Bridging the Domain Gap: Transfer Learning and Aggressive Fine-Tuning for Robust Plant Disease Detection in Low-Resource African Agriculture

Sunday Aspita Abraham GODWINSUNDAY11223@GMAIL.COM
ML Collective, Remote / Global

Olanloye Dauda Odunayo ODUN.OLANLOYE@BOWEN.EDU.NG
Department of Computer Science,
Bowen University, Iwo, Osun State, Nigeria

Abidemi Emmanuel Adeniyi AEADENIYI@OAUIFE.EDU.NG
Department of Information Systems,
Faculty of Computing Science and Engineering,
Obafemi Awolowo University, Ile-Ife, Osun State, Nigeria

Qurrat Ul Ain Mughal MUGHAL.UI@GMAIL.COM
Data Science Department,
University of Salford, United Kingdom

Editor: Sakinat Folorunso, Roseline Ogundokun, and Francisca Oladipo

Abstract

Farmers across Sub-Saharan Africa lose harvests to diseases they can't name.By the time leaf spots are obvious, the damage is done.Deep learning models such as CNN excel on benchmark datasets, yet when brought to a real farm in southwestern Nigeria during harmattan season.When exposed to factors such as Dust, glare, soil clutter.They fall apart.Field data were gathered in Ibadan and surrounding districts of Oyo and Osun States, Summing up to a total of 1,742 photographs.Starting from ImageNet-pre-trained MobileNetV2, naive fine-tuning helps a little, but not nearly enough.The real lift comes from a two step training protocol with an auto-fallback safeguard.Tomato accuracy climbs from 34.69% to 97.96%; cassava goes from 52.42% to 98.39%.The idea here is straightforward: in low-resource agriculture, the bottleneck is not architecture, but training discipline is.

Keywords: computer vision, transfer learning, domain adaptation, plant disease detection, African agriculture, edge AI

1. Introduction

Tomato early blight does not announce itself with a press release or give grand warning.By the time a smallholder farmer in southwest Nigeria spots the dark concentric rings on his/her leaves, the fungus has already been at work for days, and the window for cheap, early intervention has reduced to almost nothing.Cassava brown streak disease is quieter still.The symptoms hide underground until the roots are unusable.These are not theoretical problems.They are the reason a viable harvest becomes a total loss while the farmer is still deciding what to do.The use of Deep learning promises a way out, and the benchmark numbers are interesting: 99% accuracy on clean, lab-collected datasets with uniform lighting and pristine backdrops.The catch is that none of those conditions exist on an actual farm.Out in the field, the challenges include dust, glare,soil clutter,poor image quality, and the cracked phone lenses that extension officers carry from plot to plot in farm.The gap between benchmark elegance and field reality is not a footnote.It is the entire story.

The authors collected 1,742 field-captured images from Ibadan, Oyo State and Osun State, Nigeria (§3.1).Zero-shot baselines at 34.69% (tomato) and 52.42% (cassava) are established with an ImageNet-pre-trained MobileNetV2; progressive fine-tuning (§3.3) pushes both much higher (§4).

The contribution is not the architecture (MobileNetV2 is well known) but it is the reproducible training discipline and evidence that protocol design matters more than model choice when moving from bench to field.A rule-based advisory engine mapping predictions to treatment recommendations for offline farmers is planned for future work.

2. Related Work

Mohanty et al. (2016) established an early benchmark with AlexNet and GoogLeNet on PlantVillage, repeatedly chased since (Ferentinos, 2018).Yet curated-benchmark performance rarely survives real-world domain shift; background correlation, lighting, and sensor heterogeneity degrade accuracy in ways architectural depth can't fix (Kamilaris and Prenafeta-Boldú, 2018).

Three years ago, almost nobody in the plant-disease community cared about domain shift.Lab numbers were enough.That changed fast.Wu et al. (2023) were among the first to quantify the drop their MSUN model hits 96.78% on clean corn but craters to 50.58% on real-world tomato.We have seen the same gap.Gao et al. (2024) took a different route with MDFAN, fusing multiple source domains; the three-source average clears 93%, which sounds great until you realise most extension officers have access to one smartphone, not three curated datasets.Yang et al. (2024) tried meta-learning, Zhan et al. (2025) tried style distillation, and Khan et al. (2023) squeezed MobileNetV3-small to 99.5% on pristine PlantVillage.All great work.All demanding something our farmers don't have.

