated detection systems based on deep learning. While Convolutional Neural Networks (CNNs) and Vision Transformers (ViTs) are widely used for plant disease classification, a direct comparison between ViT-B/16 and MambaOut Tiny under identical training conditions remains rare. This study compares the two models using the PlantVillage dataset, which comprises 4,188 maize leaf images across four classes. Both models were fine-tuned using ImageNet-1K pre-trained weights, the AdamW optimizer, a learning rate of 1e-4, a batch size of 16, and 10 epochs. The results demonstrate that MambaOut Tiny achieved an accuracy of 97% and a macro-average F1-score of 0.96, outperforming ViT-B/16, which achieved 95% accuracy and an F1-score of 0.93. Additionally, MambaOut Tiny features a lower parameter count (30 million versus 86 million) and faster inference time (0.01262 seconds per image). These findings indicate that MambaOut Tiny is more efficient for precision agriculture systems with limited computational resources.

Keywords: Vision Transformer, MambaOut Tiny, Deep Learning, Corn Leaf Disease Classification, Computer Vision

1. INTRODUCTION

Among the crops that underpin global food security, maize (Zea mays L.) ranks second only to rice, serving as a core input for both food processing and livestock feed [1], [2]. That importance is constantly tested by foliar disease, which drags down plant vigor and final yield; pathogen damage to crops worldwide is estimated to cost close to two trillion US dollars annually [2].

A large share of corn leaf damage traces back to three fungal infections in particular Gray Leaf Spot, Common Rust, and Northern Leaf Blight [2], [3], [4]. These spread quickly once conditions turn favorable, making an outbreak genuinely hard to contain once it starts [2]. Diagnosis in the field still depends heavily on an agronomist's trained eye, and that reliance on individual judgment makes the process slow to scale and inconsistent across inspectors [5], [6]. In areas where that expertise is scarce, farmers risk applying the wrong treatment after a mistaken diagnosis, which compounds both the ecological and financial damage [2].

Deep learning has substantially raised what automated image classification can achieve over the past several years. Convolutional Neural Networks (CNNs) dominated computer-vision work, plant pathology included, for a long stretch of that period [3], [4], but their dependence on local receptive fields makes them poorly suited to relating features positioned far apart in an image

e-ISSN 2827-8356
p-ISSN 2827-9999

[7], [8]. The Vision Transformer (ViT), introduced by Dosovitskiy et al. [9], gets around that limitation by applying self-attention across image patches so the model can capture dependencies at a global scale, and it has since delivered strong results across a range of classification tasks [10], [11], [12].

Subsequent research has applied ViT-style models directly to plant imagery with encouraging outcomes: accuracy in the high 90s on PlantVillage-type benchmarks shows up repeatedly, alongside variants built for multispectral input, attention-driven multi-crop pipelines, and hybrid encoders that borrow convolutional blocks to keep parameter counts manageable [10], [13], [14], [15], [16]. Taken together, this body of work positions ViT and its variants among the more capable tools currently available for plant disease recognition.

A separate line of work blends convolution and attention rather than picking one over the other. Dual-branch CNN-ViT designs, transformer-CNN hybrids built for messy field backgrounds, and architectures pairing transformers with Fourier-domain operators have all reported accuracy ranging from the high 80s to high 90s while keeping parameter budgets modest [6], [7], [17]. The shared insight across these efforts is that combining local and global feature extraction can preserve accuracy while trimming computational overhead.

Transformers, meanwhile, have begun losing ground to a newer family of sequence models. State Space Models (SSMs) and Mamba [18] specifically replace quadratic self-attention with a selective-scan mechanism that scales linearly and, by some estimates, runs several times faster than comparable transformer setups. Bidirectional adaptations built for vision tasks have matched or surpassed transformer baselines on standard benchmarks while substantially cutting GPU memory demand [19], and 2D-scan variants have pushed ImageNet accuracy past 80% while retaining that linear-complexity advantage [20]. Interestingly, not every SSM variant keeps the state-space core: MambaOut [21] removes it in favor of a gated convolutional design, yet still outperforms genuine Mamba-based vision models on ImageNet a counterintuitive finding worth noting.

SSM-style architectures have also started appearing outside natural-image benchmarks. In medical imaging they have matched ViT's accuracy at lower compute cost, and in remote sensing they have outperformed both CNN and transformer baselines [14], [22]. Closer to the present topic, Zhang et al. [23] found that VMamba outperformed Vision Transformer on plant disease identification, reaching high-90s accuracy on PlantVillage and low-90s accuracy on a smaller, more difficult dataset.

