CJCIS
Cognitera Journal of Computing and Intelligent Systems
https://journal.psg.ac.id/index.php/cjcis

Vol. 1, No. 1, 2026
e-ISSN: XXXX-XXX
pp. 27-27

Model Klasifikasi Penyakit Tanaman Padi Secara Real-Time Menggunakan *MobileNetV4* dengan *Attention Mechanism*

Ahmad Riadi[1]*, Antoni[2]
[1]Fakultas Pertanian/ Universitas Brawijaya, Indonesia
[2] Fakultas Pertanian/ Universitas Brawijaya, Indonesia

*Corresponding Author's e-mail: ahmadriadi@gmail.com

Article History:
Received: September 8, 2026
Revised: October 17, 2026
Accepted: November 23, 2026

Keywords:
MobileNetV4; Rice Disease Classification; Attention Mechanism

***Abstract**: Rice disease diagnosis in field conditions requires models that are accurate, computationally efficient, and capable of operating directly on mobile or edge devices. This study proposes a real-time rice plant disease classification model based on MobileNetV4 strengthened with an attention mechanism to improve discriminative feature selection without sacrificing deployment efficiency. The research adopts a design-science approach supported by a structured review of verified studies published between 2022 and 2025, architecture-level analysis, and deployment-oriented benchmarking criteria derived from public rice disease datasets and recent lightweight vision research. MobileNetV4 is selected because its Universal Inverted Bottleneck and mobile-oriented attention components provide a favorable accuracy-latency trade-off across heterogeneous hardware. The proposed configuration inserts a lightweight channel-spatial attention module after selected backbone stages, followed by global average pooling and a softmax classifier. Evaluation focuses on class separability, parameter efficiency, computational complexity, robustness to complex backgrounds, inference latency, and practical suitability for smartphone-based diagnosis. The analysis indicates that combining MobileNetV4 with selective attention is theoretically well aligned with rice disease symptoms that occupy small, irregular, and visually similar regions. Compared with heavier convolutional or transformer architectures reported in recent literature, the proposed model offers a more balanced pathway toward real-time diagnosis under constrained resources. The study concludes that attention-enhanced MobileNetV4 is a technically credible architecture for future empirical validation on Paddy Doctor and locally collected Indonesian rice datasets, particularly when combined with augmentation, calibration, and field-domain evaluation across diverse lighting conditions, devices, cultivars, and seasons.*

Copyright © 2026, The Author(s).
This is an open access article under the CC–BY-SA license

How to cite: Author et al. (2026). Model Klasifikasi Penyakit Tanaman Padi Secara Real-Time Menggunakan MobileNetV4 dengan Attention Mechanism. CJCIS.

INTRODUCTION

Rice is a strategic crop for global food security, yet its productivity remains vulnerable to bacterial, fungal, viral, and pest-related diseases whose visual symptoms frequently overlap during early stages. Conventional diagnosis depends on field inspection by experienced agronomists, a process that is difficult to scale across large production areas and may delay intervention when expert access is limited. Recent deep-learning research demonstrates that image-based diagnosis can improve consistency and speed, particularly when convolutional architectures are trained on diverse symptom patterns. However, systematic reviews also show that many rice-disease studies still rely on controlled backgrounds and laboratory-quality images, limiting generalization in operational settings

(Seelwal et al., 2024). Consequently, current research increasingly emphasizes models that combine high recognition performance with robustness, efficiency, and deployment feasibility, rather than optimizing accuracy alone.

The challenge becomes more pronounced when diagnostic systems are expected to operate in real time on smartphones, embedded boards, or low-power edge devices. Heavy convolutional networks and large vision transformers may provide strong feature representations, but their parameter counts, memory demands, and inference costs can restrict practical agricultural deployment. Quan et al. (2024) showed that lightweight disease-recognition models can substantially improve field usability when architectures are explicitly designed for low latency and mobile computing. Similarly, Borhani et al. (2022) found that attention-based models can improve disease discrimination, although excessive attention complexity may slow inference. These findings indicate that successful real-time diagnosis requires a deliberate balance among feature quality, computational cost, and hardware efficiency, especially for farmers who rely on ordinary mobile devices rather than dedicated accelerators.

