International Journal of Computer Information Systems and Industrial Management Applications
ISSN 2150-7988 Volume 18(1s) 2026 pp. 1-8
https://doi.org/10.70917/ijcisim-2026-2111
Article

Swin Transformer algorithm for Multiple Disease detection on Paddy Crops: A hierarchical Vision Method with shifted window

Ambrayya*[1], William Thomas H M[2], S Prabhavathi[3], Sadyojatha K M[4]

[1*]Department of Electronics and Communication Engineering, Ballari Institute of Technology and Management, Ballari, Affiliated to Visvesvaraya Technological University, Belagavi 590018, India. Orcid ID: [0000-0003-0694-4662]. ambrayyabs@gmail.com
[2]Department of Electronics and Communication Engineering, Ballari Institute of Technology and Management, Ballari, Affiliated to Visvesvaraya Technological University, Belagavi 590018, India. Orcid ID: [0000-0002-4880-8485]. williamthomas.hm@gmail.com
[3]Department of Electronics and Communication Engineering, Rao Bahadur Y. Mahabaleshwarappa Engineering College, Ballari, Affiliated to Visvesvaraya Technological University, Belagavi-590018, Karnataka, India. Orcid ID: [0000-0002-9787-0242]. prabkir@gmail.com
[4]Department of Electronics and Communication Engineering, Ballari Institute of Technology and Management, Ballari, Affiliated to Visvesvaraya Technological University, Belagavi 590018, India. Orcid ID: [0000-0002-7736-5601]. saddukm@gmail.com

Abstract: The health monitoring of paddy is a significant research issue since outbreaks of diseases may decrease the yield, quality of grains, and the profitability of the farms. The process of field diagnosis is time-consuming and can be inaccurate at times, which drives automated image-based recognition systems. This paper explores a hierarchical vision architecture that is founded on the Swin Transformer to detect multiple paddy health-classes with shifted-window attention. The experiments were done on a publicly available 10-class paddy leaf image dataset with 10,407 labeled images that were stratified into 7,280 training images, 1,557 validation images and 1,570 test images. To overcome the issue of class imbalance and enhance generalization, a transfer-learning setup of Swin-Tiny was trained using weighted cross-entropy loss, AdamW optimization, and data augmentation. The final model had the best validation accuracy of 96.98% and the final test accuracy of 96.50%. The model also achieved weighted precision, weighted recall, and weighted F1-score values of 96.56, 96.50, and 96.51, respectively, and a macro F1-score of 96.22. Analysis by classes indicated very high performance in dead heart, bacterial panicle blight, hispa and normal classes whereas downy mildew was the most difficult category. The results suggest that a hierarchical vision approach with Swin Transformer can offer a high level of competitiveness in multi-class paddy leaf health recognition.
Keywords: paddy disease classification; Swin Transformer; shifted-window attention; hierarchical vision; rice leaf image analysis; deep learning.

1. Introduction

Rice is a significant staple crop in the world, and the loss in yield due to diseases continues to be a significant issue to food security and agricultural sustainability (Fukagawa and Ziska, 2019; Seelwal et al., 2024). Visual inspection is commonly used in disease recognition in routine practice, and is labor-intensive, subjective, and cannot be easily scaled to large cultivation areas. These constraints have contributed to the growing popularity of machine learning and deep learning in the diagnosis of crops by images (Mohanty et al., 2016; Ferentinos, 2018; Saleem et al., 2019).

In the rice domain, both conventional image-processing pipelines and state-of-the-art deep neural networks have been investigated in the past. Classical work involved the use of segmentation, handcrafted features, and support vector machines in the classification of rice disease (Prajapati et al., 2017). Subsequent research embraced convolutional neural networks, transfer learning, and detection structures to paddy disease detection in the field (Rahman et al., 2020; Sethy et al., 2020; Bari et al., 2021; Wang et al., 2021). Recent reviews, however, observe that there are still persistent issues with class imbalance, field variability, subtle symptom boundaries, as well as generalization across acquisition conditions (Seelwal et al., 2024; Simhadri et al., 2025; Isma'il et al., 2025).

