Supplementary Material

Training and Hyperparameters Details

Table 9 provides comprehensive experimental documentation for reproducibility purposes with complete specification of hyperparameters, implementation framework details, and data availability information. All datasets used in our benchmarking analysis are publicly available: PlantVillage (Mohanty et al., 2016), PlantDoc and Cropped PlantDoc (Singh et al., 2020), FieldPlant (Singh and Kaur, 2019), FieldPlant (Moupojou et al., 2023)and Cassava Leaf Diseases datasets (Kaggle, 2024).

Table 9 Comprehensive Reproducibility and Experimental Configuration Documentation

Configuration Category	Specification
Training Hyperparameters	**Optimizer:** AdamW (lr=2e-5, weight_decay=0.01, default betas) **Learning Rate:** Fixed 2e-5 (no scheduling applied) **Weight Decay:** 0.01 for L2 regularization **Batch Size:** 16 (memory-optimized for single GPU) **Epochs:** Maximum 20 with early stopping (patience=3 on validation F1-score) **Loss Function:** CrossEntropyLoss **Random Seeds:** Base seed 42, incremented per run (42, 43, 44) **Early Stopping:** Best model saved based on validation F1-score
Data Processing	**Image Resolution:** 224×224 pixels **Training Augmentation:** RandomResizedCrop(224), RandomHorizontalFlip(0.5), RandomRotation(10°) **Validation/Test:** Resize to 224×224 with no augmentation **Data Splits:** Stratified 80% train, 5% validation, 15% test **Split Strategy:** train_test_split with fixed random_state=42
Implementation Framework	**Deep Learning Framework:** PyTorch **Model Library:** HuggingFace **Hardware - CPU:** Intel Core i7-12700K (12 cores, 20 threads, up to 5.0 GHz) **Hardware - RAM:** 128GB DDR4 system memory **Hardware - GPU:** NVIDIA GeForce RTX 4090 (24GB VRAM) **Software:** CUDA 12.2, Python 3.12
Model Architectures	**ResNet50:** microsoft/resnet-50 **ConvNeXt:** facebook/convnext-base-224 **ViT:** google/vit-base-patch16-22 **SWIN:** microsoft/swin-base-patch4-window7-224

Computational Complexity Analysis

To provide complete transparency regarding model efficiency and deployment feasibility, Table 10 presents comprehensive computational complexity metrics for all evaluated architectures, including parameter counts, memory requirements, computational loads, and inference performance measured on our experimental hardware.

Key Computational Insights:

– **Parameter Efficiency:** ResNet50 demonstrates superior parameter efficiency with 563.2 img/s using only 23.5M parameters, achieving an efficiency ratio of 23.96 compared to 1.47-4.59 for transformer-based architectures.

Table 10 Computational Complexity and Performance Analysis of Deep Learning Architectures

Model	Parameters (Millions)	Model Size (MB)	GFLOPs	Inference Time (ms)	Throughput (img/s)	Efficiency Ratio*
ResNet50	23.5	89.8	4.1	1.78±0.05	563.2	23.96
ViT	85.8	327.3	0.2	2.54±0.08	394.0	4.59
ConvNeXt	87.6	334.1	91.1	2.78±0.12	360.2	4.11
SWIN	86.8	330.9	0.2	7.81±0.15	128.0	1.47

Measurement Conditions: Intel Core i7-12700K, NVIDIA RTX 4090, CUDA 12.2, batch size=1, input resolution=224×224.
Statistical Analysis: Inference times reported as mean±std over 100 runs with 10 warmup iterations.
***Efficiency Ratio:** Throughput per million parameters (higher values indicate better parameter efficiency).

- **Memory Requirements:** Transformer models (ViT, ConvNeXt, SWIN) require 3.7× more memory (327-334 MB) compared to ResNet50 (89.8 MB), impacting deployment feasibility on resource-constrained devices.
- **Computational Load Limitation:** Despite having fewer parameters, ResNet50 requires 4.1 GFLOPs compared to 0.2 GFLOPs for ViT and SWIN, highlighting architectural differences in computational patterns between CNNs and attention mechanisms.
- **Inference Performance:** SWIN exhibits significantly slower inference (7.81 ms) despite competitive accuracy, indicating potential deployment challenges for real-time agricultural applications requiring rapid disease detection.