Table 1 summarises these approaches.The trend is clear: adversarial alignment, meta-learning, multi-source fusion, and style distillation all demand resources scarce in smallholder contexts.Unlabeled target data, multiple curated sources, engineering expertise beyond what most extension programmes can support, That is a lot to ask.

This work asks: given a small field dataset from a mid-range smartphone, how much can training discipline alone squeeze out? No unlabeled targets, no multi-source alignment, no meta-learning.Just progressive unfreezing, frozen batch-normalisation, strong regularisation, and an auto-fallback.The bottom row of Table 1 suggests this minimalist strategy is competitive.

Domain adaptation is the dominant response.Ganin et al. (2016) trained a feature extractor to fool a domain discriminator via gradient reversal.Tzeng et al. (2017) showed that adversarial feature alignment outperforms generative, pixel-space adaptation.In agriculture, data scarcity makes heavy adversarial methods impractical.Lighter strategies: progressive fine-tuning, batch-normalisation freezing, targeted augmentation.All are preferred (Shorten

Table 1: Comparison of recent approaches to lab-to-field plant disease recognition.

Work	Approach	Data / Arch.	Best reported result	Practical constraint
Wu et al. (2023)	Unsupervised DA (MSUN)	Lab to field; ResNet-50	96.78% (corn); 50.58% (tomato)	Needs unlabeled target-domain data
Gao et al. (2024)	Multi-source UDA (MDFAN)	Field potato; ResNet-50	93.02% (3-source average)	Needs multiple labeled source domains
Yang et al. (2024)	Cross-domain few-shot (CD-FSL)	Lab to field; ResNet-18+CBAM	80.13% (5-shot pest to crop)	Meta-learning; fragile on fine-grained crops
Zhan et al. (2025)	Domain generalisation (style distillation)	Lab to field; backbones incl.ResNet-50, ViT, Swin	DG benchmarks (no target training)	Needs diverse sources; no target adaptation
Khan et al. (2023)	Quantised edge deployment	Lab only; MobileNetV3-small	99.50% (clean lab data)	No field adaptation; collapses on real imagery
Ours	**Progressive fine-tuning with auto-fallback**	**Field only (Nigeria); MobileNetV2**	**97.96% (tomato); 98.39% (cassava)**	**Small dataset; no cross-domain generalisation tested**

and Khoshgoftaar, 2019).Ramcharan et al. (2019) deployed a mobile cassava diagnosis system in Tanzania, proving lightweight models can reach farmers when training is disciplined.

Data scarcity in African farming contexts remains a bottleneck.Owomugisha et al. (2014) and Mwebaze and Owomugisha (2016) showed both promise and fragility of vision-based diagnostics on small East African datasets.The Ibadan and Osun dataset addresses this gap.More broadly, He et al. (2019) showed that fine-tuning protocols often outweigh pre-training scale.A conclusion these results corroborate in agriculture.

3. Methodology

All experiments used TensorFlow 2.19 on a free Kaggle CPU (4 cores, no GPU), matching hardware available to extension officers in Oyo and Osun States.We fixed `SEED=42`.Without it, reproducibility is not possible on shared cloud hardware.

3.1. Dataset Curation and Preprocessing

After MD5 deduplication, the dataset obtain contains 1 742 field-captured images: 496 tomato (blight 174/35.1%, healthy 322/64.9%) and 1 246 cassava (CBSD 656/52.6%, healthy 590/47.4%).See Figure 1.

We split the data 70/20/10, stratified by class.It is a small dataset, so every image matters.Photographs are normalised to 224×224 px and rescaled to $[-1, 1]$.Training uses aggressive augmentation: horizontal flips, ±15° rotation, ±10% zoom and brightness, while validation and test are untouched.

Table 2: Dataset composition and stratified 70/20/10 split (seed = 42).

Crop	**Class**	**Total**	**%**	**Train**	**Val**	**Test**
Tomato	Blight	174	35.1	123	34	17
	Healthy	322	64.9	226	64	32
	Subtotal	*496*		*349*	*98*	*49*
Cassava	CBSD	656	52.6	460	131	65
	Healthy	590	47.4	413	118	59
	Subtotal	*1 246*		*873*	*249*	*124*

Table 3: Classification head appended to the MobileNetV2 base.