Transformer-based vision architectures are an exciting alternative since they are able to represent long-range dependencies more explicitly than traditional convolutional networks. Vision Transformer proposed the patch-based paradigm of transformer-based image recognition (Dosovitskiy et al., 2021), and Swin Transformer expanded this concept with a hierarchical architecture with shifted-window self-attention, which allows the learning of multi-scale features efficiently (Liu et al., 2021). There are also studies of rice-specific transformers that have demonstrated competitive results, indicating that transformer representations are well adapted to the analysis of paddy leaf symptoms (Zhou et al., 2023; Yang et al., 2023).

It is against this backdrop that the current paper discusses the title-fixed framework, Swin Transformer algorithm Multiple Disease detection on Paddy Crops A hierarchical Vision Method with shifted window, on a publicly available benchmark of paddy leaf health classes. Even though the benchmark used incorporates disease, pest, symptom, and normal classes, it is a realistic multi-class paddy health recognition environment that has been extensively utilized in the literature and in real-world benchmarking processes (Kaggle, 2022; Petchiammal et al., 2023). The main assumption is that shifted-window attention is able to maintain the detail of local lesions, whilst being able to model contextual relationships between adjacent image areas.

1.1 Research objectives

- To create a well-organized paddy leaf image data that can be used in multi-class machine-learning-based health recognition.
- To implement image pre-processing and augmentation methods to enhance generalization and minimize the effects of class imbalance.
- To apply Swin Transformer model as a hierarchical vision approach with shifted-window attention to predict paddy classes.
- To assess the obtained model in terms of overall and class-based performance measures and examine its advantages and disadvantages.

2. Related Work

Plant disease recognition on a large scale took off with the publication of open image datasets like PlantVillage, which allowed experiments with deep learning on a large scale (Hughes and Salathé, 2015; Mohanty et al., 2016). AlexNet, ResNet, MobileNet, DenseNet and EfficientNet CNNs later evolved into standard baselines in agricultural vision problems (Krizhevsky et al., 2012; He et al., 2016; Howard et al., 2017; Tan and Le, 2019).

Specific studies of rice have developed beyond traditional classification to real-time detection and transfer-learning pipelines. Convolutional neural networks were used to detect rice diseases and pests in field images, deep features with support vector machines, and Faster R-CNN were used to diagnose rice leaf diseases in real-time, respectively (Rahman et al., 2020; Sethy et al., 2020; Bari et al., 2021). Other works added attention mechanisms, dense connectivity, or transfer learning to enhance robustness (Wang et al., 2021; Feng et al., 2021; Jiang et al., 2023; Simhadri and Kondaveeti, 2023; Yuan et al., 2024).

The quality of the data sets is also crucial. The Paddy Doctor resource that was reported by Petchiammal et al. (2023) serves as a good reference point in the automated paddy disease classification and demonstrated the competitiveness of deep residual models on paddy images. The more recent reviews note that the accuracy is not the only aspect of the rice disease research that needs to be improved in the future, but also robustness, fair evaluation, explainability, and deployment realism (Seelwal et al., 2024; Simhadri et al., 2025; Isma'il et al., 2025).

The Swin Transformer is a more suitable model to image recognition tasks than the standard ViT, which needs both local detail and hierarchical representation learning. Its shifted-window mechanism decreases the computational cost and enables information exchange between adjacent windows, which is beneficial to paddy symptoms which can be small, diffuse, or spatially fragmented (Liu et al., 2021). Zhou et al. (2023) proved the usefulness of transformer-based rice disease detection, which encouraged the current analysis of a Swin-based model on a publicly available paddy leaf dataset.

