Natural Resources for Human Health
eISSN: 2583-1194

Original Research

Received 17 Aug 2026
Revised 10 Sep 2026
Accepted 28 Sep 2026

Natural Resources for Human Health 2026, Vol. 6 (23s): 158–165

KEYWORDS:
Ginger leaf disease; severity staging; deep learning; EfficientNet-B0; Grad-CAM; precision agriculture.

Deep Learning-Based Severity Staging of Ginger Leaf Diseases Using EfficientNet-B0: Early, Middle, and Later Stage Classification

Dhananjaya Kumar H S[1], Dr. K Satyanarayan Reddy[2]

[1]Research Scholar, Institute of Computer Science and Information Science, Srinivas University, Mangalore, India. ORCID: 0009-0004-7929-7106; Email: dhanunccsuo@gmail.com
[2]Research Professor, Institute of Computer Science and Information Science, Srinivas University, Mangalore, India. ORCID: 0000-0003-3765-3198; Email: vicechancellor@srinivasuniversity.edu.in

ABSTRACT: Timely and accurate assessment of disease severity in ginger (*Zingiber officinale*) crops is critical for informed agronomic decision-making, yet most existing plant disease detection systems focus solely on disease identification without quantifying progression stage. This study presents a deep learning-based framework for classifying ginger leaf disease severity into three stages — Early, Middle, and Later across three disease categories: Damage-Pest, Dehydrated, and Leaf-blight. A dataset of 8,712 field-collected and annotated leaf images, approximately balanced across classes and stages, was used to train an EfficientNet-B0 convolutional neural network. Model performance was evaluated using stratified 5-fold cross-validation, achieving a mean classification accuracy of 76.68% ± 0.50% and a mean macro-F1 score of 76.50% ± 0.53%, with a narrow 95% confidence interval [0.760, 0.770] confirming consistent performance across folds. Gradient-weighted Class Activation Mapping (Grad-CAM) was employed to visualize model attention, confirming that predictions were driven by lesion-affected leaf regions. These results demonstrate the feasibility of automated severity staging as a practical decision-support tool for precision ginger cultivation and disease management.

1. INTRODUCTION

Ginger (*Zingiber officinale*) is an economically important spice and medicinal crop cultivated widely across tropical and subtropical regions. Its yield and quality are frequently compromised by biotic and abiotic stresses, including pest damage, dehydration stress, and fungal leaf blight infections. Conventional disease assessment relies on manual visual inspection by agronomists or farmers, a process that is time-consuming, subjective, and difficult to scale across large cultivation areas.

While deep learning-based plant disease detection has received substantial research attention in recent years, the majority of existing approaches address binary or multi-class disease identification without characterizing the severity or progression stage of the disease. Severity information, however, is often more actionable for farmers than a bare diagnosis, since it directly informs the urgency and type of intervention required — for instance, distinguishing an Early-stage infection that may respond to preventive treatment from a Later-stage infection that may require more aggressive management or crop-loss mitigation.

This study addresses this gap by proposing a severity-staging framework that classifies ginger leaf disease images into three progression stages — Early, Middle, and Later — across three disease categories. The primary contributions of this work are: (1) construction and use of a severity-annotated ginger leaf disease dataset spanning three disease classes and three severity stages; (2) training and rigorous cross-validated evaluation of an EfficientNet-B0-based severity classifier; and (3) interpretability analysis using Grad-CAM to validate that model predictions are grounded in biologically meaningful leaf regions.

2. RELATED WORK

2.1 Deep Learning for Plant Disease Detection

Convolutional neural network (CNN) architectures have been widely applied to plant disease classification. EfficientNet-based transfer learning has repeatedly been shown to outperform alternative architectures on leaf disease classification benchmarks: a recent comparative study evaluated CNN and hybrid ConvNet-ViT architectures on banana, cherry, and tomato leaf disease datasets and used Grad-CAM to validate model attention across all evaluated architectures [1]. More broadly, EfficientNet variants have consistently ranked among the strongest performers in leaf disease classification comparisons against ResNet, MobileNet, and VGG-family architectures, motivating its selection as the base architecture for the present study [2].