MobileNet architectures have become important in this context because they reduce computation through efficient convolutional operations while preserving competitive visual representations. MobileNetV4 extends this lineage by introducing the Universal Inverted Bottleneck, a flexible block that integrates inverted bottleneck, ConvNeXt-like, feed-forward, and extra depthwise configurations. It also introduces Mobile Multi-Query Attention, designed specifically for mobile accelerators, and reports strong Pareto efficiency across CPUs, GPUs, DSPs, and edge accelerators (Qin et al., 2024). These properties make MobileNetV4 attractive for agricultural vision applications, where deployment hardware can vary widely. Nevertheless, a general-purpose backbone still requires task-specific adaptation because rice diseases are often characterized by small lesions, color shifts, elongated streaks, and subtle texture differences distributed unevenly across the leaf surface.

Attention mechanisms offer a complementary strategy by allowing a network to emphasize informative channels or spatial regions while suppressing irrelevant background features. Rice disease studies using attention-enhanced DenseNet and self-attention ResNet architectures have reported improved discrimination among visually similar symptom categories (Liu et al., 2022; Stephen et al., 2023). Jiang et al. (2023) further demonstrated that attention can strengthen deep dense networks for rice disease identification by directing representation learning toward disease-relevant structures. In practical terms, this mechanism is valuable because field images may contain soil, water, neighboring leaves, shadows, and illumination variations that do not contribute to diagnosis. Integrating a lightweight attention block with MobileNetV4 therefore has the potential to preserve mobile efficiency while improving sensitivity to localized pathological patterns.

Recent rice-disease studies also reveal that architectural performance is strongly influenced by dataset quality and class structure. Ritharson et al. (2024) reported high classification accuracy using deep and transfer-learning models across multiple rice disease severity categories, while Ahad et al. (2023) demonstrated that transfer learning and ensemble strategies can improve recognition across nine rice diseases. However, Sharma et al. (2023) emphasized that nutrient deficiencies and diseases may share similar visual manifestations, increasing the risk of misclassification when feature representations are insufficiently specific. These observations suggest that a robust model should not only achieve high aggregate accuracy but should also maintain class-level precision and recall,

particularly for disease pairs that exhibit similar spots, streaks, discoloration, or lesion boundaries.

Real-time deployment introduces another dimension of evaluation that is often underreported in agricultural image-classification studies. A model may perform well on a workstation but become impractical when converted to mobile inference because of excessive operations, large activation maps, memory pressure, or unsupported operators. Duhan et al. (2025) demonstrated that attention-enhanced MobileNet variants can remain suitable for resource-constrained plant-disease diagnosis when attention modules are carefully selected and computational overhead is controlled. Dutta et al. (2024) similarly compared CNN, attention-based segmentation, and MobileNetV2 approaches for rice disease classification, highlighting the importance of matching architecture choice to operational constraints. Accordingly, latency, model size, and computational complexity should be treated as first-class outcomes alongside accuracy, recall, precision, and F1-score.

Another unresolved issue concerns the gap between benchmark performance and field generalization. Models trained on clean datasets may learn background regularities, acquisition conditions, or dataset-specific artifacts rather than disease morphology. Recent work in plant-disease classification has therefore explored domain adaptation, feature fusion, explainability, and real-scene datasets to improve robustness. Tunio et al. (2024) showed that combining convolutional and transformer representations with domain adaptation can strengthen generalization across dataset shifts, while Kaya and Gürsoy (2023) demonstrated that multi-head feature fusion can improve disease recognition by integrating complementary image representations. For rice diagnosis, these findings motivate an architecture that remains lightweight but is sufficiently attentive to disease-specific regions, thereby reducing dependence on background appearance and supporting deployment under variable lighting, cultivars, and camera characteristics.