Corn-specific research has its own track record: transfer learning across an ensemble of CNN backbones, a lightweight attention-augmented CNN purpose-built for corn disease, and a detection pipeline pairing Faster-RCNN with a ResNet50 backbone and CBAM attention have all reached roughly 98% accuracy on corn leaf datasets [2], [3], [4]. Collectively, these results have made corn leaf imagery something of a standard proving ground for new disease-classification models.

A broader comparative study on grape leaves, covering 14 CNN and 17 ViT variants, drives home a related point: no single architecture wins across the board, and only systematic side-by-side testing can reveal which model genuinely performs best for a given crop and dataset [24]. The ViT-B/16-versus-MambaOut Tiny comparison in this paper follows that same logic, applied here to corn.

What is largely missing from the existing literature is a controlled comparison between ViT and MambaOut specifically for corn leaf disease. Most ViT-centered studies chase peak accuracy without weighing the computational cost involved, and while VMamba's advantage over ViT on plant imagery has already been documented [23], MambaOut a distinct and more recent architecture has yet to be benchmarked against ViT under matched training conditions for this particular crop.

This study addresses that gap: ViT-B/16 [19] and MambaOut Tiny [21] are trained and

JOURNAL OF COMPUTER SCIENCE AND INFORMATICS ENGINEERING
VOL. 05, NO. 3, JULY 2026: 436-447
Doi: https://doi.org/10.55537/cosie.v5i3.1834

e-ISSN 2827-8356
p-ISSN 2827-9999

evaluated on corn leaf images under identical conditions the same ImageNet-1K pretrained starting weights [25], optimizer, learning rate, batch size, and epoch budget and judged on accuracy, precision, recall, F1-score, and per-image inference time to determine which model better suits automated corn disease detection.

Beyond the direct comparison, this work aims to give other researchers a practical reference point when selecting an architecture for plant-image classification, and to contribute, in a modest way, to Indonesia's broader push toward smarter agriculture specifically, tools that help farmers spot and manage corn leaf disease quickly through AI running on ordinary mobile hardware or edge devices [19].

2. RESEARCH METHODOLOGY

3.

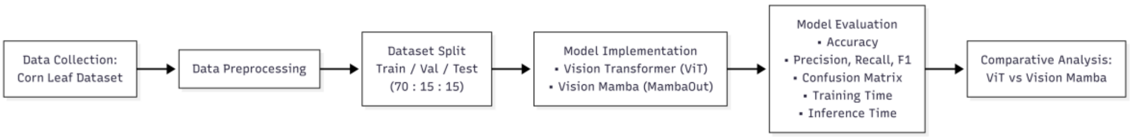

Figure 1. Research Flowchart

2.1 Dataset Collection

The research used the dataset used in this study is the Plant Village Corn Plant Database, containing 4,188 corn leaf images divided into four classes: Cercospora Leaf Spot, Common Rust, Northern Leaf Blight, and Healthy. This dataset has previously been used as a standard benchmark in previous corn disease classification studies built on CNN and classical machine learning [2], [4].

2.2 Data Preprocessing

Every image was resized to 224×224 pixels so its dimensions would match the input requirements of both Vision Transformer (ViT-B/16) [9] and MambaOut Tiny [21]. Normalization followed the conventional ImageNet statistics, with mean values of [0.485, 0.456, 0.406] and standard deviation of [0.229, 0.224, 0.225] [25]. Training images were further augmented with Random Horizontal Flip, Random Rotation (±15°), and Color Jitter (brightness = 0.2, contrast = 0.2) to help the models generalize better. Validation and test images, in contrast, went through only resizing and normalization with no augmentation applied, keeping the evaluation results unbiased.

2.3 Dataset Splitting

A stratified 70:15:15 split divided the dataset into 2,931 training images, 628 validation images, and 629 testing images. Class proportions were kept consistent across all three subsets so the representation stayed balanced.

2.4 Model Training

Both ViT-B/16 and MambaOut Tiny were retuned starting from ImageNet-1K pre-trained weights [25]. Training relied on the AdamW optimizer with a learning rate of 1e-4, a batch size of 16, CrossEntropyLoss, and a total of 10 epochs. Experiments were conducted using the Google Colaboratory equipped with NVIDIA Tesla T4 GPUs (16 GB VRAM), Intel Xeon CPUs, and 12 GB RAM. All models were implemented using the PyTorch framework.