2.2 Ginger Leaf Disease Detection and Disease Severity Estimation

Ginger leaf disease detection using deep learning remains comparatively underexplored relative to other crops. An early study collected 7,014 field images of healthy and bacterial-wilt-infected ginger leaves from farms in Ethiopia and trained a custom five-layer CNN, achieving 95.2% test accuracy, though the work was limited to binary healthy/diseased classification [3]. More recently, a comparative study evaluated several pre-trained architectures on a self-collected ginger leaf dataset spanning four classes — healthy, leaf blight, dehydrated, and pest damage — reporting EfficientNetB3 as the best-performing model at 94.3% accuracy after augmenting the dataset to 10,910 images [4]. Notably, the three disease classes and per-class image distribution used in the present study closely match those reported for this dataset, indicating a shared or closely related data source; unlike that study, which addressed binary disease-type classification, the present work focuses specifically on severity staging within each disease class.

Beyond disease-type classification, severity estimation has received comparatively limited research attention across crops generally. Early work applying deep learning to fine-grained severity classification showed that CNNs could estimate discrete severity levels for apple black rot with 90.4% accuracy, establishing CNN-based severity classification as viable [5]. Subsequent multi-task approaches have combined species identification, disease classification, and severity estimation within a single network, achieving strong performance across all three sub-tasks simultaneously [6]. A recent systematic review of CNN-based plant disease severity assessment methods found that severity assessment remains substantially less studied than disease-type identification, and identified dataset scarcity and class imbalance as persistent challenges in this sub-field [7] — a gap the present study's severity-staged ginger dataset directly addresses. In a related crop, a multi-task framework combining segmentation and attention-based classification was recently used to stage rice bacterial leaf blight severity while applying Grad-CAM for interpretability, underscoring both the growing interest in severity-stage classification beyond disease-type identification and the pairing of severity staging with explain ability methods across crops [11].

2.3 Interpretability in Agricultural Deep Learning

Grad-CAM and related class-activation approaches, originally developed to produce gradient-based visual explanations for CNN predictions without requiring architectural modification or retraining [9], have increasingly been adopted in agricultural deep learning to build trust in model predictions by visually confirming that models attend to disease-relevant leaf regions rather than background artefacts. Grad-CAM has been applied to validate attention patterns across multiple architectures on multi-crop leaf disease datasets, with heat maps shown to concentrate on streaked, spotted, or otherwise symptomatic leaf regions across disease classes [1]. Similarly, Grad-CAM applied to a ResNet152-based corn leaf disease classifier confirmed that the model's attention aligned with diseased leaf regions rather than background, supporting the practical value of interpretability for building end-user trust among farmers and agronomists [8]. The present study applies the same principle to ginger leaf severity staging, using Grad-CAM not only to build confidence in correct predictions but also to diagnose likely sources of misclassification.

3. MATERIALS AND METHODS

3.1 Dataset

A severity-staged ginger leaf disease dataset comprising 8,712 labelled images was used for this study, spanning three disease categories — Damage-Pest, Dehydrated, and Leaf-blight — and three severity stages: Early, Middle, and Later. Table 1 summarizes the class-wise and stage-wise distribution of the dataset, which was approximately balanced across both dimensions, reducing the risk of class-imbalance bias during training. The per-class image counts closely match those reported for a self-

collected, augmented ginger leaf disease dataset described in related work on ginger disease-type classification [4], suggesting a shared or closely related underlying data source; this study extends that line of work by staging severity within each disease class rather than classifying disease type alone.

Table 1. Class-wise and stage-wise distribution of the ginger leaf disease dataset.

Class	Early	Middle	Later	Total
Damage-Pest	681	688	707	2,076
Dehydrated	1,125	1,130	1,117	3,372
Leaf-blight	1,087	1,093	1,084	3,264
Total	**2,893**	**2,911**	**2,908**	**8,712**

3.2 Model Architecture

An EfficientNet-B0 convolutional neural network was employed as the base architecture for severity-stage classification (Early / Middle / Later). EfficientNet-B0 was selected for its favourable accuracy-to-parameter-count trade-off and its established effectiveness in prior agricultural image classification studies. The final classification layer was replaced with a 3-class output layer corresponding to the three severity stages.