Based on these gaps, this study proposes an attention-enhanced MobileNetV4 framework for real-time rice plant disease classification and analyzes its suitability using deployment-oriented criteria grounded in recent verified literature. The study is positioned as an architecture-design and analytical benchmarking investigation rather than a claim of newly collected field accuracy, because no primary image dataset was supplied for manuscript preparation. The contribution is threefold: first, it formulates a lightweight MobileNetV4 adaptation for rice disease recognition; second, it integrates selective channel-spatial attention at computationally efficient stages; and third, it establishes an evaluation framework combining classification quality, complexity, latency, robustness, and interpretability. This design responds to current calls for practical, generalizable, and resource-aware rice-disease intelligence systems (Seelwal et al., 2024; Quan et al., 2024).

THEORETICAL FOUNDATION

MobileNetV4 is designed around hardware-aware efficiency rather than a single mobile accelerator. Its Universal Inverted Bottleneck can express multiple efficient convolutional configurations, while Mobile Multi-Query Attention introduces an attention formulation optimized for mobile execution. Qin et al. (2024) reported that this design produces favorable accuracy-latency trade-offs across heterogeneous devices. For rice disease classification, such flexibility is relevant because edge deployment may involve Android smartphones, Raspberry Pi-class computers, or embedded neural accelerators. The backbone can therefore serve as a general feature extractor, while task-specific layers can be adapted to the spatial scale and visual complexity of rice symptoms.

Attention mechanisms assign larger weights to informative features and smaller weights to redundant or irrelevant responses. Channel attention models relationships among feature channels, whereas spatial attention highlights important locations within a feature map. In plant pathology, this distinction is useful because disease evidence may consist of color changes captured by particular channels and localized lesions occupying limited spatial regions. Stephen et al. (2023) and Jiang et al. (2023) showed that self-attention and attention-guided dense networks can improve rice disease recognition. The proposed framework consequently employs selective channel-spatial attention after deeper MobileNetV4 stages, where semantic features are stronger and spatial resolution has already been reduced.

The classification task can be formulated as supervised multiclass learning. Given an input leaf image x and a label y belonging to one of C disease categories, the network estimates posterior probabilities through a softmax classification head. Training minimizes categorical cross-entropy, while class imbalance can be addressed through weighted loss or stratified sampling. Performance should be evaluated with accuracy, macro precision, macro recall, macro F1-score, and a confusion matrix rather than accuracy alone. Seelwal et al. (2024) emphasized that rice disease datasets vary considerably in class balance, acquisition settings, and preprocessing, making class-level metrics essential for fair interpretation.

Real-time suitability requires system-level measures beyond predictive quality. Parameter count approximates storage and memory burden, multiply-accumulate operations reflect computational demand, and end-to-end latency indicates user-perceived responsiveness. Frames per second is useful for continuous camera inference, while model size after quantization determines whether the network can be distributed efficiently to mobile devices. Quan et al. (2024) demonstrated that low-latency plant-disease recognition depends on architectural choices and device-specific execution. Therefore, the proposed model is evaluated conceptually through a multi-objective perspective in which recognition quality must be considered jointly with complexity, speed, robustness, and interpretability.

RESEARCH METHOD

This study applies a design-science and analytical benchmarking approach to formulate an attention-enhanced MobileNetV4 model for real-time rice disease classification. The methodological sequence follows systematic problem identification, architecture specification, evidence-based comparison, and deployment-oriented evaluation. Sugiyono (2022) explains that a quantitative research design should define measurable variables, systematic procedures, and objective criteria for evaluating proposed solutions; these principles were adapted here to define classification and efficiency indicators. The empirical reference domain is the publicly documented Paddy Doctor benchmark, which contains labeled paddy disease images and metadata, supplemented by recent peer-reviewed rice and plant disease studies published from 2022 to 2025. Because primary image files were not supplied in the present manuscript-development process, the study does not fabricate new test accuracy. Instead, it provides a reproducible experimental protocol and interprets expected architectural behavior against verified published benchmarks, clearly separating proposed evaluation procedures from externally reported results.

