Interpretable Lightweight Model for Rice Leaf Disease Detection Using Knowledge Distillation

Mrs. D.N.B.T. Sundari[1], Ch. Meghana[2], K. Abhisree Reddy[3], Sara Syed[4], G. Bindhu[5]
[1,2,3,4,5] *Sphoorthy Engineering College*

***Abstract*—Rice is one of the most important food crops in the world, and rice production is greatly influenced by diseases like Bacterial Leaf Blight, Brown Spot, Blast, Sheath Blight, and Tungro. It is often difficult to detect diseases in remote areas due to the unavailability of experts. In this paper, a novel rice leaf disease detection system with minimal resources is suggested by using knowledge distillation. In this paper, a teacher model with a ResNet50 backbone is utilized to train a lightweight CNN student model with 91.8K parameters and a model size of 0.35 MB, which achieved an accuracy of 99.29%. The model is robust to brightness changes and has fewer training samples, ensuring the reliability of the model in real-world scenarios. Moreover, the model is enhanced by the addition of Grad-CAM, which helps to increase the transparency of the model.**

I. INTRODUCTION

Rice is one of the primary food crops in the world. Rice cultivation is affected by several leaf diseases, including bacterial blight, blast, and brown spot. These can cause serious damage if not identified in the early stages. Conventional methods of rice leaf disease detection involve manual inspection by experts. This is a tedious and impractical approach, especially in a smart agricultural environment.

In recent times, rice leaf disease detection using computer vision has gained more attention due to the rapid development of deep learning techniques. Convolutional Neural Networks (CNNs) have been reported to deliver high accuracy in image classification problems. However, deep learning architectures, including ResNet and Inception, require high computational power, which makes it difficult to implement in real-time environments, especially on mobile devices. Therefore, a lightweight rice leaf disease detection system using knowledge distillation is proposed in this work. The proposed system uses a high-accuracy teacher model, which can be used to train a lightweight student model. The use of Grad-CAM also provides a visual explanation of the system, which can be used to increase trust in the system.

II. RELATED WORK

In recent years, there has been a rapid pace of work on applying deep learning techniques to agricultural applications, especially with reference to rice leaf disease detection. In this area, Bijoy et al. (2024) have introduced a novel technique using a deep convolutional neural network along with an enhanced dataset to detect rice leaf diseases. The technique has been successful because it was designed with sustainability as a prime objective. The approach has been successful because of the robust feature extraction capabilities along with a well-curated dataset. This technique has been successful because it has been able to detect diseases at an early stage, which helps to prevent losses to the agricultural sector as well as to prevent the overuse of pesticides, thereby supporting sustainability.

In a similar vein, Babu et al. (2022) studied the efficacy of standard deep learning architectures such as VGG16 and ResNet in identifying rice leaf diseases. Their primary aim was to propose efficient approaches that can run on mobile platforms to facilitate their usage in the field. This aligns with the concept of smart farming. Although their approaches were less advanced compared to recent techniques, their focus on efficiency and ease of use makes their contribution significant in the development of recent approaches.

In addition to recent developments in agricultural-related artificial intelligence approaches, knowledge distillation has also been recognized as a significant technique in developing efficient variants of complex artificial intelligence approaches. Moslemi et al. (2024) presented a comprehensive overview of recent developments in knowledge distillation. Their contribution was significant in that they presented a classification of knowledge distillation approaches based on student-teacher configurations, learning modes such as offline, online, and self-knowledge distillation, and application domains. They also discussed significant challenges in knowledge distillation approaches and future directions.

In this area, Mehnaz and Islam (2025) have introduced a comparative study on using CNNs, Transformers, as well as classical machine learning techniques for rice leaf disease detection. From this study, it has been ascertained that although using Transformers has been successful because of the attention mechanisms, CNNs are advantageous because they are computationally efficient.