3. Materials and Methods

3.1 Dataset

The Kaggle paddy disease classification dataset (publicly available) was used in the experiments and consists of 10,407 labeled paddy leaf images belonging to 10 categories: bacterial_leaf_blight, bacterial_leaf_streak, bacterial_panicle_blight, blast, brown_spot, dead_heart, downy_mildew, hispa, normal, and tungro (Kaggle The competition release also contains an unlabeled hold-out test set but in the current study only the labeled image set was used and a stratified 70:15:15 split was made to train, validate and test. This generated 7,280 training images, 1,557 validation images and 1,570 test images. Since the benchmark is class-imbalanced, weighted loss was used during training. The summary of the class distribution in the experiments is presented in Figure 1 and Table 1.

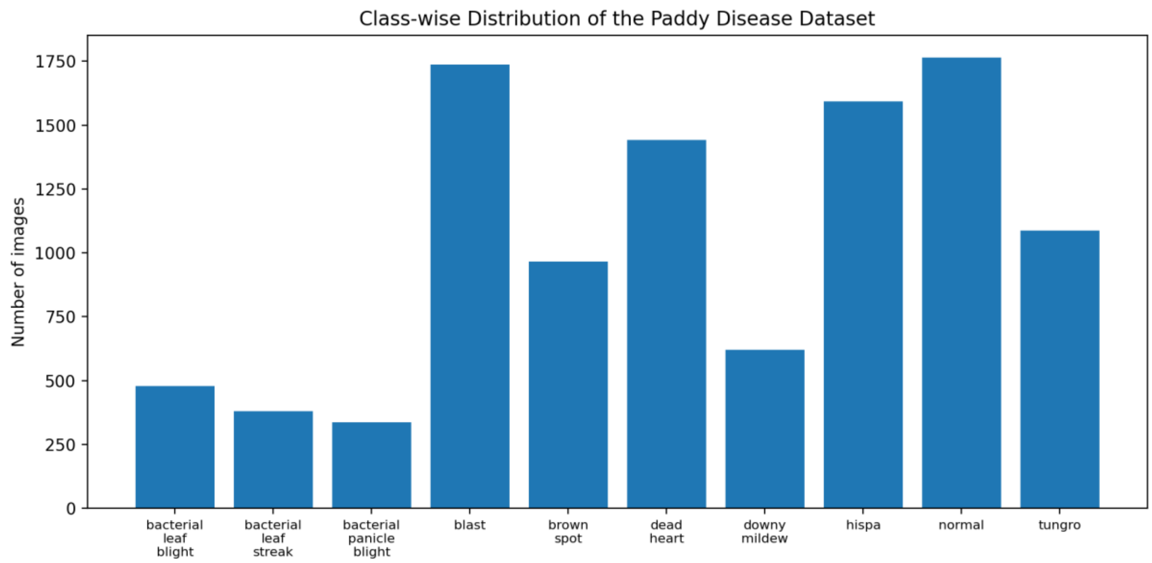

Figure 1. Class-wise distribution of the 10-class paddy health dataset used in the experiments.

Table 1. Stratified dataset split by class.

Class	Train	Validation	Test	Total
Bacterial Leaf Blight	335	71	73	479
Bacterial Leaf Streak	266	57	57	380
Bacterial Panicle Blight	235	50	52	337
Blast	1216	260	262	1738
Brown Spot	675	144	146	965
Dead Heart	1009	216	217	1442
Downy Mildew	434	93	93	620
Hispa	1115	239	240	1594
Normal	1234	264	266	1764
Tungro	761	163	164	1088
Total	7280	1557	1570	10407

3.2 Pre-processing and augmentation

All images were down sampled to 224 × 224 pixels and normalized with ImageNet channel statistics and then fed into the model. In training, data augmentation involved random horizontal flipping, random rotation, and color jittering to change the brightness, contrast, and saturation. This augmentation policy was meant to enhance model robustness to variability of field-level imaging without changing the class semantics.