2.5 Model Evaluation

Model performance is measured using accuracy, precision, recall, and F1-score [26], supplemented by confusion matrix analysis. Average per-image inference time was also recorded so computational efficiency could be compared between the two architectures.

2.6 Analysis and Comparison

Both models were evaluated side-by-side based on classification accuracy and computational efficiency for corn leaf disease detection, including accuracy, precision, recall, F1-

JOURNAL OF COMPUTER SCIENCE AND INFORMATICS ENGINEERING
VOL. 05, NO. 3, JULY 2026: 436-447
Doi: https://doi.org/10.55537/cosie.v5i3.1834
e-ISSN 2827-8356
p-ISSN 2827-9999

score, inference time, and number of parameters. This comparison highlights the strengths and limitations of each model's architecture and its suitability for real-world agricultural applications. [19], [23].

2.7 Vision Transformer (ViT-B/16) Architecture

Vision Transformer (ViT-B/16) [9] approaches image classification by splitting an input image into 16×16-pixel patches and passing them through layers of multi-head self-attention. Its architecture consists of 12 encoder layers, a hidden size of 768, and 12 attention heads [9]. For this study, the final classification layer was modified to output four corn leaf disease categories.

2.8 MambaOut Tiny Architecture

MambaOut Tiny [21] is built around a State Space Model (SSM) design focused on computational efficiency. Rather than ViT's quadratic self-attention, it depends on linear-complexity operations, which reduces memory use and speeds up inference [18], [19]. Its architecture combines depthwise convolution, gating mechanisms, and state-space representation, giving it fewer parameters and faster inference than ViT [19], [21].

2.9 Training Configuration

To ensure a fair comparison, both models were trained with identical settings using the AdamW optimizer, a learning rate of le-4, a batch size of 16, CrossEntropyLoss, and 10 training epochs. All experiments were conducted on the Google Colaboratory using GPU acceleration with the PyTorch framework. Both networks were initialized from pre-trained ImageNet-1K weights [25]. to accelerate convergence and improve generalization, a strategy that has been widely adopted in previous plant disease classification studies [2], [3], [23], [25]. For reproducibility, the dataset was using divided a stratified 70:15:15 split with a fixed random seed (random_state = 42), ensuring consistent class distributions across the training, validation, and testing subsets.

4. RESULTS AND DISCUSSION

3.1 Dataset

The Plant Village Corn Database used here contains 4,188 leaf images in four categories: Cercospora Leaf Spot, Common Rust, Northern Leaf Blight, and Healthy, and has previously supported machine learning and CNN-based corn disease research [2], [4]. All images were resized to 224×224 pixels to meet the input requirements of ViT and MambaOut Tiny. Table 1 shows how the images were distributed across classes before splitting.

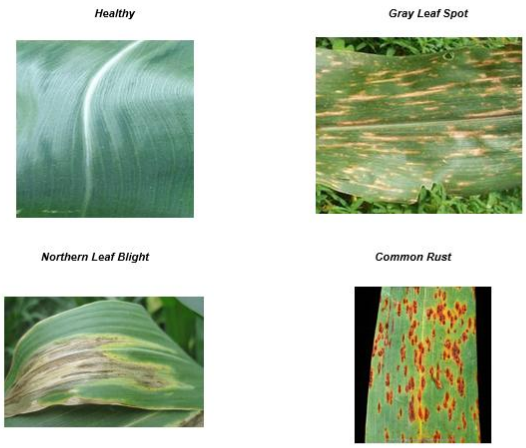

Figure 2. Example images from the dataset

JOURNAL OF COMPUTER SCIENCE AND INFORMATICS ENGINEERING
VOL. 05, NO. 3, JULY 2026: 436-447
Doi: https://doi.org/10.55537/cosie.v5i3.1834
e-ISSN 2827-8356
p-ISSN 2827-9999

Table 1. Dataset Distribution

Class	**Total**
Cercospora Leaf Spot	574
Common Rust	1.306
Northern Leaf Blight	1.146
Healthy	1.162
Total	4188

3.2 Data Preprocessing Results

Every image was resized to 224×224 pixels and normalized using standard ImageNet parameters [24]. During training, Random Horizontal Flip, Random Rotation (±15°), and Color Jitter (brightness = 0.2, contrast = 0.2) were applied to broaden generalization [25]. Validation and test datasets only underwent resizing and normalization, with additional data deliberately omitted to keep the evaluation unbiased. [26].