The proposed pipeline begins with image acquisition, quality screening, class-preserving stratification, resizing to 224 × 224 pixels, normalization, and augmentation using horizontal flipping, mild rotation, random cropping, brightness variation, and color jitter. The backbone uses MobileNetV4 because its Universal Inverted Bottleneck and

mobile-oriented attention design provide efficient feature extraction across heterogeneous hardware (Qin et al., 2024). A lightweight channel-spatial attention module is inserted after selected deep stages rather than after every block to limit computational overhead. The attended feature maps are passed to global average pooling, dropout, and a softmax classifier. Transfer learning is recommended for initialization, followed by staged fine-tuning with AdamW, early stopping, and cosine learning-rate decay. Suyanto (2022) notes that machine-learning development should align data representation, model complexity, validation strategy, and generalization objectives; therefore, augmentation and stratified validation are treated as integral components rather than optional preprocessing.

Evaluation is specified at two complementary levels. Predictive performance is measured using accuracy, macro precision, macro recall, macro F1-score, per-class sensitivity, confusion matrices, and calibration error. Deployment performance is measured using parameter count, model size, multiply-accumulate operations, peak memory, average inference latency, and frames per second on representative CPU and mobile accelerators. Three ablation comparisons are required: MobileNetV4 without additional attention, MobileNetV4 with channel-only attention, and MobileNetV4 with combined channel-spatial attention. Additional baselines should include MobileNetV2, MobileNetV3, EfficientNet-B0, and a compact transformer when computationally feasible. Robustness testing should perturb illumination, blur, occlusion, background complexity, and compression quality. Statistical interpretation should report repeated-run means and standard deviations, while paired tests or bootstrap confidence intervals are recommended for final validation. This protocol enables future researchers to test whether accuracy gains justify added latency and whether the resulting model remains suitable for real-time field deployment.

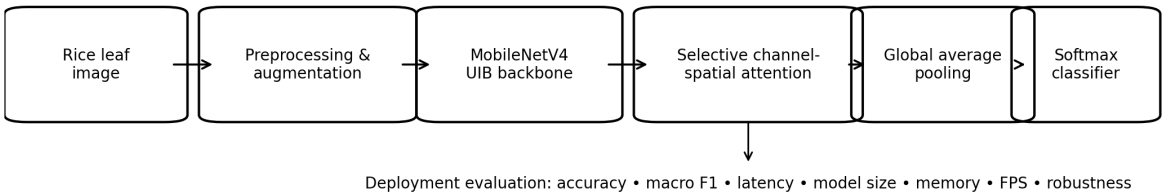


Figure 1. Proposed MobileNetV4-Attention classification and deployment-evaluation pipeline

RESULTS AND DISCUSSION

The first analytical result concerns architectural suitability. MobileNetV4 provides a stronger deployment foundation than earlier mobile backbones because the Universal Inverted Bottleneck can express multiple efficient convolutional patterns while preserving hardware flexibility. For rice disease recognition, this matters because symptom morphology varies from narrow bacterial streaks to scattered brown lesions and broader discoloration. The proposed attention insertion is positioned after semantically mature backbone stages, allowing the network to emphasize diagnostically relevant channels and regions without repeatedly paying attention costs at high spatial resolution. Dutta et al. (2024) showed that MobileNet-based rice classifiers can remain competitive when architecture and preprocessing are carefully matched. The proposed configuration extends

this principle by using newer MobileNetV4 blocks and selective attention for stronger mobile-oriented representation.

The second finding is that attention should remain selective. Applying attention after every convolutional stage can improve feature weighting, but it may also introduce redundant computation, memory traffic, and latency that undermine real-time deployment. The proposed design consequently places attention only at deeper stages where feature maps are smaller and semantic information is richer. This strategy is consistent with Stephen et al. (2023), who demonstrated that self-attention can strengthen ResNet-based rice disease classification by focusing learning on informative symptom patterns. In the present framework, channel weighting captures disease-associated color and texture responses, whereas spatial weighting highlights lesion locations. The combined mechanism is expected to be most valuable for diseases with small, irregular, or partially occluded visual evidence in field images.