3.3 Proposed Swin Transformer framework

The suggested classifier employs Swin-Tiny, a hierarchical vision transformer that divides the input image into non-overlapping patches and processes them in a series of Swin stages (Liu et al., 2021). In every stage, window-based multi-head self-attention (W-MSA) is preceded by shifted-window multi-head self-attention (SW-MSA), which allows information to be exchanged across windows at a computational cost manageable. Patch-merging layers lower the spatial resolution and enhance representational depth, thus aiding multi-scale feature learning. In the current experiments, a Swin-Tiny backbone, which was pretrained on ImageNet, was applied in a transfer-learning setup with a trainable classification head. This design maintains the title-prescribed hierarchical shifted-window structure and is still feasible to classify paddy according to the public-benchmark.

3.4 Training configuration

The 15 epochs of training were done with AdamW and learning rate of 1 x 10 -4 and a StepLR scheduler with step size 5 and gamma 0.5. Class-weighted cross-entropy loss was employed to reduce the imbalance in classes. The most successful model was chosen based on the accuracy of the validation and the reported final results were calculated on the independent test split.

Table 2. Main training configuration used in the final experiment.

Parameter	Setting
Model	Swin-Tiny (swin_tiny_patch4_window7_224)
Input resolution	224 × 224
Number of classes	10
Training epochs	15
Optimizer	AdamW
Initial learning rate	1 × 10^-4
Learning-rate scheduler	StepLR (step size = 5, gamma = 0.5)
Loss function	Class-weighted cross-entropy
Data augmentation	Horizontal flip, random rotation, color jitter
Model selection	Best validation accuracy

3.5 Evaluation metrics

The performance was measured using accuracy, weighted precision, weighted recall, weighted F1-score, macro precision, macro recall, macro F1-score and the confusion matrix. Weighted metrics are based on the prevalence of classes, but macro metrics consider all classes in the same way and thus are more accurate in showing whether minority classes are being overlooked.

4. Results and Discussion

4.1 Overall quantitative performance

The last Swin Transformer model had a highest validation accuracy of 96.98%. It achieved 96.50% accuracy, 96.56% weighted precision, 96.50% weighted recall, and 96.51% weighted F1-score on the independent test set. The macro precision, macro recall, and macro F1-score were 96.29, 96.21 and 96.22 respectively. These values show that the model not only did well in general, but also behaved equally well across classes.

Table 3. Overall performance of the proposed Swin Transformer model.

Metric	Value
Best validation accuracy	96.98%
Test accuracy	96.50%
Weighted precision	96.56%
Weighted recall	96.50%
Weighted F1-score	96.51%
Macro precision	96.29%
Macro recall	96.21%
Macro F1-score	96.22%

4.2 Class-wise behavior

The performance in terms of classes was also high in nearly all categories. Dead heart was identified with a perfect score of 1 and bacterial panicle blight scored F1-score of 0.99. Bacterial leaf streak, hispa, normal and tungro also scored strong F1-scores. The class-wise F1-score was the smallest in the case of downy mildew (0.89), indicating that this class is still more visually ambiguous and confusing with blast or other stress-related patterns. Brown spot also had slightly low recall compared to the strongest classes which imply that minority error modes are still present.

Table 4. Class-wise precision, recall, F1-score, and support on the test set.

Class	Precision	Recall	F1-score	Support
Bacterial Leaf Blight	0.95	0.97	0.96	73
Bacterial Leaf Streak	0.98	0.96	0.97	57
Bacterial Panicle Blight	1.00	0.98	0.99	52
Blast	0.95	0.96	0.95	262
Brown Spot	0.99	0.93	0.96	146
Dead Heart	1.00	1.00	1.00	217
Downy Mildew	0.86	0.91	0.89	93
Hispa	0.98	0.96	0.97	240
Normal	0.97	0.98	0.97	266
Tungro	0.96	0.95	0.96	164

4.3 Training dynamics

The learning curves affirm stable convergence. The accuracy of validation increased quickly in the initial half of training, and reached over 94% in epoch 6 and 96.98 percent in epoch 14. The training loss was steadily reducing, with a loss of 1.3700 in the initial epoch, and 0.0356 in the fifteenth epoch, and validation loss was low in the subsequent epochs. The minimal difference between the training and validation performance shows that the regularization and augmentation strategy that was adopted was effective.

