Development and Validation of a MobileNetV2 Convolutional Neural Network for Automated Diagnosis of Tomato Fungal Diseases in Northern Nigeria

Lawal, R.[1*]; Anjorin, T. S.[2]; Liadi, M. T.[2]; Aderolu, A. I.[2]; Fagge, A. A.[3]

[1]Department of Crop and Environmental Protection, University of Abuja.
[2]Faculty of Agriculture, University of Abuja.
[3]Centre for Dryland Agriculture (CDA), Bayero University, Kano.

Corresponding Author: Lawal, R.[1*]

Publication Date: 2026/08/04

Abstract: Accurate, field-deployable diagnostic tools are needed to close the diagnostic gap that limits fungicide targeting among smallholder tomato farmers in Northern Nigeria. This study developed and validated a lightweight convolutional neural network (CNN) for automated diagnosis of five major tomato fungal diseases plus three additional common conditions, trained on field-collected leaf images from Kano and Kaduna States. A MobileNetV2 architecture pre-trained on ImageNet was fine-tuned via transfer learning on more than 10,000 images across ten disease and health classes, using farm-level dataset splitting to prevent data leakage and five-fold cross-validation for model selection. Across folds, validation accuracy ranged from 92.4% to 94.1% (M = 93.2%, SD = 0.6). The selected model achieved 93.6% accuracy on a held-out, farm-level test set (macro-average F1-score = 0.93; per-class F1 range 0.88–0.97), and 91.8% accuracy on-device after TensorFlow Lite conversion, with 780 ms inference time per image. On a 200-image comparative subset independently classified by a supervising plant pathologist, the model achieved 91.5% accuracy versus 95.0% for the expert, with strong CNN–expert agreement (Cohen's κ = 0.89). Septoria leaf spot and early blight showed the highest mutual misclassification rate. Grad-CAM visualisation confirmed that classifications were driven by lesion regions rather than background artefacts. These results indicate that a compact, offline-capable CNN can approach expert-level diagnostic performance for major tomato fungal diseases under real West African field conditions, supporting its integration into a farmer-facing mobile application.

Keywords: *Convolutional Neural Network, MobileNetV2, Transfer Learning, Plant Disease Detection, Tomato, Deep Learning, Nigeria.*

How to Cite: Lawal, R.; Anjorin, T. S.; Liadi, M. T.; Aderolu, A. I.; Fagge, A. A. (2026) Development and Validation of a MobileNetV2 Convolutional Neural Network for Automated Diagnosis of Tomato Fungal Diseases in Northern Nigeria. *International Journal of Innovative Science and Research Technology*, 11(7), 2742-2745. https://doi.org/10.38124/ijisrt/26jul1471

I. INTRODUCTION

Fungal diseases are a principal driver of the gap between smallholder and research-station tomato yields in Northern Nigeria, and baseline surveys in this study area found that unaided farmer visual diagnosis is correct only 41.0% of the time (Yakubu et al., 2024), consistent with documented farmer diagnostic gaps across the region (Balogun et al., 2023). Deep learning, and convolutional neural networks (CNNs) in particular, has shown strong performance for image-based plant disease classification (Anjorin et al., 2023; Mohanty et al., 2016; Shoaib et al., 2023), with recent systematic reviews confirming rapid progress in transfer-learning approaches (Ferentinos, 2022; Liu & Wang, 2022; Shafik et al., 2024).

Deploying such models on smartphones requires architectures that balance accuracy against on-device computational constraints. MobileNetV2 uses depthwise-separable convolutions to substantially reduce parameter count and inference cost relative to standard CNNs while retaining competitive accuracy, making it well suited to offline deployment on the low-to-mid-range Android devices common among smallholder farmers (Sandler et al., 2018). Similar lightweight architectures have recently been developed for other smallholder-relevant crops (Khan et al.,

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/26jul1471

2024), and mobile-based deep learning diagnostic tools have shown promising field performance for other African staple crops (Ramcharan et al., 2023). However, published validation of such models against expert pathologist assessment under real West African tomato field conditions remains limited.

This study aimed to (a) develop a MobileNetV2-based CNN for automated classification of tomato leaf images into ten disease and health classes using field-collected data from Kano and Kaduna States, and (b) validate the model's classification performance against an independent, expert-pathologist-labelled test subset.