3.3 Data Pre-processing and Augmentation

Input images were resized to 224×224 pixels and normalized using standard Image Net mean and standard deviation values. To improve model generalization and reduce over fitting, training images were augmented using random horizontal flipping, random rotation (±15°), and colour jitter (brightness, contrast, and saturation).

3.4 Training Protocol

Model performance was evaluated using stratified 5-fold cross-validation, stratified on the severity-stage label to preserve class balance across folds. For each fold, the model was trained using the Adam optimizer (learning rate = 3×10^{-4}, weight decay = 1×10^{-4}) with a cosine annealing learning-rate schedule, and cross-entropy loss, using a batch size of 32. The best-performing epoch within each fold, selected by validation macro-F1 score, was retained for final evaluation.

3.5 Evaluation Metrics

Model performance was assessed using accuracy, macro-averaged F1 score, macro-averaged precision, and macro-averaged recall, computed per fold and aggregated as mean ± standard deviation across the five folds. Statistical reliability was further quantified using 95% confidence intervals on the macro-F1 score across folds.

3.6 Interpretability Analysis

To assess model interpretability and validate that classification decisions were driven by biologically meaningful leaf regions, Gradient-weighted Class Activation Mapping (Grad-CAM) was applied to the final convolutional layer of the trained model, generating class-discriminative localization heat maps for representative test images.

4. RESULTS AND DISCUSSION

4.1 Classification Performance

The EfficientNet-B0 severity-staging model achieved a mean classification accuracy of 76.68% ± 0.50% and a mean macro-F1 score of 76.50% ± 0.53% across five-fold cross-validation, with mean macro-precision of 76.91% ± 0.49% and mean macro-recall of 76.69% ± 0.50% (Table 2). The low standard deviations across folds indicate stable and reproducible model performance rather than fold-dependent variance.

Natural Resources for Human Health
eISSN: 2583-1194

Table 2. Five-fold cross-validation performance of the EfficientNet-B0 severity-staging model.

Model	Accuracy (%)	F1-macro (%)	Precision-macro (%)	Recall-macro (%)
EfficientNet-B0	76.68 ± 0.50	76.50 ± 0.53	76.91 ± 0.49	76.69 ± 0.50

4.2 Statistical Validation

A 95% confidence interval computed on the macro-F1 score across the five folds yielded a mean of 0.7650, with a narrow interval of [0.7597, 0.7702] (Table 3), further supporting the consistency of model performance across data partitions.

Table 3. 95% confidence interval on macro-F1 score across folds.

Model	Mean Macro-F1	95% CI Lower	95% CI Upper
EfficientNet-B0	0.7650	0.7597	0.7702