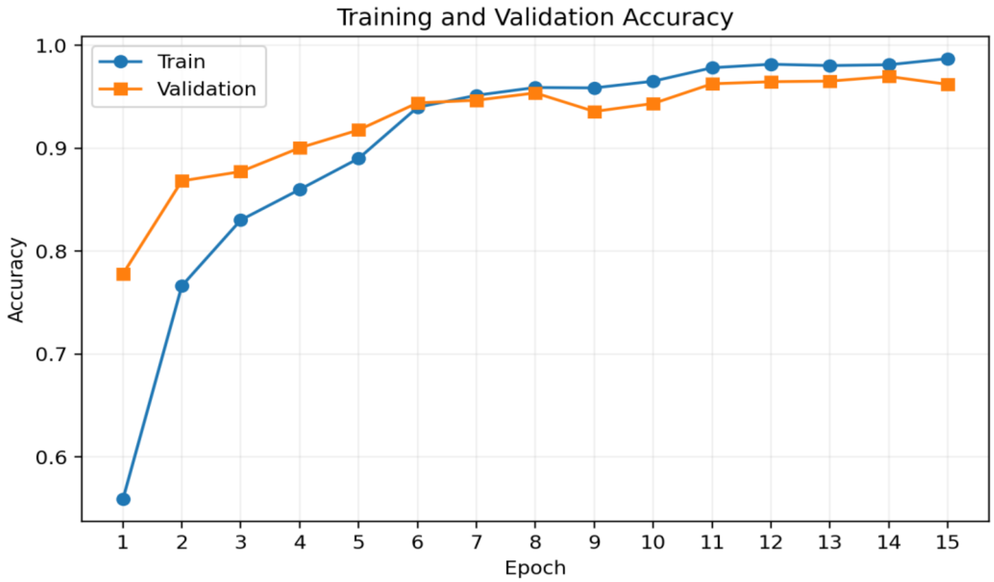

Figure 2. Training and validation accuracy over 15 epochs.

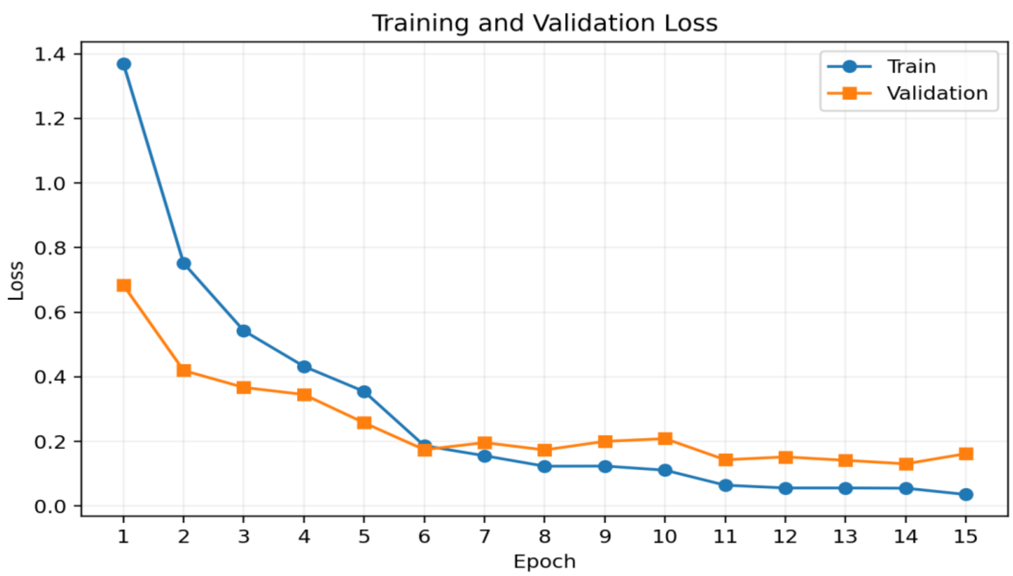

Figure 3. Training and validation loss over 15 epochs.