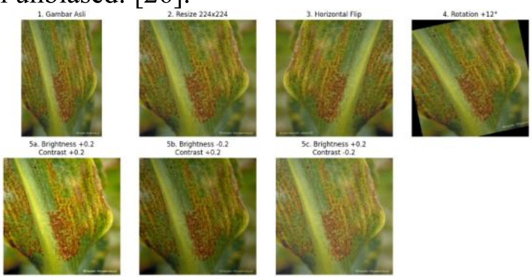

Figure 3. Example of data preprocessing for corn leaves affected by Common Rust

3.3 Dataset Splitting Results

Applying the stratified 70:15:15 split produced 2,931 training images, 628 validation images, and 629 test images [2], [26]. Class balance held steady throughout: Common Rust contributed the most images (1,306), while Cercospora Leaf Spot contributed the fewest (574). This split was chosen specifically to keep the evaluation representative of how well each model generalizes [26].

Table 2. Dataset Splitting

Class	**Total**	**Training**	**Validation**	**Testing**
Cercospora Leaf Spot	574	402	86	86
Common Rust	1.306	914	196	196
Northern Leaf Blight	1.146	802	172	172
Healthy	1.162	813	174	175
Total	4188	2931	628	629

3.4 Model Training Results

ViT-B/16 and MambaOut Tiny were both trained under matched conditions: the AdamW optimizer, a 1e-4 learning rate, batch size of 16, and 10 epochs. Accuracy and loss for the training and validation sets were tracked across epochs, and both models showed a consistent pattern of increasing accuracy as loss decreased.

MambaOut Tiny converged more smoothly than ViT and finished with a higher validation accuracy in the final epoch. This stability likely stems from architectural differences: the gated convolution mechanism in MambaOut Tiny [21] tends to propagate gradients more consistently than ViT's self-attention mechanism, which is more sensitive to initialization and optimization choices given its quadratic complexity [9], [18].

JOURNAL OF COMPUTER SCIENCE AND INFORMATICS ENGINEERING
VOL. 05, NO. 3, JULY 2026: 436-447
Doi: https://doi.org/10.55537/cosie.v5i3.1834 e-ISSN 2827-8356
p-ISSN 2827-9999

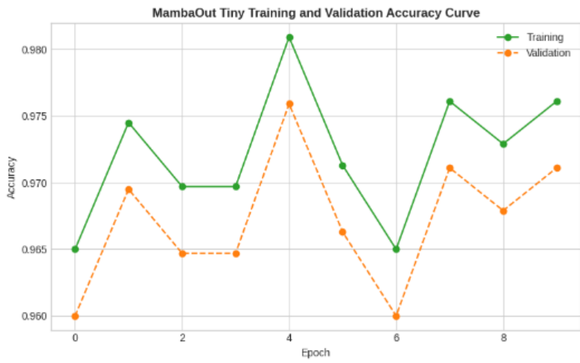

Figure 4. MambaOut Tiny Training and Validation Accuracy Curve

3.5 Model Evaluation Results

Both Vision Transformer (ViT-B/16) and MambaOut Tiny were evaluated after training under the same settings the AdamW optimizer, a 1e-4 learning rate, and a batch size of 16 across 10 epochs.

3.6 Vision Transformer (ViT-B/16) Evaluation Results

ViT-B/16 achieved 95% accuracy on the separated test set. The Healthy class was classified with perfect precision, posting an F1-score of 1.00 an expected result, since healthy leaves look visually distinct from diseased ones. Cercospora Leaf Spot proved the weakest class, largely because its visual symptoms overlap with Northern Leaf Blight [2], [3]. Table 3 lists the detailed metrics.

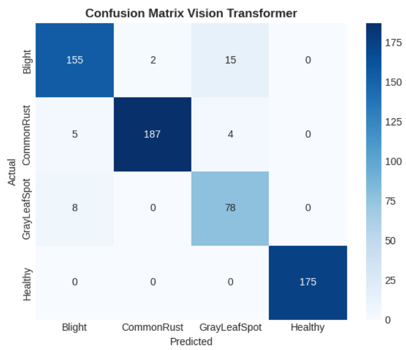

Figure 5. Confusion Matrix Vision Transformer

Table 3. Vision Transformer (ViT-B/16) Results

Class	Precision	Recall	F1-Score	Support
Cercospora Leaf Spot	0.80	0.91	0.85	86
Common Rust	0.99	0.95	0.97	196
Northern Leaf Blight	0.92	0.90	0.91	172
Healthy	1.00	1.00	1.00	175
Accuracy	-	-	0.95	629
Macro Avg	0.93	0.94	0.93	629
Weighted Avg	0.95	0.95	0.95	629

3.7 MambaOut Tiny Evaluation Results

 e-ISSN 2827-8356
p-ISSN 2827-9999

MambaOut Tiny reached 97% test accuracy, edging out ViT-B/16 on nearly every metric. Its clearest gain showed up on Cercospora Leaf Spot, where the F1-score rose from 0.85 to 0.92, an indication that it handles visually similar disease patterns more effectively [20], [21]. It also produced fewer overall errors 18 misclassifications compared with 31 for ViT-B/16. Table 4 presents the full results.