4.3 Confusion Matrix Analysis

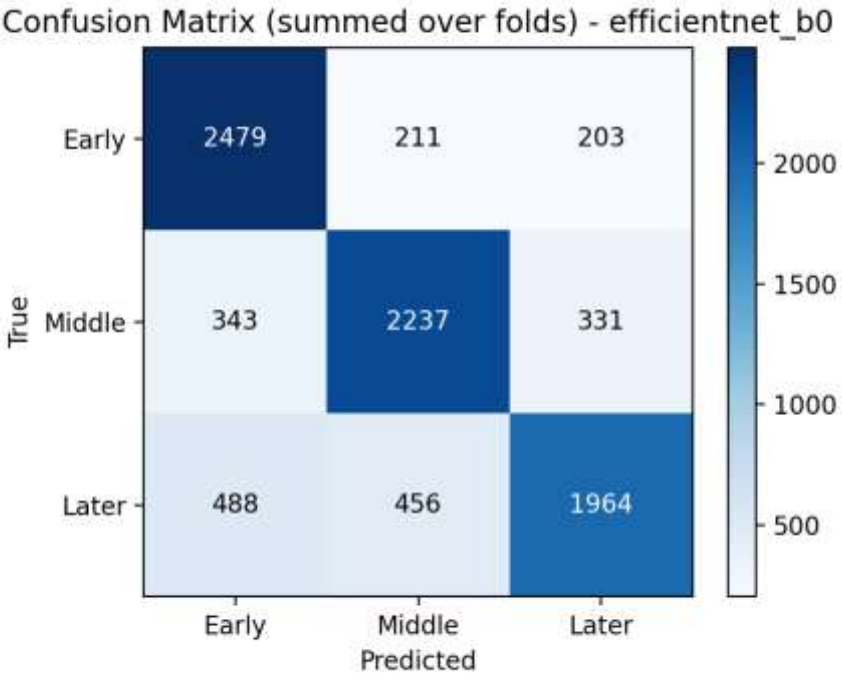


Figure 1. Aggregated confusion matrix across folds for the EfficientNet-B0 severity-staging model.

Per-class recall computed from the aggregated confusion matrix (Figure 1) was 85.7% for Early, 76.8% for Middle, and 67.5% for Later, indicating that classification accuracy decreased progressively with disease severity stage. The Later stage was the most frequently misclassified, with 488 instances predicted as Early and 456 as Middle, out of 2,908 total Later-stage samples. Notably, the confusion between Early and Later stages (488 instances) was comparable in magnitude to the confusion between the adjacent Middle and Later stages (456 instances), which is a less intuitive pattern than would be expected if misclassification were driven purely by visual similarity between adjacent stages of a continuous progression. This suggests that some Later-stage images may share visual characteristics with Early-stage presentations — for instance, if severe leaf damage or dehydration produces localized regions that resemble early-stage symptoms — and warrants further examination of the specific misclassified samples in future work. The Early stage, by contrast, was classified with the highest reliability, likely reflecting more visually distinctive and consistent symptom presentation at that stage.

4.4 Grad-CAM Interpretability Analysis

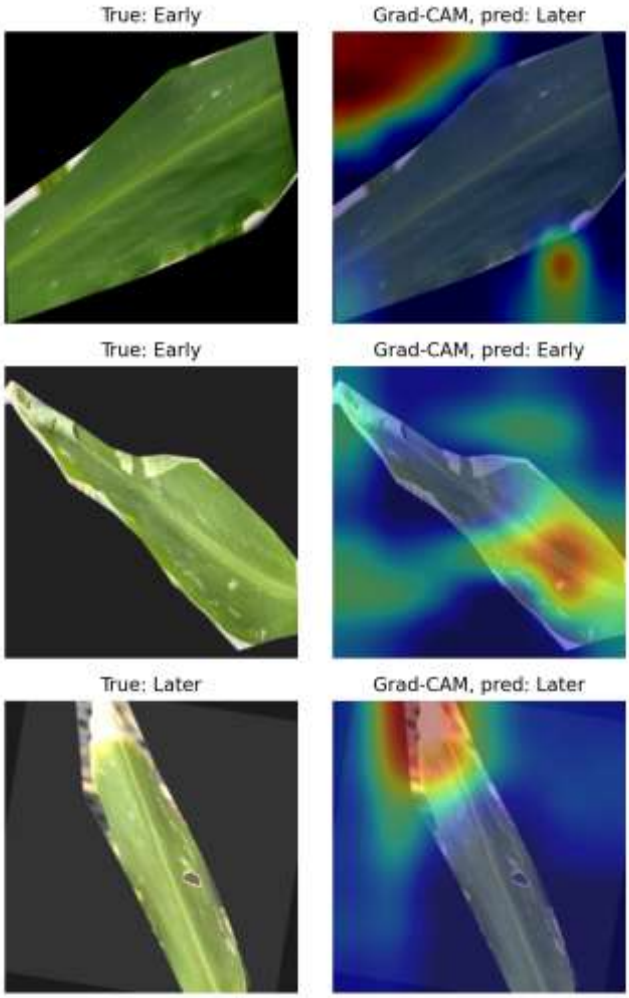


Figure 2. Grad-CAM visualizations for representative test images across severity stages (left: original image; right: Grad-CAM overlay with predicted stage).