Expanding the application of KD beyond traditional domains, Abbasi et al. (2024) explored its use in IoT traffic classification. By distilling a large teacher model into a lightweight student model, the study demonstrated that high classification accuracy could be maintained even with reduced computational demands. This approach is especially valuable in edge computing scenarios where resource constraints are prominent. Their findings validate the broader applicability of KD in real-world systems, particularly in low-latency and energy-sensitive environments such as IoT networks.

Together, these studies illustrate the diverse applicability and growing importance of deep learning and knowledge distillation. In agriculture, tailored deep models like dCNNs and attention-driven architectures are revolutionizing disease detection, while in systems and networking, KD is proving critical for developing scalable and efficient AI solutions. These dual trajectories underscore a broader trend in AI research—towards sustainability, efficiency, and practical deployment.

III. METHOD/APPROACH

This section presents the end-to-end methodology of our rice leaf disease detection system, encompassing dataset preparation, model design using knowledge distillation, robustness evaluation, interpretability analysis, and deployment strategy.

3.1 Dataset Preparation and Augmentation The foundational dataset for our study is based on an enhanced dataset proposed in the paper titled "Towards Sustainable Agriculture: A Novel Approach for Rice Leaf Disease Detection Using dCNN and Enhanced Dataset". In that paper, the authors addressed the problem of insufficient and unreliable public datasets for rice leaf disease detection by curating a new dataset through the combination of existing sources and manual image collection. The Enhanced Rice Leaf dataset used in this study is available at .

Table 1: Dataset Composition and Splits

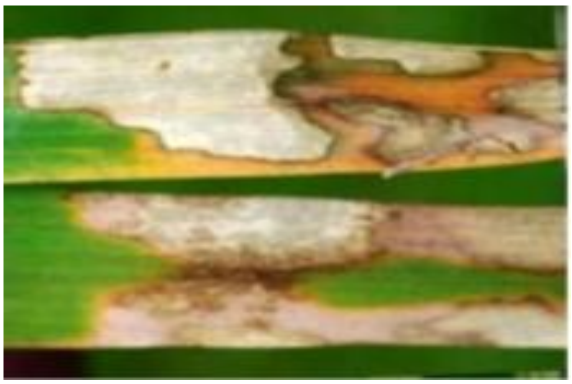
Class	Train	Validation	Test
Bacterial Leaf Blight	1228	351	175
Brown Spot	1077	308	153
Blast	508	145	73
Sheath Blight	430	123	61
Tungro	673	192	96

The final enhanced dataset comprises 5,593 images categorized into five disease classes: bacterial leaf blight, brown spot, blast, sheath blight, and Tungro. Table 2 presents the class-wise distribution. The dataset was divided into three subsets for training, validation, and testing as outlined in Table 3.

a) Bacterial Leaf Blight

(b) Brown Spot

(c) Blast



(d) Sheath Blight

(e) Tungro

Fig. 1: Representative disease samples from the enhanced dataset.

Table 2: Class-wise Distribution of Final Dataset

Disease Class	Number of Images
Bacterial Leaf Blight	1754
Brown Spot	1538
Blast	726
Sheath Blight	614
Tungro	961
Total	**5593**

Table 3: Dataset Split

Split	Number of Images	Notes
Training	3158	Slightly imbalanced
Validation	1277	Balanced across all classes
Testing	850	Balanced across all classes

Knowledge Distillation Framework Knowledge Distillation (KD), first introduced by Hinton et al., is a powerful model compression and transfer learning technique designed to address the limitations of deploying large-scale deep learning models in resource-constrained environments. It involves training a compact, low-complexity student model to replicate the behavior of a large, high-performing teacher model. This is typically done by minimizing the divergence between the output distributions (soft targets) produced by the teacher and those of the student[11]. Unlike traditional training methods that rely solely on hard labels (one-hot encoded vectors), KD utilizes the softmax outputs of the teacher model, which are softened using a temperature parameter T:

$$q_i = \frac{\exp(z_i/T)}{\sum_j \exp(z_j/T)}$$

Here, zi represents the logit (pre-softmax activation) for class i, and T is the temperature parameter. A higher T produces a softer probability distribution, expos ing relationships between classes and encoding the teacher's confidence. During distillation, the student is trained to minimize a combined loss

$$\mathcal{L}_{\mathrm{KD}} = (1-\alpha)\cdot\mathcal{L}_{\mathrm{CE}}(y, p_s) + \alpha\cdot T^2\cdot\mathcal{L}_{\mathrm{KL}}(q_t, q_s)$$