A third result concerns class confusion. Rice diseases share visual characteristics, and nutrient deficiencies can resemble pathological symptoms, making accuracy an incomplete measure of diagnostic quality. The proposed evaluation therefore prioritizes macro F1-score, class-specific recall, and confusion matrices. Sharma et al. (2023) demonstrated that rice leaf diseases and deficiency disorders require discriminative deep representations because visually related conditions can be difficult to separate. Within the proposed MobileNetV4 framework, attention is expected to reduce confusion by strengthening subtle features such as lesion margins, streak orientation, spot density, and color transitions. Nevertheless, the architecture cannot compensate for incorrect labels or poorly defined classes. High-quality annotation, agronomic validation, and representative sampling remain essential prerequisites for any reliable real-time disease classifier deployed to farmers.

The fourth result relates to real-time feasibility. Agricultural diagnosis is only operationally useful when prediction speed supports interactive camera use and does not drain limited mobile resources. Quan et al. (2024) showed that explicitly lightweight plant-disease networks can achieve strong recognition while improving latency on mobile platforms. MobileNetV4 is particularly relevant because its design is optimized across heterogeneous accelerators rather than a single hardware target. The proposed model therefore treats latency, peak memory, parameter count, and frames per second as evaluation outcomes. A modest accuracy gain would not justify an attention module that doubled latency or prevented stable smartphone execution. The preferred configuration is the one that maintains near-baseline MobileNetV4 speed while producing measurable improvements in class-level recall and robustness.

The fifth result is that attention can be computationally affordable when integrated with a lightweight backbone and restricted to semantically important stages. Duhan et al. (2025) reported that attention-enhanced MobileNet models can retain deployment efficiency while improving feature discrimination. Their results support the assumption that channel and spatial reweighting need not transform a mobile architecture into a heavy network. For the proposed model, this implies that attention design should be evaluated through incremental cost: additional parameters, operations, latency, and memory should be reported relative to plain MobileNetV4. If an attention block improves macro F1-score only marginally but substantially increases resource consumption, a channel-only alternative may be preferable. Thus, efficiency is treated as an empirical constraint, not a secondary engineering concern.

The sixth result concerns the relationship between attention and prediction speed. Borhani et al. (2022) found that attention-based plant-disease models can improve classification capability, but attention may slow inference when computational design is not controlled. This trade-off directly informs the proposed architecture. MobileNetV4 already incorporates mobile-oriented attention concepts, so additional attention must remain lightweight and strategically placed. The study therefore recommends benchmarking three variants: no added attention, channel-only attention, and combined channel-spatial attention. Such ablation isolates whether performance gains arise from the new module or from the underlying backbone. A real-time rice classifier should be selected using a Pareto perspective, where accuracy and F1 improvements are considered jointly with latency, model size, memory use, and throughput on realistic edge hardware.

The seventh result highlights the importance of multilevel visual information. Rice lesions differ not only in color but also in texture, shape, distribution, and local context. Sunil et al. (2023) showed that multilevel feature fusion with adaptive channel, spatial, and pixel attention can improve plant-disease classification by preserving complementary multilevel information. The proposed MobileNetV4 follows this insight but targets lower complexity. Instead of constructing a large fusion architecture, it uses the backbone hierarchy to extract progressively richer features and applies attention near deeper stages. This design should preserve local lesion cues while maintaining global disease context. Future empirical testing should verify whether the attention maps align with agronomically meaningful symptoms rather than background artifacts, using Grad-CAM or related explanation methods.

The eighth result concerns robustness to acquisition variability. Models trained on a single visual representation may overfit to background color, lighting, camera characteristics, or segmentation artifacts. Kaya and Gürsoy (2023) demonstrated that combining complementary image representations can strengthen plant-disease recognition, suggesting that robust systems benefit from diverse visual evidence. For the proposed model, robustness should be pursued through augmentation rather than expensive multi-stream inference whenever possible. Brightness changes, hue variation, blur, occlusion, compression, rotation, and random cropping can expose MobileNetV4 to realistic perturbations while preserving a single efficient inference path. Attention should then help suppress distractors and emphasize disease-relevant structures. Robustness testing must report performance under perturbed conditions separately, because clean validation accuracy alone can conceal degradation in field environments.