Grad-CAM visualizations (Figure 2) revealed mixed but informative attention patterns. In the second and third examples — a correctly classified Early-stage sample and a correctly classified Later-stage sample — model attention was concentrated on discoloured or damaged regions of the leaf blade and leaf tip, consistent with biologically meaningful, lesion-driven decision-making. In the first example, a true Early-stage image that was misclassified as Later, the model's attention was partly drawn to a bright region near the image corner outside the leaf boundary, alongside a smaller hotspot on a leaf-edge notch; this suggests that background or edge artefacts may have contributed to at least some misclassifications, and highlights a concrete direction for improvement — such as tighter leaf segmentation or background masking during pre-processing — to reduce the influence of non-leaf regions on the model's predictions.

4.5 Per-Fold and Per-Class Performance Visualization

To complement the aggregate statistics reported in Table 2 and Table 3, Figures 3 and 4 visualize per-fold and per-class performance directly from the cross-validation results.

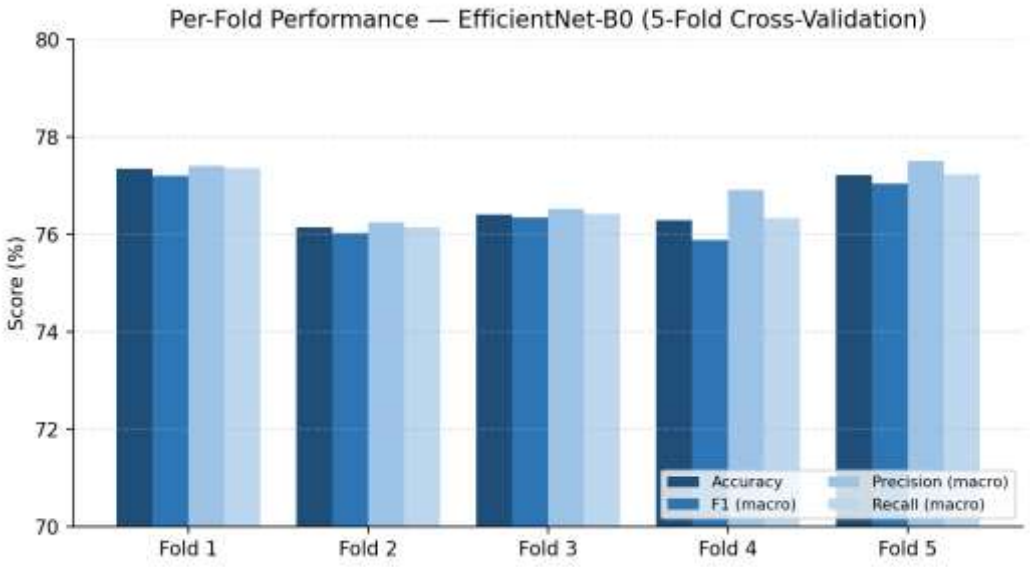


Figure 3. Accuracy, macro-F1, macro-precision, and macro-recall for each of the five cross-validation folds.

As shown in Figure 3, all four metrics remained within a narrow band of approximately 76% to 77.5% across folds, with no single fold showing a marked outlier, further corroborating the low standard deviations reported in Table 2 and indicating that the model's performance is not driven by a particular favourable train/validation split.