4.4 Confusion-matrix analysis and discussion

The confusion matrix indicates that most of the predictions were along the diagonal, which supports a high level of separability of the 10 classes. Dead heart was flawlessly classified. The largest residual errors were in downy mildew which was sometimes predicted as blast and in a few cases tungro and hispa images which were predicted as neighboring classes. The plausibility of these patterns is that there are overlapping texture, discoloration, or lesion-distribution patterns in some of the leaf symptoms. Overall, the confusion profile confirms that the shifted-window attention can extract fine-grained local evidence and extended contextual relationships required to recognize the paddy leaf. The attained performance is also competitive to the larger scope of rice disease recognition studies documented in the literature, but direct numerical comparison should be viewed with caution when datasets, splits, and protocols are dissimilar (Petchiammal et al., 2023; Zhou et al., 2023; Simhadri et al., 2025).

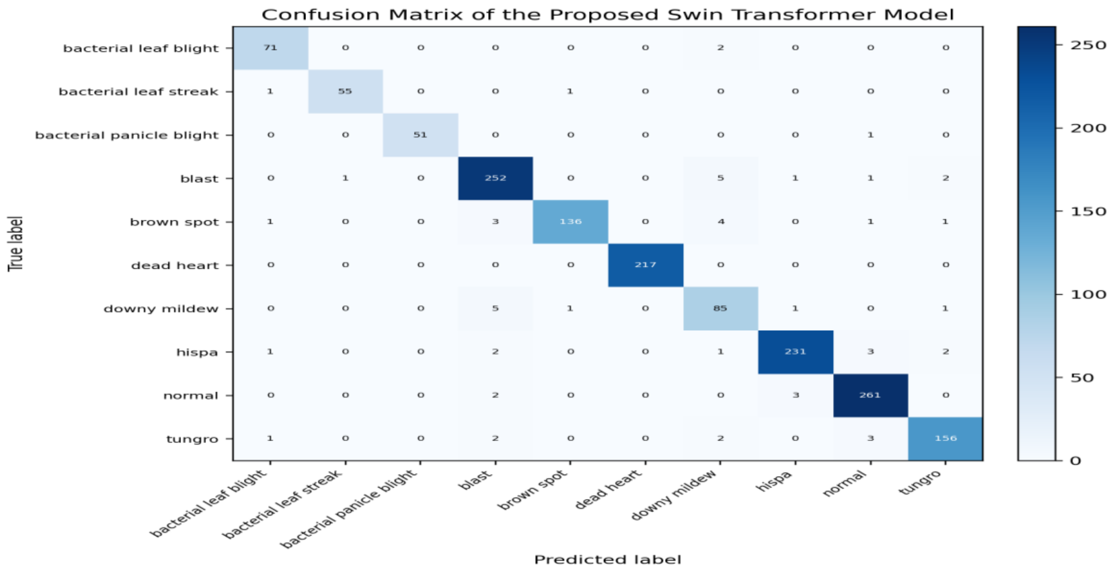

Figure 4. Confusion matrix of the final Swin Transformer model on the test set.

4.5 Limitations

Several limitations must be admitted. To begin with, the experiment adopted a leaf-centric community benchmark as opposed to a multi-organ field experiment, therefore the results can be viewed as paddy leaf health recognition and not a system-wide diagnosis. Second, the benchmark incorporates disease, pest, symptom, and normal classes within a single multi-class protocol, which is realistic in operational screening but not pathology-only recognition. Third, the present manuscript presents a high-quality Swin Transformer experiment, but is not yet a complete in-house comparison of baseline to a variety of CNN and ViT models with the same split. Future work can consider these points with a wider data collection, cross-dataset analysis, model comparison, and explainability analysis.