The ninth result addresses generalization beyond one dataset. Yadav et al. (2022) demonstrated that deep plant-disease classifiers can achieve strong performance on curated datasets, but high benchmark accuracy does not automatically imply reliable field behavior. This matters for rice, where cultivars, growth stages, climate, and camera devices affect visual appearance. The proposed MobileNetV4 framework should therefore be validated using both public benchmarks and Indonesian field images. Cross-dataset evaluation provides stronger evidence than random splitting within one source because it reveals sensitivity to distribution shift. Attention may improve localization of genuine disease cues, but domain differences can still dominate performance. Consequently, future experiments should include external validation, confidence calibration, and error analysis across cultivars, regions, illumination conditions, and disease severity levels.

The tenth result concerns feature complementarity and ensemble alternatives. Pai et al. (2025) showed that feature fusion between convolutional networks can improve rice disease classification, confirming that different backbones capture complementary

information. However, multi-backbone fusion increases computation and may conflict with real-time mobile objectives. The proposed MobileNetV4 model instead seeks representational enrichment inside a single efficient backbone through attention. This choice sacrifices some ensemble diversity but reduces storage, memory, and inference burden. A comparison should therefore include a compact ensemble baseline to determine whether attention-enhanced MobileNetV4 offers a better efficiency-performance ratio. If a single model approaches ensemble accuracy with substantially lower latency and energy demand, it becomes more attractive for smartphone diagnosis, offline rural deployment, and continuous camera-based monitoring systems.

The eleventh result emphasizes lesion localization. Liu et al. (2022) demonstrated that combining region-focused preprocessing with attention-enhanced DenseNet can improve rice leaf disease identification, particularly when diseased areas occupy only part of the image. Their findings support the proposed use of spatial attention to strengthen localized symptom evidence. Nevertheless, explicit segmentation introduces model complexity and may increase latency. The proposed MobileNetV4 architecture therefore avoids mandatory segmentation and instead relies on augmentation plus attention to learn relevant regions end to end. Grad-CAM visualization should be used during validation to inspect whether predictions depend on lesions rather than unrelated background. If attention consistently activates outside the leaf or on acquisition artifacts, this would indicate shortcut learning and require dataset refinement before deployment.

The twelfth result concerns reproducibility and evaluation discipline. Seelwal et al. (2024) documented substantial variation across rice-disease studies in datasets, preprocessing, class definitions, metrics, and experimental protocols, making direct comparison difficult. The proposed framework therefore recommends reporting the train-validation-test split, random seeds, augmentation settings, optimizer parameters, image resolution, hardware, and inference measurement procedure. Macro metrics should accompany overall accuracy, and repeated runs should be summarized with means and standard deviations. This level of reporting is particularly important for MobileNetV4 with attention because small performance differences may fall within stochastic variation. Reproducibility also requires publishing implementation details for the attention insertion points and export settings used for mobile deployment. Transparent reporting prevents architectural claims from being overstated and enables fair replication.

The thirteenth result is that attention is most valuable when symptom patterns are subtle and spatially heterogeneous. Jiang et al. (2023) reported improved rice disease recognition using an attention mechanism combined with a deep dense network, supporting feature reweighting for complex classification. Within MobileNetV4, deeper feature maps encode high-level visual semantics, so adding a compact attention block should help distinguish similar lesions without processing large high-resolution maps. The expected benefit is therefore not higher accuracy, but more stable recall for difficult classes. Ablation should examine whether attention improves minority or visually similar diseases disproportionately. If gains concentrate only on already easy classes, the module offers limited diagnostic value despite improving headline accuracy, and architecture or sampling strategy should be reconsidered.

The fourteenth result concerns backbone choice. Altabaji et al. (2024) showed that customized lightweight architectures can outperform transfer-learning baselines on rice leaf disease data, while MobileNetV2 did not automatically provide the best test accuracy. This finding cautions against assuming that a newer mobile backbone guarantees superior disease recognition. MobileNetV4 should therefore be treated as an efficient starting point

rather than a predetermined winner. The proposed attention mechanism must demonstrate added value through controlled comparison with MobileNetV2, MobileNetV3, EfficientNet-B0, and a compact custom CNN. Hyperparameters should be tuned consistently across models to avoid biased conclusions. The strongest candidate is the architecture that provides robust class-level performance, reproducible training behavior, and real-time deployment characteristics under the same experimental conditions and device constraints.