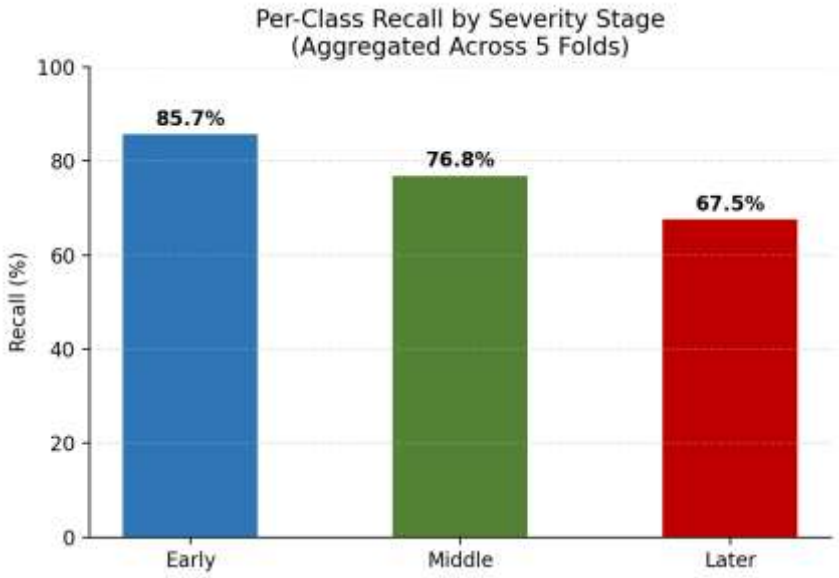


Figure 4. Per-class recall by severity stage, aggregated across all five folds.

Figure 4 visualizes the progressive decline in per-class recall from Early (85.7%) to Middle (76.8%) to Later (67.5%) stages already noted in Section 4.3, highlighting that the Later stage is disproportionately responsible for the model's overall error rate and should be prioritized in future data collection and augmentation efforts.

4.6 Discussion

The proposed EfficientNet-B0 severity-staging model achieved consistent classification performance across five-fold cross-validation, with a mean accuracy of 76.68% and a narrow 95% confidence interval on macro-F1, indicating that the model's performance is stable and not driven by a favourable train/validation split. This level of performance is achieved despite the

model being trained without ImageNet-pretrained initialization; transfer learning from pretrained weights is a well-established technique for improving convergence speed and final accuracy in image classification tasks — for instance, fine-tuning a pretrained EfficientNet-B0 rather than training it from a frozen or randomly initialized state has been shown to yield double-digit accuracy gains on leaf-disease classification benchmarks [10] and its omission here represents a methodological limitation rather than a ceiling on achievable performance for this architecture and dataset.

Compared to disease-type classification tasks, which often distinguish visually distinct categories, severity staging is inherently a more challenging task because disease progression is continuous and boundaries between adjacent stages can be gradual rather than discrete. This is reflected in the per-class recall results, which declined progressively from 85.7% (Early) to 76.8% (Middle) to 67.5% (Later), indicating that the model finds later-stage samples harder to classify reliably than early-stage ones. The Grad-CAM analysis additionally suggests that a subset of misclassifications may be attributable to attention on background or edge artifacts rather than purely to genuine visual ambiguity between stages, pointing to leaf segmentation or background masking as a concrete avenue for improving robustness in future iterations.

The Grad-CAM interpretability analysis supports the biological plausibility of the model's decisions, indicating that predictions are grounded in disease-relevant leaf regions rather than spurious background correlations, which is an important consideration for eventual deployment in real-world agricultural settings where model trust and explain ability are critical for farmer adoption.

4.7 Limitations

This study has several limitations that should be acknowledged. First, the model was trained without pretrained weight initialization due to a network-access constraint in the training environment, which likely limits the achievable accuracy relative to a fully transfer-learned model. Second, only a single architecture (EfficientNet-B0) was evaluated in this study; comparison against alternative architectures such as ResNet-50 and Vision Transformer variants, as conducted in related work on ginger leaf disease detection, was not performed here and is left as future work. Third, the dataset, while reasonably sized and balanced, originates from a specific geographic and imaging context, and generalization to other regions, cultivars, or imaging conditions has not been evaluated.=

5. CONCLUSION

This study presented a deep learning-based framework for severity staging of ginger leaf diseases into Early, Middle, and Later stages using an EfficientNet-B0 classifier, evaluated via stratified 5-fold cross-validation on a dataset of 8,712 labelled images. The model achieved a mean accuracy of 76.68% ± 0.50% and mean macro-F1 of 76.50% ± 0.53%, with consistent performance across folds confirmed by a narrow 95% confidence interval. Grad-CAM visualizations further confirmed that model predictions were grounded in disease-relevant leaf regions, supporting the interpretability and practical trustworthiness of the approach. Future work will focus on incorporating pretrained transfer learning to improve accuracy, extending the comparison to additional architectures (ResNet-50, Vision Transformer), and validating the framework across broader geographic and cultivar contexts to support real-world deployment as a precision agriculture decision-support tool.