Where,

- LCE is the cross-entropy loss between the true label y and the student's prediction ps,
- LKL is the Kullback-Leibler divergence between the teacher's soft targets qt and the student's soft predictions qs,
- α controls the balance between hard and soft loss components,
- T is the temperature that controls the smoothness of the probability distributions.

3.2 Teacher Model Architecture The teacher model is based on the ResNet50 architecture, known for its deep residual connections and strong feature extraction capability. Training is done in two stages:

- Phase 1: Train top layers with frozen base.
- Phase 2: Fine-tune last 30 layers with a lower learning rate.

3.3 Student Model Design The student model is a compact 5-layer CNN designed for deployment on edge devices. It incorporates

- Depth wise Separable Convolutions to reduce complexity.
- Residual Connections to retain feature richness and aid gradient flow.

With only 91.8K parameters and a model size of 0.35 MB, it offers high efficiency with minimal performance drop. The model architecture for Custom CNN is given in Figure 3. 7

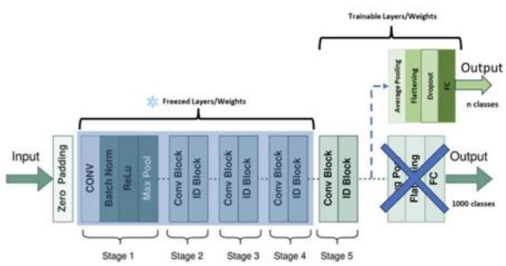


Fig. 2: ResNet50 architecture used as the teacher model

3.4 Robustness Evaluation We assessed the model's reliability under:

• Lighting Variations: ±20% brightness to simulate different sunlight conditions.

• Limited Training Data: Performance evaluated with 40%, 50%, and 60% of the training data.

3.5 Interpretability Analysis For model interpretability, we employed Grad-CAM to visualize class-specific attention. These heatmaps highlight which image regions influenced the model's prediction, aiding user trust and explainability. 4.7 Training Protocol The following configuration was used for training both models:

• Optimizer: Adam with learning rate 1 ×10−4
• Batch size: 32
• Early stopping: Patience = 15
• Temperature (T): 2
• Loss weight (α): 0.3

3.6 End-to-End Pipeline

The full pipeline integrates image preprocessing, training, evaluation, and deployment. A web and mobile interface allows users to upload leaf images for real-time disease classification using the student model via a backend API