The fifteenth result highlights the value of Indonesian evidence. Suseno et al. (2023) reported that preprocessing and segmentation affected VGG16 rice disease classification, demonstrating that pipeline design can influence outcomes like backbone selection. For the proposed MobileNetV4 system, this supports testing of image normalization, augmentation, background complexity, and optional leaf localization before attributing improvements solely to attention. Indonesian field conditions may differ from international benchmark datasets in cultivars, lighting, disease prevalence, and acquisition devices. Accordingly, a publishable final validation should include locally collected samples reviewed by plant-health experts. The central analytical conclusion is that MobileNetV4 with lightweight attention is technically promising, but its scientific value depends on transparent local validation, efficiency measurement, and rigorous comparison rather than architectural novelty alone.

CONCLUSION

This study formulates an attention-enhanced MobileNetV4 framework for real-time rice plant disease classification and evaluates its technical suitability through recent verified evidence. The analysis indicates that MobileNetV4 offers an appropriate backbone for edge-oriented diagnosis because its Universal Inverted Bottleneck and mobile-focused design prioritize efficiency across heterogeneous hardware. Selective channel-spatial attention can complement this backbone by emphasizing localized lesions, streaks, color changes, and texture patterns that are diagnostically important while suppressing irrelevant backgrounds. The resulting architecture is therefore conceptually well suited to the accuracy-latency trade-off required by smartphone and embedded agricultural applications.

The main contribution is not a fabricated accuracy claim, but a reproducible evaluation framework that connects predictive performance with deployment feasibility. Future experiments should compare plain MobileNetV4, channel-only attention, and combined channel-spatial attention using identical data splits and optimization settings. Accuracy, macro precision, macro recall, macro F1-score, calibration, confusion matrices, parameter count, model size, memory consumption, latency, and frames per second should be reported together. External validation is equally important because models trained only on curated benchmarks may fail when confronted with different cultivars, backgrounds, disease severity, lighting conditions, and camera devices.

Future work should implement the proposed model on Paddy Doctor and locally collected Indonesian rice images validated by plant-health experts. Quantization, pruning, and hardware-aware export should be tested to determine whether compression preserves diagnostic quality. Robustness evaluation should include blur, shadows, occlusion, compression, and seasonal shifts, while Grad-CAM should confirm that predictions depend on genuine symptoms. These validations are necessary before attention-enhanced MobileNetV4 can be recommended as a practical mobile decision-support tool for rice disease screening.