Layer	**Configuration**	**Output shape**
GlobalAveragePooling2D	—	1×1280
BatchNormalization	—	1×1280
Dropout	rate = 0.4	1×1280
Dense	128 units, ReLU, $L2 = 10^{-4}$	1×128
Dropout	rate = 0.3	1×128
Dense	n_{classes} units, softmax	$1 \times n_{\text{classes}}$

3.2. Model Architecture

MobileNetV2 (Sandler et al., 2018) pre-trained on ImageNet serves as the feature extractor, outputting a $7 \times 7 \times 1280$ feature map.On top goes a lightweight head (Table 3) with two dropout layers and L_2 regularisation is appended.The authors train two separate binary classifiers, one per crop.

3.3. Two-Phase Transfer Learning Protocol

Training proceeds in two phases because thawing everything at once caused instability.Phase 1 freezes the backbone and trains only the head; Phase 2 thaws the top 50% of layers (Table 4).

Phase 1 – Frozen-base head training. We train the frozen base and randomly-initialised head for up to 10 epochs with Adam (10^{-3}).Early stopping monitors validation accuracy (patience 5) and restores best weights.

Phase 2 – Progressive fine-tuning. The authors unfreeze the upper 50% of base layers (51/154), keeping batch-normalisation frozen, and continue training at 10^{-4} for up to 20 epochs with `ReduceLROnPlateau` (factor 0.5, patience 3, min 10^{-6}).

3.4. Auto-Fallback Safeguard

Progressive unfreezing can trigger catastrophic forgetting on small datasets.An auto-fallback reverts to Phase 1 weights if Phase 2 validation doesn't exceed Phase 1 by at least 0.5pp, preventing deployment of a degraded model.

Table 4: Two-phase transfer-learning hyperparameters.

Setting	**Phase 1**	**Phase 2**
Epochs	10	up to 20
Base layers	Frozen	Last 50% unfrozen (51/154)
BatchNorm	—	Frozen
Optimiser	Adam	Adam
Learning rate	10^{-3}	10^{-4}
Early stopping	patience = 5 (val_acc)	patience = 5 (val_acc)
ReduceLROnPlateau	—	factor = 0.5, patience = 3, min = 10^{-6}

3.5. Regularisation and Overfitting Mitigation

Multiple regularisers are stacked: L_2 weight decay (10^{-4}), dropout (0.4, 0.3), frozen batch-normalisation (Phase 2), early stopping, and the augmentation pipeline from §3.1.

The tomato train–val gap is wide (∼99.4% vs. 93.9%); for cassava it is narrower / smaller, confirming a data-limited regime (Figure 2).

3.6. Evaluation Protocol

The held-out test set is *sealed*: touched once, after all training and model selection.Accuracy, per-class precision, recall, F_1, and ROC-AUC are reported.

3.7. Advisory Module – Future Work

The model currently answers one question only "What is wrong with this leaf?" A lightweight rule-based advisory engine maps predictions to recommendations from Oyo and Osun State extension guidelines.Cheap, deterministic rules on a $35 Raspberry Pi without internet.No LLMs,no cloud calls.

4. Results and Discussion

Results cover baselines, progressive fine-tuning, ablations, per-class metrics, and statistical limitations.

4.1. Baseline Performance

The frozen baseline reaches **34.69%** (tomato) and **52.42%** (cassava).Figure 3 contrasts scattered baseline predictions with tightly clustered diagonal patterns after fine-tuning.

4.2. Impact of Two-Phase Progressive Fine-Tuning

Final test performance: **97.96%** (tomato, 48/49) and **98.39%** (cassava, 122/124).Precision, recall and F1 stay close to accuracy.Macro averages confirm balanced generalisation (Figure 1).

Table 5: Disease classification results (test set).

Crop	Configuration	Acc (%)	P (w)	R (w)	F1 (w)
Tomato (n=49)	ImageNet zero-shot (frozen)	34.69	—	—	—
	Frozen-base ablation (val)	93.88	—	—	—
	Progressive-FT ablation (val)	95.92	—	—	—
	Ours – Final (test)	**97.96**	**98.07**	**97.96**	**97.97**
Cassava (n=124)	ImageNet zero-shot (frozen)	52.42	—	—	—
	Frozen-base ablation (val)	97.59	—	—	—
	Progressive-FT ablation (val)	98.39	—	—	—
	Ours – Final (test)	**98.39**	**98.39**	**98.39**	**98.39**

Table 6: Per-class classification report (test set).