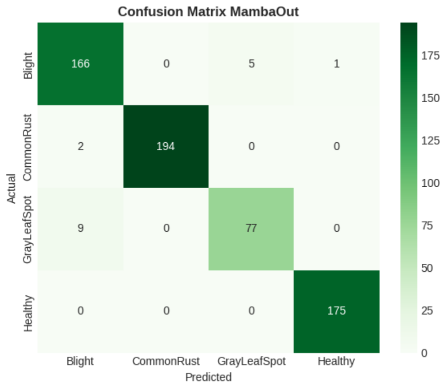

Figure 6. Confusion Matrix MambaOut Tiny

Table 4. MambaOut Tiny Results

Class	Precision	Recall	F1-Score	Support
Cercospora Leaf Spot	0.94	0.90	0.92	86
Common Rust	1.00	0.99	0.99	196
Northern Leaf Blight	0.94	0.97	0.95	172
Healthy	1.00	1.00	1.00	175
Accuracy	-	-	0.97	629
Macro Avg	0.97	0.96	0.96	629
Weighted Avg	0.97	0.97	0.97	629

3.8 Model Performance Comparison

Overall, MambaOut Tiny outperformed Vision Transformer both in accuracy (97% versus 95%) and computational cost, delivering faster inference (0.01262 s/image compared with 0.01946 s/image for ViT-B/16) while using roughly a third as many parameters (about 30M against 86M) an advantage for resource-limited devices [19], [21].

This outcome lines up with what other SSM studies have reported: VMamba's edge over ViT on plant-disease imagery, a wider effective receptive field than typical CNNs offer, throughput gains of roughly five times over comparable transformer models, and, for bidirectional vision variants, inference speeds around 2.8 times faster alongside GPU memory savings above 85% [18], [19], [20], [23].

Several architectural traits plausibly explain MambaOut Tiny's edge here. Its gated convolution component [21] handles local feature extraction efficiently, while the underlying state-space representation still captures dependencies spanning a wider portion of the leaf a combination that matters for disease classification, since both fine-grained lesion texture and the broader shape of leaf damage carry diagnostic value [23]. Beyond that, the linear computational scaling built into SSM-style designs [18] is inherently cheaper to train and run than ViT's quadratic self-attention [9]. The selective-scan mechanism further lets the model weight information adaptively, allowing it to lean on disease-relevant regions of an image while discounting background clutter.

Compared with earlier corn-disease work, results in the high-97 to 98.6% accuracy range have been achieved using ensembles of CNN backbones, a lightweight attention-based CNN, and a Faster-RCNN detector paired with CBAM attention [2], [3], [4]. The 97% accuracy MambaOut Tiny reached here falls within that same range, and it gets there with a noticeably leaner

 e-ISSN 2827-8356
p-ISSN 2827-9999

computational footprint. The gap that remains is easy to explain: those higher-scoring studies tuned across as many as ten separate CNN variants with extensive hyperparameter search, whereas this experiment used a single fixed configuration and only 10 training epochs. Given that, MambaOut Tiny's parameter count around 30M, versus the tens or hundreds of millions typical of CNN pipelines still makes it the more practical choice for edge deployment [19], [23].

The broader case for SSM-based architectures in efficient plant-disease systems continues to strengthen [19], [23]: hybrid designs have already proven useful for identification under messy field conditions [6], and cross-architecture benchmarking work has confirmed that no single model type wins in every setting [24] a conclusion this comparison reinforces rather than contradicts.

One caveat worth noting is that the dataset used here spans only four disease categories, short of the full range of corn leaf conditions encountered in the field [6]. Other research points to concrete ways of closing that gap: feeding the model multiple leaves at once for richer disease context [15], capturing spatial patterns that single-leaf input tends to miss [17], and extending SSM techniques into new agricultural applications the way they have already spread into other domains [14], [22]. A natural next step, then, is to broaden the dataset both in disease categories and environmental variation and to test whether multi-leaf input further improves classification accuracy [15].