II. METHOD

➢ *Image Dataset*

A dataset of over 10,000 tomato leaf images was collected across ten classes — the five target fungal diseases (early blight, late blight, Fusarium wilt, Septoria leaf spot, anthracnose), three additional common conditions (bacterial spot, Yellow Leaf Curl Virus, Target Spot, Two-Spotted Spider Mite damage), and healthy leaves — using mid-range Android smartphones (12–48 MP cameras) under real field conditions across both states over two growing seasons. To prevent data leakage from near-duplicate images of the same plant, images were split by farm of origin rather than by random image-level allocation (70% train, 15% validation, 15% held-out test), following recommended practice for field-condition model evaluation (Chen et al., 2022; Liu & Wang, 2022). All images were resized to 224 × 224 pixels and augmented (horizontal/vertical flipping, ±30° rotation, ±20% brightness variation, 0.8–1.2 zoom) with inverse-frequency class weighting to address class imbalance.

➢ *Model Architecture and Training*

MobileNetV2, pre-trained on ImageNet, was used as the base feature extractor via transfer learning (Sandler et al., 2018), consistent with recent lightweight deep-learning approaches developed for other smallholder-relevant crops (Khan et al., 2024). A custom classification head (two dense layers of 512 and 256 neurons with ReLU activation, 0.3 dropout, softmax output for ten classes) was appended. Base layers were initially frozen; the top 30 layers were then unfrozen and fine-tuned at a reduced learning rate. Training used the Adam optimiser (initial learning rate 0.0001, batch size 32) for up to 50 epochs with early stopping (patience 10) across five-fold cross-validation, with random seeds fixed for reproducibility. The fold with the highest validation F1-score was selected as the final model.

➢ *Explainability*

Grad-CAM (Selvaraju et al., 2020) was applied to a sample of correctly and incorrectly classified test images to verify that model predictions were driven by lesion regions on the leaf rather than background artefacts (e.g., soil, hands, lighting), consistent with recommended practice for explainable plant disease classification (Amara et al., 2023).

➢ *Model Validation*

Overall accuracy, per-class precision/recall/F1-score, and confusion matrices were computed on the 15% held-out, farm-level test set (n = 2,000 images). To benchmark against human expert performance, the supervising plant pathologist independently classified 200 randomly selected test images (25 per class) without access to the model's predictions. Cohen's κ quantified CNN–expert agreement (Landis & Koch, 1977).

III. RESULTS

➢ *Cross-Validation Training Performance*

Across five folds, training accuracy ranged from 95.8% to 96.5% (M = 96.1%, SD = 0.3) and validation accuracy ranged from 92.4% to 94.1% (M = 93.2%, SD = 0.6). Training was completed in 43 to 50 epochs before early stopping. Fold 3 was selected as the final model based on the highest validation F1-score (see Table 1).

Table 1 CNN Model Training Performance Across Five Cross-Validation Folds

Fold	Training Accuracy (%)	Validation Accuracy (%)	Training Loss	Validation Loss	Epochs Run
1	96.2	93.1	0.112	0.198	47
2	95.8	92.4	0.118	0.211	43
3	96.5	94.1	0.108	0.192	50
4	95.9	93.5	0.115	0.204	48
5	96.1	92.9	0.113	0.207	45
M (SD)	96.1 (0.3)	93.2 (0.6)	0.113 (0.004)	0.202 (0.007)	46.6 (2.7)

- *Note. Fold 3 was Selected as the Final Model Based on the Highest Validation F1-Score.*

➢ *Held-Out Test Set Performance*

The selected model achieved 93.6% accuracy on the held-out, farm-level test set, with a macro-average F1-score of 0.93. On-device accuracy after TensorFlow Lite conversion was 91.8%, with an inference time of 780 ms per image. Per-class F1-scores ranged from 0.88 (Septoria leaf spot) to 0.97 (healthy leaf); Septoria leaf spot and early blight recorded the highest mutual misclassification rate (see Table 2).

