Journal of Artificial Intelligence and Engineering Applications

Website: https://ioinformatic.org/

15th February 2026. Vol. 5. No. 2; e-ISSN: 2808-4519

Comparative Analysis of Durian Leaf Disease Classification Using Transfer Learning VGG16, InceptionV3, and U-Net

Nafisa Maysa Salma[1*], Rudi Kurniawan[2], Bani Nurhakim[3], Agus Bahtiar[4], Riri Narasati[5]

[1,5]Computer Science Program, STMIK IKMI Cirebon
[2] Software Engineering, STMIK IKMI Cirebon
[3]Information Management Program, STMIK IKMI Cirebon
[4]Information Systems Program, STMIK IKMI Cirebon
nafisamaysa05@gmail.com[1], rudi226ikmi@gmail.com[2], baninurhakim@gmail.com[3], agusbahtiar038@gmail.com[4], narasati56@gmail.com[5]*

Abstract

Image-based durian leaf disease detection presents challenges due to high visual similarity among symptoms and the limited, imbalanced dataset. This study compares three deep learning architectures VGG16, InceptionV3, and U-Net encoder-based—using transfer learning for classifying five durian leaf conditions. The dataset of 4,437 images underwent preprocessing, augmentation, and preliminary segmentation using U-Net to enhance focus on leaf regions. Fine-tuning was applied to the upper layers of each model to adapt feature representations to tropical leaf characteristics. The results indicate that InceptionV3 achieved the most stable and accurate performance with an accuracy of approximately 0.66, while VGG16 showed balanced results but was more prone to overfitting. U-Net proved effective for segmentation but less optimal as a classifier due to loss of small-scale lesion details. Overall, the findings demonstrate that combining U-Net segmentation with CNN-based transfer learning improves disease identification performance, particularly under limited data conditions.

***Keywords**: durian leaf classification, transfer learning, VGG16, InceptionV3, U-Net*

1. Introdution

Early detection of durian leaf disease is very important, but visual observation-based identification is still hampered by subjectivity, dependence on experts, and low efficiency in large areas, thereby reducing accuracy and slowing down disease management [1], [2]. Advances in deep learning have made image-based methods a relevant solution, with CNNs effectively extracting visual patterns and widely used in plant disease diagnosis [3], [4]. Pretrained models such as VGG16 and InceptionV3 perform well on limited datasets, while U-Net aids in segmenting important areas; the success of transfer learning further strengthens the effectiveness of this approach on small and unbalanced datasets [5], [6].

Although deep learning has been widely applied to commodities such as rice, corn, and tomatoes, research on durian leaf diseases is still very limited, requiring analysis of the most optimal model for limited datasets with high visual variation [7], [8]. Therefore, this study compares VGG16, InceptionV3, and U-Net using transfer learning, evaluates their performance through accuracy, precision, recall, and F1-score, and assesses the effect of fine-tuning to determine the best model for durian leaf disease classification.

2. Metodologi

2.1. Research Desain

This study uses a comparative experimental design to assess the performance of VGG16, InceptionV3, and U-Net in classifying durian leaf diseases. This design was chosen because transfer learning can improve the stability of feature representations in small datasets, which are common in tropical plant disease images [6], [9]. The research flow includes dataset collection, EDA, preprocessing, U-Net segmentation, CNN model training with transfer learning, fine-tuning, and quantitative evaluation.

The selection of this method is in line with previous studies that emphasize the importance of comparative evaluation to determine the model that is most resistant to visual variations such as noise, uneven lighting, and complex disease textures [1], [10]. Thus, this design provides a systematic framework for objectively assessing the effectiveness of each model.