Statistical Analysis

To address reproducibility concerns and provide rigorous statistical validation, Table 11 presents comprehensive performance metrics with statistical uncertainty quantification across all model-dataset combinations. Each result represents the mean performance across three independent experimental runs using different random seeds (42, 43, 44), with standard deviations and 95% confidence intervals calculated using the t-distribution appropriate for small sample sizes.

Key Statistical Observations:

- **Laboratory vs. Field Variance:** PlantVillage (laboratory conditions) exhibits minimal performance variance (std < 0.76%), while real-world datasets show substantially higher variance, with PlantDoc demonstrating standard deviations up to 3.00% for ResNet50, indicating reduced model stability under challenging conditions.
- **Model-Specific Robustness:** SWIN consistently demonstrates lower standard deviations across most datasets (0.12-1.50%), suggesting superior training stability, while ResNet50 shows the highest variance on challenging datasets (PlantDoc: 3.00%, cPD: 2.08%), indicating sensitivity to environmental complexity.
- **Statistical Significance:** Non-overlapping confidence intervals between top-performing models (ViT, SWIN, ConvNeXt) and ResNet50 on challenging datasets confirm statistically significant performance differences, validating architectural superiority claims.

Table 11 Statistical Analysis of Model Performance with Standard Deviations and 95% Confidence Intervals (3 runs)

Dataset	Model	Accuracy	F1-Score	Accuracy CI	F1 CI
PV	ViT	99.44±0.05	99.43±0.05	[99.39, 99.49]	[99.38, 99.48]
	ResNet50	98.19±0.70	98.12±0.76	[97.49, 98.89]	[97.36, 98.88]
	ConvNeXt	99.47±0.11	99.47±0.11	[99.36, 99.58]	[99.36, 99.58]
	SWIN	99.32±0.12	99.32±0.12	[99.20, 99.44]	[99.20, 99.44]
PD	ViT	76.42±1.96	74.96±2.28	[74.46, 78.38]	[72.68, 77.24]
	ResNet50	41.80±3.00	34.37±2.93	[38.80, 44.80]	[31.44, 37.30]
	ConvNeXt	71.76±1.44	70.46±1.99	[70.32, 73.20]	[68.47, 72.45]
	SWIN	77.63±1.05	75.96±1.50	[76.58, 78.68]	[74.46, 77.46]
cPD	ViT	86.18±1.95	85.79±1.95	[84.23, 88.13]	[83.84, 87.74]
	ResNet50	52.74±1.57	48.06±2.08	[51.17, 54.31]	[45.98, 50.14]
	ConvNeXt	78.97±1.39	78.59±1.48	[77.58, 80.36]	[77.11, 80.07]
	SWIN	88.07±0.63	87.74±0.67	[87.44, 88.70]	[87.07, 88.41]
FP	ViT	96.00±0.18	96.01±0.19	[95.82, 96.18]	[95.82, 96.20]
	ResNet50	85.98±1.95	83.86±2.42	[84.03, 87.93]	[81.44, 86.28]
	ConvNeXt	95.59±0.50	95.68±0.42	[95.09, 96.09]	[95.26, 96.10]
	SWIN	95.53±0.24	95.48±0.46	[95.29, 95.77]	[95.02, 95.94]
CLD	ViT	85.91±0.15	85.84±0.24	[85.76, 86.06]	[85.60, 86.08]
	ResNet50	83.98±0.52	84.03±0.51	[83.46, 84.50]	[83.52, 84.54]
	ConvNeXt	87.60±0.31	87.68±0.26	[87.29, 87.91]	[87.42, 87.94]
	SWIN	88.01±0.64	88.00±0.67	[87.37, 88.65]	[87.33, 88.67]

Note: Results from 3 independent runs with different random seeds. Values as Mean±Std Dev. **Datasets:** PV=PlantVillage, PD=PlantDoc, cPD=Cropped PlantDoc, FP=FieldPlant, CLD=Cassava Leaf Disease Dataset.