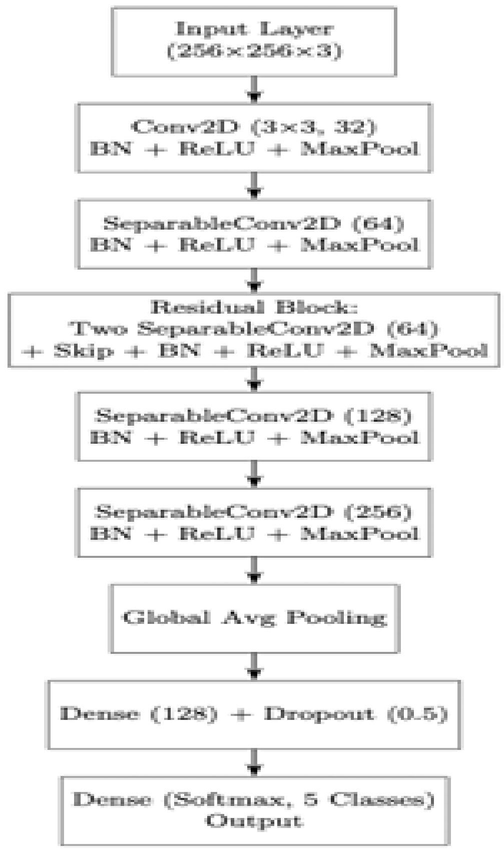


Fig. 3: Custom CNN Student Model Architecture

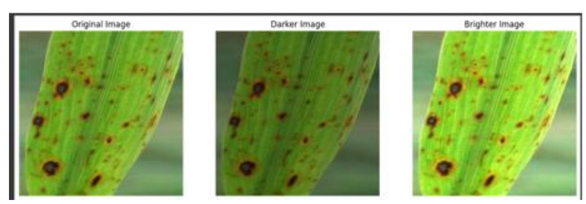


Fig. 4: Lighting variations: (a) Original (b)-20% brightness (c) +20% brightness

IV. DISCUSSION

This section discusses the interpretation of the experimental results and evaluation of the suggested model in comparison to existing models.

Based on the experimental results, it is observed that the knowledge-distilled student model has achieved high accuracy in classification, with a value of 99.29%, while maintaining a reduced model size in comparison to existing models like ResNet50 and VGG16.

Moreover, the robustness analysis has shown the consistency of the model in terms of brightness variation and limited training conditions, which is important in real-time agricultural scenarios where brightness is not fixed.

In comparison to existing models, the suggested model has achieved better results in terms of accuracy and efficiency, since some models have achieved higher accuracy, but with a large model size, which is not suitable in real-time scenarios.

However, certain limitations are observed. In some cases, Grad-CAM visualizations indicate attention outside the exact lesion regions, suggesting potential dataset bias or misinterpretation by the model. Additionally, performance slightly degrades under extreme lighting conditions, indicating scope for improvement.

Overall, the proposed approach successfully bridges the gap between high-performance models and real-world deployment constraints, though further enhancements in robustness and interpretability can improve reliability.

V. CONCLUSION

This project aims to develop a comprehensive solution to design a lightweight, interpretable, and deployable model to carry out rice leaf disease detection. This is a crucial problem in agriculture. This project was able to successfully train a model to approximate the performance of a high-accuracy model by employing a knowledge distillation framework. The system utilizes Convolutional Neural Networks and depth-wise separable convolution to design a student model that is computationally efficient and highly accurate. Additionally, the system utilizes Gradient-weighted Class Activation Mapping to meet the rising need to ensure model explainability. This allows farmers to understand the predictions made by the model. The model was extensively tested on the standard set of images. Additionally, images with low and high lighting conditions were used to evaluate the performance of the model.. Experiments with reduced training data demonstrated that knowledge distillation could effectively compensate for limited data availability, thereby offering a scalable and data-efficient solution. In sum, this work contributes to the field of smart agriculture by bridging the gap between high-performance AI and field-level practicality. The combination of lightweight model design, data-efficient training, and interpretability makes the proposed solution particularly well-suited for real-world deployment in edge environments. Moving forward, the framework provides a solid foundation for future advance ments involving multi-teacher learning, multimodal fusion, and advanced distillation strategies

REFERENCES

[1] Shrivastava, V.K., Pradhan, M.K.: Rice plant disease classification using color features: a machine learning paradigm. Journal of Plant Pathology 103(1), 17–26 (2021)

[2] Azim, M.A., Islam, M.K., Rahman, M.M., Jahan, F.: An effective feature extraction method for rice leaf disease classification. TELKOMNIKA (Telecommunication Computing Electronics and Control) 19(2), 463–470 (2021)

[3] Hossain, S.M.M., Tanjil, M.M.M., Ali, M.A.B., Islam, M.Z., Islam, M.S., Mobassirin, S., Sarker, I.H., Islam, S.R.: Rice leaf diseases recognition using convolutional neural networks. In: Advanced Data Mining and Applications: 16th International Conference, ADMA 2020, Foshan, China, November 12–14, 2020, Proceedings 16, pp. 299–314 (2020). Springer