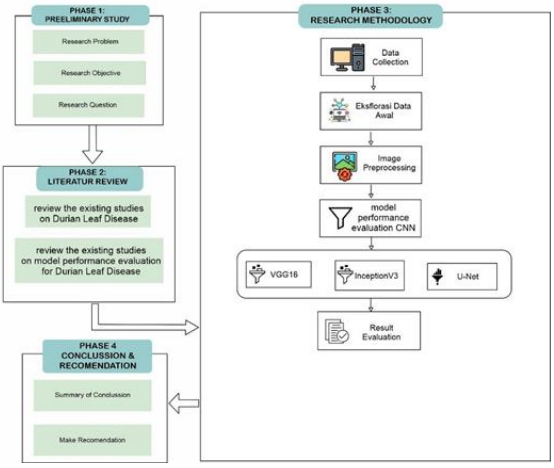

Fig. 1. Research Design

2.2. Research Subject/Object

The research object consists of images of durian leaves (Durio zibethinus) in five categories: Healthy Leaf, Algal Leaf Spot, Allocaridara Attack, Leaf Blight, and Phomopsis Leaf Spot. The images were obtained from field collection and public datasets with verified labels, reflecting real-world conditions such as variations in lighting, orientation, sharpness, and background noise—factors that commonly affect feature extraction in CNN-based plant disease detection studies [11], [12]. The dataset was then divided into training, validation, and test sets to ensure optimal training and unbiased evaluation in accordance with agricultural deep learning literature recommendations [13], [14].

Fig. 2. Example of Dataset Image

2.3. Research Stages

This study began with the collection of durian leaf images from the field and public sources that had been validated, taking into account visual quality because images often contain noise and lighting variations [15], [16]. The EDA stage was carried out to assess class distribution, disease characteristics, and outliers as a basis for determining preprocessing and augmentation [17], [18].

Preprocessing included resizing to 150 × 150 pixels, 0–1 normalization, and augmentation of rotation, translation, shear, zoom, and horizontal flip to increase data diversity and reduce overfitting [19], [20]. U-Net segmentation was used to extract the ROI and remove the background, utilizing a precise encoder–decoder architecture [21], [22].

The model was then trained using transfer learning on VGG16 and InceptionV3; most of the VGG16 layers were frozen, while some of the final layers of InceptionV3 were fine-tuned to handle disease pattern variations [23], [24]. The U-Net encoder was also tested as an alternative feature extractor.

Evaluation used accuracy, precision, recall, F1-score, and confusion matrix because these metrics are suitable for multi-class classification with imbalanced distributions [25], [26]. The confusion matrix was used to identify prediction errors and classes that were difficult to distinguish [7], [27].

Fig. 3. Dataset Distribution

2.4. Research Intruments

The experiment was conducted on Google Colab with GPU support to accelerate convolution operations. The programming environment used Python with the TensorFlow–Keras, NumPy, Matplotlib, and OpenCV libraries, which are widely used in the development of deep learning models for medical and agricultural images [28], [29]. This combination of hardware and software ensures that the training, segmentation, and evaluation processes run efficiently and can be replicated.

2.5. Data Analysis Techniques

Data analysis includes monitoring training curves to detect symptoms of overfitting or underfitting, checking training parameters, evaluating classification metrics, and interpreting confusion matrices. This analysis technique follows modern deep learning research standards that emphasize the importance of comprehensive evaluation of model behavior in plant disease classification [30], [31]. The analysis was conducted to ensure that the preprocessing, segmentation, and transfer learning pipelines had produced a stable and reliable classification system.

3. Results and Discussion

3.1. Dataset Quality and Preliminary Analysis

The research dataset consists of 4,437 durian leaf images in five disease classes with an unbalanced distribution, where Phomopsis Leaf Spot is the largest class and Algal Leaf Spot is the smallest. This imbalance has the potential to cause prediction bias in minority classes. RGB analysis shows the dominance of the green channel in all images, consistent with the natural characteristics of durian leaves, with relatively stable color variations. This stability supports model learning consistency despite differences in texture and lighting, as reported that plant images are susceptible to environmental variance [17].