Although MambaOut Tiny consistently outperformed ViT-B/16 across all evaluation metrics, confidence intervals and statistical significance tests were not performed because the comparison was based on a single stratified train validation test split. Therefore, the reported performance reflects the observed results under identical experimental settings. Future studies should include repeated experimental runs and statistical analyses to further validate the robustness of the observed performance differences.

3.9 Limitations

This study has several limitations. First, the experiments were conducted only using the PlantVillage dataset [1],[2] , collected under controlled conditions and may not fully represent real-world agricultural environments [1],[6]. Second, the comparison was limited to two deep learning architectures, namely ViT-B/16 and MambaOut Tiny. Third, the experiments were performed using a single stratified train validation test split without repeated training runs, confidence interval estimation, or statistical significance testing. Future work should evaluate additional transformer and state space based architectures[7],[8],[21] on more diverse field datasets and perform repeated experiments with statistical validation [2]to improve the reliability and generalizability of the findings.

JOURNAL OF COMPUTER SCIENCE AND INFORMATICS ENGINEERING
VOL. 05, NO. 3, JULY 2026: 436-447
Doi: https://doi.org/10.55537/cosie.v5i3.1834
e-ISSN 2827-8356
p-ISSN 2827-9999

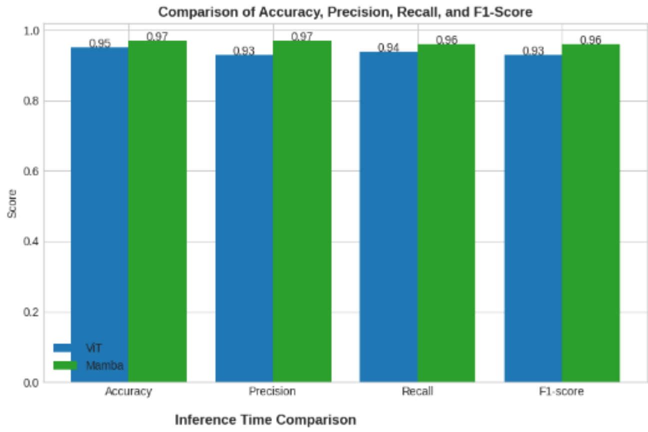

Figure 7. Comparison of Accuracy, Precision, Recall, and F1-Score

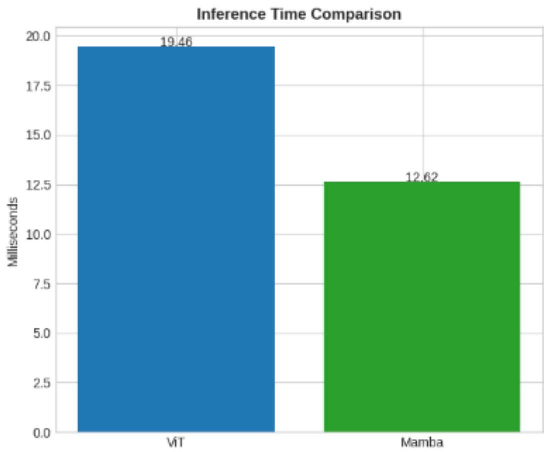

Figure 8. Inference Time Comparison

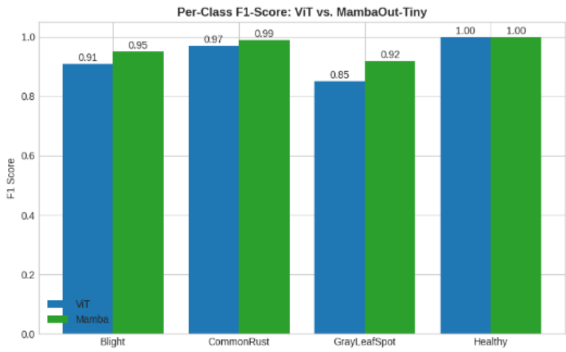

Figure 9. Per-Class F1-Score: ViT vs. MambaOut-Tiny

5. CONCLUSION

Both ViT-B/16 and MambaOut Tiny performed well at classifying corn leaf diseases across the four categories examined here, and comparing them under identical experimental conditions offers useful insight into each architecture's relative strengths and weaknesses for plant disease classification.

MambaOut Tiny came out ahead of Vision Transformer overall, reaching 97% accuracy versus 95%, with precision, recall, and F1-score also consistently higher and more stable across

e-ISSN 2827-8356
p-ISSN 2827-9999

test classes. These findings echo what Yu and Wang [21] and Liu et al. [20] reported regarding the advantages of SSM- and gated-CNN-based designs for image classification. MambaOut Tiny's stronger performance can largely be traced to its gated convolution mechanism and state-space representation, which together support efficient feature extraction and effective modeling of long-range dependencies [18], [20], [21].