5. Conclusion

In this work, a hierarchical vision system based on Swin Transformer and shifted-window attention was applied and tested to classify paddy health on a public-benchmark. The final model, using a 10-class stratified split of 10,407 labeled paddy leaf images, had a best validation and test accuracy of 96.98 and 96.50 respectively, along with 96.51 and 96.22 weighted F1-score and macro F1-score respectively. The findings show that hierarchical shifted-window representation learning is quite effective in paddy leaf recognition and can be used as a solid base to future and more robust paddy disease detection systems. Future studies are encouraged to generalize the methodology to multi-organ imagery, external validation, interpretability analysis and direct comparison with other high-performing baselines.

References

1. Bari, B. S., Islam, M. N., Rashid, M., Hasan, M. J., Razman, M. A. M., Musa, R. M., Nasir, A. F. A., & Majeed, A. P. P. A. (2021). A real-time approach of diagnosing rice leaf disease using deep learning-based Faster R-CNN framework. PeerJ Computer Science, 7, e432. https://doi.org/10.7717/peerj-cs.432
2. Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., Dehghani, M., Minderer, M., Heigold, G., Gelly, S., Uszkoreit, J., & Houlsby, N. (2021). An image is worth 16×16 words: Transformers for image recognition at scale. International Conference on Learning Representations.
3. Feng, L., Wu, B., He, Y., & Zhang, C. (2021). Hyperspectral imaging combined with deep transfer learning for rice disease detection. Frontiers in Plant Science, 12, 693521. https://doi.org/10.3389/fpls.2021.693521
4. Ferentinos, K. P. (2018). Deep learning models for plant disease detection and diagnosis. Computers and Electronics in Agriculture, 145, 311–318. https://doi.org/10.1016/j.compag.2018.01.009
5. Fukagawa, N. K., & Ziska, L. H. (2019). Rice: Importance for global nutrition. Journal of Nutritional Science and Vitaminology, 65(Supplement), S2–S3. https://doi.org/10.3177/jnsv.65.S2
6. He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 770–778. https://doi.org/10.1109/CVPR.2016.90