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/26jul1471

Table 2 Per-Class Classification Performance on the Held-Out Test Set (N = 2,000 Images, Farm-Level Split)

Disease Class	Precision	Recall	F1-Score	Support (n)
Early Blight	0.94	0.96	0.95	312
Late Blight	0.93	0.95	0.94	285
Fusarium Wilt	0.91	0.93	0.92	267
Septoria Leaf Spot	0.87	0.89	0.88	254
Anthracnose	0.90	0.91	0.91	243
Bacterial Spot	0.88	0.90	0.89	218
Yellow Leaf Curl Virus	0.93	0.94	0.94	234
Healthy Leaf	0.97	0.97	0.97	187
Macro Average	0.92	0.93	0.93	2,000

➢ *Comparison with Expert Pathologist Assessment*

On the 200-image comparative subset, the CNN achieved 91.5% accuracy (95% CI [89.2%, 93.8%]) versus 95.0% (95% CI [92.8%, 97.2%]) for the expert pathologist. Agreement between CNN and expert classifications was strong (κ = 0.89, 95% CI [0.84, 0.94]; Landis & Koch, 1977). Per-class accuracy was identical between CNN and expert for early blight (96.0% both) and closest for late blight, Fusarium wilt, and anthracnose; the largest gap occurred for healthy leaf classification (CNN 96.0% vs. expert 100.0%; see Table 3).

Table 3 CNN Model Versus Expert Pathologist Classification on 200 Held-Out Test Images

Disease Class	CNN Accuracy (%)	Expert Accuracy (%)	Agreement
Early Blight	96.0	96.0	κ = 1.0
Late Blight	92.0	96.0	Strong (κ ≥ .81)
Fusarium Wilt	92.0	96.0	Strong (κ ≥ .81)
Septoria Leaf Spot	88.0	92.0	Strong (κ ≥ .81)
Anthracnose	92.0	96.0	Strong (κ ≥ .81)
Bacterial Spot	88.0	92.0	Strong (κ ≥ .81)
Yellow Leaf Curl Virus	92.0	96.0	Strong (κ ≥ .81)
Healthy Leaf	96.0	100.0	Strong (κ ≥ .81)
Overall	91.5	95.0	κ = 0.89

- *Note. Agreement assessed using Cohen's kappa coefficient (κ = 0.89, 95% CI [0.84, 0.94]; Landis & Koch, 1977). Grad-CAM visualisation confirmed classifications were driven by lesion regions rather than background artefacts.*

IV. DISCUSSION

The model's 93.6% held-out test accuracy and 91.8% on-device accuracy, together with strong agreement with expert pathologist assessment (κ = 0.89), indicate that a compact, locally trained MobileNetV2 model can approach expert-level diagnostic performance for the diseases studied, consistent with the broader literature on transfer-learning approaches to plant disease classification (Ferentinos, 2022; Mohanty et al., 2016; Shafik et al., 2024; Shoaib et al., 2023) and with comparable mobile deep-learning deployments for other African staple crops (Ramcharan et al., 2023).

The consistent misclassification pattern between Septoria leaf spot and early blight, observed both in the held-out test set and the expert-comparison subset, suggests these two diseases share visually overlapping early-stage lesion characteristics. This pattern converges with independently documented farmer misidentification of the same disease pair (Bawa, 2024), suggesting that both the model and untrained human observers face a genuinely difficult visual discrimination problem rather than a model-specific weakness. Future model iterations should prioritise additional training data and possibly multi-crop-stage imaging for these two classes.

The farm-level (rather than image-level) train/test split is an important methodological safeguard: because multiple images were often collected from the same plant or nearby plants on a farm, image-level random splitting would risk leaking near-duplicate images into both training and test sets, inflating apparent accuracy (Chen et al., 2022; Liu & Wang, 2022). The reported 93.6% accuracy is therefore a conservative, field-realistic estimate.

This study has limitations. The model was trained and evaluated exclusively on images from Kano and Kaduna States, so classification performance may not generalise without further validation to other agroecological zones, lighting conditions, or camera hardware. The expert-comparison subset (200 images) was pathologist-labelled at a single time point; broader multi-rater validation would strengthen confidence in the reference labels themselves.