MambaOut Tiny was also faster at inference (0.01262 s/image versus 0.01946 s/image for ViT-B/16) and needed fewer parameters (roughly 30M compared with 86M), consistent with the general pattern of state-space architectures scaling more efficiently than self-attention mechanisms [18]. The lower inference time and smaller number of parameters indicate that MambaOut Tiny is a promising architecture for applications with limited computational resources [19].

Overall, MambaOut Tiny demonstrated better classification performance and computational efficiency than ViT-B/16 under the experimental settings used in this study. These findings indicate that MambaOut Tiny is a promising alternative for corn leaf disease classification, particularly in applications with limited computational resources [1]. Pairing solid accuracy with fast inference and a lean parameter count gives it a genuine practical edge in settings where compute is scarce but quick diagnosis is still essential.

Future work should test these models on datasets with broader environmental variation [6], expand the volume of training data to improve generalization, incorporate multiple leaves per sample for richer disease information [15], and build mobile-based applications for real-time field diagnosis. Extending MambaOut's evaluation to larger datasets, additional disease categories, and other crop species would further validate its usefulness for agricultural applications.

6. ACKNOWLEDGMENTS

The authors would like to express their sincere gratitude to the Information Systems Study Program, Faculty of Information Technology, Mercu Buana University, Yogyakarta, for their guidance and academic support throughout this research. They also thank the PlantVillage project for providing the publicly available maize leaf disease dataset used in this study.

REFERENCES

[1] F. Khan, N. Zafar, M. N. Tahir, M. Aqib, H. Waheed, and Z. Haroon, "A mobile-based system for maize plant leaf disease detection and classification using deep learning," *Front. Plant Sci.*, vol. 14, no. May, pp. 1–18, May 2023, doi: 10.3389/fpls.2023.1079366.
[2] M. Fraiwan, E. Faouri, and N. Khasawneh, "Classification of Corn Diseases from Leaf Images Using Deep Transfer Learning," *Plants*, vol. 11, no. 20, pp. 1–14, 2022, doi: 10.3390/plants11202668.
[3] Y. Chen *et al.*, "DFCANet: A Novel Lightweight Convolutional Neural Network Model for Corn Disease Identification," *Agriculture (Switzerland)*, vol. 12, no. 12, 2022, doi: 10.3390/agriculture12122047.
[4] M. Masood *et al.*, "MaizeNet: A Deep Learning Approach for Effective Recognition of Maize Plant Leaf Diseases," *IEEE Access*, vol. 11, no. May, pp. 52862–52876, 2023, doi: 10.1109/ACCESS.2023.3280260.
[5] A. Tabbakh and S. S. Barpanda, "A Deep Features Extraction Model Based on the Transfer Learning Model and Vision Transformer 'TLMViT' for Plant Disease Classification," *IEEE Access*, vol. 11, no. May, pp. 45377–45392, 2023, doi: 10.1109/ACCESS.2023.3273317.
[6] S. Jia, G. Wang, H. Li, Y. Liu, L. Shi, and S. Yang, "ConvTransNet-S: A CNN-Transformer Hybrid Disease Recognition Model for Complex Field Environments," *Plants*, vol. 14, no. 15, pp. 1–17, 2025, doi: 10.3390/plants14152252.

e-ISSN 2827-8356
p-ISSN 2827-9999

[7] Q. Meng, J. Guo, H. Zhang, Y. Zhou, and X. Zhang, "A dual-branch model combining convolution and vision transformer for crop disease classification," *PLoS One*, vol. 20, no. 4 April, pp. 1–23, 2025, doi: 10.1371/journal.pone.0321753.

[8] X. Li and S. Li, "Transformer Help CNN See Better: A Lightweight Hybrid Apple Disease Identification Model Based on Transformers," *Agriculture (Switzerland)*, vol. 12, no. 6, 2022, doi: 10.3390/agriculture12060884.

[9] A. Dosovitskiy *et al.*, "An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale," *ICLR 2021 - 9th International Conference on Learning Representations*, Jun. 2021, [Online]. Available: http://arxiv.org/abs/2010.11929

[10] Y. Borhani, J. Khoramdel, and E. Najafi, "A deep learning based approach for automated plant disease classification using vision transformer," *Sci. Rep.*, vol. 12, no. 1, pp. 1–10, 2022, doi: 10.1038/s41598-022-15163-0.