[4] Ghosal, S., Sarkar, K.: Rice leaf diseases classification using cnn with transfer learning. In: 2020 IEEE Calcutta Conference (Calcon), pp. 230–236 (2020). IEEE

[5] Chen, J., Zhang,D., Nanehkaran, Y.A., Li, D.: Detection of rice plant diseases based on deep transfer learning. Journal of the Science of Food and Agriculture 100(7), 3246–3256 (2020)

[6] Bijoy, M.H., Hasan, N., Biswas, M., Mazumdar, S., Jimenez, A., Ahmed, F., Rasheduzzaman, M., Momen, S.: Towards sustainable agriculture: A novel approach for rice leaf disease detection using dcnn and enhanced dataset. IEEE Access 12, 34174–34191 (2024) https://doi.org/10.1109/ACCESS.2024.3371511 14

[7] Bijoy, M.H., Hasan, N., Biswas, M., Mazumdar, S., Jimenez, A., Ahmed, F., Rasheduzzaman, M., Momen, S.: Enhanced Rice Leaf Dataset. https://drive. google.com/drive/folders/1ZYldfJSbqCEEJfmvq

FMvn37WRlik9-1Z. Accessed: 2025-05-06 (2025)

[8] Su, N.T., Hung, P.D., Vinh, B.T., Diep, V.T.: Rice leaf disease classification using deep learning and target for mobile devices. In: International Conference on Emerging Technologies and Intelligent Systems, pp. 136–148 (2021). Springer

[9] Hinton, G., Vinyals, O., Dean, J.: Distilling the knowledge in a neural network. In: NIPS Deep Learning and Representation Learning Workshop (2015)

[10] Mirzadeh, S.I., Farajtabar, M., Li, A., Levine, N., Matsukawa, A., Ghasemzadeh, H.: Improved knowledge distillation via teacher assistant. In: Proceedings of the AAAI Conference on Artificial Intelligence, vol. 34, pp. 5191–5198 (2020)

[11] Cheng, X., Rao, Z., Chen, Y., Zhang, Q.: Explaining knowledge distillation by quantifying the knowledge. In: 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pp. 12922–12932 (2020). https://doi.org/10. 1109/CVPR42600.2020.01294

[12] Mehnaz, S., Islam, M.T.: Rice leaf disease detection: A comparative study between cnn, transformer and non-neural network architectures. arXiv preprint arXiv:2501.06740 (2025)

[13] Babu, S., Maravarman, M., Pitchai, R.: Detection of rice plant disease using deep learning techniques. Journal of Mobile Multimedia 18(3), 757–770 (2022)

[14] Moslemi, A., Briskina, A., Dang, Z., Li, J.: A survey on knowledge distillation: Recent advancements. Machine Learning with Applications 18, 100605 (2024)

[15] Abbasi, M., Shahraki, A., Prieto, J., Arrieta, A.G., Corchado, J.M.: Unleashing the potential of knowledge distillation for iot traffic classification. IEEE Trans actions on Machine Learning in Communications and Networking 2, 221–235 (2024)

[16] UCI Machine Learning Repository: Rice Leaf Diseases Data Set. https://archive. ics.uci.edu/ml/datasets/Rice. Accessed: 2025-05-06 (2023)

[17] Sethy, P.K., Barpanda, N.K., Rath, A.K., Behera, S.K.: Deep featurebased rice leaf disease identification using support vector machine. Computers and Electronics in Agriculture 175, 105527 (2020)

[18] Bhattacharya, S., Mukherjee, A., Phadikar, S.: A deep learning approach for the classification of rice leaf diseases. Intelligence Enabled Research: DoSIER 2019, 61–69 (2020) 15

[19] Ahad, M.T., Li, Y., Song, B., Bhuiyan, T.: Comparison of cnn-based deep learning architectures for rice diseases classification. Artificial Intelligence in Agriculture 9, 22–35 (2023)

[20] Rawat, P., Pandey, A., et al.: Rice leaf diseases classification using deep learning techniques.