7. Howard, A. G., Zhu, M., Chen, B., Kalenichenko, D., Wang, W., Weyand, T., Andreetto, M., & Adam, H. (2017). MobileNets: Efficient convolutional neural networks for mobile vision applications. arXiv:1704.04861.
8. Hughes, D. P., & Salathé, M. (2015). An open access repository of images on plant health to enable the development of mobile disease diagnostics. arXiv:1511.08060.
9. Isma'il, U. I., Chua, H. N., Nordin, R., & Ahmed, M. K. (2025). A comprehensive review of deep learning approaches for rice disease detection: Datasets, methodologies, and future directions. Smart Agricultural Technology, 11, 100976. https://doi.org/10.1016/j.atech.2025.100976
10. Jiang, M., Feng, C., Fang, X., Huang, Q., Zhang, C., & Shi, X. (2023). Rice disease identification method based on attention mechanism and deep dense network. Electronics, 12(3), 508. https://doi.org/10.3390/electronics12030508
11. Kaggle. (2022). Paddy Disease Classification. Kaggle Competitions. https://www.kaggle.com/competitions/paddy-disease-classification
12. Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). ImageNet classification with deep convolutional neural networks. Advances in Neural Information Processing Systems, 25.
13. Liu, Z., Lin, Y., Cao, Y., Hu, H., Wei, Y., Zhang, Z., Lin, S., & Guo, B. (2021). Swin Transformer: Hierarchical vision transformer using shifted windows. Proceedings of the IEEE/CVF International Conference on Computer Vision, 10012–10022. https://doi.org/10.1109/ICCV48922.2021.00986
14. Mohanty, S. P., Hughes, D. P., & Salathé, M. (2016). Using deep learning for image-based plant disease detection. Frontiers in Plant Science, 7, 1419. https://doi.org/10.3389/fpls.2016.01419
15. Ning, H., Liu, S., Zhu, Q., & Zhou, T. (2023). Convolutional neural network in rice disease recognition: Accuracy, speed and lightweight. Frontiers in Plant Science, 14, 1269371. https://doi.org/10.3389/fpls.2023.1269371
16. Petchiammal, A., Briskline Kiruba, S., Murugan, D., & Pandarasamy, A. (2023). Paddy Doctor: A visual image dataset for automated paddy disease classification and benchmarking. Proceedings of the 6th Joint International Conference on Data Science & Management of Data, 203–207. https://doi.org/10.1145/3570991.3570994
17. Prajapati, H. B., Shah, J. P., & Dabhi, V. K. (2017). Detection and classification of rice plant diseases. Intelligent Decision Technologies, 11(3), 357–373. https://doi.org/10.3233/IDT-170301
18. Rahman, C. R., Arko, P. S., Ali, M. E., Khan, M. A. I., Apon, S. H., Nowrin, F., & Wasif, A. (2020). Identification and recognition of rice diseases and pests using convolutional neural networks. Biosystems Engineering, 194, 112–120. https://doi.org/10.1016/j.biosystemseng.2020.03.020
19. Ren, S., He, K., Girshick, R., & Sun, J. (2015). Faster R-CNN: Towards real-time object detection with region proposal networks. Advances in Neural Information Processing Systems, 28.
20. Saleem, M. H., Potgieter, J., & Arif, K. M. (2019). Plant disease detection and classification by deep learning. Plants, 8(11), 468. https://doi.org/10.3390/plants8110468
21. Seelwal, P., Dhiman, P., Gulzar, Y., Kaur, A., Wadhwa, S., & Onn, C. W. (2024). A systematic review of deep learning applications for rice disease diagnosis: Current trends and future directions. Frontiers in Computer Science, 6, 1452961. https://doi.org/10.3389/fcomp.2024.1452961
22. Sethy, P. K., Barpanda, N. K., Rath, A. K., & Behera, S. K. (2020). Deep feature based rice leaf disease identification using support vector machine. Computers and Electronics in Agriculture, 175, 105527. https://doi.org/10.1016/j.compag.2020.105527
23. Simhadri, C. G., & Kondaveeti, H. K. (2023). Automatic recognition of rice leaf diseases using transfer learning. Agronomy, 13(4), 961. https://doi.org/10.3390/agronomy13040961
24. Simhadri, C. G., Kondaveeti, H. K., Vatsavayi, V. K., Mitra, A., & Ananthachari, P. (2025). Deep learning for rice leaf disease detection: A systematic literature review on emerging trends, methodologies and techniques. Information Processing in Agriculture, 12(2), 151–168. https://doi.org/10.1016/j.inpa.2024.04.006
25. Tan, M., & Le, Q. V. (2019). EfficientNet: Rethinking model scaling for convolutional neural networks. Proceedings of the 36th International Conference on Machine Learning, 6105–6114.
26. Wang, Y., Wang, H., & Peng, Z. (2021). Rice diseases detection and classification using attention based neural network and Bayesian optimization. Expert Systems with Applications, 178, 114770. https://doi.org/10.1016/j.eswa.2021.114770
27. Yang, H., Deng, X., Shen, H., Lei, Q., Zhang, S., & Liu, N. (2023). Disease detection and identification of rice leaf based on improved detection transformer. Agriculture, 13(7), 1361. https://doi.org/10.3390/agriculture13071361
28. Yuan, P., Xia, Y., Tian, Y., & Xu, H. (2024). TRiP: A transfer learning based rice disease phenotype recognition platform using SENet and microservices. Frontiers in Plant Science, 14, 1255015. https://doi.org/10.3389/fpls.2023.1255015
29. Zhou, C., Zhong, Y., Zhou, S., Song, J., & Xiang, W. (2023). Rice leaf disease identification by residual-distilled transformer. Engineering Applications of Artificial Intelligence, 121, 106020. https://doi.org/10.1016/j.engappai.2023.106020