[11] S. Aboelenin, F. A. Elbasheer, M. M. Eltoukhy, W. M. El-Hady, and K. M. Hosny, "A hybrid Framework for plant leaf disease detection and classification using convolutional neural networks and vision transformer," *Complex and Intelligent Systems*, vol. 11, no. 2, 2025, doi: 10.1007/s40747-024-01764-x.

[12] S. Murugavalli and R. Gopi, "Plant leaf disease detection using vision transformers for precision agriculture," *Sci. Rep.*, vol. 15, no. 1, pp. 1–17, 2025, doi: 10.1038/s41598-025-05102-0.

[13] M. De Silva and D. Brown, "Multispectral Plant Disease Detection with Vision Transformer-Convolutional Neural Network Hybrid Approaches," *Sensors (Basel)*, vol. 23, no. 20, 2023, doi: 10.3390/s23208531.

[14] K. Chen, B. Chen, C. Liu, W. Li, Z. Zou, and Z. Shi, "RSMamba: Remote Sensing Image Classification with State Space Model," *IEEE Geoscience and Remote Sensing Letters*, vol. 21, pp. 1–5, 2024, doi: 10.1109/LGRS.2024.3407111.

[15] E. T. Baek, "Attention Score-Based Multi-Vision Transformer Technique for Plant Disease Classification," *Sensors*, vol. 25, no. 1, 2025, doi: 10.3390/s25010270.

[16] P. S. Thakur, P. Khanna, T. Sheorey, and A. Ojha, "Explainable vision transformer enabled convolutional neural network for plant disease identification: PlantXViT," no. Dl, Jul. 2022, [Online]. Available: http://arxiv.org/abs/2207.07919

[17] B. Hu, W. Jiang, J. Zeng, C. Cheng, and L. He, "FOTCA: hybrid transformer-CNN architecture using AFNO for accurate plant leaf disease image recognition," *Front. Plant Sci.*, vol. 14, no. September, pp. 1–12, 2023, doi: 10.3389/fpls.2023.1231903.

[18] A. Gu and T. Dao, "Mamba: Linear-Time Sequence Modeling with Selective State Spaces," no. Kalman 1960, pp. 1–36, May 2024, [Online]. Available: http://arxiv.org/abs/2312.00752

[19] L. Zhu, B. Liao, Q. Zhang, X. Wang, W. Liu, and X. Wang, "Vision Mamba: Efficient Visual Representation Learning with Bidirectional State Space Model," *Proc. Mach. Learn. Res.*, vol. 235, pp. 62429–62442, Nov. 2024, [Online]. Available: http://arxiv.org/abs/2401.09417

[20] Y. Liu *et al.*, "VMamba: Visual State Space Model," in *Advances in Neural Information Processing Systems 37*, San Diego, California, USA: Neural Information Processing Systems Foundation, Inc. (NeurIPS), 2024, pp. 103031–103063. doi: 10.52202/079017-3273.

[21] W. Yu and X. Wang, "MambaOut: Do We Really Need Mamba for Vision?," in *2025 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, IEEE, Jun. 2025, pp. 4484–4496. doi: 10.1109/CVPR52734.2025.00423.

[22] Y. Yue and Z. Li, "MedMamba: Vision Mamba for Medical Image Classification," Sep. 2024, [Online]. Available: http://arxiv.org/abs/2403.03849

e-ISSN 2827-8356
p-ISSN 2827-9999

[23] H. Zhang, S. Li, J. Xie, Z. Chen, J. Chen, and J. Guo, "VMamba for plant leaf disease identification: design and experiment," *Front. Plant Sci.*, vol. 16, no. April, pp. 1–17, 2025, doi: 10.3389/fpls.2025.1515021.

[24] I. Kunduracioglu and I. Pacal, "Advancements in deep learning for accurate classification of grape leaves and diagnosis of grape diseases," *Journal of Plant Diseases and Protection*, vol. 131, no. 3, pp. 1061–1080, 2024, doi: 10.1007/s41348-024-00896-z.

[25] F. Zhuang *et al.*, "A Comprehensive Survey on Transfer Learning," *Proceedings of the IEEE*, vol. 109, no. 1, pp. 43–76, Jan. 2021, doi: 10.1109/JPROC.2020.3004555.

[26] A. Tharwat, "Classification assessment methods," *Applied Computing and Informatics*, vol. 17, no. 1, pp. 168–192, 2018, doi: 10.1016/j.aci.2018.08.003.