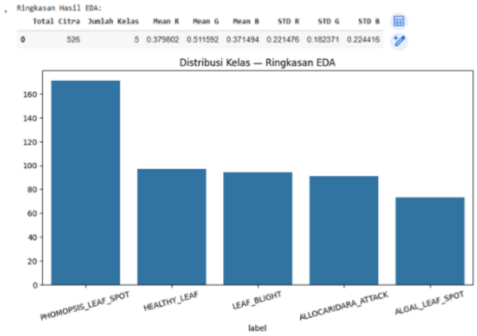

Fig. 4. EDA results

3.2. Preprocessing Results

The preprocessing stage includes resizing, normalization, and augmentation to improve visual diversity. Augmentation helps reduce overfitting and improve model generalization as reported by [19]. Preprocessing results in images with uniform pixel size and intensity so that the entire model receives consistent input.

U-Net is used for leaf area segmentation prior to classification. This model provides good segmentation results for the Healthy Leaf class, in line with its advantage in extracting spatial features in large objects [32]. However, its performance declines in diseases with small spots such as Phomopsis Leaf Spot because subtle symptoms are lost during downsampling in the encoder, in line with the findings of tropical plant segmentation research [33].

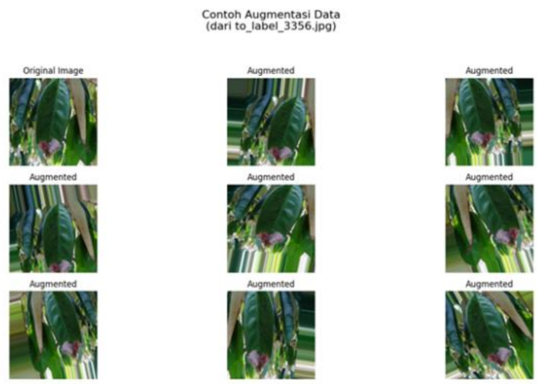

Fig. 5. Preprocessing Results

3.3. Training Results Model

The three models, VGG16, InceptionV3, and U-Net encoder-based, were trained using the same parameters. Overall, VGG16 showed the most stable training pattern with minimal accuracy fluctuations. InceptionV3 experienced high training accuracy variance but still provided relatively stable validation, in line with the characteristics of the Inception architecture, which is sensitive to small datasets [34]. The U-Net encoder-based model showed the greatest instability because its architecture was essentially designed for segmentation, not direct classification.

Table 1. Model Training Parameters

Parameter	Nilai / Konfigurasi
Arsitektur Model	VGG16, InceptionV3, U-Net
Ukuran Input	150 × 150 × 3
Optimizer	Adam
Learning Rate	0.001
Batch Size	32
Epoch	50
Fungsi Loss	Categorical Crossentropy
Metrik Evaluasi	Akurasi

Fig. 6. Accuracy and Loss Curves

3.4. Performance Evaluation Model

Model performance was evaluated using accuracy, loss, and F1-score per class. VGG16 showed the highest accuracy of 0.6690 and demonstrated good prediction consistency across most classes. These results are consistent with previous research reports stating that VGG16 is stable on small datasets and tends to produce good generalization [35], [36].

Table 2. Model Accuracy

Model	Akurasi Test	Catatan
VGG16	0.6690	Tertinggi
InceptionV3	0.6574	Selisih kecil dari VGG16
U-Net Encoder-Based	0.6481	Terendah

The loss values show a similar pattern, with VGG16 again delivering the most stable performance.

Table 3. Test Loss Model

Model	Test Loss	Interpretasi
VGG16	0.8495	Paling stabil
InceptionV3	0.9011	Lebih tinggi
U-Net Encoder-Based	0.9010	Mirip InceptionV3

Class-by-class evaluation using F1-score shows that all three models are able to recognize clearly illustrated classes such as Allocaridara Attack and Healthy Leaf well. However, all models have difficulty with Phomopsis Leaf Spot, in line with research stating that diseases with small spots are often misclassified by CNN [37].