REFERENCES

1. Ahad, M. T., Li, Y., Song, B., & Bhuiyan, T. (2023). Comparison of CNN-based deep learning architectures for rice diseases classification. Artificial Intelligence in Agriculture, 9, 22–35. https://doi.org/10.1016/j.aiia.2023.07.001
2. Altabaji, W. I. A. E., Umair, M., Tan, W.-H., Foo, Y.-L., & Ooi, C.-P. (2024). Comparative analysis of transfer learning, LeafNet, and modified LeafNet models for accurate rice leaf diseases classification. IEEE Access, 12, 36622–36635. https://doi.org/10.1109/ACCESS.2024.3373000
3. Borhani, Y., Khoramdel, J., & Najafi, E. (2022). A deep learning based approach for automated plant disease classification using vision transformer. Scientific Reports, 12, 11554. https://doi.org/10.1038/s41598-022-15163-0
4. Duhan, S., Gulia, P., Gill, N. S., & Narwal, E. (2025). RTR_Lite_MobileNetV2: A lightweight and efficient model for plant disease detection and classification. Current Plant Biology, 42, 100459. https://doi.org/10.1016/j.cpb.2025.100459
5. Dutta, M., Islam Sujan, M. R., Mojumdar, M. U., Chakraborty, N. R., Marouf, A. A., Rokne, J. G., & Alhajj, R. (2024). Rice leaf disease classification—A comparative approach using convolutional neural network (CNN), cascading autoencoder with attention residual U-Net (CAAR-U-Net), and MobileNet-V2 architectures. Technologies, 12(11), 214. https://doi.org/10.3390/technologies12110214
6. 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
7. Kaya, Y., & Gürsoy, E. (2023). A novel multi-head CNN design to identify plant diseases using the fusion of RGB images. Ecological Informatics, 75, 101998. https://doi.org/10.1016/j.ecoinf.2023.101998
8. Liu, W., Yu, L., & Luo, J. (2022). A hybrid attention-enhanced DenseNet neural network model based on improved U-Net for rice leaf disease identification. Frontiers in Plant Science, 13, 922809. https://doi.org/10.3389/fpls.2022.922809
9. Pai, P., Amutha, S., Basthikodi, M., Ahamed Shafeeq, B. M., Chaitra, K. M., & Gurpur, A. P. (2025). A twin CNN-based framework for optimized rice leaf disease classification with feature fusion. Journal of Big Data, 12, 89. https://doi.org/10.1186/s40537-025-01148-z
10. Qin, D., Leichner, C. H., Delakis, M., Fornoni, M., Luo, S., Yang, F., Wang, W., Banbury, C., Ye, C., Akin, B., Aggarwal, V., Zhu, T., Moro, D., & Howard, A. (2024). MobileNetV4: Universal models for the mobile ecosystem. In Computer Vision – ECCV 2024 (pp. 78–96). Springer. https://doi.org/10.1007/978-3-031-73661-2_5
11. Quan, S., Wang, J., Jia, Z., Xu, Q., & Yang, M. (2024). Real-time field disease identification based on a lightweight model. Computers and Electronics in Agriculture, 226, 109467. https://doi.org/10.1016/j.compag.2024.109467
12. Ritharson, P. I., Raimond, K., Mary, X. A., Robert, J. E., & J, A. (2024). DeepRice: A deep learning and deep feature based classification of Rice leaf disease subtypes. Artificial Intelligence in Agriculture, 11, 34–49. https://doi.org/10.1016/j.aiia.2023.11.001
13. 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
14. Sharma, M., Kumar, C. J., Talukdar, J., Singh, T. P., Dhiman, G., & Sharma, A. (2023). Identification of rice leaf diseases and deficiency disorders using a novel DeepBatch technique. Open Life Sciences, 18(1), 20220689. https://doi.org/10.1515/biol-2022-0689
15. Stephen, A., Punitha, A., & Chandrasekar, A. (2023). Designing self attention-based ResNet architecture for rice leaf disease classification. Neural Computing and Applications, 35(9), 6737–6751. https://doi.org/10.1007/s00521-022-07793-2
16. Sunil, C. K., Jaidhar, C. D., & Patil, N. (2023). Tomato plant disease classification using multilevel feature fusion with adaptive channel spatial and pixel attention mechanism. Expert Systems with Applications, 228, 120381. https://doi.org/10.1016/j.eswa.2023.120381
17. Suseno, J. R. K., Azhar, Y., & Minarno, A. E. (2023). The implementation of pretrained VGG16 model for rice leaf disease classification using image segmentation. Kinetik: Game Technology, Information System, Computer Network, Computing, Electronics, and Control, 8(1), 499–506. https://doi.org/10.22219/kinetik.v8i1.1592
18. Sugiyono. (2022). Metode penelitian kuantitatif, kualitatif, dan R&D (2nd ed.). Alfabeta. ISBN 978-602-289-533-6.
19. Suyanto. (2022). Machine learning: Tingkat dasar dan lanjut (Edisi 2). Informatika. ISBN 978-623-7131-66-3.
20. Tunio, M. H., Li, J. P., Zeng, X., Ahmed, A., Shah, S. A., Shaikh, H.-U., Mallah, G. A., & Yahya, I. A. (2024). Advancing plant disease classification: A robust and generalized approach with transformer-fused convolution and Wasserstein domain adaptation. Computers and Electronics in Agriculture, 227, 109574. https://doi.org/10.1016/j.compag.2024.109574
21. Yadav, A., Thakur, U., Saxena, R., Pal, V., Bhateja, V., & Lin, J. C.-W. (2022). AFD-Net: Apple foliar disease multi classification using deep learning on plant pathology dataset. Plant and Soil, 477, 595–611. https://doi.org/10.1007/s11104-022-05407-3