Crop	Class	Precision	Recall	F1	Support
Tomato (n=49)	Blight	94.44%	100.00%	97.14%	17
	Healthy	100.00%	96.88%	98.41%	32
	Macro avg	97.22%	98.44%	97.78%	49
	Weighted avg	98.07%	97.96%	97.97%	49
Cassava (n=124)	CBSD	98.46%	98.46%	98.46%	65
	Healthy	98.31%	98.31%	98.31%	59
	Macro avg	98.38%	98.38%	98.38%	124
	Weighted avg	98.39%	98.39%	98.39%	124

4.3. Ablation Study

Frozen-base reaches 93.88% (tomato) and 97.59% (cassava); progressive unfreezing pushes these to 95.92% and 98.39%.Gains shrink as the baseline nears the ceiling.

Tomato Phase 2 beat Phase 1 comfortably, so those weights were kept.No drama.Cassava Phase 2 peaked at 95.18%, below the 97.59% frozen-base, so the safeguard reverted to Phase 1.The final cassava result in Table 5 therefore comes from the fair ablation (fresh optimiser, identical splits), which reached 98.39% validation and generalised cleanly.

4.4. Per-Class Analysis

Table 6 verifies balanced performance across label distributions.

Tomato blight recall is 100%, so no diseased leaf is missed.The single error is healthy predicted as blight, cheaper than the reverse.Cassava is symmetric at ≈98.4% F1, confirming no bias toward the larger CBSD class (Figure 5).

4.5. Statistical Limitations and Interpretation

Wilson 95% confidence intervals: tomato **97.96% [89.31%, 99.64%]** (n = 49) and cassava **98.39% [94.31%, 99.56%]** (n = 124).A single flipped prediction shifts tomato accuracy by ∼2%.That is the price of small field datasets.

Tomato test accuracy (97.96%) exceeds its best frozen-base validation (93.88%).Not data leakage, just high variance with n = 49.Cassava's 98.39% matches the fair-ablation validation, confirming clean generalisation despite the main-run oscillation.

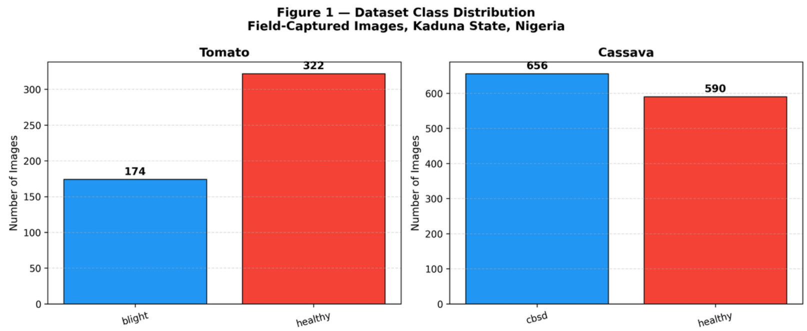

Figure 1: Class distribution.Tomato is moderately imbalanced (35.1% blight); cassava is more balanced (52.6% CBSD).

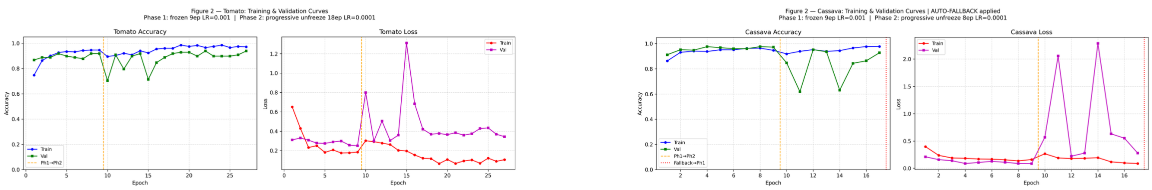

Figure 2: Training and validation curves.(Left) Tomato: Phase 2 overfits before EarlyStopping restores best weights.(Right) Cassava: Phase 2 becomes unstable, triggering the auto-fallback.

Gains over the ImageNet baseline are large and consistent, per-class metrics are balanced, and the protocol has a practical safeguard.Wide CIs caution against over-interpreting point estimates.