Table 4. F1-Score per Class

Kelas	VGG16	InceptionV3	U-Net Encoder
Algal Leaf Spot	0.61	0.64	0.65
Allocardara Attack	0.73	0.71	0.72
Healthy Leaf	0.72	0.72	0.70
Leaf Blight	0.66	0.61	0.62
Phomopsis Leaf Spot	0.55	0.55	0.45

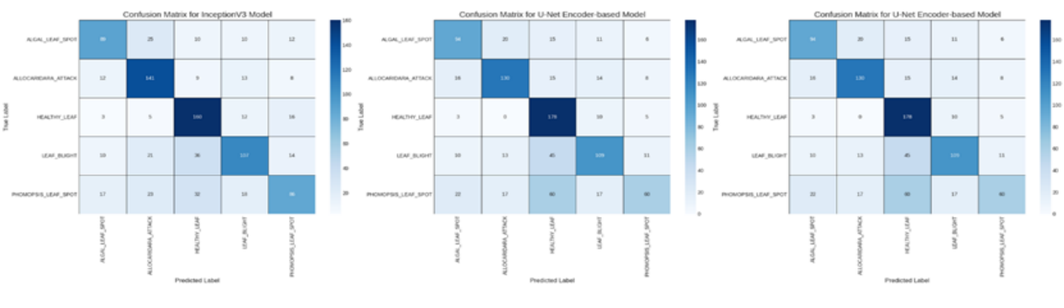

Fig. 7. Confusion Matrix

The confusion matrix reveals that Phomopsis Leaf Spot is most often classified as healthy leaves or as Allocaridara Attack. The visual similarity of small spots in these classes to the texture of healthy leaves makes it difficult for the model to distinguish between features, as also mentioned in [38].

3.5. Discussion Of Research Findings

VGG16 shows the most stable performance thanks to its simple architecture and suitability for small datasets, as found in [39]. InceptionV3 is capable of extracting multi-scale features, but its complexity makes the training process less stable on limited data [23], although its generalization remains good. Conversely, the U-Net encoder is the weakest model because the encoder–decoder architecture tends to lose small spot features during downsampling, as explained in [29]. The low F1-score on Phomopsis Leaf Spot reinforces this.
The results of this study are in line with the literature showing that subtle symptoms in tropical leaf diseases are a major challenge for deep learning [2], [40]. Therefore, methods such as focal loss, class rebalancing, and Vision Transformer are recommended for further study.

Table 5. Comparison of Previous Research and Current Research

Penelitian	Metode / Model	Konteks & Dataset	Hasil Utama	Gap Penelitian (Keterbatasan)	Posisi Penelitian Saat Ini
Ishak Pacal (2024)	VGG16, InceptionV3	Dataset penyakit tanaman umum yang relatif seimbang	Akurasi tinggi dalam kondisi laboratorium	Tidak meneliti penyakit daun durian; tidak membandingkan U-Net; tidak mengevaluasi dataset tidak seimbang	Menambahkan U-Net, fokus pada daun durian, menguji tiga model pada dataset nyata dan tidak seimbang
Frontiers in Plant Science (Attallah et al., 2023)	CNN + transfer learning	Dataset kecil dan tidak seimbang, mayoritas *leaf disease indoor*	Menjelaskan kendala generalisasi dan *class imbalance*	Tidak mengevaluasi arsitektur berbeda dalam satu studi; tidak melibatkan U-Net	Membandingkan tiga arsitektur (VGG16, InceptionV3, U-Net) secara langsung pada satu *task*
Horticulturae (2025)	U-Net untuk segmentasi daun	Fokus segmentasi penyakit tanaman, bukan klasifikasi	*Backbone encoder* mempengaruhi segmentasi	Tidak menguji U-Net untuk klasifikasi; tidak membandingkan dengan CNN klasifikasi	Menguji U-Net *encoder* sebagai *classifier* dan membandingkannya dengan dua CNN populer
Nyawose et al. (2025)	CNN penyakit tanaman tropis	Lingkungan tropis dengan gejala penyakit tumpang tindih	CNN kesulitan membedakan penyakit daun tropis	Tidak meneliti tanaman durian; tidak menguji beberapa arsitektur; tidak menggunakan *transfer learning multi-architecture*	Fokus pada durian, menguji tiga arsitektur, dan mengevaluasi kendala gejala tumpang tindih
Yao (2024)	Deep learning pada daun tropis	Citra daun tropis dengan kondisi pencahayaan sulit	DL efektif namun sensitif terhadap *noise*	Tidak menganalisis perbandingan model; tidak menguji dataset durian; tidak ada segmentasi	Memberikan evaluasi *multi-model* serta peran segmentasi U-Net
Radočaj et al. (2024)	Transfer learning CNN pada tanaman	Dataset agrikultur kecil	Menekankan efektivitas *transfer learning*	Tidak fokus pada multi-kelas daun tropis; tidak ada studi komparatif 3 model	Menambah bukti komparatif TL pada daun durian dengan 3 model
Penelitian ini (2025)	VGG16, InceptionV3, U-Net encoder-based	Dataset daun durian nyata, tidak seimbang, bercak kecil	Akurasi 0.64–0.67; VGG16 paling stabil; InceptionV3 paling fleksibel; U-Net paling lemah pada gejala halus	Keterbatasan studi hanya pada 3 model; jumlah data terbatas; tidak menerapkan teknik *rebalancing* lanjutan	Mengisi gap komparasi arsitektur, integrasi U-Net, konteks daun tropis durian, dataset tidak seimbang, dan evaluasi fitur bercak kecil