ACKNOWLEDGEMENTS

The authors thank the Institute of Computer Science and Information Science, Srinivas University, Mangalore, for the facilities provided to carry out this research.

AUTHOR CONTRIBUTIONS

Dhananjaya Kumar H S: Conceptualization, Methodology, Software, Data Curation, Formal Analysis, Investigation, Writing – Original Draft.

Dr. K Satyanarayan Reddy: Supervision, Validation, Writing – Review & Editing.

CONFLICT OF INTEREST

The authors declare that there is no conflict of interest.

ETHICS STATEMENT

This study did not involve human participants, human data, or animal subjects. Ethical approval was not required.

DATA AVAILABILITY STATEMENT

The datasets generated and/or analyzed during the current study are available from the corresponding author upon reasonable request.

REFERENCES

[1] Murugesan, S., Chinnadurai, J., Srinivasan, S., Mathivanan, S.K., Chandan, R.R., Moorthy, U., 2025. Robust multiclass classification of crop leaf diseases using hybrid deep learning and Grad-CAM interpretability. *Scientific Reports* 15, 29955. https://doi.org/10.1038/s41598-025-14847-7

[2] Tan, M., Le, Q., 2019. EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks. In: *Proceedings of the 36th International Conference on Machine Learning (ICML)*, Long Beach, CA, USA, 9–15 June 2019.

[3] Yigezu, M.G., Woldeyohannis, M.M., Tonja, A.L., 2022. Early Ginger Disease Detection Using Deep Learning Approach. In: *Lecture Notes of the Institute for Computer Sciences, Social-Informatics and Telecommunications Engineering (LNICST)*, vol. 411. Springer, Cham, Switzerland, pp. 480–488. https://doi.org/10.1007/978-3-030-93709-6_32

[4] Wong, W.Z., Tew, Y., Tan, C.W., 2026. Ginger Leaf Diseases Detection Using Deep Learning: A Comparative Study of Pre-Trained Models. *Engineering Proceedings* 128(1), 1. https://doi.org/10.3390/engproc2026128001

[5] Wang, G., Sun, Y., Wang, J., 2017. Automatic Image-Based Plant Disease Severity Estimation Using Deep Learning. *Computational Intelligence and Neuroscience* 2017, 2917536. https://doi.org/10.1155/2017/2917536

[6] Liang, Q., Xiang, S., Hu, Y., Coppola, G., Zhang, D., Sun, W., 2019. PD2SE-Net: Computer-assisted plant disease diagnosis and severity estimation network. *Computers and Electronics in Agriculture* 157, 518–529. https://doi.org/10.1016/j.compag.2019.01.034

[7] Shi, T., Liu, Y., Zheng, X., Hu, K., Huang, H., Liu, H., Huang, H., 2023. Recent advances in plant disease severity assessment using convolutional neural networks. *Scientific Reports* 13, 2336. https://doi.org/10.1038/s41598-023-29230-7

[8] Gopalan, K., Srinivasan, S., Agarwal, P., Singh, M., Mathivanan, S.K., Moorthy, U., 2025. Corn leaf disease diagnosis: enhancing accuracy with resnet152 and grad-cam for explainable AI. *BMC Plant Biology*. https://doi.org/10.1186/s12870-025-06386-0

[9] 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, 336–359. https://doi.org/10.1007/s11263-019-01228-7

[10] Ali, H., Shifa, N., Benlamri, R., Farooque, A.A., Yaqub, R., 2025. A fine-tuned EfficientNet-B0 convolutional neural network for accurate and efficient classification of apple leaf diseases. *Scientific Reports* 15, 25732. https://doi.org/10.1038/s41598-025-04479-2

[11] Sudhesh, K.M., Aarthi, R., Sainamole Kurian, P., Sikha, O.K., 2025. Interpretable multitask deep learning model for detecting and analyzing severity of rice bacterial leaf blight. *Scientific Reports* 15, 27313. https://doi.org/10.1038/s41598-025-12276-0