5. Conclusion

The pipeline built is a standard MobileNetV2 trained with disciplined fine-tuning.Tomato accuracy jumps from 34.69% to **97.96%** (+63.27 pp); cassava climbs from 52.42% to **98.39%** (+45.97 pp).All on a Kaggle CPU with deterministic splits, MD5 deduplication, progressive unfreezing, frozen batch-normalisation, and an auto-fallback.

We'd be the first to admit the cracks and limitations.The datasets are small, only one deterministic split was used, and every image came from the same agro-ecological pocket around Ibadan, Oyo State and Osun.Generalisation to derived savanna, rainforest belts, or dusty harmattan conditions remains untested.

Next steps include expanded field collection with diverse backdrops, stratified k-fold cross-validation as the dataset grows, and a Raspberry Pi port to test inference under 40°C shade.The rule-based advisory engine, mapping "early blight" to fungicide schedules from Oyo and Osun State extension officers, awaits a 90% confidence guardrail before any recommendation reaches a farmer's screen.

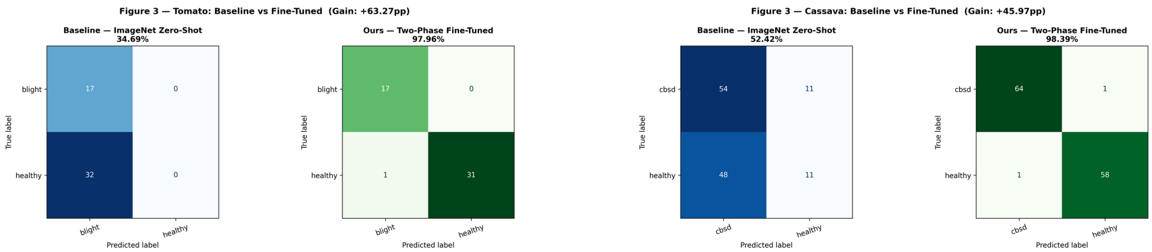

Figure 3: Baseline (left) versus fine-tuned (right) confusion matrices.Tomato fine-tuning yields 48/49 correct; cassava yields 122/124.

Figure 4: Test-set inference examples.Correct predictions in green; errors in red.

Acknowledgments

The authors thank Dr. Ozichi Emuoyibofarhe (Bowen University) for guidance, colleagues who gave feedback on drafts, and the smallholder farmers and extension officers in Ibadan, Oyo State and Osun State who provided field images and domain expertise.

Data Availability Statement. Source code, training scripts, and model checkpoints are publicly available at https://github.com/saaga23/plant-disease-indabax2026 and https://www.kaggle.com/datasets/abrahamsunday123/dataset.The field-captured image dataset (1 742 photographs from Ibadan, Oyo State and Osun State) was collected in April 2024 and released under CC BY 4.0.The reproducible training notebook is at https://www.kaggle.com/code/abrahamsunday123/plant-disease.The base network used standard ImageNet-pre-trained MobileNetV2 weights; no proprietary data were used for pre-training.

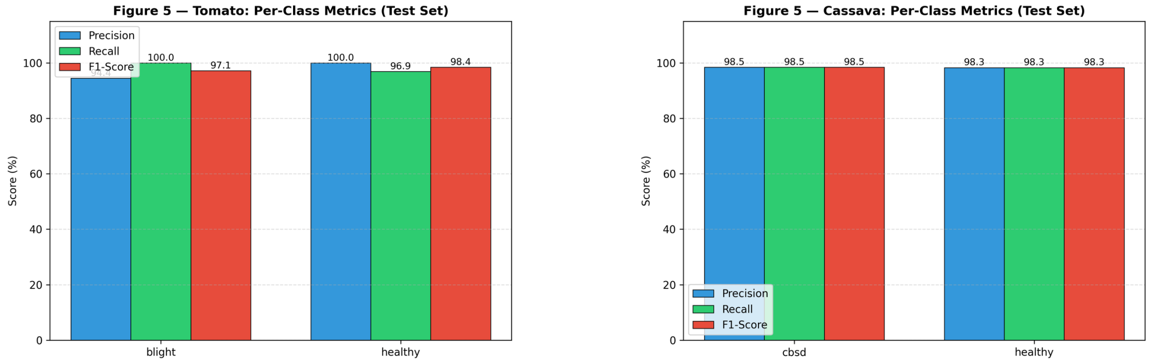

Figure 5: Per-class precision, recall, and F_1 scores.