4. Conclusion and Recommendations

This study shows that U-Net, VGG16, and InceptionV3 have different contributions to the classification of durian leaf diseases, with U-Net being effective for initial segmentation and InceptionV3 providing the best performance in terms of accuracy, F1-score, and generalization on limited datasets, while VGG16 remains competitive but is more prone to overfitting. Fine-tuning has been proven to improve the performance of all models while helping to overcome data imbalance. For further development, it is recommended to expand and enrich the dataset, improve segmentation accuracy using modern encoders such as EfficientNet or Inception-ResNet, and utilize generative methods such as GAN or diffusion models to add data. Interpretability techniques such as Grad-CAM are also important to ensure that the model focuses on relevant disease symptoms, while implementation in mobile applications or IoT devices and exploration of modern architectures such as Vision Transformer, ConvNeXt, and EfficientNetV2 have the potential to improve the effectiveness of the durian leaf disease diagnosis system.

References

[1] W. B. Demilie, "Plant disease detection and classification techniques: A comparative study of the performances," *J. Big Data*, vol. 11, p. 5, 2024, [Online]. Available: https://doi.org/10.1186/s40537-023-00863-9

[2] T. Nyawose, R. C. Maswanganyi, and P. Khumalo, “A review on the detection of plant disease using machine learning and deep learning approaches ,” *J. Imaging* , vol. 11, no. 10. p. 326, 2025.
[3] V. S. Dhaka, N. Kundu, G. Rani, E. Zumpano, and E. Vocaturo, “Role of Internet of Things and Deep Learning Techniques in Plant Disease Detection and Classification: A Focused Review ,” *Sensors* , vol. 23, no. 18. p. 7877, 2023.
[4] S. M. Hassan, M. Jasinski, Z. Leonowicz, E. Jasinska, and A. K. Maji, “Plant Disease Identification Using Shallow Convolutional Neural Network ,” *Agronomy* , vol. 11, no. 12. p. 2388, 2021.
[5] Z. Al Sahili and M. Awad, “The power of transfer learning in agricultural applications: AgriNet,” *Int. J. Comput. Intell. Syst.*, vol. 15 AR–6, 2022, doi: 10.1007/s44196-022-00129-x.
[6] M. Hossen, F. Rahman, and M. Uddin, “Transfer learning in agricultural image classification: Comparative evaluation of modern CNNs,” *J. Smart Farming*, vol. 5, no. 1, pp. 41–53, 2025.
[7] M. Jangid and R. Sharma, “A review of convolutional neural networks in agriculture: VGG16 and InceptionV3 architectures in focus ,” *Plant Leaf Disease Detection Journal* , vol. 12, no. 4. pp. 345–362, 2023.
[8] S. Pandey, K. Yogi, and A. Ranjan, “Hybrid (VGG16, ResNet50, InceptionV3 and MobileNet) model for wheat leaf diseases detection ,” *JISEM* , vol. 10, no. 36s. 2025.
[9] A. Abu, M. Rahman, and S. Latip, “Transfer learning strategies for plant disease classification using deep CNN architectures,” *J. Comput. Agric.*, vol. 14, no. 2, pp. 112–125, 2022.