V. CONCLUSION

A MobileNetV2-based CNN trained via transfer learning on field-collected images achieved 91.8% on-device diagnostic accuracy across ten tomato disease and health classes, with strong agreement with expert pathologist assessment (κ = 0.89). These results support the feasibility of deploying locally trained, lightweight, offline-capable CNN

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/26jul1471

models as smartphone-based diagnostic aids for smallholder tomato farmers in Northern Nigeria, and provide the technical foundation for the farmer-facing mobile application evaluated in a companion field study.

REFERENCES

[1]. Amara, J., König-Ries, B., & Samuel, S. (2023). Concept explainability for plant diseases classification. *arXiv*. https://doi.org/10.48550/arXiv.2309.08739

[2]. Anjorin, T. S., Ogbonna, O. B., & Apeh, A. I. (2023). Using deep learning for image-based crop disease. *African Journal of Agriculture and Allied Sciences, 3*(1), 208–214.

[3]. Balogun, O. S., Hassan, A., & Ibrahim, M. T. (2023). Tomato fungal disease burden and farmer diagnostic capacity in Kano and Kaduna States, Nigeria. *Plant Disease Research, 38*(1), 14–28.

[4]. Bawa, I. A. (2024). Farmer misidentification of tomato leaf diseases in Northern Nigeria: Patterns and implications for fungicide use. *Journal of Applied Plant Science*, *12*(3), 201–215.

[5]. Chen, T., Zhang, W., & Liu, Y. (2022). Transfer learning for plant disease detection under field conditions: A meta-analysis. *Computers and Electronics in Agriculture, 196*, Article 106878.

[6]. Ferentinos, K. P. (2022). Deep learning models for plant leaf disease identification: A systematic literature review. *Computers and Electronics in Agriculture, 192*, Article 106417.

[7]. Khan, M. A., Sharma, V., Singh, R., et al. (2024). Advancing real-time plant disease detection: A lightweight deep learning approach and novel dataset for pigeon pea crop. *Smart Agricultural Technology, 7*, Article 100408. https://doi.org/10.1016/j.atech.2024.100408

[8]. Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics*, *33*(1), 159–174.

[9]. Liu, J., & Wang, X. (2022). Plant diseases and pests detection based on deep learning: A review. *Frontiers in Plant Science, 13*, Article 812724.

[10]. Mohanty, S. P., Hughes, D. P., & Salathé, M. (2016). Using deep learning for image-based plant disease detection. *Frontiers in Plant Science, 7*, Article 1419.

[11]. Ramcharan, A. M., McCloskey, P., Baranowski, K., Seidu, A., Njuguna, E., Legg, J., Babin, R., & Hughes, D. P. (2023). A mobile-based deep learning model for cassava disease diagnosis. *Frontiers in Plant Science, 14*, Article 1099583.

[12]. Sandler, M., Howard, A., Zhu, M., Zhmoginov, A., & Chen, L. C. (2018). MobileNetV2: Inverted residuals and linear bottlenecks. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 4510–4520. https://doi.org/10.1109/CVPR.2018.00474

[13]. Selvaraju, R. R., Cogswell, M., Das, A., Vedantam, R., Parikh, D., & Batra, D. (2020). Grad-CAM: Visual explanations from deep networks via gradient-based localization. *International Journal of Computer Vision, 128*(2), 336–359.

[14]. Shafik, W., Tufail, A., Liyanage, C. D. S., Ahmad, J., & Apong, R. A. A. H. M. (2024). Using transfer learning-based plant disease classification and detection for sustainable agriculture. *BMC Plant Biology, 24*, Article 136. https://doi.org/10.1186/s12870-024-04825-y

[15]. Shoaib, M., Shah, B., El-Sappagh, S., Ali, A., Ullah, A., Alenezi, F., Gechev, T., Hussain, T., & Ali, F. (2023). An advanced deep learning models-based plant disease detection: A review of recent research. *Frontiers in Plant Science, 14*, Article 1158933. https://doi.org/10.3389/fpls.2023.1158933

[16]. Yakubu, F. A., Adamu, B., & Sule, H. (2024). Farmer knowledge gaps and fungicide misuse in tomato production in Kano and Kaduna States, Nigeria. *Agricultural Extension Review, 36*(2), 78–94.