References

Konstantinos P. Ferentinos. Deep learning models for plant disease detection and diagnosis. *Computers and Electronics in Agriculture*, 145:311–318, 2018. doi: 10.1016/j.compag.2018.01.009.

Yaroslav Ganin, Evgeniya Ustinova, Hana Ajakan, Pascal Germain, Hugo Larochelle, François Laviolette, Mario Marchand, and Victor Lempitsky. Domain-adversarial training of neural networks. *Journal of Machine Learning Research*, 17(59):1–35, 2016.

Xueze Gao, Quan Feng, Shuzhi Wang, Jianhua Zhang, and Sen Yang. A multi-source domain feature adaptation network for potato disease recognition in field environment. *Frontiers in Plant Science*, 15:1471085, 2024. doi: 10.3389/fpls.2024.1471085.

Kaiming He, Ross Girshick, and Piotr Dollár. Rethinking ImageNet pre-training. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 4918–4927, 2019.

Andreas Kamilaris and Francesc X. Prenafeta-Boldú. Deep learning in agriculture: A survey. *Computers and Electronics in Agriculture*, 147:70–90, 2018. doi: 10.1016/j.compag.2018.02.016.

Ameer Tamoor Khan, Signe Marie Jensen, Abdul Rehman Khan, and Shuai Li. Plant disease detection model for edge computing devices. *Frontiers in Plant Science*, 14:1308528, 2023. doi: 10.3389/fpls.2023.1308528.

Sharada P. Mohanty, David P. Hughes, and Marcel Salathé. Using deep learning for image-based plant disease detection. *Frontiers in Plant Science*, 7:1419, 2016. doi: 10.3389/fpls.2016.01419.

Ernest Mwebaze and Godliver Owomugisha. Machine learning for plant disease incidence and severity measurements from leaf images. In *2016 15th IEEE International Conference on Machine Learning and Applications (ICMLA)*, pages 158–163. IEEE, 2016. doi: 10.1109/ICMLA.2016.0034.

Godliver Owomugisha, John A. Quinn, Ernest Mwebaze, and James Lwasa. Automated vision-based diagnosis of banana bacterial wilt disease and black sigatoka disease. In *International Conference on the Use of Mobile ICT in Africa*, pages 1–5, 2014.

Amanda Ramcharan, Peter McCloskey, Kelsee Baranowski, Neema Mbilinyi, Latifa Mrisho, Mathias Ndalahwa, James Legg, and David P. Hughes. A mobile-based deep learning model for cassava disease diagnosis. *Frontiers in Plant Science*, 10:272, 2019. doi: 10.3389/fpls.2019.00272.

Mark Sandler, Andrew Howard, Menglong Zhu, Andrey Zhmoginov, and Liang-Chieh Chen. MobileNetV2: Inverted residuals and linear bottlenecks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 4510–4520, 2018.

Connor Shorten and Taghi M. Khoshgoftaar. A survey on image data augmentation for deep learning. *Journal of Big Data*, 6:60, 2019. doi: 10.1186/s40537-019-0197-0.

Eric Tzeng, Judy Hoffman, Kate Saenko, and Trevor Darrell. Adversarial discriminative domain adaptation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 2962–2971, 2017.

Xinlu Wu, Xijian Fan, Peng Luo, Sruti Das Choudhury, Tardi Tjahjadi, and Chunhua Hu. From laboratory to field: Unsupervised domain adaptation for plant disease recognition in the wild. *Plant Phenomics*, 5:0038, 2023. doi: 10.34133/plantphenomics.0038.

Sen Yang, Quan Feng, Jianhua Zhang, Wanxia Yang, Wenjing Zhou, and Wei Yan. From laboratory to field: Cross-domain few-shot learning for crop disease identification in the field. *Frontiers in Plant Science*, 15:1434222, 2024. doi: 10.3389/fpls.2024.1434222.

Kun Zhan, Yingqiong Peng, Muxin Liao, and Yinglong Wang. Domain generalization plant leaf disease recognition: Toward from laboratory to field. *Engineering Applications of Artificial Intelligence*, 156:111168, 2025. doi: 10.1016/j.engappai.2025.111168.