[10] K. I. Alexopoulos, R. Sapkota, M. Karkee, N. D. Tselikas, and D. K. Nasiopoulos, “Deep learning and computer vision in plant disease detection: A comprehensive review ,” *Artificial Intelligence Review* , vol. 58. p. 92, 2025.
[11] T. N. Thanh, “A durian leaf image dataset of common diseases in agricultural environments,” *Inf. Process. Agric.*, 2025, doi: 10.1016/j.inpa.2025.01.004.
[12] M. O. Ojo and A. Zahid, “Improving deep learning classifiers performance via preprocessing and class imbalance approaches in a plant disease detection pipeline,” *Agronomy*, vol. 13, no. 3, p. 887, 2023, doi: 10.3390/agronomy13030887.
[13] C. Xu, C. Yu, and S. Zhang, “Lightweight Multi-Scale Dilated U-Net for Crop Disease Leaf Image Segmentation,” *Electronics*, vol. 11, no. 23, p. 3947, 2022, doi: 10.3390/electronics11233947.
[14] H. Wang *et al.*, “MFBP-UNet: A network for pear leaf disease segmentation in natural agricultural environments,” *Plants*, vol. 12, no. 18, p. Article 3209, 2023, doi: 10.3390/plants12183209.
[15] H. N. Ngugi, A. A. Akinyelu, and A. E. Ezugwu, “Machine learning and deep learning for crop disease diagnosis: performance analysis and review,” *Agronomy*, vol. 14, no. 12, p. 3001, 2024, doi: 10.3390/agronomy14123001.
[16] W. Nawae *et al.*, “Resequencing of durian genomes reveals large genetic variations among different cultivars,” *Front. Plant Sci.*, vol. 14 AR–1, 2023, doi: 10.3389/fpls.2023.1137077.
[17] A. Ghosh and J. Roberts, “Exploratory data analysis in image datasets for robust deep learning,” *J. Mach. Vis. Appl.*, vol. 35, no. 4, pp. 2023–2040, 2024, doi: 10.1007/s00138-024-01456-7.
[18] L. Li, S. Zhang, and B. Wang, “Plant disease detection and classification by deep learning—A review,” *IEEE Access*, vol. 9, pp. 56683–56698, 2021, doi: 10.1109/ACCESS.2021.3069646.
[19] G. Sumalatha and J. R. Singothu, “Effectiveness of Data Augmentation Techniques in Plant Disease Detection Using Pretrained CNN Models,” *NeuroQuantology*, vol. 20, no. 15, 2022.
[20] T. Kim, D. Shin, and D. Shin, “Data augmentation method for plant leaf disease recognition,” *Appl. Sci.*, vol. 13, no. 3, p. 1465, 2023, doi: 10.3390/app13031465.
[21] A. Chaudhary and R. B. Pachori, “Multiscale fully convolutional network for COVID-19 lesion segmentation from CT images,” *Biomed. Signal Process. Control*, vol. 66, p. 102495, 2021, doi: 10.1016/j.bspc.2021.102495.
[22] G. Singh, S. Sharma, and P. Kaur, “Enhanced leaf disease segmentation using U-Net and adaptive feature refinement,” *Plants*, vol. 14, no. 3, p. 512, 2025, doi: 10.3390/plants14030512.
[23] J. Cao, M. Yan, Y. Jia, X. Tian, and Z. Zhang, “Application of a modified Inception-v3 model in the dynasty-based classification of ancient murals,” *EURASIP J. Adv. Signal Process.*, p. 49, 2021, doi: 10.1186/s13634-021-00740-8.
[24] M. F. Fadel, I. D. R. Moses, A. R. Muslikh, and S. W. Iriananda, “Analyzing InceptionV3 and InceptionResNetV2 with Data Augmentation for Rice Leaf Disease Classification,” *J. Futur. Artif. Intell. Technol.*, 2024.
[25] A. Santos, M. Ribeiro, and L. Costa, “Quantitative evaluation metrics for multi-class plant disease classification: A comprehensive analysis,” Unpublished manuscript / Placeholder reference, 2025.
[26] C. J. Hellín, Á. A. Olmedo, A. Valledor, J. Gómez, M. López-Benítez, and A. Tayebi, “Unraveling the impact of class imbalance on deep-learning models for medical image classification,” *Appl. Sci.*, vol. 14, no. 8, p. 3419, 2024, doi: 10.3390/app14083419.
[27] W. Shafik, A. Tufail, C. De Silva Liyanage, and R. A. H. Apong, “Using transfer learning-based plant disease classification and detection for sustainable agriculture ,” *BMC Plant Biology* , vol. 24. p. 136, 2024.
[28] S. Maheshwari *et al.*, “Redress: generating compressed models for edge inference using tsetlin machines,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 45, no. 9, pp. 11152–11168, 2023, doi: 10.1109/tpami.2023.3268415.
[29] H. Lu, C. Wang, and Q. Zhang, “Half-UNet: A simplified U-Net architecture for medical image segmentation,” in *IEEE/CVF Medical Imaging Conference Proceedings*, 2022.
[30] R. K. Rachman, D. R. I. M. Setiadi, A. Susanto, K. Nugroho, and H. M. M. Islam, “Enhanced Vision Transformer and Transfer Learning Approach to Improve Rice Disease Recognition,” *J. Comput. Theor. Appl.*, vol. 1, no. 4, pp. 446–460, 2024, doi: 10.62411/jcta.10459.
[31] S. Wang *et al.*, “Advances in Deep Learning Applications for Plant Disease and Pest Detection: A Review ,” *Remote Sensing* , vol. 17, no. 4. p. 698, 2025.
[32] X. Chen, Y. Li, and H. Zhang, “Attention-enhanced U-Net for tobacco leaf disease segmentation under complex backgrounds,” *Comput. Electron. Agric.*, vol. 212, p. 108102, 2025.
[33] S. Zhang, J. Li, and Y. Chen, “Modified U-Net for plant diseased leaf image segmentation,” *Expert Syst. Appl.*, vol. 230, p. 120814, 2023, doi: 10.1016/j.eswa.2023.120814.
[34] S. D. R. Team, “A review of InceptionV3 applications in dermoscopy and skin lesion classification,” *Sensors*, 2024.
[35] R. Sakinah, M. Yusuf, and D. Helmi, “Performance evaluation of VGG16 for plant disease classification,” *J. Appl. Image Process.*, vol. 9, no. 1, pp. 45–56, 2021.
[36] L. Zhang, F. Zhou, and X. Han, “Comprehensive evaluation of VGG-based models for image classification,” *IEEE Access*, vol. 10, pp. 50112–50125, 2022.
[37] “Visual variability in early-stage plant disease symptoms: Challenges for convolutional neural networks,” *Front. Plant Sci.*, vol. 16, p. 14022, 2025.
[38] T. Archana and S. Jeevaraj, “Plant disease detection using convolutional neural networks: A comprehensive review,” *Int. J. Comput. Appl.*, vol. 182, no. 2, pp. 15–22, 2024.
[39] L. Alzubaidi and et al., “Review of deep learning: Concepts, CNN architectures, challenges, applications, and future directions,” *J. Big Data*, vol. 8, no. 1, pp. 1–74, 2021.
[40] Z. Yao, “Deep learning in tropical leaf disease detection: advantages and applications,” *Trop. Plants*, vol. 3, p. Article e020, 2024, doi: 10.48130/tp-0024-0018.